9

Variable-Coefficient Linear Equations

In the preceding chapters we have primarily studied linear problems with constant coefficients. Many interesting problems involve spatially-varying coefficients. This chapter is devoted to exploring some of the issues that arise, both analytically and numerically, in this case.

There are several distinct forms that a variable-coefficient hyperbolic system might take, each of which arises naturally in some contexts. One possibility is that the coefficient matrix A multiplying q_x varies with x, so the system is

$$q_t + A(x)q_x = 0. \tag{9.1}$$

This system is hyperbolic in some domain if $A(x)$ is diagonalizable with real eigenvalues at each x in the domain. Such problems still model wave propagation, and the finite volume methods developed in previous chapters can be applied fairly directly, in spite of the fact that this system (9.1) is not in conservation form and there is no flux function.

Another form of variable-coefficient problem that often arises is

$$q_t + (A(x)q)_x = 0, \tag{9.2}$$

in which the matrix $A(x)$ appears within the x-derivative. For the constant-coefficient problem with $A(x) \equiv A$, the two forms (9.1) and (9.2) are identical, but with variable coefficients they are distinct and have different solutions. The equation (9.2) is a conservation law, with the flux function

$$f(q, x) = A(x)q. \tag{9.3}$$

In this case the flux function depends explicitly on the location x as well as on the value of the conserved quantities q. Again this equation is hyperbolic whenever $A(x)$ is diagonalizable with real eigenvalues.

For a given physical problem it may be possible to derive an equation of either form (9.1) or (9.2), depending on how the vector q is chosen. For example, in Section 9.1 we will see how flow through a pipe can lead to an advection equation of either form, depending on how q is defined.

We can go back and forth between the forms at the expense of introducing source terms. By applying the product rule to the x-derivative in (9.2), we could rewrite this equation in

the form

$$q_t + A(x)q_x = -A'(x)q,$$

which has the form of (9.1) with the addition of a source term. Conversely, we could rewrite (9.1) as

$$q_t + (A(x)q)_x = A'(x)q,$$

which now has the form of a conservation law with a source term. Normally we wish to avoid adding source terms to the equation if they can be avoided by a better choice of variables.

Another form of variable coefficients that often arises naturally is a *capacity function* $\kappa(x)$, as described in the context of heat capacity in Section 2.3, giving systems of the form

$$\kappa(x)q_t + (A(x)q)_x = 0, \tag{9.4}$$

for example. Again this could be manipulated into other forms, e.g.,

$$q_t + \kappa^{-1}(x)A(x)q_x = -\kappa^{-1}(x)A'(x)q,$$

but for problems where κq is the proper conserved quantity it may be preferable to work directly with the form (9.4), using the capacity-form differencing algorithms introduced in Section 6.16.

9.1 Advection in a Pipe

Consider an incompressible fluid flowing through a pipe with variable cross-sectional area $\kappa(x)$, and suppose we want to model the advection of some tracer down the length of the pipe. We denote the area by κ, since we will see that this is a natural capacity function. There are several ways we might model this, depending on how we choose our variables, leading to different forms of the variable-coefficient advection equation.

We are assuming here that a one-dimensional formulation is adequate, i.e., that all quantities vary only with x (distance along the pipe) and t, and are essentially constant across any given cross section. In reality the flow through a pipe with varying diameter cannot be truly one-dimensional, since there must be a velocity component in the radial direction in regions where the walls converge or diverge. But we will assume this is sufficiently small that it can be safely ignored, which is true if the variation in κ is sufficiently smooth. The fluid velocity is then given by a single axial velocity $u(x)$ that varies only with x. (See Sections 9.4.1 and 9.4.2 for other contexts where this makes sense even if u is discontinuous.)

Note that if $\kappa(x)$ is measured in square meters and $u(x)$ in meters per second, say, then the product $\kappa(x)u(x)$ has units of cubic meters per second and measures the volume of fluid passing any point per unit time. Since the fluid is assumed to be incompressible, this must be the same at every point in the pipe, and we will denote the flow rate by U,

$$U = \kappa(x)u(x) \equiv \text{constant}. \tag{9.5}$$

If we know U and the cross-sectional area $\kappa(x)$, then we can compute

$$u(x) = U/\kappa(x). \tag{9.6}$$

(In Section 9.4 we consider a formulation in which $\kappa(x)u(x)$ need not be constant.)

Now suppose that we introduce a tracer into the fluid (e.g., some food coloring or chemical contaminant into water, with very small concentration so that it does not affect the fluid dynamics). We wish to study how a given initial concentration profile advects downstream. There are two distinct ways we might choose to measure the concentration, leading to different forms of the advection equation.

One approach would be to measure the concentration in grams per cubic meter (mass per unit volume of fluid), which is what we would probably actually measure if we took a small sample of fluid from the pipe and determined the concentration. Call this variable $\bar{q}(x, t)$.

However, since we are solving a one-dimensional problem with quantities assumed to vary in only the x-direction, another natural possibility would be to measure the concentration in units of mass per unit length of the pipe (grams per meter). If we call this variable $q(x, t)$, then

$$\int_{x_1}^{x_2} q(x, t)\, dx \tag{9.7}$$

measures the total mass[1] of tracer in the section of pipe from x_1 to x_2. Since this mass can only change due to tracer moving past the endpoints, q is the proper conserved quantity for this problem. The one-dimensional velocity $u(x)$ measures the rate at which tracer moves past a point and the product $u(x)q(x, t)$ is the flux in grams per second. Hence with this choice of variables we obtain the conservative advection equation

$$q_t + (u(x)q)_x = 0. \tag{9.8}$$

There is a simple relation between q and $\bar{q}$, given by

$$q(x, t) = \kappa(x)\bar{q}(x, t), \tag{9.9}$$

since multiplying the mass per unit volume by the cross-sectional area gives the mass per unit length. Hence the total mass of tracer in $[x_1, x_2]$, given by (9.7), can be rewritten as

$$\int_{x_1}^{x_2} \kappa(x)\bar{q}(x, t)\, dx.$$

With this choice of variables we see that the cross-sectional area acts as a capacity function, which is quite natural in that this area clearly determines the fluid capacity of the pipe at each point.

Using the relation (9.9) in (9.8) gives an advection equation for $\bar{q}$,

$$\kappa(x)\bar{q}_t + (u(x)\kappa(x)\bar{q})_x = 0. \tag{9.10}$$

[1] In problems where chemical kinetics is involved, we should measure the "mass" in moles rather than grams, and the density in moles per meter or moles per cubic meter.

But now notice that applying (9.5) allows us to rewrite this as

$$\kappa(x)\bar{q}_t + U\bar{q}_x = 0. \tag{9.11}$$

Dividing by κ and using (9.6) gives

$$\bar{q}_t + u(x)\bar{q}_x = 0 \tag{9.12}$$

as another form of the advection equation.

Comparing the advection equations (9.8) and (9.12) shows that they have the form of (9.2) and (9.1), respectively. The velocity $u(x)$ is the same in either case, but depending on how we measure the concentration, we obtain either the nonconservative or the conservative form of the equation. The form (9.11) is a third distinct form, in which the capacity function and flow rate appear instead of the velocity.

The nonconservative form (9.12) of the advection equation is often called the *transport equation* or the *color equation*. If we think of food coloring in water, then it is $\bar{q}$, the mass per unit volume, that determines the color of the water. If we follow a parcel of water through the pipe, we expect the color to remain the same even as the area of the pipe and velocity of the water change. This is easily verified, since $\bar{q}(x, t)$ is constant along characteristic curves (recall the discussion of Section 2.1.1). By contrast the conserved quantity q is not constant along characteristic curves. The mass per unit length varies with the cross-sectional area even if the color is constant. From (9.8) we obtain $q_t + u(x)q_x = -u'(x)q$, so that along any characteristic curve $X(t)$ satisfying $X'(t) = u(X(t))$ we have

$$\frac{d}{dt}q(X(t), t) = -u'(X(t))\, q(X(t), t). \tag{9.13}$$

9.2 Finite Volume Methods

Natural upwind finite difference methods for the equations (9.8) and (9.12) are easy to derive and take the following form in the case $U > 0$:

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left[u(x_i)Q_i^n - u(x_{i-1})Q_{i-1}^n\right] \tag{9.14}$$

and

$$\bar{Q}_i^{n+1} = \bar{Q}_i^n - \frac{\Delta t}{\Delta x}u(x_i)\left(\bar{Q}_i^n - \bar{Q}_{i-1}^n\right) \tag{9.15}$$

respectively. Variations of each formula are possible in which we use $u(x_{i-1/2})$ in place of $u(x_i)$. Either choice gives a first-order accurate approximation. In this section we will see how to derive these methods using the finite volume framework of Godunov's method, in a manner that allows extension to high-resolution methods.

The Riemann problem at the cell interface $x_{i-1/2}$ now takes the form of an advection equation with piecewise constant coefficients as well as piecewise constant initial data. The solution depends on what form of the advection equation we are solving. It also depends on how we discretize the velocity $u(x)$. Two natural possibilities are:

1. Associate a velocity u_i with each grid cell. We might define u_i as the pointwise value $u_i = u(x_i)$, or we could compute the cell average of $u(x)$, or we could view the pipe as having a piecewise constant cross-sectional area with value κ_i in the ith cell, and then take $u_i = U/\kappa_i$.
2. Associate a velocity $u_{i-1/2}$ with each cell interface $x_{i-1/2}$.

There are certain advantages and disadvantages of each form, and which is chosen may depend on the nature of the problem being solved. To begin, we will concentrate on the first choice. For some problems, however, it is more natural to use edge velocities $u_{i-1/2}$, and this choice is discussed in Section 9.5. This is true in particular when we extend the algorithms to more than one space dimension; see Section 18.2.

