4

Finite Volume Methods

In this chapter we begin to study finite volume methods for the solution of conservation laws and hyperbolic systems. The fundamental concepts will be introduced, and then we will focus on first-order accurate methods for linear equations, in particular the *upwind method* for advection and for hyperbolic systems. This is the linear version of *Godunov's method*, which is the fundamental starting point for methods for nonlinear conservation laws, discussed beginning in Chapter 15. These methods are based on the solution to Riemann problems as discussed in the previous chapter for linear systems.

Finite volume methods are closely related to finite difference methods, and a finite volume method can often be interpreted directly as a finite difference approximation to the differential equation. However, finite volume methods are derived on the basis of the integral form of the conservation law, a starting point that turns out to have many advantages.

4.1 General Formulation for Conservation Laws

In one space dimension, a finite volume method is based on subdividing the spatial domain into intervals (the "finite volumes," also called *grid cells*) and keeping track of an approximation to the integral of q over each of these volumes. In each time step we update these values using approximations to the flux through the endpoints of the intervals.

Denote the ith grid cell by

$$\mathcal{C}_i = \left(x_{i-1/2}, x_{i+1/2}\right),$$

as shown in Figure 4.1. The value Q_i^n will approximate the average value over the ith interval at time t_n:

$$Q_i^n \approx \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} q(x, t_n)\, dx \equiv \frac{1}{\Delta x}\int_{\mathcal{C}_i} q(x, t_n)\, dx, \tag{4.1}$$

where $\Delta x = x_{i+1/2} - x_{i-1/2}$ is the length of the cell. For simplicity we will generally assume a uniform grid, but this is not required. (Nonuniform grids are discussed in Section 6.17.)

If $q(x, t)$ is a smooth function, then the integral in (4.1) agrees with the value of q at the midpoint of the interval to $\mathcal{O}(\Delta x^2)$. By working with cell averages, however, it is easier to use important properties of the conservation law in deriving numerical methods. In particular, we can insure that the numerical method is *conservative* in a way that mimics the

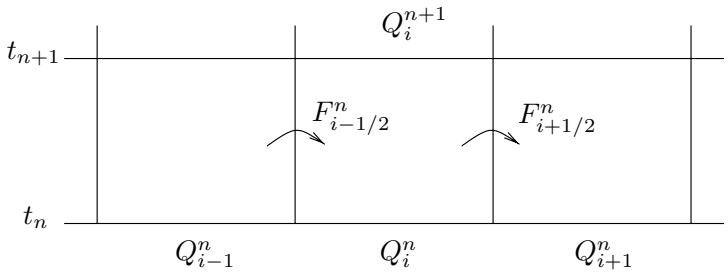

Fig. 4.1. Illustration of a finite volume method for updating the cell average Q_i^n by fluxes at the cell edges. Shown in x–t space.

true solution, and this is extremely important in accurately calculating shock waves, as we will see in Section 12.9. This is because $\sum_{i=1}^N Q_i^n \, \Delta x$ approximates the integral of q over the entire interval $[a, b]$, and if we use a method that is in *conservation form* (as described below), then this discrete sum will change only due to fluxes at the boundaries $x = a$ and $x = b$. The total mass within the computational domain will be preserved, or at least will vary correctly provided the boundary conditions are properly imposed.

The integral form of the conservation law (2.2) gives

$$\frac{d}{dt}\int_{\mathcal{C}_i} q(x,t)\,dx = f\big(q(x_{i-1/2},t)\big) - f\big(q(x_{i+1/2},t)\big). \tag{4.2}$$

We can use this expression to develop an explicit time-marching algorithm. Given Q_i^n, the cell averages at time t_n, we want to approximate Q_i^{n+1}, the cell averages at the next time t_{n+1} after a time step of length $\Delta t = t_{n+1} - t_n$. Integrating (4.2) in time from t_n to t_{n+1} yields

$$\int_{\mathcal{C}_i} q(x,t_{n+1})\,dx - \int_{\mathcal{C}_i} q(x,t_n)\,dx = \int_{t_n}^{t_{n+1}} f\big(q(x_{i-1/2},t)\big)\,dt - \int_{t_n}^{t_{n+1}} f\big(q(x_{i+1/2},t)\big)\,dt.$$

Rearranging this and dividing by Δx gives

$$\frac{1}{\Delta x}\int_{\mathcal{C}_i} q(x,t_{n+1})\,dx = \frac{1}{\Delta x}\int_{\mathcal{C}_i} q(x,t_n)\,dx - \frac{1}{\Delta x}\left[\int_{t_n}^{t_{n+1}} f\big(q(x_{i+1/2},t)\big)\,dt - \int_{t_n}^{t_{n+1}} f\big(q(x_{i-1/2},t)\big)\,dt\right]. \tag{4.3}$$

This tells us exactly how the cell average of q from (4.1) should be updated in one time step. In general, however, we cannot evaluate the time integrals on the right-hand side of (4.3) exactly, since $q(x_{i\pm1/2}, t)$ varies with time along each edge of the cell, and we don't have the exact solution to work with. But this does suggest that we should study numerical methods of the form

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\big(F_{i+1/2}^n - F_{i-1/2}^n\big), \tag{4.4}$$

where $F^n_{i-1/2}$ is some approximation to the average flux along $x = x_{i-1/2}$:

$$F^n_{i-1/2} \approx \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f\big(q(x_{i-1/2}, t)\big)\,dt. \tag{4.5}$$

If we can approximate this average flux based on the values Q^n, then we will have a fully discrete method. See Figure 4.1 for a schematic of this process.

For a hyperbolic problem information propagates with finite speed, so it is reasonable to first suppose that we can obtain $F^n_{i-1/2}$ based only on the values Q^n_{i-1} and Q^n_i, the cell averages on either side of this interface (see Section 4.4 for some discussion of this). Then we might use a formula of the form

$$F^n_{i-1/2} = \mathcal{F}\big(Q^n_{i-1}, Q^n_i\big) \tag{4.6}$$

where $\mathcal{F}$ is some *numerical flux function*. The method (4.4) then becomes

$$Q^{n+1}_i = Q^n_i - \frac{\Delta t}{\Delta x}\Big[\mathcal{F}\big(Q^n_i, Q^n_{i+1}\big) - \mathcal{F}\big(Q^n_{i-1}, Q^n_i\big)\Big]. \tag{4.7}$$

The specific method obtained depends on how we choose the formula $\mathcal{F}$, but in general any method of this type is an explicit method with a *three-point stencil*, meaning that the value Q^{n+1}_i will depend on the three values Q^n_{i-1}, Q^n_i, and Q^n_{i+1} at the previous time level. Moreover, it is said to be in *conservation form*, since it mimics the property (4.3) of the exact solution. Note that if we sum $\Delta x\, Q^{n+1}_i$ from (4.4) over any set of cells, we obtain

$$\Delta x \sum_{i=I}^{J} Q^{n+1}_i = \Delta x \sum_{i=I}^{J} Q^n_i - \frac{\Delta t}{\Delta x}\big(F^n_{J+1/2} - F^n_{I-1/2}\big). \tag{4.8}$$

The sum of the flux differences cancels out except for the fluxes at the extreme edges. Over the full domain we have exact conservation except for fluxes at the boundaries. (Numerical boundary conditions are discussed later.)

The method (4.7) can be viewed as a direct finite difference approximation to the conservation law $q_t + f(q)_x = 0$, since rearranging it gives

$$\frac{Q^{n+1}_i - Q^n_i}{\Delta t} + \frac{F^n_{i+1/2} - F^n_{i-1/2}}{\Delta x} = 0. \tag{4.9}$$

Many methods can be equally well viewed as finite difference approximations to this equation or as finite volume methods.

4.2 A Numerical Flux for the Diffusion Equation

The above derivation was presented for a conservation law in which the flux $f(q)$ depends only on the state q. The same derivation works more generally, however, for example if the flux depends explicitly on x or if it depends on derivatives of the solution such as q_x. As an example consider the diffusion equation (2.22), where the flux (2.20) is

$$f(q_x, x) = -\beta(x) q_x.$$

Given two cell averages Q_{i-1} and Q_i, the numerical flux $\mathcal{F}(Q_{i-1}, Q_i)$ at the cell interface between can very naturally be defined as

$$\mathcal{F}(Q_{i-1}, Q_i) = -\beta_{i-1/2}\left(\frac{Q_i - Q_{i-1}}{\Delta x}\right), \tag{4.10}$$

where $\beta_{i-1/2} \approx \beta(x_{i-1/2})$. This numerical flux has the natural physical interpretation that the conserved quantity measured by q flows from one grid cell to its neighbor at a rate proportional to the difference in Q-values in the two cells, with $\beta_{i-1/2}$ measuring the conductivity of the interface between these cells. This is a macroscopic version of Fick's law or Fourier's law (or Newton's law of cooling).