9.3 The Color Equation

In the case of the color equation (9.12), the Riemann problem models a jump in color from $\bar{Q}_{i-1}$ to $\bar{Q}_i$ at $x_{i-1/2}$. Since color is constant along particle paths, this jump in color will simply propagate into cell $\mathcal{C}_i$ at the velocity u_i in this cell, and so in the notation of Chapter 4 we have a wave $\mathcal{W}_{i-1/2}$ (omitting the superscript 1, since there is only one wave) with jump

$$\mathcal{W}_{i-1/2} = \bar{Q}_i - \bar{Q}_{i-1},$$

and speed

$$s_{i-1/2} = u_i.$$

Again we are assuming positive velocities u_i in each cell $\mathcal{C}_i$. If $U < 0$, then we would have $s_{i-1/2} = u_{i-1}$, since the wave would enter cell $i-1$. In spite of the fact that the color equation is not in conservation form, we can still view $\bar{Q}_i^n$ as an approximation to the cell average of $\bar{q}(x,t)$ at time t_n. In this case $\bar{Q}_i^n$ is simply the "average color" over this cell, and $\Delta x\, \bar{Q}_i^n$ is not the total mass of any conserved quantity. It is no longer true that the change in this cell average is given by a flux difference, but it *is* true that the change can be computed from the wave and speed. Note that (for $U > 0$) only the wave $\mathcal{W}_{i-1/2}$ enters cell $\mathcal{C}_i$, and the value of $\bar{q}$ in this cell is changed by $\bar{Q}_i - \bar{Q}_{i-1}$ at each point the wave has reached. The wave propagates a distance $s_{i-1/2}\,\Delta t$ over the time step and hence has covered the fraction $s_{i-1/2}\,\Delta t/\Delta x$ of the cell. The cell average is thus modified by

$$\bar{Q}_i^{n+1} = \bar{Q}_i^n - \frac{\Delta t}{\Delta x} s_{i-1/2} \mathcal{W}_{i-1/2}.$$

This gives the upwind method (9.15). Also note that this has exactly the form (4.43),

$$\bar{Q}_i^{n+1} = \bar{Q}_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{A}^+ \Delta \bar{Q}_{i-1/2} + \mathcal{A}^- \Delta \bar{Q}_{i+1/2}\right),$$

if we define

$$\begin{aligned} \mathcal{A}^+ \Delta \bar{Q}_{i-1/2} &= s_{i-1/2} \mathcal{W}_{i-1/2}, \\ \mathcal{A}^- \Delta \bar{Q}_{i-1/2} &= 0. \end{aligned} \tag{9.16}$$

These formulas have been presented for the case $U > 0$, so that all velocities u_i are positive. The more general formulas that can also be used in the case $U < 0$ are

$$\begin{aligned} \mathcal{W}_{i-1/2} &= \bar{Q}_i^n - \bar{Q}_{i-1}^n, \\ s_{i-1/2} &= \begin{cases} u_i & \text{if } U > 0, \\ u_{i-1} & \text{if } U < 0, \end{cases} \\ &= u_i^+ + u_{i-1}^-, \\ \mathcal{A}^+ \Delta \bar{Q}_{i-1/2} &= s_{i-1/2}^+ \mathcal{W}_{i-1/2}, \\ \mathcal{A}^- \Delta \bar{Q}_{i-1/2} &= s_{i-1/2}^- \mathcal{W}_{i-1/2}, \end{aligned} \tag{9.17}$$

using the notation (4.40).

9.3.1 High-Resolution Corrections

The first-order upwind method for the color equation can be greatly improved by including high-resolution corrections as developed in Chapter 6. The improved method again has the form (5.7),

$$\bar{Q}_i^{n+1} = \bar{Q}_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{A}^+ \Delta Q_{i-1/2} + \mathcal{A}^- \Delta Q_{i+1/2}\right) - \frac{\Delta t}{\Delta x}\left(\tilde{F}_{i+1/2} - \tilde{F}_{i-1/2}\right), \tag{9.18}$$

where

$$\tilde{F}_{i-1/2} = \frac{1}{2}|s_{i-1/2}|\left(1 - \frac{\Delta t}{\Delta x}|s_{i-1/2}|\right)\widetilde{\mathcal{W}}_{i-1/2}. \tag{9.19}$$

The modified wave $\widetilde{\mathcal{W}}_{i-1/2}$ is obtained by applying a wave limiter as described in Chapter 6.

It is important to note, however, that this method is *not* formally second-order accurate when applied to a smooth solution, even when no limiter function is used. The local truncation error is found to be (see Exercise 9.1)

$$\text{local truncation error} = -\frac{1}{2}[\Delta x - u(x)\,\Delta t]u'(x)q_x(x,t) + \mathcal{O}(\Delta x^2). \tag{9.20}$$

The method is formally second-order accurate only in the case of constant coefficients, where $u'(x) \equiv 0$. However, this truncation error has a quite different form than the truncation error for the first-order upwind method (9.15), which is

$$\Delta x\, u(x) q_{xx}(x,t) + \mathcal{O}(\Delta x^2). \tag{9.21}$$

This truncation error depends on q_{xx} and leads to substantial numerical dissipation, as with any first-order upwind method. The truncation error (9.20), on the other hand, is a multiple of q_x instead. This can be viewed as corresponding to a small error in the propagation speed.

The modified equation (see Section 8.6) for the high-resolution method is

$$q_t + u(x)q_x = -\frac{1}{2}[\Delta x - u(x)\,\Delta t]u'(x)q_x,$$

which we can rewrite (ignoring higher-order terms) as

$$q_t + u\left(x + \frac{1}{2}[\Delta x - u(x)\,\Delta t]\right)q_x = 0 \tag{9.22}$$

This is again the color equation, but with the velocity evaluated at a point that is shifted by less than a grid cell. The solution to this equation looks essentially the same as for the correct color equation, but slightly offset. In particular there is no diffusion or smearing of the solution (at least not at the $\mathcal{O}(\Delta x)$ level), and hence we expect the solution obtained with this method to look much better than what would be obtained from the first-order upwind method.

A true second-order accurate method can be derived using the Lax–Wendroff approach (Exercise 9.2), but results are typically similar to what is produced by the method above when the solution is smooth. The high-resolution method with limiters included works very well in practice, and much better than the formally second-order method in cases where the solution is discontinuous.

9.3.2 Discontinuous Velocities

For flow in a pipe, reducing the problem to a one-dimensional advection equation is only valid if the diameter of the pipe, and hence the velocity, is very slowly varying. Otherwise the true fluid velocity may have a substantial radial component. This is particularly true at a point where the diameter is discontinuous. Here the motion will be fully multidimensional and perhaps even turbulent. However, one-dimensional approximations are frequently used even in this case in many engineering applications involving transport through complicated networks of pipes, for example.

There are also other applications where a one-dimensional advection equation with rapidly varying or even discontinuous velocities is physically relevant. One example, discussed in Section 9.4.2, arises in traffic flow along a one-lane road with a discontinuous speed limit. Another example is discussed in Section 9.4.1.

An advantage of the high-resolution finite volume methods is that they typically perform very well even when the coefficients in the equation are discontinuous. You may wish to experiment with the examples in `[claw/book/chap9/color/discontu]`.

9.4 The Conservative Advection Equation

Now suppose we are solving the conservative equation (9.8), $q_t + (u(x)q)_x = 0$, with velocities u_i given in each grid cell. Again we can use the form (5.7) after defining the fluctuations $\mathcal{A}^+\Delta Q_{i-1/2}$ and $\mathcal{A}^-\Delta Q_{i-1/2}$ in terms of waves via the solution to the Riemann problem at $x_{i-1/2}$. To do so properly we must carefully consider the Riemann problem, which now consists of using piecewise constant data defined by Q_{i-1} and Q_i and also piecewise constant coefficients in the advection equation,

$$\begin{aligned} q_t + u_{i-1}q_x &= 0 \qquad \text{if } x < x_{i-1/2}, \\ q_t + u_i q_x &= 0 \qquad \text{if } x > x_{i-1/2}. \end{aligned} \tag{9.23}$$

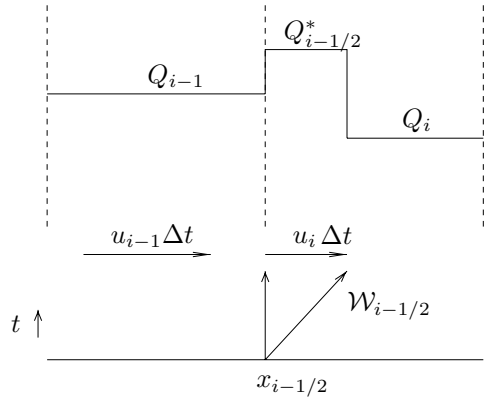

Fig. 9.1. Solution to the Riemann problem for the conservative advection equation when $u_{i-1} > u_i$.

The solution is *not* simply a single wave propagating at velocity u_i with strength $Q_i - Q_{i-1}$. The correct solution is indicated in Figure 9.1, and consists of *two* discontinuities in the solution: a stationary one at $x_{i-1/2}$ where Q_{i-1} jumps to some new value $Q^*_{i-1/2}$, and the wave

$$\mathcal{W}_{i-1/2} = Q_i - Q^*_{i-1/2} \tag{9.24}$$

propagating at the expected speed

$$s_{i-1/2} = u_i \tag{9.25}$$

(again we assume $U > 0$). The discontinuity at $x_{i-1/2}$ arises from the fact that the jump in u at $x_{i-1/2}$ corresponds to a jump in the cross-sectional area of the pipe. (A jump to a larger area $\kappa_i > \kappa_{i-1}$ is shown in Figure 9.1, in which case $u_i < u_{i-1}$.) Consequently, the density of tracer q, which is measured in the one-dimensional manner as mass per unit length, must increase, since there is suddenly more liquid and hence more tracer per unit length of pipe. The "color" $\bar{q}$ measured in mass per unit volume, on the other hand, does not experience any discontinuity at $x_{i-1/2}$, and so we did not have to worry about this effect in solving the color equation.