Using (4.10) in (4.7) gives a standard finite difference discretization of the diffusion equation,

$$Q_i^{n+1} = Q_i^n + \frac{\Delta t}{\Delta x^2}\left[\beta_{i+1/2}(Q_{i+1}^n - Q_i^n) - \beta_{i-1/2}(Q_i^n - Q_{i-1}^n)\right]. \tag{4.11}$$

If $\beta \equiv$ constant, then this takes the simpler form

$$Q_i^{n+1} = Q_i^n + \frac{\Delta t}{\Delta x^2}\beta(Q_{i-1}^n - 2Q_i^n + Q_{i+1}^n) \tag{4.12}$$

and we recognize the centered approximation to q_{xx}.

For parabolic equations, explicit methods of this type are generally not used, since they are only stable if $\Delta t = \mathcal{O}(\Delta x^2)$. Instead an implicit method is preferable, such as the standard *Crank–Nicolson method*,

$$\begin{aligned} Q_i^{n+1} = Q_i^n + \frac{\Delta t}{2\,\Delta x^2}\big[&\beta_{i+1/2}(Q_{i+1}^n - Q_i^n) - \beta_{i-1/2}(Q_i^n - Q_{i-1}^n) \\ &+ \beta_{i+1/2}(Q_{i+1}^{n+1} - Q_i^{n+1}) - \beta_{i-1/2}(Q_i^{n+1} - Q_{i-1}^{n+1})\big]. \end{aligned} \tag{4.13}$$

This can also be viewed as a finite volume method, with the flux

$$F_{i-1/2}^n = -\frac{1}{2\,\Delta x}\left[\beta_{i-1/2}(Q_i^n - Q_{i-1}^n) + \beta_{i-1/2}(Q_i^{n+1} - Q_{i-1}^{n+1})\right].$$

This is a natural approximation to the time-averaged flux (4.5), and in fact has the advantage of being a second-order accurate approximation (since it is centered in both space and time) as well as giving an unconditionally stable method.

The stability difficulty with explicit methods for the diffusion equation arises from the fact that the flux (4.10) contains Δx in the denominator, leading to stability restrictions involving $\Delta t/(\Delta x)^2$ after multiplying by $\Delta t/\Delta x$ in (4.4). For first-order hyperbolic equations the flux function involves only q and not q_x, and *explicit* methods are generally more efficient. However, some care must be taken to obtain stable methods in the hyperbolic case as well.

4.3 Necessary Components for Convergence

Later in this chapter we will introduce various ways to define the numerical flux function of (4.6) for hyperbolic equations, leading to various different finite volume methods. There

are several considerations that go into judging how good a particular flux function is for numerical computation. One essential requirement is that the resulting method should be *convergent*, i.e., the numerical solution should converge to the true solution of the differential equation as the grid is refined (as $\Delta x, \Delta t \to 0$). This generally requires two conditions:

- The method must be *consistent* with the differential equation, meaning that it approximates it well locally.
- The method must be *stable* in some appropriate sense, meaning that the small errors made in each time step do not grow too fast in later time steps.

Stability and convergence theory are discussed in more detail in Chapter 8. At this stage we simply introduce some essential ideas that are useful in discussing the basic methods.

4.3.1 Consistency

The numerical flux should approximate the integral in (4.5). In particular, if the function $q(x, t) \equiv \bar{q}$ is constant in x, then q will not change with time and the integral in (4.5) simply reduces to $f(\bar{q})$. As a result, if $Q_{i-1}^n = Q_i^n = \bar{q}$, then we expect the numerical flux function $\mathcal{F}$ of (4.6) to reduce to $f(\bar{q})$, so we require

$$\mathcal{F}(\bar{q}, \bar{q}) = f(\bar{q}) \tag{4.14}$$

for any value $\bar{q}$. This is part of the basic consistency condition. We generally also expect continuity in this function as Q_{i-1} and Q_i vary, so that $\mathcal{F}(Q_{i-1}, Q_i) \to f(\bar{q})$ as $Q_{i-1}, Q_i \to \bar{q}$. Typically some requirement of *Lipschitz continuity* is made, e.g., there exists a constant L so that

$$|\mathcal{F}(Q_{i-1}, Q_i) - f(\bar{q})| \le L \max(|Q_i - \bar{q}|, |Q_{i-1} - \bar{q}|). \tag{4.15}$$

4.4 The CFL Condition

Stability analysis is considered in detail in Chapter 8. Here we mention only the CFL condition, which is a *necessary* condition that must be satisfied by any finite volume or finite difference method if we expect it to be stable and converge to the solution of the differential equation as the grid is refined. It simply states that the method must be used in such a way that information has a chance to propagate at the correct physical speeds, as determined by the eigenvalues of the flux Jacobian $f'(q)$.

With the explicit method (4.7) the value Q_i^{n+1} depends only on three values Q_{i-1}^n, Q_i^n, and Q_{i+1}^n at the previous time step. Suppose we apply such a method to the advection equation $q_t + \bar{u} q_x = 0$ with $\bar{u} > 0$ so that the exact solution simply translates at speed $\bar{u}$ and propagates a distance $\bar{u}\,\Delta t$ over one time step. Figure 4.2(a) shows a situation where $\bar{u}\,\Delta t < \Delta x$, so that information propagates less than one grid cell in a single time step. In this case it makes sense to define the flux at $x_{i-1/2}$ in terms of Q_{i-1}^n and Q_i^n alone. In Figure 4.2(b), on the other hand, a larger time step is used with $\bar{u}\,\Delta t > \Delta x$. In this case the true flux at $x_{i-1/2}$ clearly depends on the value of Q_{i-2}^n, and so should the new cell average Q_i^{n+1}. The method (4.7) would certainly be unstable when applied with such a large time

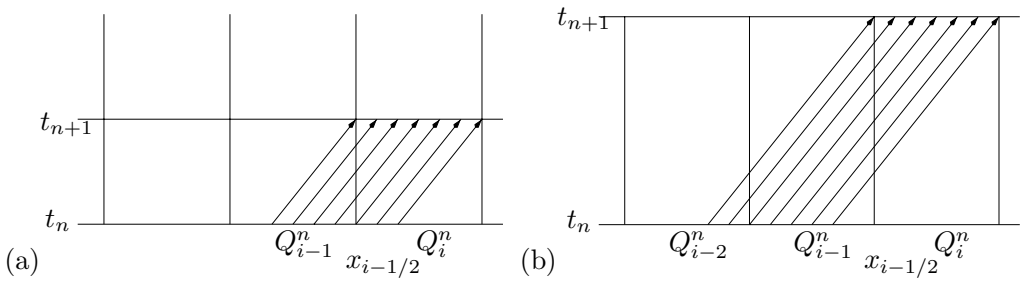

Fig. 4.2. Characteristics for the advection equation, showing the information that flows into cell $\mathcal{C}_i$ during a single time step. (a) For a small enough time step, the flux at $x_{i-1/2}$ depends only on the values in the neighboring cells – only on Q^n_{i-1} in this case where $\bar{u} > 0$. (b) For a larger time step, the flux should depend on values farther away.

step, no matter how the flux (4.6) was specified, if this numerical flux depended only on Q^n_{i-1} and Q^n_i.

This is a consequence of the *CFL condition*, named after Courant, Friedrichs, and Lewy. They wrote one of the first papers on finite difference methods for partial differential equations [93] in 1928. (There is an English translation in [94].) They used finite difference methods as an analytic tool for proving the existence of solutions of certain PDEs. The idea is to define a sequence of approximate solutions (via finite difference equations), prove that they converge as the grid is refined, and then show that the limit function must satisfy the PDE, giving the existence of a solution. In the course of proving convergence of this sequence (which is precisely what we are interested in numerically), they recognized the following necessary stability condition for any numerical method:

CFL Condition: *A numerical method can be convergent only if its numerical domain of dependence contains the true domain of dependence of the PDE, at least in the limit as Δt and Δx go to zero.*

It is very important to note that the CFL condition is only a *necessary* condition for stability (and hence convergence). It is not always *sufficient* to guarantee stability. In the next section we will see a numerical flux function yielding a method that is unstable even when the CFL condition is satisfied.