The new value $Q^*_{i-1/2}$ that arises in solving the Riemann problem can be found using the fact that the flux $f = uq$ must be continuous at $x_{i-1/2}$ in the Riemann solution. The same amount of tracer is leaving cell $\mathcal{C}_{i-1}$ as entering cell $\mathcal{C}_i$. This yields

$$u_{i-1} Q_{i-1} = u_i Q^*_{i-1/2}$$

and hence

$$Q^*_{i-1/2} = \frac{u_{i-1} Q_{i-1}}{u_i}. \tag{9.26}$$

This value is used in the wave (9.24). The fluctuation is computed in terms of this wave and

the speed (9.25) as

$$\begin{aligned}\mathcal{A}^+\Delta Q_{i-1/2} &= s_{i-1/2}\mathcal{W}_{i-1/2}\\ &= u_i\left(Q_i - \frac{u_{i-1}Q_{i-1}}{u_i}\right)\\ &= u_i Q_i - u_{i-1}Q_{i-1}.\end{aligned} \tag{9.27}$$

Again this is for the case $U > 0$, and in this case we have no fluctuation to the left:

$$\mathcal{A}^-\Delta Q_{i-1/2} = 0. \tag{9.28}$$

Note that the sum of the fluctuations has the form of a flux difference, as we expect for this conservative equation. The finite volume method

$$Q_i^{n+1} = Q_i - \frac{\Delta t}{\Delta x}\left(A^+\Delta Q_{i-1/2} + A^-\Delta Q_{i+1/2}\right) \tag{9.29}$$

reduces to the natural upwind method (9.14).

If $U < 0$, then we find that

$$\begin{aligned}Q^*_{i-1/2} &= \frac{u_i Q_i}{U_{i-1}},\\ \mathcal{W}_{i-1/2} &= Q^*_{i-1/2} - Q_{i-1},\\ s_{i-1/2} &= u_{i-1},\\ \mathcal{A}^-\Delta Q_{i-1/2} &= s_{i-1/2}\mathcal{W}_{i-1/2} = u_i Q_i - u_{i-1}Q_{i-1},\\ \mathcal{A}^+\Delta Q_{i-1/2} &= 0.\end{aligned} \tag{9.30}$$

In this case (9.29) again reduces to the natural upwind method. High-resolution correction terms can be added using (9.18) and (9.19).

Since the Riemann solution now involves two jumps, one might wonder why we don't need two waves $\mathcal{W}^1_{i-1/2}$ and $\mathcal{W}^2_{i-1/2}$ instead of just one, as introduced above. The answer is that the second wave moves at speed $s = 0$. We could include it in the Riemann solution, but it would drop out of the expression for $\mathcal{A}^+\Delta Q_{i-1/2}$, since we sum $s^p\mathcal{W}^p$. The corresponding high-resolution corrections also drop out when $s = 0$. So nothing would be gained by including this wave except additional work in applying the updating formulas.

9.4.1 Conveyer Belts

The solution illustrated in Figure 9.1 can perhaps be better understood by considering a different context than fluid flow in a pipe. Consider a series of conveyer belts, each going for the length of one grid cell with velocity u_i, and let $q(x, t)$ measure the density of packages traveling down this series of belts. See Figure 9.2. If the velocity u_i is less than u_{i-1}, then the density will increase at the junction $x_{i-1/2}$ just as illustrated in Figure 9.1. See also Figure 9.3 for another interpretation, discussed in the next subsection.

The structure illustrated in Figure 9.1 with two waves can also be derived by solving the Riemann problem for a system of two conservation laws, obtained by introducing u as a second conserved variable; see Exercise 13.11.

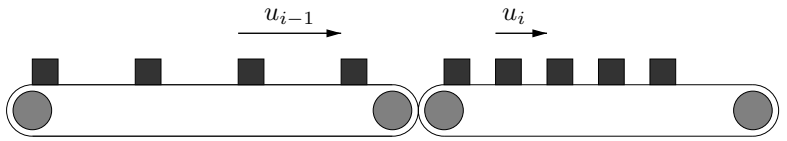

Fig. 9.2. Packages moving from a conveyer belt with speed u_{i-1} to a slower belt with speed u_i. The density of packages increases at the junction.

9.4.2 Traffic Flow

Similar advection equations arise in studying the flow of cars along a one-lane highway. To obtain a linear equation, we assume traffic is light enough that cars are always traveling at the speed limit $u(x)$, which can vary with x, the distance along the highway. If $u(x)$ is piecewise constant, then this is exactly the same problem as the conveyer-belt problem mentioned in Section 9.4.1, with cars replacing packages.

For heavier traffic the speed typically also depends on the density of cars and the problem becomes nonlinear. This more interesting case is examined in detail in Chapter 11.

Let $q(x, t)$ be the density of cars at point x at time t. Of course, at any particular point x there is either a car or no car, so the concept of density at a point appears to be meaningless. What we mean should be clear, however: over a reasonably long stretch of road from x_1 to x_2 the number of cars present at time t should be roughly

$$\text{number of cars between } x_1 \text{ and } x_2 \approx \int_{x_1}^{x_2} q(x, t)\, dx. \tag{9.31}$$

It will be convenient to measure the density in units of cars per car length, assuming all cars have the same length. Then empty highway corresponds to $q = 0$, bumper-to-bumper traffic corresponds to $q = 1$, and in general $0 \le q \le 1$. Then for (9.31) to make sense we must measure distance x in car lengths. Also we measure the speed limit $u(x)$ in car lengths per time unit. The flux of cars is given by $f(q) = u(x)q$ (in cars per time unit) and we obtain the conservation law (9.8).

One way to simulate traffic flow is to track the motion of individual vehicles, assuming a finite set of m cars on the highway. Let $X_k(t)$ be the location of the kth car at time t. Then the motion of each car can be obtained by solving the ordinary differential equation

$$X_k'(t) = u(X_k(t))$$

with initial conditions corresponding to its initial location $X_k(0)$. Note that cars move along characteristics of the advection equation. (This is not true more generally if the velocity depends on the density; see Chapter 11.)

Figure 9.3 illustrates the simulation of a line of cars with an initial density variation corresponding to a Riemann problem, traveling along a highway with a discontinuous speed limit

$$u(x) = \begin{cases} 2 & \text{if } x < 0, \\ 1 & \text{if } x > 0. \end{cases}$$

The paths of individual cars are plotted over time $0 \le t \le 20$. The *density* $q_k(t)$ observed by each driver is also plotted at both the initial and final times. For this discrete simulation

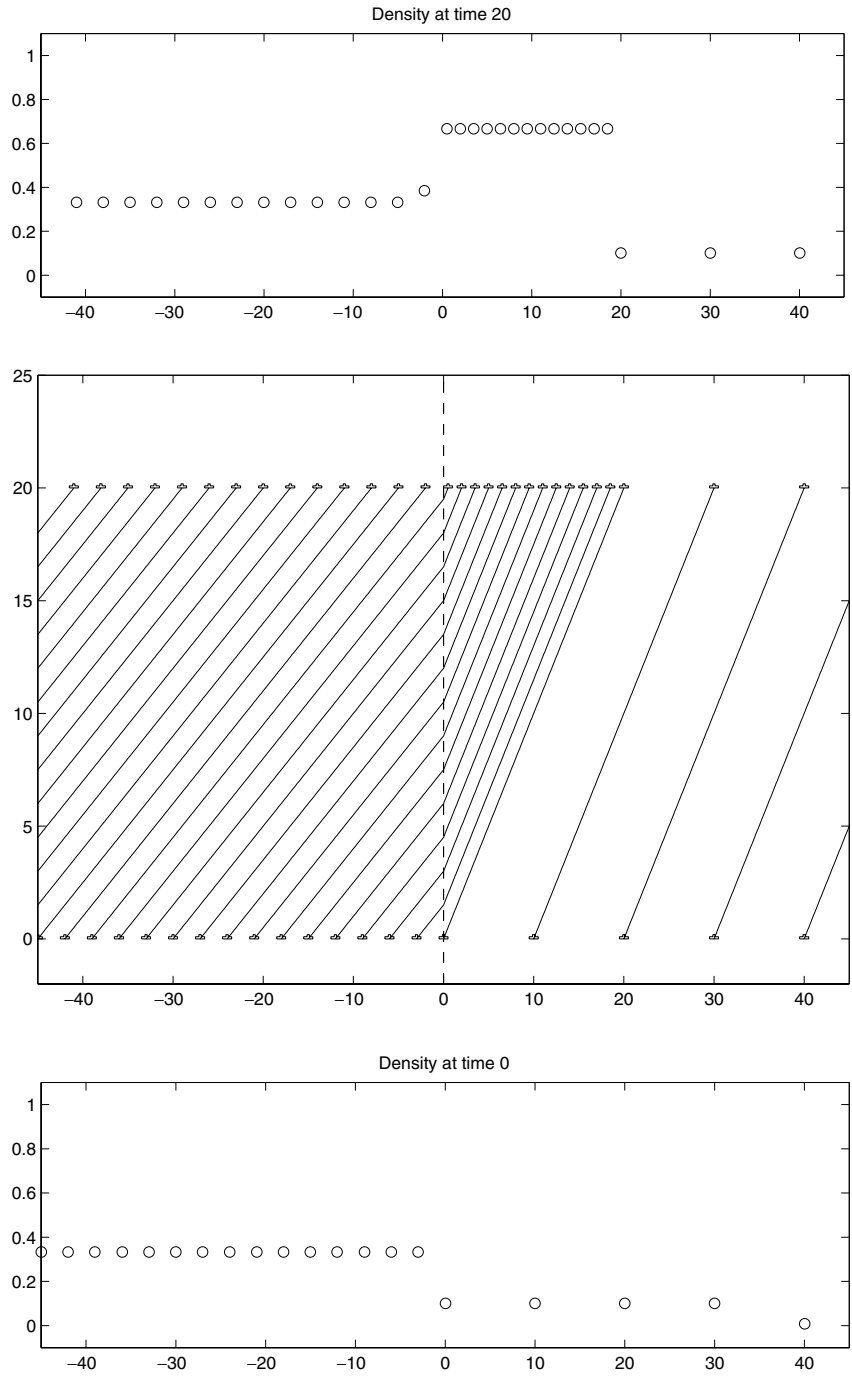

Fig. 9.3. Traffic flow model corresponding to the Riemann problem of Figure 9.1, with a change in speed limit at $x = 0$. Vehicle trajectories are shown, along with the density observed by each driver. We can also view the cars as representing packages moving from one conveyer belt to a slower one as in Figure 9.2. `[claw/book/chap9/traffic]`