The domain of dependence $\mathcal{D}(X, T)$ for a PDE has been defined in Section 3.6. The numerical domain of dependence of a method can be defined in a similar manner as the set of points where the initial data can possibly affect the numerical solution at the point (X, T). This is easiest to illustrate for a finite difference method where pointwise values of Q are used, as shown in Figure 4.3 for a three-point method. In Figure 4.3(a) we see that Q^2_i depends on Q^1_{i-1}, Q^1_i, Q^1_{i+1} and hence on $Q^0_{i-2}, \ldots, Q^0_{i+2}$. Only initial data in the interval $X - 2\,\Delta x^a \le x \le X + 2\,\Delta x^a$ can affect the numerical solution at $(X, T) = (x_i, t_2)$. If we now refine the grid by a factor of 2 in both space and time ($\Delta x^b = \Delta x^a/2$), but continue to focus on the same physical point (X, T), then we see in Figure 4.3(b) that the numerical approximation at this point now depends on initial data at more points in the interval $X - 4\,\Delta x^b \le x \le X + 4\,\Delta x^b$. But this is the same interval as before. If we continue to refine the grid with the ratio $\Delta t/\Delta x \equiv r$ fixed, then the numerical domain of dependence of a general point (X, T) is $X - T/r \le x \le X + T/r$.

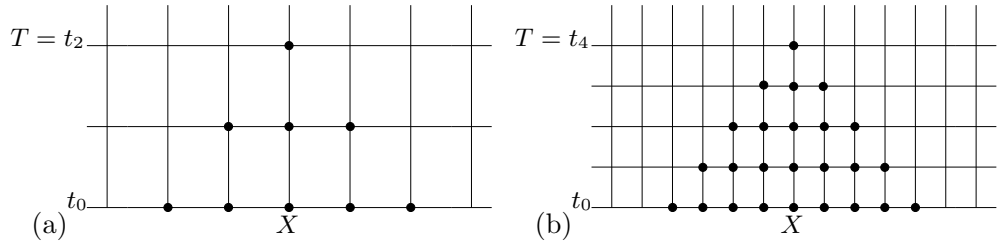

Fig. 4.3. (a) Numerical domain of dependence of a grid point when using a three-point explicit finite difference method, with mesh spacing Δx^a. (b) On a finer grid with mesh spacing $\Delta x^b = \frac{1}{2}\Delta x^a$. Similar figures can be drawn for finite volume methods.

In order for the CFL condition to be satisfied, the domain of dependence of the true solution must lie within this interval. For the advection equation $q_t + \bar{u}q_x = 0$, for example, $\mathcal{D}(X, T)$ is the single point $X - \bar{u}T$, since $q(X, T) = \mathring{q}(X - \bar{u}T)$. The CFL condition then requires

$$X - T/r \le X - \bar{u}T \le X + T/r$$

and hence

$$\nu \equiv \left| \frac{\bar{u}\,\Delta t}{\Delta x} \right| \le 1. \tag{4.16}$$

If this condition is not satisfied, then a change in the initial data $\mathring{q}$ at $X - \bar{u}T$ would change the true solution at (X, T) but could have no effect on the numerical solution at this point. Clearly the method cannot converge to the proper solution for all choices of initial data under these circumstances.

The ratio ν in (4.16) is sometimes called the CFL number, or more frequently the *Courant number*. Returning to the finite volume method illustrated in Figure 4.2, note that the Courant number measures the fraction of a grid cell that information propagates through in one time step. For a hyperbolic system of equations there are generally a set of m wave speeds $\lambda^1, \ldots, \lambda^m$ as described in Chapter 3, and the true domain of dependence is given by (3.14). In this case we define the Courant number by

$$\nu = \frac{\Delta t}{\Delta x} \max_p |\lambda^p|. \tag{4.17}$$

For a three-point method the CFL condition again leads to a necessary condition $\nu \le 1$.

Note that if the method has a wider stencil, then the CFL condition will lead to a more lenient condition on the time step. For a centered five-point stencil in which Q_i^{n+1} depends also on Q_{i-2}^n and Q_{i+2}^n, the CFL condition gives $\nu \le 2$. Again this will only be a necessary condition, and a more detailed analysis of stability would be required to determine the actual stability constraint needed to guarantee convergence.

For hyperbolic equations we typically use explicit methods and grids for which the Courant number is somewhat smaller than 1. This allows keeping $\Delta t/\Delta x$ fixed as the grid is refined, which is sensible in that generally we wish to add more resolution at the same rate in both space and in time in order to improve the solution.

For a parabolic equation such as the diffusion equation, on the other hand, the CFL condition places more severe constraints on an explicit method. The domain of dependence of any point (X, T) for $T > 0$ is now the whole real line, $\mathcal{D}(X, T) = (-\infty, \infty)$, and data at every point can affect the solution everywhere. Because of this infinite propagation speed, the CFL condition requires that the numerical domain of dependence must include the whole real line, at least in the limit as $\Delta t, \Delta x \to 0$. For an explicit method this can be accomplished by letting Δt approach zero more rapidly than Δx as we refine the grid, e.g., by taking $\Delta t = \mathcal{O}(\Delta x^2)$ as required for the method (4.11). A better way to satisfy the CFL condition in this case is to use an implicit method. In this case the numerical domain of dependence is the entire domain, since all grid points are coupled together.

4.5 An Unstable Flux

We now return to the general finite volume method (4.4) for a hyperbolic system and consider various ways in which the numerical flux might be defined. In particular we consider flux functions $\mathcal{F}$ as in (4.6). We wish to define the average flux at $x_{i-1/2}$ based on the data Q^n_{i-1} and Q^n_i to the left and right of this point. A first attempt might be the simple arithmetic average

$$F^n_{i-1/2} = \mathcal{F}(Q^n_{i-1}, Q^n_i) = \frac{1}{2}\big[f(Q^n_{i-1}) + f(Q^n_i)\big]. \tag{4.18}$$

Using this in (4.4) would give

$$Q^{n+1}_i = Q^n_i - \frac{\Delta t}{2\,\Delta x}\big[f(Q^n_{i+1}) - f(Q^n_{i-1})\big]. \tag{4.19}$$

Unfortunately, this method is generally unstable for hyperbolic problems and cannot be used, even if the time step is small enough that the CFL condition is satisfied. (See Exercise 8.1.)

4.6 The Lax–Friedrichs Method

The classical Lax–Friedrichs (LxF) method has the form

$$Q^{n+1}_i = \frac{1}{2}(Q^n_{i-1} + Q^n_{i+1}) - \frac{\Delta t}{2\,\Delta x}\big[f(Q^n_{i+1}) - f(Q^n_{i-1})\big]. \tag{4.20}$$

This is very similar to the unstable method (4.19), but the value Q^n_i is replaced by the average $\frac{1}{2}(Q^n_{i-1} + Q^n_{i+1})$. For a linear hyperbolic equation this method is stable provided $\nu \le 1$, where the Courant number ν is defined in (4.17).

At first glance the method (4.20) does not appear to be of the form (4.4). However, it can be put into this form by defining the numerical flux as

$$\mathcal{F}(Q^n_{i-1}, Q^n_i) = \frac{1}{2}\big[f(Q^n_{i-1}) + f(Q^n_i)\big] - \frac{\Delta x}{2\,\Delta t}(Q^n_i - Q^n_{i-1}). \tag{4.21}$$

Note that this flux looks like the unstable centered flux (4.18) with the addition of another term similar to the flux (4.10) of the diffusion equation. By using this flux we appear to be modeling the advection–diffusion equation $q_t + f(q)_x = \beta q_{xx}$ with $\beta = \frac{1}{2}(\Delta x)^2/\Delta t$. But

if we fix $\Delta t/\Delta x$, then we see that this coefficient vanishes as the grid is refined, so in the limit the method is still consistent with the original hyperbolic equation. This additional term can be interpreted as *numerical diffusion* that damps the instabilities arising in (4.19) and gives a method that can be shown to be stable for Courant number up to 1 (which is also the CFL limit for this three-point method). However, the Lax–Friedrichs method introduces much more diffusion than is actually required, and gives numerical results that are typically badly smeared unless a very fine grid is used.

4.7 The Richtmyer Two-Step Lax–Wendroff Method

The Lax–Friedrichs method is only first-order accurate. Second-order accuracy can be achieved by using a better approximation to the integral in (4.5). One approach is to first approximate q at the midpoint in time, $t_{n+1/2} = t_n + \frac{1}{2}\Delta t$, and evaluate the flux at this point. The *Richtmyer method* is of this form with

$$F^n_{i-1/2} = f\bigl(Q^{n+1/2}_{i-1/2}\bigr), \tag{4.22}$$

where

$$Q^{n+1/2}_{i-1/2} = \frac{1}{2}\bigl(Q^n_{i-1} + Q^n_i\bigr) - \frac{\Delta t}{2\,\Delta x}\bigl[f\bigl(Q^n_i\bigr) - f\bigl(Q^n_{i-1}\bigr)\bigr]. \tag{4.23}$$

Note that $Q^{n+1/2}_{i-1/2}$ is obtained by applying the Lax–Friedrichs method at the cell interface with Δx and Δt replaced by $\frac{1}{2}\Delta x$ and $\frac{1}{2}\Delta t$ respectively.

For a linear system of equations, $f(q) = Aq$, the Richtmyer method reduces to the standard Lax–Wendroff method, discussed further in Section 6.1. As we will see, these methods often lead to spurious oscillations in solutions, particularly when solving problems with discontinuous solutions. Additional numerical diffusion (or *artificial viscosity*) can be added to eliminate these oscillations, as first proposed by von Neumann and Richtmyer [477]. In Chapter 6 we will study a different approach to obtaining better accuracy that allows us to avoid these oscillations more effectively.

4.8 Upwind Methods

The methods considered above have all been centered methods, symmetric about the point where we are updating the solution. For hyperbolic problems, however, we expect information to propagate as waves moving along characteristics. For a system of equations we have several waves propagating at different speeds and perhaps in different directions. It makes sense to try to use our knowledge of the structure of the solution to determine better numerical flux functions. This idea gives rise to *upwind* methods in which the information for each characteristic variable is obtained by looking in the direction from which this information should be coming.

For the scalar advection equation there is only one speed, which is either positive or negative, and so an upwind method is typically also a *one-sided* method, with Q^{n+1}_i determined based on values only to the left or only to the right. This is discussed in the next section.

For a system of equations there may be waves traveling in both directions, so an upwind method must still use information from both sides, but typically uses characteristic decomposition (often via the solution of Riemann problems) to select *which* information to use from each side. Upwind methods for systems of equations are discussed beginning in Section 4.10.

4.9 The Upwind Method for Advection

For the constant-coefficient advection equation $q_t + \bar{u} q_x = 0$, Figure 4.2(a) indicates that the flux through the left edge of the cell is entirely determined by the value Q_{i-1}^n in the cell to the left of this cell. This suggests defining the numerical flux as

$$F_{i-1/2}^n = \bar{u} Q_{i-1}^n. \tag{4.24}$$

This leads to the standard *first-order upwind method* for the advection equation,

$$Q_i^{n+1} = Q_i^n - \frac{\bar{u}\,\Delta t}{\Delta x}\left(Q_i^n - Q_{i-1}^n\right). \tag{4.25}$$

Note that this can be rewritten as

$$\frac{Q_i^{n+1} - Q_i^n}{\Delta t} + \bar{u}\left(\frac{Q_i^n - Q_{i-1}^n}{\Delta x}\right) = 0,$$

whereas the unstable centered method (4.19) applied to the advection equation is

$$\frac{Q_i^{n+1} - Q_i^n}{\Delta t} + \bar{u}\left(\frac{Q_{i+1}^n - Q_{i-1}^n}{2\,\Delta x}\right) = 0.$$

The upwind method uses a one-sided approximation to the derivative q_x in place of the centered approximation.

Another interpretation of the upwind method is suggested by Figure 4.4(a). If we think of the Q_i^n as being values at grid points, $Q_i^n \approx q(x_i, t_n)$, as is standard in a finite difference method, then since $q(x, t)$ is constant along characteristics we expect

$$Q_i^{n+1} \approx q(x_i, t_{n+1}) = q(x_i - \bar{u}\,\Delta t, t_n).$$

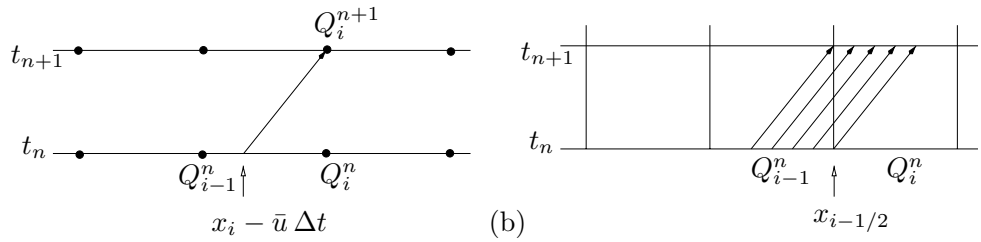

Fig. 4.4. Two interpretations of the upwind method for advection. (a) If Q_i^n represents the value at a grid point, then we can trace the characteristic back and interpolate. (b) If Q_i^n represents the cell average, then the flux at the interface is determined by the cell value on the upwind side.

If we approximate the value on the right by a linear interpolation between the grid values Q_{i-1}^n and Q_i^n, we obtain the method

$$Q_i^{n+1} = \frac{\bar{u}\,\Delta t}{\Delta x} Q_{i-1}^n + \left(1 - \frac{\bar{u}\,\Delta t}{\Delta x}\right) Q_i^n. \tag{4.26}$$

This is simply the upwind method, since a rearrangement gives (4.25).

Note that we must have

$$0 \le \frac{\bar{u}\,\Delta t}{\Delta x} \le 1 \tag{4.27}$$

in order for the characteristic to fall between the neighboring points so that this interpolation is sensible. In fact, (4.27) must be satisfied in order for the upwind method to be stable, and also follows from the CFL condition. Note that if (4.27) is satisfied then (4.26) expresses Q_i^{n+1} as a *convex combination* of Q_i^n and Q_{i-1}^n (i.e., the weights are both nonnegative and sum to 1). This is a key fact in proving stability of the method. (See Section 8.3.4.)

We are primarily interested in finite volume methods, and so other interpretations of the upwind method will be more valuable. Figure 4.4(b) and Figure 4.5 show the finite volume viewpoint, in which the value Q_i^n is now seen as a cell average of q over the ith grid cell $\mathcal{C}_i$. We think of mixing up the tracer within this cell so that it has this average value at every point in the cell, at time t_n. This defines a piecewise constant function at time t_n with the

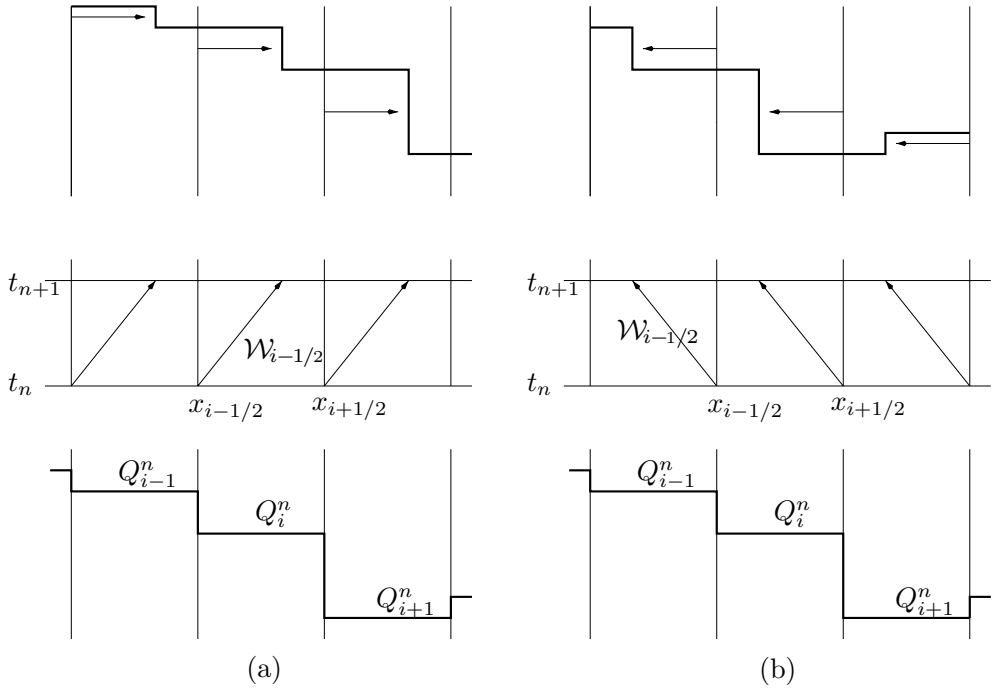

Fig. 4.5. Wave-propagation interpretation of the upwind method for advection. The bottom pair of graphs shows data at time t_n, represented as a piecewise constant function. Over time Δt this function shifts by a distance $\bar{u}\,\Delta t$ as indicated in the middle pair of graphs. We view the discontinuity that originates at $x_{i-1/2}$ as a wave $\mathcal{W}_{i-1/2}$. The top pair shows the piecewise constant function at the end of the time step after advecting. The new cell averages Q_i^{n+1} in each cell are then computed by averaging this function over each cell. (a) shows a case with $\bar{u} > 0$, while (b) shows $\bar{u} < 0$.