a discrete density is defined for each car by

$$q_k(t) = \frac{1}{X_{k+1}(t) - X_k(t)}, \tag{9.32}$$

based on the distance to the car immediately ahead. This can be shown to be consistent with the previous notion of density; see Exercise 9.4.

The density at intermediate times can also be observed from the car paths in Figure 9.3. Taking a slice at any fixed time, the density is inversely proportional to the distance between paths.

Note that the density of cars changes discontinuously at the point $x = 0$ where the speed limit changes. In this simple model we assume cars and drivers can adjust speed instantaneously in response to the speed limit. This viewpoint may help to understand Figure 9.1. Of course if u_{i-1}/u_i is sufficiently large, then $Q^*_{i-1/2}$ in (9.26) may be greater than 1 even if Q_{i-1} is not. Although this causes no problem mathematically, with our definition of density this shouldn't happen physically. This is a case where cars are approaching the discontinuity in speed limit too rapidly to be accommodated in the slower region even at the bumper-to-bumper density of $q = 1$. In this case the linear model is definitely inadequate, and instead a nonlinear shock wave would arise in the real world. See Section 16.4.1 for a discussion of nonlinear traffic flow with a discontinuous speed limit.

For the linear traffic flow problem, the density naturally satisfies a conservative advection equation of the form (9.8). One can, however, define a new variable $\bar{q}(x, t)$ that instead satisfies the color equation (9.12), by setting

$$\bar{q}(x, t) = u(x)q(x, t).$$

This is simply the flux of cars. We have

$$\bar{q}_t = (uq)_t = uq_t = -u(uq)_x = -u\bar{q}_x,$$

and so (9.12) is satisfied. You should convince yourself that it makes sense for the flux of cars to be constant along characteristics, as must hold for the color equation.

9.5 Edge Velocities

So far we have assumed that the variable velocity $u(x)$ is specified by a constant value u_i within the ith grid cell. In some cases it is more natural to instead assume that a velocity $u_{i-1/2}$ is specified at each cell interface. This can be viewed as a transfer rate between the cells $\mathcal{C}_{i-1}$ and $\mathcal{C}_i$.

9.5.1 The Color Equation

Solving the color equation $\bar{q}_t + u(x)\bar{q}_x = 0$ with edge velocities specified is very simple. We need only set

$$\begin{aligned} \mathcal{W}_{i-1/2} &= \bar{Q}_i - \bar{Q}_{i-1}, \\ s_{i-1/2} &= u_{i-1/2}, \end{aligned} \tag{9.33}$$

and use the usual formula

$$\begin{aligned} \mathcal{A}^+ \Delta Q_{i-1/2} &= \big(s_{i-1/2}\big)^+ \mathcal{W}_{i-1/2}, \\ \mathcal{A}^- \Delta Q_{i-1/2} &= \big(s_{i-1/2}\big)^- \mathcal{W}_{i-1/2}. \end{aligned} \tag{9.34}$$

9.5.2 The Conservative Equation

The conservative advection equation $q_t + (u(x)q)_x = 0$ can also be solved using the waves and speeds

$$\begin{aligned} \mathcal{W}_{i-1/2} &= Q_i - Q_{i-1}, \\ s_{i-1/2} &= u_{i-1/2}, \end{aligned} \tag{9.35}$$

to define second-order corrections (perhaps modified by limiters). However, we cannot use the expressions (9.34) to define the fluctuations. This would not lead to a conservative method, since (4.54) would not be satisfied. There are several different approaches that can be used to define fluctuations that do satisfy (4.54) and lead to successful methods. One approach is to notice that an upwind flux $F_{i-1/2}$ is naturally defined by

$$F_{i-1/2} = u^+_{i-1/2} Q_{i-1} + u^-_{i-1/2} Q_i$$

at the edge between the cells. Recall that the goal in defining $\mathcal{A}^\pm \Delta Q_{i-1/2}$ is to implement the flux-differencing algorithm

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\big(F^n_{i+1/2} - F^n_{i-1/2}\big) \tag{9.36}$$

via the formula

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\big(\mathcal{A}^+ \Delta Q_{i-1/2} + \mathcal{A}^- \Delta Q_{i+1/2}\big). \tag{9.37}$$

We can accomplish this by setting

$$\begin{aligned} \mathcal{A}^+ \Delta Q_{i-1/2} &= F_i - F_{i-1/2}, \\ \mathcal{A}^- \Delta Q_{i-1/2} &= F_{i-1/2} - F_{i-1}, \end{aligned} \tag{9.38}$$

where the cell-centered flux values F_i are chosen in an arbitrary way. When inserted into (9.37) the value F_i cancels out and (9.36) results. For simplicity one could even use $F_i = 0$ in each cell.

Aesthetically it is nicer to use some approximation to the flux in cell i, for example

$$F_i = \big(u^+_{i-1/2} + u^-_{i+1/2}\big) Q_i, \tag{9.39}$$

so that $\mathcal{A}^\pm \Delta Q_{i-1/2}$ have the physical interpretation of flux differences. Note that if all velocities are positive, then these formulas reduce to

$$F_{i-1/2} = u_{i-1/2} Q_{i-1}, \qquad F_i = u_{i-1/2} Q_i,$$

so that

$$\begin{aligned} \mathcal{A}^+ \Delta Q_{i-1/2} &= u_{i-1/2}(Q_i - Q_{i-1}), \\ \mathcal{A}^- \Delta Q_{i-1/2} &= \big(u_{i-1/2} - u_{i-3/2}\big) Q_{i-1}. \end{aligned} \tag{9.40}$$

The right-going fluctuation is $s_{i-1/2}\mathcal{W}_{i-1/2}$ (as for the color equation), while the left-going fluctuation accounts for the variation of u across cell $\mathcal{C}_{i-1}$. Note that even though all flow is to the right, some information must flow to the left, since the accumulation rate in cell $\mathcal{C}_{i-1}$ depends on the outflow velocity $u_{i-1/2}$ as well as the inflow velocity $u_{i-3/2}$. These fluctuations result in the first-order method

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\big(u_{i+1/2} Q_i^n - u_{i-1/2} Q_{i-1}^n\big). \tag{9.41}$$

Note also that the splitting of (9.40) gives distinct approximations to the two terms that arise when we rewrite the conservative equation in the form

$$q_t + u(x) q_x + u'(x) q = 0. \tag{9.42}$$

Another way to derive the method (9.37) with the splitting (9.40) is to start with the form (9.42) and view this as the color equation with a source term $-u'(x)q$. Then $\mathcal{W}_{i-1/2}$, $s_{i-1/2}$, and $\mathcal{A}^+ \Delta Q_{i-1/2}$ all come from the color equation. The term $\mathcal{A}^- \Delta Q_{i-1/2}$ from (9.40) is different from the value 0 that would be used for the color equation, and can be viewed as a device for introducing the appropriate source term into cell $\mathcal{C}_{i-1}$ (rather than using a fractional-step method as discussed in Chapter 17). This approach has the advantage of maintaining conservation, while a fractional-step method might not.

If second-order correction terms are added to this algorithm using the waves and speeds (9.35) and the formula (9.18), then it can be shown that the method is formally second-order accurate if no limiters are used (Exercise 9.6).

9.6 Variable-Coefficient Acoustics Equations

Previously we have studied the equations of linear acoustics in a uniform medium where the density ρ_0 and the bulk modulus K_0 are the same at every point. We now consider acoustics in a one-dimensional heterogeneous medium in which the values $\rho(x)$ and $K(x)$ vary with x. The theory and numerical methods developed here also apply to elasticity or electromagnetic waves in heterogeneous media, and to other hyperbolic systems with variable coefficients.

An important special case is a *layered medium* in which ρ and K are piecewise constant. In applying finite volume methods we will assume the medium has this structure, at least at the level of the grid cells. We assume the ith cell contains a uniform material with density ρ_i and bulk modulus K_i. Smoothly varying material parameters can be approximated well in this manner on a sufficiently fine grid. Our discussion will focus primarily on the case of piecewise constant media, since we need to understand this case in detail in order to implement methods based on Riemann solvers. See Section 9.14 for discussion of how to choose the ρ_i and K_i if the coefficients vary on the subgrid scale.