value Q_i^n in cell $\mathcal{C}_i$. As time evolves, this piecewise constant function advects to the right with velocity $\bar{u}$, and the jump between states Q_{i-1}^n and Q_i^n shifts a distance $\bar{u}\,\Delta t$ into cell $\mathcal{C}_i$. At the end of the time step we can compute a new cell average Q_i^{n+1} in order to repeat this process. To compute Q_i^{n+1} we must average the piecewise constant function shown in the top of Figure 4.5 over the cell. Computing this average results in the same convex combination (4.26) as was motivated by the characteristic-based approach of Figure 4.4(a), as the reader should verify.

We can also take a wave-propagation viewpoint, which will prove useful in extending and implementing the upwind method. The jump $\mathcal{W}_{i-1/2} \equiv Q_i^n - Q_{i-1}^n$ can be viewed as a wave that is moving into cell $\mathcal{C}_i$ at velocity $\bar{u}$. This wave modifies the value of q by $-\mathcal{W}_{i-1/2}$ at each point it passes. Over the time step it moves a distance $\bar{u}\,\Delta t$ and passes through a fraction $\bar{u}\,\Delta t/\Delta x$ of the grid cell, and hence the cell average is modified by this fraction of $-\mathcal{W}_{i-1/2}$:

$$Q_i^{n+1} = Q_i^n + \frac{\bar{u}\,\Delta t}{\Delta x}\bigl(-\mathcal{W}_{i-1/2}\bigr). \tag{4.28}$$

This again results in the upwind method (4.25).

In the above discussion we have assumed that $\bar{u} > 0$. On the other hand if $\bar{u} < 0$ then the upwind direction is to the right and so the numerical flux at $x_{i-1/2}$ is

$$F_{i-1/2}^n = \bar{u}\,Q_i^n. \tag{4.29}$$

The upwind method then has the form

$$Q_i^{n+1} = Q_i^n - \frac{\bar{u}\,\Delta t}{\Delta x}\bigl(Q_{i+1}^n - Q_i^n\bigr). \tag{4.30}$$

This can also be written in wave-propagation form as

$$Q_i^{n+1} = Q_i^n - \frac{\bar{u}\,\Delta t}{\Delta x}\mathcal{W}_{i+1/2}, \tag{4.31}$$

with $\mathcal{W}_{i+1/2} = Q_{i+1}^n - Q_i^n$. All the interpretations presented above carry over to this case $\bar{u} < 0$, with the direction of flow reversed. The method (4.31) is stable provided that

$$-1 \le \frac{\bar{u}\,\Delta t}{\Delta x} \le 0. \tag{4.32}$$

The two formulas (4.24) and (4.29) can be combined into a single upwind formula that is valid for $\bar{u}$ of either sign,

$$F_{i-1/2}^n = \bar{u}^- Q_i^n + \bar{u}^+ Q_{i-1}^n, \tag{4.33}$$

where

$$\bar{u}^+ = \max(\bar{u}, 0), \qquad \bar{u}^- = \min(\bar{u}, 0). \tag{4.34}$$

The wave-propagation versions of the upwind method in (4.28) and (4.31) can also be combined to give the more general formula

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left(\bar{u}^+ \mathcal{W}_{i-1/2} + \bar{u}^- \mathcal{W}_{i+1/2}\right). \tag{4.35}$$

This formulation will be useful in extending this method to more general hyperbolic problems. Not all hyperbolic equations are in conservation form; consider for example the variable-coefficient linear equation (2.78) or the quasilinear system (2.81) with suitable coefficient matrix. Such equations do not have a flux function, and so numerical methods of the form (4.7) cannot be applied. However, these hyperbolic problems can still be solved using finite volume methods that result from a simple generalization of the high-resolution methods developed for hyperbolic conservation laws. The unifying feature of all hyperbolic equations is that they model waves that travel at finite speeds. In particular, the solution to a Riemann problem with piecewise constant initial data (as discussed in Chapter 3) consists of waves traveling at constant speeds away from the location of the jump discontinuity in the initial data.

In Section 4.12 we will generalize (4.35) to obtain an approach to solving hyperbolic systems that is more general than the flux-differencing form (4.4). First, however, we see how the upwind method can be extended to systems of equations.

4.10 Godunov's Method for Linear Systems

The upwind method for the advection equation can be derived as a special case of the following approach, which can also be applied to systems of equations. This will be referred to as the *REA algorithm*, for reconstruct–evolve–average. These are one-word summaries of the three steps involved.

Algorithm 4.1 (REA).

1. Reconstruct *a piecewise polynomial function* $\tilde{q}^n(x, t_n)$ *defined for all* x*, from the cell averages* Q_i^n*. In the simplest case this is a piecewise constant function that takes the value* Q_i^n *in the* i*th grid cell, i.e.,*

$$\tilde{q}^n(x, t_n) = Q_i^n \qquad \text{for all } x \in \mathcal{C}_i.$$

2. Evolve *the hyperbolic equation exactly (or approximately) with this initial data to obtain* $\tilde{q}^n(x, t_{n+1})$ *a time* Δt *later.*
3. Average *this function over each grid cell to obtain new cell averages*

$$Q_i^{n+1} = \frac{1}{\Delta x}\int_{\mathcal{C}_i} \tilde{q}^n(x, t_{n+1})\, dx.$$

This whole process is then repeated in the next time step.

In order to implement this procedure, we must be able to solve the hyperbolic equation in step 2. Because we are starting with piecewise constant data, this can be done using the

theory of Riemann problems as introduced for linear problems in Chapter 3. When applied to the advection equation, this leads to the upwind algorithm, as illustrated in Figure 4.5.

The general approach of Algorithm 4.1 was originally proposed by Godunov [157] as a method for solving the nonlinear Euler equations of gas dynamics. Application in that context hinges on the fact that, even for this nonlinear system, the Riemann problem with piecewise constant initial data can be solved and the solution consists of a finite set of waves traveling at constant speeds, as we will see in Chapter 13.

Godunov's method for gas dynamics revolutionized the field of computational fluid dynamics, by overcoming many of the difficulties that had plagued earlier numerical methods for compressible flow. Using the wave structure determined by the Riemann solution allows shock waves to be handled in a properly "upwinded" manner even for systems of equations where information propagates in both directions. We will explore this for linear systems in the remainder of this chapter.

In step 1 we reconstruct a function $\tilde{q}^n(x, t_n)$ from the discrete cell averages. In Godunov's original approach this reconstruction is a simple piecewise constant function, and for now we concentrate on this form of reconstruction. This leads most naturally to Riemann problems, but gives only a first-order accurate method, as we will see. To obtain better accuracy one might consider using a better reconstruction, for example a piecewise linear function that is allowed to have a nonzero slope σ_i^n in the ith grid cell. This idea forms the basis for the *high-resolution* methods that are considered starting in Chapter 6.

Clearly the exact solution at time t_{n+1} can be constructed by piecing together the Riemann solutions, provided that the time step Δt is short enough that the waves from two adjacent Riemann problems have not yet started to interact. Figure 4.6 shows a schematic diagram of this process for the equations of linear acoustics with constant sound speed c, in which case this requires that

$$c\,\Delta t \le \frac{1}{2}\Delta x,$$

so that each wave goes at most halfway through the grid cell. Rearranging gives

$$\frac{c\,\Delta t}{\Delta x} \le \frac{1}{2}. \tag{4.36}$$

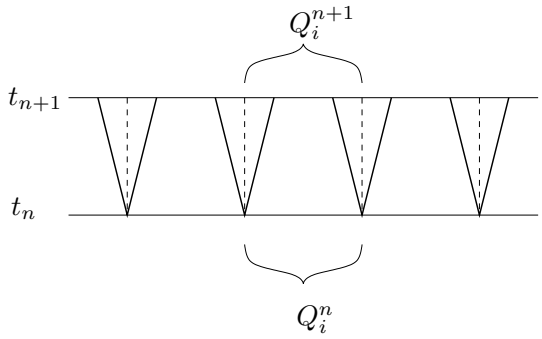

Fig. 4.6. An illustration of the process of Algorithm 4.1 for the case of linear acoustics. The Riemann problem is solved at each cell interface, and the wave structure is used to determine the exact solution time Δt later. This solution is averaged over the grid cell to determine Q_i^{n+1}.