It is possible to solve the variable-coefficient acoustics equations in conservation form by writing them in terms of momentum and strain; see [273] for example. Here we explore the nonconservative formulation that arises when pressure and velocity are used as the dependent variables. These are the quantities that must be continuous at an interface between different materials and are also physically more intuitive. This example provides a nice case study in solving a nonconservative variable-coefficient system, which can form a basis for developing algorithms for other hyperbolic systems that perhaps cannot be rewritten in conservation form.

The form (2.52) of the constant-coefficient acoustic equations generalizes naturally to

$$q_t + A(x)q_x = 0, \tag{9.43}$$

where

$$q(x,t) = \begin{bmatrix} p(x,t) \\ u(x,t) \end{bmatrix}, \qquad A(x) = \begin{bmatrix} 0 & K(x) \\ 1/\rho(x) & 0 \end{bmatrix}. \tag{9.44}$$

At each point x we can diagonalize this matrix as in Section 2.8,

$$A(x) = R(x)\Lambda(x)R^{-1}(x). \tag{9.45}$$

If we define the *sound speed*

$$c(x) = \sqrt{K(x)/\rho(x)} \tag{9.46}$$

and the *impedance*

$$Z(x) = \rho(x)c(x) = \sqrt{K(x)\rho(x)}, \tag{9.47}$$

then the eigenvector and eigenvalue matrix of (9.45) are

$$R(x) = \begin{bmatrix} -Z(x) & Z(x) \\ 1 & 1 \end{bmatrix}, \qquad \Lambda(x) = \begin{bmatrix} -c(x) & 0 \\ 0 & c(x) \end{bmatrix}. \tag{9.48}$$

9.7 Constant-Impedance Media

An interesting special case arises if $\rho(x)$ and $K(x)$ are related in such a way that $Z(x) = Z_0$ is constant, which happens if $K(x) = Z_0^2/\rho(x)$ everywhere. Then the eigenvector matrix $R(x) \equiv R$ is constant in space. In this case we can diagonalize the system (9.43) by multiplying by R^{-1} to obtain

$$(R^{-1}q)_t + [R^{-1}A(x)R](R^{-1}q)_x = 0.$$

Defining

$$w(x,t) = R^{-1}q(x,t)$$

as in Section 2.9, we can write the above system as

$$w_t + \Lambda(x)w_x = 0. \tag{9.49}$$

This diagonal system decouples into two scalar advection equations (color equations in the terminology of Section 9.1)

$$w_t^1 - c(x)w_x^1 = 0 \quad \text{and} \quad w_t^2 + c(x)w_x^2 = 0.$$

Left-going and right-going sound waves propagate independently of one another with the variable sound speed $c(x)$.

Example 9.1. Consider a piecewise constant medium with a single discontinuity at $x = 0$,

$$\rho(x) = \begin{cases} \rho_l & \text{if } x < 0, \\ \rho_r & \text{if } x > 0, \end{cases} \qquad K(x) = \begin{cases} K_l & \text{if } x < 0, \\ K_r & \text{if } x > 0. \end{cases} \tag{9.50}$$

Take $\rho_l = 1$, $\rho_r = 2$ and $K_l = 1$, $K_r = 0.5$, so that the impedance is $Z_0 = 1$ everywhere although the sound speed jumps from $c_l = 1$ to $c_r = 0.5$. Figure 9.4 shows a right-going acoustic wave having $p(x, t) = Z_0 u(x, t)$ everywhere. As it passes through the interface the pulse is compressed due to the change in velocity (just as would happen with the scalar color equation), but the entire wave is transmitted and there is no reflection at the interface. Compare with Figure 9.5, where there is also a jump in the impedance.

9.8 Variable Impedance

Note that the diagonalization (9.49) is possible only because R is constant in space. If $R(x)$ varies with x then $R^{-1}(x)q_x = (R^{-1}(x)q)_x - R_x^{-1}(x)q$, where $R_x^{-1}(x)$ is the derivative of $R^{-1}(x)$ with respect to x. In this case multiplying (9.43) by $R^{-1}(x)$ yields

$$\begin{aligned} w_t + \Lambda(x)w_x &= R_x^{-1}(x)q \\ &= B(x)w, \end{aligned} \tag{9.51}$$

where the matrix $B(x)$ is given by $B(x) = R_x^{-1}(x)R(x)$ and is not a diagonal matrix. In this case we still have advection equations for w^1 and w^2, but they are now coupled together by the source terms. The left-going and right-going acoustic waves no longer propagate independently of one another.

Rather than attempting to work with this system, we will see that the Riemann problem can be posed and solved at a general interface between two different materials. By assuming the medium is layered (so the impedance is piecewise constant), we can reduce the general problem with variable coefficients to solving such Riemann problems.

Before considering the general Riemann problem, consider what happens when a wave hits an interface in a medium of the form (9.50) if there is a jump in impedance at the interface.

Example 9.2. Consider the medium (9.50) with $\rho_l = 1$, $\rho_r = 4$ and $K_l = 1$, $K_r = 1$. As in Example 9.1, the velocity jumps from $c_l = 1$ to $c_r = 0.5$, but now the impedance is also discontinuous with $Z_l = 1$ and $Z_r = 2$. Figure 9.5 shows what happens as a right-going wave hits the interface. Part of the wave is transmitted, but part is reflected as a left-going wave.

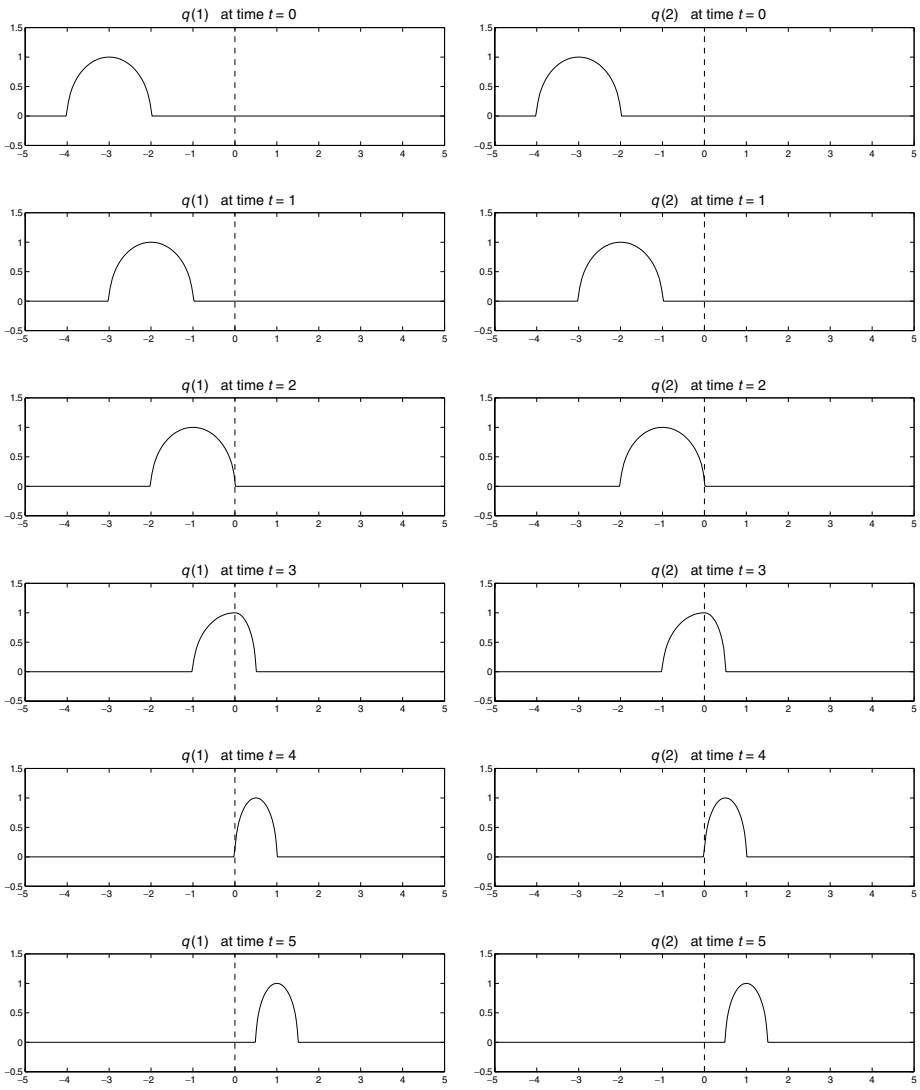

Fig. 9.4. Right-going acoustic pulse hitting a material interface (dashed line) where the sound speed changes from 1 to 0.5 but the impedance is the same. Left column: pressure. Right column: velocity. `[claw/book/chap9/acoustics/interface]`

If the incident pressure pulse has magnitude p_0, then the transmitted pulse has magnitude $C_T p_0$ and the reflected pulse has magnitude $C_R p_0$, where the *transmission* and *reflection coefficients* are given by

$$C_T = \frac{2Z_r}{Z_l + Z_r}, \qquad C_R = \frac{Z_r - Z_l}{Z_l + Z_r}. \tag{9.52}$$

For the example in Figure 9.5, we have $C_T = 4/3$ and $C_R = 1/3$. Partial reflection always occurs at any interface where there is an *impedance mismatch* with $Z_l \neq Z_r$. There are several ways to derive these coefficients. We will do so below by solving an appropriate Riemann problem, as motivated by the next example.