The quantity $c\,\Delta t/\Delta x$ is simply the Courant number, so it appears that we are limited in (4.36) to a Courant number less than 1/2. But we will see below that this method is easily extended to Courant numbers up to 1.

4.11 The Numerical Flux Function for Godunov's Method

We now develop a finite volume method based on Algorithm 4.1 that can be easily implemented in practice. As presented, the algorithm seems cumbersome to implement. The exact solution $\tilde{q}^n(x, t_{n+1})$ will typically contain several discontinuities and we must compute its integral over each grid cell in order to determine the new cell averages Q_i^{n+1}. However, it turns out to be easy to determine the numerical flux function $\mathcal{F}$ that corresponds to Godunov's method.

Recall the formula (4.5), which states that the numerical flux $F_{i-1/2}^n$ should approximate the time average of the flux at $x_{i-1/2}$ over the time step,

$$F_{i-1/2}^n \approx \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f(q(x_{i-1}, t))\,dt.$$

In general the function $q(x_{i-1/2}, t)$ varies with t, and we certainly don't know this variation of the exact solution. However, we can compute this integral *exactly* if we replace $q(x, t)$ by the function $\tilde{q}^n(x, t)$ defined in Algorithm 4.1 using Godunov's piecewise constant reconstruction. The structure of this function is shown in Figure 3.3, for example, and so clearly $\tilde{q}^n(x_{i-1/2}, t)$ is constant over the time interval $t_n < t < t_{n+1}$. The Riemann problem centered at $x_{i-1/2}$ has a similarity solution that is constant along rays $(x - x_{i-1/2})/(t - t_n) =$ constant, and looking at the value along $(x - x_{i-1/2})/t = 0$ gives the value of $\tilde{q}^n(x_{i-1/2}, t)$. Denote this value by $Q_{i-1/2}^{\Downarrow} = q^{\Downarrow}(Q_{i-1}^n, Q_i^n)$. This suggests defining the numerical flux $F_{i-1/2}^n$ by

$$\begin{aligned} F_{i-1/2}^n &= \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f\big(q^{\Downarrow}(Q_{i-1}^n, Q_i^n)\big)\,dt \\ &= f\big(q^{\Downarrow}(Q_{i-1}^n, Q_i^n)\big). \end{aligned} \tag{4.37}$$

This gives a simple way to implement Godunov's method for a general system of conservation laws:

- Solve the Riemann problem at $x_{i-1/2}$ to obtain $q^{\Downarrow}(Q_{i-1}^n, Q_i^n)$.
- Define the flux $F_{i-1/2}^n = \mathcal{F}(Q_{i-1}^n, Q_i^n)$ by (4.37).
- Apply the flux-differencing formula (4.4).

Godunov's method is often presented in this form.

4.12 The Wave-Propagation Form of Godunov's Method

By taking a slightly different viewpoint, we can also develop simple formulas for Godunov's method on linear systems of equations that are analogous to the form (4.35) for the upwind

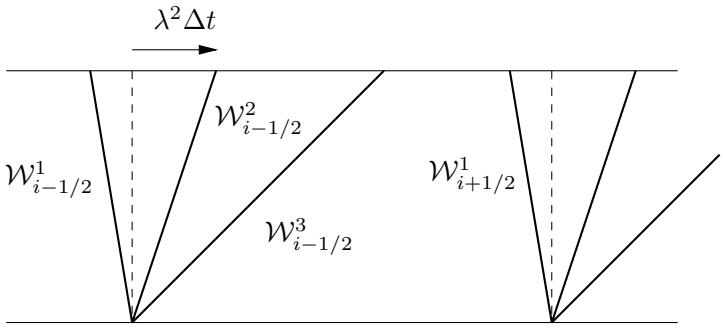

Fig. 4.7. An illustration of the process of Algorithm 4.1 for the case of a linear system of three equations. The Riemann problem is solved at each cell interface, and the wave structure is used to determine the exact solution time Δt later. The wave $\mathcal{W}^2_{i-1/2}$, for example, has moved a distance $\lambda^2 \Delta t$ into the cell.

method on the advection equation. This viewpoint is particularly useful in extending Godunov's method to hyperbolic systems that are not in conservation form.

Figure 4.7 shows a more complicated version of Figure 4.6, in which a linear system of three equations is solved assuming $\lambda^1 < 0 < \lambda^2 < \lambda^3$. The function $\tilde{q}^n(x, t_{n+1})$ will typically have three discontinuities in the grid cell $\mathcal{C}_i$, at the points $x_{i-1/2} + \lambda^2 \Delta t$, $x_{i-1/2} + \lambda^3 \Delta t$, and $x_{i+1/2} + \lambda^1 \Delta t$.

Instead of trying to work with this function directly to compute the new cell average, recall from Section 3.8 that for a linear system the solution to the Riemann problem can be expressed as a set of waves,

$$Q_i - Q_{i-1} = \sum_{p=1}^{m} \alpha^p_{i-1/2} r^p \equiv \sum_{p=1}^{m} \mathcal{W}^p_{i-1/2}. \tag{4.38}$$

Let's investigate what effect each wave has on the cell average. Consider the wave denoted by $\mathcal{W}^2_{i-1/2}$ in Figure 4.7, for example. It consists of a jump in q given by

$$\mathcal{W}^2_{i-1/2} = \alpha^2_{i-1/2} r^2,$$

propagating at speed λ^2, and hence after time Δt it has moved a distance $\lambda^2 \Delta t$. This wave modifies the value of q over a fraction of the grid cell given by $\lambda^2 \Delta t / \Delta x$. It follows that the effect of this wave on the *cell average* of q is to change the average value by the amount

$$-\frac{\lambda^2 \Delta t}{\Delta x} \mathcal{W}^2_{i-1/2}.$$

The minus sign arises because the value $\mathcal{W}^2_{i-1/2}$ measures the jump from right to left, and is analogous to the minus sign in (4.28).

Each of the waves entering the grid cell has an analogous effect on the cell average, and the new cell average can be found by simply adding up these independent effects. For the

case shown in Figure 4.7, we thus find that

$$\begin{aligned} Q_i^{n+1} &= Q_i^n - \frac{\lambda^2 \Delta t}{\Delta x} \mathcal{W}_{i-1/2}^2 - \frac{\lambda^3 \Delta t}{\Delta x} \mathcal{W}_{i-1/2}^3 - \frac{\lambda^1 \Delta t}{\Delta x} \mathcal{W}_{i+1/2}^1 \\ &= Q_i^n - \frac{\Delta t}{\Delta x}\left(\lambda^2 \mathcal{W}_{i-1/2}^2 + \lambda^3 \mathcal{W}_{i-1/2}^3 + \lambda^1 \mathcal{W}_{i+1/2}^1\right). \end{aligned} \tag{4.39}$$

Note that we use the 2- and the 3-wave originating from $x_{i-1/2}$ and the 1-wave originating from $x_{i+1/2}$, based on the presumed wave speeds. This can be written in a form that generalizes easily to arbitrary hyperbolic systems of m equations. Let

$$\lambda^+ = \max(\lambda, 0), \qquad \lambda^- = \min(\lambda, 0), \tag{4.40}$$

and suppose the solution of the Riemann problem consists of m waves $\mathcal{W}^p$ traveling at speeds λ^p, each of which may be positive or negative. Then the cell average is updated by

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left[\sum_{p=1}^m (\lambda^p)^+ \mathcal{W}_{i-1/2}^p + \sum_{p=1}^m (\lambda^p)^- \mathcal{W}_{i+1/2}^p\right]. \tag{4.41}$$

The cell average is affected by all right-going waves from $x_{i-1/2}$ and by all left-going waves from $x_{i+1/2}$. This is a generalization of (4.35). Understanding this formulation of Godunov's method is crucial to understanding many of the other algorithms presented in this book.