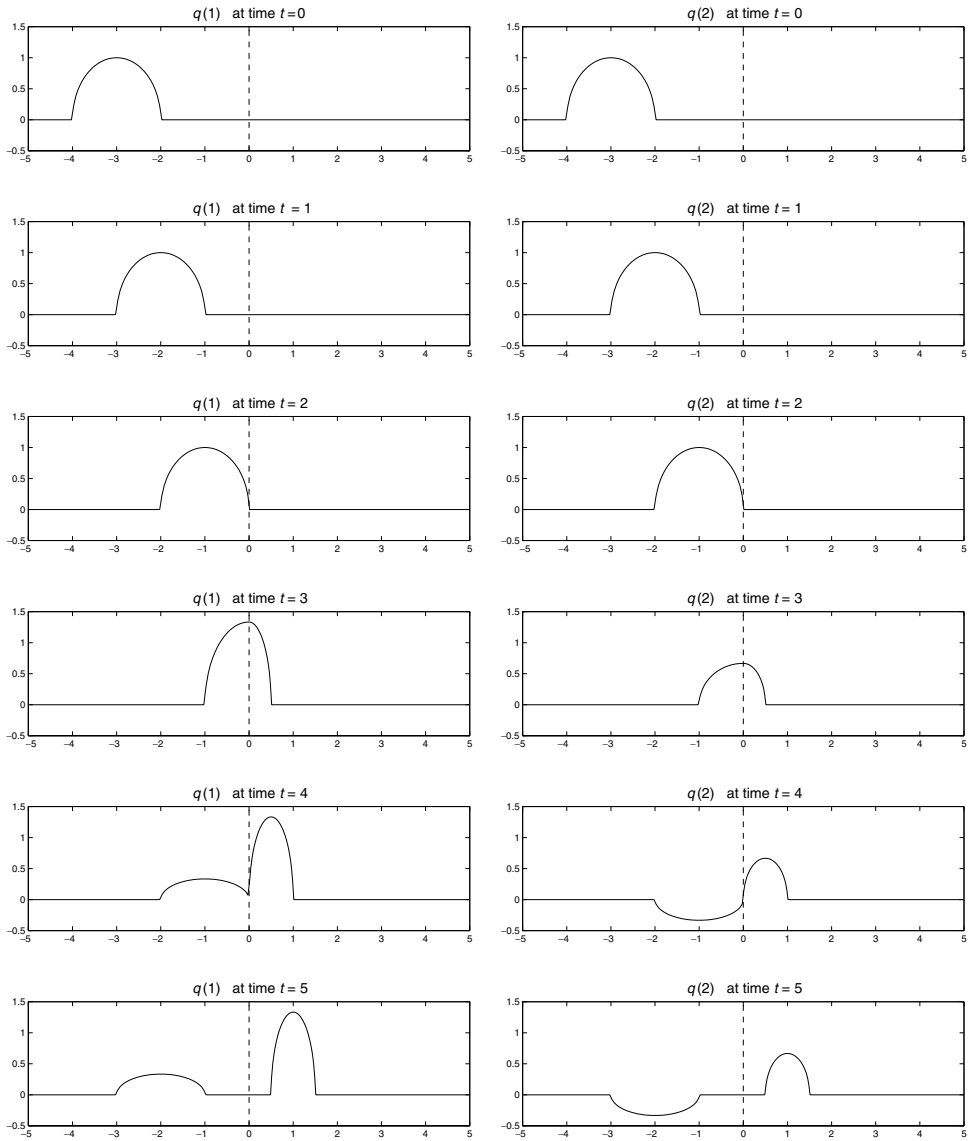

Fig. 9.5. Right-going acoustic pulse hitting a material interface (dashed line) where the sound speed changes from 1 to 0.5 and the impedance changes from 1 to 2. Part of the wave is reflected at the interface. Left column: pressure. Right column: velocity. `[claw/book/chap9/acoustics/interface]`

Example 9.3. To see how the transmission and reflection coefficients are related to a Riemann problem, consider the situation where the incident wave is a jump discontinuity as shown in Figure 9.6. When this discontinuity hits the interface, part of it is transmitted and part is reflected. Now suppose we ignore the first two frames of the time sequence shown in Figure 9.6 (at times $t = -2$ and $t = -1$) and consider what happens only from time $t = 0$ onwards. At time $t = 0$ we have data that is piecewise constant with a jump discontinuity at $x = 0$ in a medium that also has a jump discontinuity at $x = 0$. This is the proper generalization of the classic Riemann problem to the case of a heterogeneous medium. As time advances the discontinuity in the data resolves into two waves, one

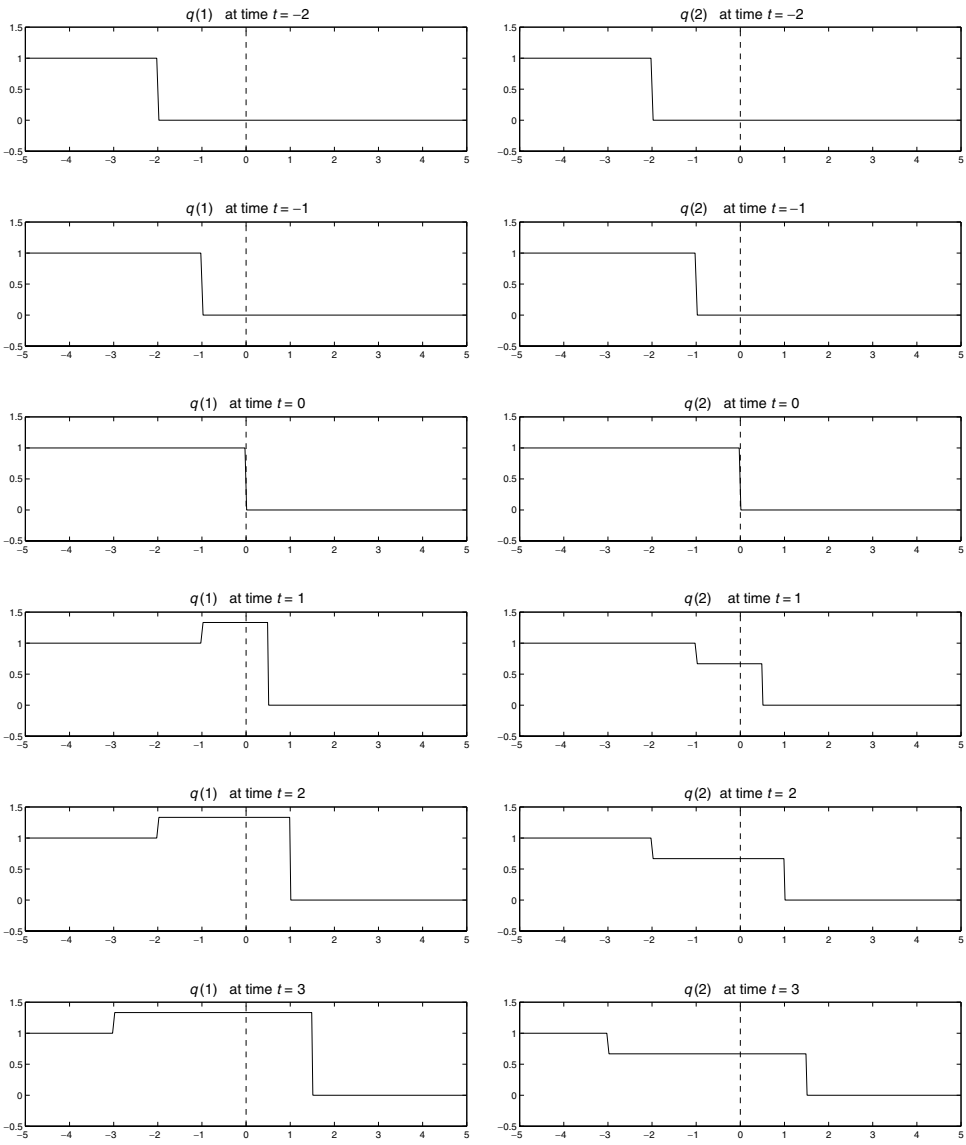

Fig. 9.6. Right-going square pulse hitting a material interface (dashed line) where the sound speed changes from 1 to 0.5 and the impedance changes from 1 to 2. The sequence from $t = 0$ onwards corresponds to the Riemann problem for variable-coefficient acoustics. Left column: pressure. Right column: velocity. `[claw/book/chap9/acoustics/interface]`

moving to the left at the sound speed of the medium on the left and the other moving to the right at the sound speed of that medium. Determining the magnitude of these two waves relative to the original jump will give the expressions (9.52) for the transmission and reflection coefficients. Being able to solve the general Riemann problem will also allow us to apply high-resolution finite volume methods to the general variable-coefficient acoustics problem.

9.9 Solving the Riemann Problem for Acoustics

The general Riemann problem for acoustics in a heterogeneous medium is defined by considering a piecewise constant medium of the form (9.50) together with piecewise constant initial data

$$q(x,0) = \begin{cases} q_l & \text{if } x < 0, \\ q_r & \text{if } x > 0. \end{cases} \tag{9.53}$$

The solution to the Riemann problem consists of two acoustic waves, one moving to the left with velocity $-c_l$ and the other to the right with velocity $+c_r$. Each wave moves at the sound speed of the material in which it propagates, as indicated in Figure 9.7. At the interface $x = 0$ the pressure and velocity perturbations are initially discontinuous, but they should be continuous for $t > 0$, after the waves have departed. Hence there should be a single state q_m between the two waves as indicated in Figure 9.7 and as seen in Figure 9.6.

We know from the theory of Section 2.8 that the jump across each wave must be an eigenvector of the coefficient matrix from the appropriate material. Hence we must have

$$q_m - q_l = \alpha^1 \begin{bmatrix} -Z_l \\ 1 \end{bmatrix} \quad \text{and} \quad q_r - q_m = \alpha^2 \begin{bmatrix} Z_r \\ 1 \end{bmatrix} \tag{9.54}$$

for some scalar coefficients α^1 and α^2. Adding these two equations together yields

$$q_r - q_l = \alpha^1 \begin{bmatrix} -Z_l \\ 1 \end{bmatrix} + \alpha^2 \begin{bmatrix} Z_r \\ 1 \end{bmatrix}.$$

This leads to a linear system of equations to solve for α^1 and α^2,

$$R_{lr}\alpha = q_r - q_l,$$

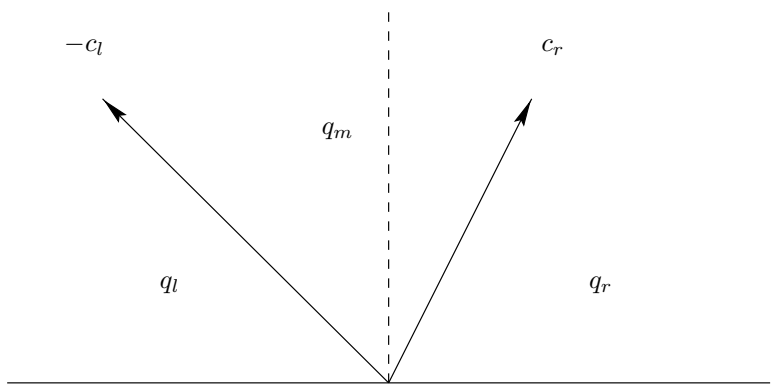

Fig. 9.7. Structure of the solution to the Riemann problem for variable-coefficient acoustics, in the x–t plane. The dashed line shows the interface between two different materials. The waves propagate at the speed of sound in each material. Between the waves is a single state q_m.

where

$$R_{lr} = \begin{bmatrix} -Z_l & Z_r \\ 1 & 1 \end{bmatrix}. \tag{9.55}$$

This is essentially the same process used to solve the Riemann problem for constant-coefficient acoustics, but with an eigenvector matrix R_{lr} that is not the matrix $R(x)$ from either side of the interface, but rather consists of the left-going eigenvector from the left medium and the right-going eigenvector from the right medium.

Since

$$R_{lr}^{-1} = \frac{1}{Z_l + Z_r} \begin{bmatrix} -1 & Z_r \\ 1 & Z_l \end{bmatrix}, \tag{9.56}$$

we find that

$$\begin{aligned} \alpha^1 &= \frac{-(p_r - p_l) + Z_r(u_r - u_l)}{Z_l + Z_r}, \\ \alpha^2 &= \frac{(p_r - p_l) + Z_l(u_r - u_l)}{Z_l + Z_r}. \end{aligned} \tag{9.57}$$

We also find that the intermediate state $q_m = q_l + \alpha^1 \mathcal{W}^1$ is given by

$$\begin{aligned} p_m &= p_l - \alpha^1 Z_l = \frac{Z_r p_l + Z_l p_r}{Z_l + Z_r} - \left(\frac{Z_r Z_l}{Z_l + Z_r}\right)(u_r - u_l), \\ u_m &= u_l + \alpha^1 \quad = \frac{Z_l u_l + Z_r u_r}{Z_l + Z_r} - \left(\frac{1}{Z_l + Z_r}\right)(p_r - p_l). \end{aligned} \tag{9.58}$$

If $Z_l = Z_r$, then these reduce to the formulas (3.31) and (3.32) for acoustics in a constant medium.

9.10 Transmission and Reflection Coefficients

Returning to Example 9.3, we can now compute the transmission and reflection coefficients for a wave hitting an interface. If the Riemann problem arises from an incident wave hitting an interface as shown in Figure 9.6, then q_l and q_r are not arbitrary but must be related by the fact that $q_r - q_l$ is the jump across the incident right-going wave, and hence

$$q_r - q_l = \beta \begin{bmatrix} Z_l \\ 1 \end{bmatrix} \tag{9.59}$$

for some scalar β. After solving the Riemann problem at the heterogeneous interface, the outgoing waves have strengths α^1 and α^2 given by

$$\begin{aligned} \alpha &= R_{lr}^{-1}(q_r - q_l) \\ &= \frac{\beta}{Z_l + Z_r} \begin{bmatrix} -1 & Z_r \\ 1 & Z_l \end{bmatrix} \begin{bmatrix} Z_l \\ 1 \end{bmatrix} \\ &= \frac{\beta}{Z_l + Z_r} \begin{bmatrix} Z_r - Z_l \\ 2Z_l \end{bmatrix}. \end{aligned} \tag{9.60}$$

We thus find that the magnitudes of the reflected and transmitted waves are

$$\alpha^1 = \left(\frac{Z_r - Z_l}{Z_l + Z_r}\right)\beta \quad \text{and} \quad \alpha^2 = \left(\frac{2Z_l}{Z_l + Z_r}\right)\beta \tag{9.61}$$

respectively. The waves are thus

$$\text{reflected wave:} \quad \alpha^1 \begin{bmatrix} -Z_l \\ 1 \end{bmatrix} = \beta \left(\frac{Z_r - Z_l}{Z_l + Z_r}\right) \begin{bmatrix} -Z_l \\ 1 \end{bmatrix},$$

$$\text{transmitted wave:} \quad \alpha^2 \begin{bmatrix} Z_r \\ 1 \end{bmatrix} = \beta \left(\frac{2Z_l}{Z_l + Z_r}\right) \begin{bmatrix} Z_r \\ 1 \end{bmatrix}.$$

Note that the pressure jump in the transmitted wave can be written as

$$\beta \left(\frac{2Z_l}{Z_l + Z_r}\right) Z_r = \left(\frac{2Z_r}{Z_l + Z_r}\right) \beta Z_l.$$

Comparing this with the pressure jump βZ_l of the incident wave (9.59) gives the transmission coefficient C_T of (9.52). The reflected wave has a pressure jump $-C_R(\beta Z_l)$, where C_R is given in (9.52). The minus sign arises from the fact that we measure the pressure jump by subtracting the value to the left of the wave from the value to the right. Since the wave is left-going, however, the gas experiences a jump in pressure of the opposite sign as the wave passes by, and it is this jump that is normally considered in defining the reflection coefficient C_R.

9.11 Godunov's Method

Having determined the solution to the Riemann problem for acoustics in a heterogeneous medium, it is easy to implement Godunov's method and other finite volume methods. Each grid cell is assigned a density ρ_i and a bulk modulus K_i, and Q_i^n represents the cell average of q over this cell at time t_n. The Riemann problem at $x_{i-1/2}$ is solved as described in Section 9.9 with $\rho_l = \rho_{i-1}$, $\rho_r = \rho_i$, etc. We obtain two waves

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

where

$$r^1_{i-1/2} = \begin{bmatrix} -Z_{i-1} \\ 1 \end{bmatrix}, \qquad r^2_{i-1/2} = \begin{bmatrix} Z_i \\ 1 \end{bmatrix}.$$

These waves propagate at speeds $s^1_{i-1/2} = -c_{i-1}$ and $s^2_{i-1/2} = +c_i$. The coefficients $\alpha^1_{i-1/2}$ and $\alpha^2_{i-1/2}$ are given by (9.57):

$$\begin{aligned} \alpha^1_{i-1/2} &= \frac{-(p_i - p_{i-1}) + Z_i(u_i - u_{i-1})}{Z_{i-1} + Z_i}, \\ \alpha^2_{i-1/2} &= \frac{(p_i - p_{i-1}) + Z_{i-1}(u_i - u_{i-1})}{Z_{i-1} + Z_i}. \end{aligned} \tag{9.62}$$

Godunov's method can be implemented in the fluctuation form

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

if we define

$$\mathcal{A}^-\Delta Q_{i-1/2} = s^1_{i-1/2}\mathcal{W}^1_{i-1/2}, \qquad \mathcal{A}^+\Delta Q_{i-1/2} = s^2_{i-1/2}\mathcal{W}^2_{i-1/2}, \tag{9.63}$$

following the general prescription (4.42). Note that the variable-coefficient acoustic equations (9.43) are not in conservation form, and so this is not a flux-difference splitting in this case, but it still leads to the proper updating of cell averages based on waves moving into the grid cells.

For constant-coefficient acoustics, we saw in Section 4.12 that we could interpret the fluctuation $\mathcal{A}^-\Delta Q_{i-1/2}$ as $A^-(Q_i - Q_{i-1})$, the "negative part" of the matrix A multiplied by the jump in Q, with A^- defined in (4.45). For variable-coefficient acoustics it is interesting to note that we can make a similar interpretation in spite of the fact that the matrix A varies. In this case we have $\mathcal{A}^-\Delta Q_{i-1/2} = A_{i-1/2}(Q_i - Q_{i-1})$, where the matrix $A_{i-1/2}$ is different than the coefficient matrices

$$A_i = \begin{bmatrix} 0 & K_i \\ 1/\rho_i & 0 \end{bmatrix} = R_i\Lambda_i R_i^{-1}$$

defined in the grid cells. Define

$$R_{i-1/2} = \begin{bmatrix} -Z_{i-1} & Z_i \\ 1 & 1 \end{bmatrix},$$

analogous to R_{lr} in (9.55). Then the wave strengths $\alpha^1_{i-1/2}$ and $\alpha^2_{i-1/2}$ are determined by

$$\alpha_{i-1/2} = R^{-1}_{i-1/2}(Q_i - Q_{i-1}).$$

The fluctuations (9.63) are thus given by

$$\begin{aligned} \mathcal{A}^-\Delta Q_{i-1/2} &= R_{i-1/2}\Lambda^-_{i-1/2}R^{-1}_{i-1/2}(Q_i - Q_{i-1}) = A^-_{i-1/2}(Q_i - Q_{i-1}), \\ \mathcal{A}^+\Delta Q_{i-1/2} &= R_{i-1/2}\Lambda^+_{i-1/2}R^{-1}_{i-1/2}(Q_i - Q_{i-1}) = A^+_{i-1/2}(Q_i - Q_{i-1}), \end{aligned} \tag{9.64}$$

where we define

$$\Lambda_{i-1/2} = \begin{bmatrix} s^1_{i-1/2} & 0 \\ 0 & s^2_{i-1/2} \end{bmatrix} = \begin{bmatrix} -c_{i-1} & 0 \\ 0 & c_i \end{bmatrix}$$

and hence

$$\begin{aligned} A_{i-1/2} &= R_{i-1/2}\Lambda_{i-1/2}R^{-1}_{i-1/2} \\ &= \frac{1}{Z_{i-1}+Z_i}\begin{bmatrix} c_iZ_i - c_{i-1}Z_{i-1} & (c_{i-1}+c_i)Z_{i-1}Z_i \\ c_{i-1}+c_i & c_iZ_{i-1} - c_{i-1}Z_i \end{bmatrix}. \end{aligned} \tag{9.65}$$

If $A_{i-1} = A_i$, then $A_{i-1/2}$ also reduces to this same coefficient matrix. In general it is different from the coefficient matrix on either side of the interface, and can be viewed as a

special average of the coefficient matrix from the two sides. If the coefficients are smoothly varying, then $A_{i-1/2} \approx A(x_{i-1/2})$.