As a shorthand notation, we will also introduce the following symbols:

$$\begin{aligned} \mathcal{A}^- \Delta Q_{i-1/2} &= \sum_{p=1}^m (\lambda^p)^- \mathcal{W}_{i-1/2}^p, \\ \mathcal{A}^+ \Delta Q_{i-1/2} &= \sum_{p=1}^m (\lambda^p)^+ \mathcal{W}_{i-1/2}^p, \end{aligned} \tag{4.42}$$

so that (4.41) can be rewritten as

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{A}^+ \Delta Q_{i-1/2} + \mathcal{A}^- \Delta Q_{i+1/2}\right). \tag{4.43}$$

The symbol $\mathcal{A}^+ \Delta Q_{i-1/2}$ should be interpreted as a single entity that measures the net effect of all right-going waves from $x_{i-1/2}$, while $\mathcal{A}^- \Delta Q_{i-1/2}$ measures the net effect of all left-going waves from this same interface. These net effects will also sometimes be called *fluctuations*. Note that within cell $\mathcal{C}_i$, it is the right-going fluctuation from the left edge, $\mathcal{A}^+ \Delta Q_{i-1/2}$, and the left-going fluctuation from the right edge, $\mathcal{A}^- \Delta Q_{i+1/2}$, that affect the cell average.

The notation introduced in (4.42) is motivated by the following observation. For the constant-coefficient linear system $q_t + Aq_x = 0$, we have

$$\mathcal{W}_{i-1/2}^p = \alpha_{i-1/2}^p r^p,$$

where r^p is the pth eigenvector of A, and the propagation speed is the corresponding

eigenvalue λ^p. Define the matrices

$$\Lambda^+ = \begin{bmatrix} (\lambda^1)^+ & & & \\ & (\lambda^2)^+ & & \\ & & \ddots & \\ & & & (\lambda^m)^+ \end{bmatrix}, \qquad \Lambda^- = \begin{bmatrix} (\lambda^1)^- & & & \\ & (\lambda^2)^- & & \\ & & \ddots & \\ & & & (\lambda^m)^- \end{bmatrix}. \tag{4.44}$$

Thus Λ^+ has only the positive eigenvalues on the diagonal, with negative ones replaced by zero, and conversely for Λ^-. Now define

$$A^+ = R\Lambda^+ R^{-1} \quad \text{and} \quad A^- = R\Lambda^- R^{-1}, \tag{4.45}$$

and note that

$$A^+ + A^- = R(\Lambda^+ + \Lambda^-)R^{-1} = R\Lambda R^{-1} = A. \tag{4.46}$$

This gives a useful splitting of the coefficient matrix A into pieces essential for right-going and left-going propagation. Now if we let $\Delta Q_{i-1/2} = Q_i - Q_{i-1}$ and multiply this vector by A^+, we obtain

$$\begin{aligned} A^+\Delta Q_{i-1/2} &= R\Lambda^+ R^{-1}(Q_i - Q_{i-1}) \\ &= R\Lambda^+ \alpha_{i-1/2} \\ &= \sum_{p=1}^{m} (\lambda^p)^+ \alpha^p_{i-1/2} r^p \\ &= \mathcal{A}^+ \Delta Q_{i-1/2}. \end{aligned} \tag{4.47}$$

Similarly, we compute that

$$\begin{aligned} A^-\Delta Q_{i-1/2} &= \sum_{p=1}^{m} (\lambda^p)^- \alpha^p_{i-1/2} r^p \\ &= \mathcal{A}^- \Delta Q_{i-1/2}. \end{aligned} \tag{4.48}$$

So in the linear constant-coefficient case, each of the fluctuations $\mathcal{A}^+\Delta Q_{i-1/2}$ and $\mathcal{A}^-\Delta Q_{i-1/2}$ can be computed by simply multiplying the matrix A^+ or A^- by the jump in Q. For variable-coefficient or nonlinear problems the situation is not quite so simple, and hence we introduce the more general notation (4.42) for the fluctuations, which can still be computed by solving Riemann problems and combining the appropriate waves. We will see that the form (4.43) of Godunov's method can still be used.

For the constant-coefficient linear problem, the wave-propagation form (4.43) of Godunov's method can be related directly to the numerical flux function (4.37). Note that the value of q in the Riemann solution along $x = x_{i-1/2}$ is

$$Q^{\Downarrow}_{i-1/2} = q^{\Downarrow}(Q_{i-1}, Q_i) = Q_{i-1} + \sum_{p:\lambda^p<0} \mathcal{W}^p_{i-1/2},$$

using the summation notation introduced in (3.19).

In the linear case $f(Q^{\Downarrow}_{i-1/2}) = AQ^{\Downarrow}_{i-1/2}$ and so (4.37) gives

$$F^n_{i-1/2} = AQ_{i-1} + \sum_{p:\lambda^p<0} A\mathcal{W}^p_{i-1/2}.$$

Since $\mathcal{W}^p_{i-1/2}$ is an eigenvector of A with eigenvalue λ^p, this can be rewritten as

$$F^n_{i-1/2} = AQ_{i-1} + \sum_{p=1}^{m} (\lambda^p)^- \mathcal{W}^p_{i-1/2}. \tag{4.49}$$

Alternatively, we could start with the formula

$$Q^{\Downarrow}_{i-1/2} = Q_i - \sum_{p:\lambda^p>0} \mathcal{W}^p_{i-1/2}$$

and obtain

$$F^n_{i-1/2} = AQ_i - \sum_{p=1}^{m} (\lambda^p)^+ \mathcal{W}^p_{i-1/2}. \tag{4.50}$$

Similarly, there are two ways to express $F^n_{i+1/2}$. Choosing the form

$$F^n_{i+1/2} = AQ_i + \sum_{p=1}^{m} (\lambda^p)^- \mathcal{W}^p_{i+1/2}$$

and combining this with (4.50) in the flux-differencing formula (4.4) gives

$$\begin{aligned} Q^{n+1}_i &= Q^n_i - \frac{\Delta t}{\Delta x}\left(F^n_{i+1/2} - F^n_{i-1/2}\right) \\ &= Q^n_i - \frac{\Delta t}{\Delta x}\left[\sum_{p=1}^{m} (\lambda^p)^- \mathcal{W}^p_{i+1/2} + \sum_{p=1}^{m} (\lambda^p)^+ \mathcal{W}^p_{i-1/2}\right], \end{aligned} \tag{4.51}$$

since the AQ_i terms cancel out. This is exactly the same expression obtained in (4.41).

For a more general conservation law $q_t + f(q)_x = 0$, we can define

$$F^n_{i-1/2} = f(Q_{i-1}) + \sum_{p=1}^{m} (\lambda^p)^- \mathcal{W}^p_{i-1/2} \equiv f(Q_{i-1}) + \mathcal{A}^- \Delta Q_{i-1/2} \tag{4.52}$$

or

$$F^n_{i-1/2} = f(Q_i) - \sum_{p=1}^{m} (\lambda^p)^+ \mathcal{W}^p_{i-1/2} \equiv f(Q_i) - \mathcal{A}^+ \Delta Q_{i-1/2}, \tag{4.53}$$

corresponding to (4.49) and (4.50) respectively, where the speeds λ^p and waves $\mathcal{W}^p$ come out of the solution to the Riemann problem.

4.13 Flux-Difference vs. Flux-Vector Splitting

Note that if we subtract (4.52) from (4.53) and rearrange, we obtain

$$f(Q_i) - f(Q_{i-1}) = \mathcal{A}^- \Delta Q_{i-1/2} + \mathcal{A}^+ \Delta Q_{i-1/2}. \tag{4.54}$$

This indicates that the terms on the right-hand side correspond to a so-called *flux-difference splitting*. The difference between the fluxes computed based on each of the cell averages Q_{i-1} and Q_i is split into a left-going fluctuation that updates Q_{i-1} and a right-going fluctuation that updates Q_i.

We can define a more general class of flux-difference splitting methods containing any method based on some splitting of the flux difference as in (4.54), followed by application of the formula (4.43). Such a method is guaranteed to be conservative, and corresponds to a flux-differencing method with numerical fluxes

$$F^n_{i-1/2} = f(Q_i) - \mathcal{A}^+ \Delta Q_{i-1/2} = f(Q_{i-1}) + \mathcal{A}^- \Delta Q_{i-1/2}. \tag{4.55}$$

For a linear system, there are other ways to rewrite the numerical flux $F^n_{i-1/2}$ that give additional insight. Using (4.47) in (4.49), we obtain

$$\begin{aligned} F^n_{i-1/2} &= (A^+ + A^-)Q_{i-1} + A^-(Q_i - Q_{i-1}) \\ &= A^+ Q_{i-1} + A^- Q_i. \end{aligned} \tag{4.56}$$

Since $A^+ + A^- = A$, the formula (4.56) gives a flux that is *consistent* with the correct flux in the sense of (4.14): If $Q_{i-1} = Q_i = \bar{q}$, then (4.56) reduces to $F_{i-1/2} = A\bar{q} = f(q)$. This has a very natural interpretation as a *flux-vector splitting*. The flux function is $f(q) = Aq$, and so AQ_{i-1} and AQ_i give two possible approximations to the flux at $x_{i-1/2}$. In Section 4.5 we considered the possibility of simply averaging these to obtain $F^n_{i-1/2}$ and rejected this because it gives an unstable method. The formula (4.56) suggests instead a more sophisticated average in which we take the part of AQ_{i-1} corresponding to right-going waves and combine it with the part of AQ_i corresponding to left-going waves in order to obtain the flux in between. This is the proper generalization to systems of equations of the upwind flux for the scalar advection equation given in (4.33).

This is philosophically a different approach from the flux-difference splitting discussed in relation to (4.54). What we have observed is that for a constant-coefficient linear system, the two viewpoints lead to exactly the same method. This is not typically the case for nonlinear problems, and in this book we concentrate primarily on methods that correspond to flux-difference splittings. However, note that given *any* flux-vector splitting, one can define a corresponding splitting of the flux difference in the form (4.54). If we have split

$$f(Q_{i-1}) = f^{(-)}_{i-1} + f^{(+)}_{i-1} \quad \text{and} \quad f(Q_i) = f^{(-)}_i + f^{(+)}_i,$$

and wish to define the numerical flux as

$$F^n_{i-1/2} = f^{(+)}_{i-1} + f^{(-)}_i, \tag{4.57}$$

then we can define the corresponding fluctuations as

$$\begin{aligned} \mathcal{A}^- \Delta Q_{i-1/2} &= f_i^{(-)} - f_{i-1}^{(-)}, \\ \mathcal{A}^+ \Delta Q_{i-1/2} &= f_i^{(+)} - f_{i-1}^{(+)}, \end{aligned} \tag{4.58}$$

to obtain a flux-difference splitting that satisfies (4.54) and again yields $F^n_{i-1/2}$ via the formula (4.55). Flux-vector splittings for nonlinear problems are discussed in Section 15.7.

4.14 Roe's Method

For a constant-coefficient linear problem there is yet another way to rewrite the flux $F^n_{i-1/2}$ appearing in (4.49), (4.50), and (4.56), which relates it directly to the unstable naive averaging of AQ_{i-1} and AQ_i given in (4.18). Averaging the expressions (4.49) and (4.50) gives

$$F^n_{i-1/2} = \frac{1}{2}\left[(AQ_{i-1} + AQ_i) - \sum_{p=1}^{m} [(\lambda^p)^+ - (\lambda^p)^-] \mathcal{W}^p_{i-1/2} \right]. \tag{4.59}$$

Notice that $\lambda^+ - \lambda^- = |\lambda|$. Define the matrix $|A|$ by

$$|A| = R|\Lambda|R^{-1}, \qquad \text{where } |\Lambda| = \text{diag}(|\lambda^p|). \tag{4.60}$$

Then (4.59) becomes

$$\begin{aligned} F^n_{i-1/2} &= \frac{1}{2}(AQ_{i-1} + AQ_i) - \frac{1}{2}|A|(Q_i - Q_{i-1}) \\ &= \frac{1}{2}[f(Q_{i-1}) + f(Q_i)] - \frac{1}{2}|A|(Q_i - Q_{i-1}). \end{aligned} \tag{4.61}$$

This can be viewed as the arithmetic average plus a correction term that stabilizes the method.

For the constant-coefficient linear problem this is simply another way to rewrite the Godunov or upwind flux, but this form is often seen in extensions to nonlinear problems based on approximate Riemann solvers, as discussed in Section 15.3. This form of the flux is often called *Roe's method* in this connection. This formulation is also useful in studying the numerical dissipation of the upwind method.

Using the flux (4.61) in the flux-differencing formula (4.4) gives the following updating formula for Roe's method on a linear system:

$$\begin{aligned} Q_i^{n+1} = Q_i^n &- \frac{1}{2}\frac{\Delta t}{\Delta x} A\left(Q^n_{i+1} - Q^n_{i-1}\right) \\ &- \frac{1}{2}\frac{\Delta t}{\Delta x} \sum_{p=1}^{m} \left(|\lambda^p| \mathcal{W}^p_{i+1/2} - |\lambda^p| \mathcal{W}^p_{i-1/2}\right). \end{aligned} \tag{4.62}$$

This can also be derived directly from (4.41) by noting that another way to express (4.40) is

$$\lambda^+ = \frac{1}{2}(\lambda + |\lambda|), \qquad \lambda^- = \frac{1}{2}(\lambda - |\lambda|). \tag{4.63}$$

We will see in Section 12.3 that for nonlinear problems it is sometimes useful to modify these definitions of $\lambda^{\pm}$.

Exercises

4.1. (a) Determine the matrices A^+ and A^- as defined in (4.45) for the acoustics equations (2.50).

(b) Determine the waves $\mathcal{W}^1_{i-1/2}$ and $\mathcal{W}^2_{i-1/2}$ that result from arbitrary data Q_{i-1} and Q_i for this system.

4.2. If we apply the upwind method (4.25) to the advection equation $q_t + \bar{u}q_x = 0$ with $\bar{u} > 0$, and choose the time step so that $\bar{u}\,\Delta t = \Delta x$, then the method reduces to

$$Q_i^{n+1} = Q_{i-1}^n.$$

The initial data simply shifts one grid cell each time step and the exact solution is obtained, up to the accuracy of the initial data. (If the data Q_i^0 is the exact cell average of $\mathring{q}(x)$, then the numerical solution will be the exact cell average for every step.) This is a nice property for a numerical method to have and is sometimes called the *unit CFL condition*.

(a) Sketch figures analogous to Figure 4.5(a) for this case to illustrate the wave-propagation interpretation of this result.

(b) Does the Lax–Friedrichs method (4.20) satisfy the unit CFL condition? Does the two-step Lax–Wendroff method of Section 4.7?

(c) Show that the exact solution (in the same sense as above) is also obtained for the constant-coefficient acoustics equations (2.50) with $u_0 = 0$ if we choose the time step so that $c\,\Delta t = \Delta x$ and apply Godunov's method. Determine the formulas for p_i^{n+1} and u_i^{n+1} that result in this case, and show how they are related to the solution obtained from characteristic theory.

(d) Is it possible to obtain a similar exact result by a suitable choice of Δt in the case where $u_0 \neq 0$ in acoustics?

4.3. Consider the following method for the advection equation with $\bar{u} > 0$:

$$\begin{aligned} Q_i^{n+1} &= Q_i^n - \left(Q_i^n - Q_{i-1}^n\right) - \left(\frac{\bar{u}\,\Delta t - \Delta x}{\Delta x}\right)\left(Q_{i-1}^n - Q_{i-2}^n\right) \\ &= Q_{i-1}^n - \left(\frac{\bar{u}\,\Delta t}{\Delta x} - 1\right)\left(Q_{i-1}^n - Q_{i-2}^n\right). \end{aligned} \tag{4.64}$$

(a) Show that this method results from a wave-propagation algorithm of the sort illustrated in Figure 4.5(a) in the case where $\Delta x \le \bar{u}\,\Delta t \le 2\,\Delta x$, so that each wave propagates all the way through the adjacent cell and part way through the next.

(b) Give an interpretation of this method based on linear interpolation, similar to what is illustrated in Figure 4.4(a).
(c) Show that this method is exact if $\bar{u}\,\Delta t/\Delta x = 1$ or $\bar{u}\,\Delta t/\Delta x = 2$.
(d) For what range of Courant numbers is the CFL condition satisfied for this method? (See also Exercise 8.6.)
(e) Determine a method of this same type that works if each wave propagates through more than two cells but less than three, i.e., if $2\,\Delta x \le \bar{u}\,\Delta t \le 3\,\Delta x$.

Large-time-step methods of this type can also be applied, with limited success, to nonlinear problems, e.g., [42], [181], [274], [275], [279], [316].