Note that implementing the method does not require working with this matrix $A_{i-1/2}$. We work directly with the waves and speeds. The Riemann solver is implemented in `[claw/book/chap9/acoustics/interface/rp1acv.f]`.

9.12 High-Resolution Methods

Godunov's method for variable-coefficient acoustics is easily extended to a high-resolution method via a natural extension of the methods presented in Section 6.13, together with appropriate wave limiters. The method has the general form

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}\left(\mathcal{A}^+ \Delta Q_{i+1/2} + \mathcal{A}^- \Delta Q_{i-1/2}\right) - \frac{\Delta t}{\Delta x}\left(\tilde{F}_{i+1/2} - \tilde{F}_{i-1/2}\right), \quad (9.66)$$

where

$$\tilde{F}_{i-1/2} = \frac{1}{2}\sum_{p=1}^{m} |s_{i-1/2}^p| \left(1 - \frac{\Delta t}{\Delta x}|s_{i-1/2}^p|\right) \widetilde{\mathcal{W}}_{i-1/2}^p. \quad (9.67)$$

Before discussing limiters, first consider the unlimited case in which $\widetilde{\mathcal{W}}_{i-1/2}^p = \mathcal{W}_{i-1/2}$. Then a calculation shows that the method (9.66) can be rewritten as

$$\begin{aligned} Q_i^{n+1} = Q_i^n &- \frac{\Delta t}{2\,\Delta x}\left[A_{i-1/2}(Q_i - Q_{i-1}) + A_{i+1/2}(Q_{i+1} - Q_i)\right] \\ &+ \frac{\Delta t^2}{2\,\Delta x^2}\left[A_{i+1/2}^2(Q_{i+1} - Q_i) - A_{i-1/2}^2(Q_i - Q_{i-1})\right]. \end{aligned} \quad (9.68)$$

This is a Lax–Wendroff-style method based on the coefficient matrices defined at the interface as in (9.65). As for the case of the variable-coefficient color equation discussed in Section 9.5, this method is not formally second-order accurate when A varies, because the final term approximates $\frac{1}{2}\Delta t^2 (A^2 q_x)_x$ while the Taylor series expansion requires $\frac{1}{2}\Delta t^2 q_{tt} = \frac{1}{2}\Delta t^2 A(Aq_x)_x$. These differ by $\frac{1}{2}\Delta t^2 A_x A q_x$. However, we have captured the $\frac{1}{2}\Delta t^2 A^2 q_{xx}$ term correctly, which is essential in eliminating the numerical diffusion of the Godunov method.

This approach yields good high-resolution results in practice in spite of the lack of formal second-order accuracy (as we've seen before, e.g., Section 8.5, Section 9.3.1). Since these high-resolution methods are of particular interest for problems where the solution or coefficients are discontinuous, formal second-order accuracy cannot be expected in any case. If the problem is in conservation form $q_t + (A(x)q)_x = 0$, then this approach in fact gives formal second-order accuracy. See Section 16.4, where this is discussed in the more general context of a conservation law with a spatially varying flux function, $q_t + f(q, x)_x = 0$.

9.13 Wave Limiters

Applying limiters to the waves $\mathcal{W}_{i-1/2}^p$ is slightly more difficult for variable-coefficient systems of equations than in the constant-coefficient case. To obtain a high-resolution

method, we wish to replace the wave $\mathcal{W}^p_{i-1/2}$ by a limited version in (9.67),

$$\widetilde{\mathcal{W}}^p_{i-1/2} = \phi\big(\theta^p_{i-1/2}\big)\mathcal{W}^p_{i-1/2}, \tag{9.69}$$

as described in Chapter 6. Again $\theta^p_{i-1/2}$ should be some measure of the smoothness of the pth characteristic component of the solution. When A is constant the eigenvectors r^1 and r^2 are the same everywhere and so $\mathcal{W}^p_{i-1/2} = \alpha^p_{i-1/2} r^p$ can easily be compared to the corresponding wave $\mathcal{W}^p_{I-1/2} = \alpha^p_{I-1/2} r^p$ arising from the neighboring Riemann problem by simply comparing the two scalar wave strengths,

$$\theta^p_{i-1/2} = \frac{\alpha^p_{I-1/2}}{\alpha^p_{i-1/2}}. \tag{9.70}$$

(Recall that $I = i \pm 1$ as in (6.61), looking in the upwind direction as determined by the sign of $s^p_{i-1/2}$.) Note that in this constant-coefficient case we have

$$\big|\theta^p_{i-1/2}\big| = \frac{\big\|\mathcal{W}^p_{I-1/2}\big\|}{\big\|\mathcal{W}^p_{i-1/2}\big\|} \tag{9.71}$$

in any vector norm.

For a variable-coefficient problem, the two waves

$$\mathcal{W}^p_{i-1/2} = \alpha^p_{i-1/2} r^p_{i-1/2} = \big(\ell^p_{i-1/2}\,\Delta Q_{i-1/2}\big) r^p_{i-1/2} \tag{9.72}$$

and

$$\mathcal{W}^p_{I-1/2} = \alpha^p_{I-1/2} r^p_{I-1/2} = \big(\ell^p_{I-1/2}\,\Delta Q_{I-1/2}\big) r^p_{I-1/2} \tag{9.73}$$

are typically not scalar multiples of one another, since the eigenvector matrices $R(x)$ and $L(x) = R^{-1}(x)$ vary with x. In this case it may make no sense to compare $\alpha^p_{i-1/2}$ with $\alpha^p_{I-1/2}$, since these coefficients depend entirely on the normalization of the eigenvectors, which may vary with x. It would make more sense to compare the magnitudes of the full waves, using the expression (9.71) for some choice of vector norm, but some way must be found to assign a sign, which we expect to be negative near extreme points. More generally, a problem with (9.71) is that it would give $|\theta| \approx 1$ whenever the two waves are of similar magnitude, regardless of whether they are actually similar vectors in phase space or not. If the solution is truly smooth, then we expect $\mathcal{W}^p_{i-1/2} \approx \mathcal{W}^p_{I-1/2}$ and not just that these vectors have nearly the same magnitude (which might be true for two vectors pointing in very different directions in phase space at some discontinuity in the medium, or in the solution for a nonlinear problem).

We would like to define $\theta^p_{i-1/2}$ in a way that reduces to (9.70) in the constant-coefficient case but that takes into account the degree of alignment of the wave vectors as well as their magnitudes. This can be accomplished by projecting the vector $\mathcal{W}^p_{I-1/2}$ onto the vector $\mathcal{W}^p_{i-1/2}$ to obtain a vector $\theta^p_{i-1/2}\mathcal{W}^p_{i-1/2}$ that is aligned with $\mathcal{W}^p_{i-1/2}$. Using the scalar coefficient of this projection as $\theta^p_{i-1/2}$ gives

$$\theta^p_{i-1/2} = \frac{\mathcal{W}^p_{I-1/2} \cdot \mathcal{W}^p_{i-1/2}}{\mathcal{W}^p_{i-1/2} \cdot \mathcal{W}^p_{i-1/2}}, \tag{9.74}$$

where $\cdot$ is the dot product in $\mathbb{R}^m$. This works well in most cases, also for nonlinear problems, and this general procedure is implemented in the `limiter.f` routine of CLAWPACK.

Another approach has been proposed by Lax & Liu [264], [313] in defining their *positive schemes*. Recall that we have used (9.72) and (9.73) to define the limiter. Lax and Liu instead use

$$\widehat{\mathcal{W}}^p_{I-1/2} = \hat{\alpha}^p_{I-1/2} r^p_{i-1/2} = \left(\ell^p_{i-1/2} \Delta Q_{I-1/2}\right) r^p_{i-1/2} \tag{9.75}$$

in place of $\mathcal{W}^p_{I-1/2}$. The vector $\widehat{\mathcal{W}}^p_{I-1/2}$ is obtained by decomposing the jump $\Delta Q_{I-1/2}$ into eigenvectors of the matrix $R_{i-1/2}$ rather than into eigenvectors of $R_{I-1/2}$. The vector $\widehat{\mathcal{W}}^p_{I-1/2}$ is now a scalar multiple of $\mathcal{W}^p_{i-1/2}$, and so we can simply use

$$\theta^p_{i-1/2} = \frac{\hat{\alpha}^p_{I-1/2}}{\alpha^p_{i-1/2}} = \frac{\widehat{\mathcal{W}}^p_{I-1/2} \cdot \mathcal{W}^p_{i-1/2}}{\mathcal{W}^p_{i-1/2} \cdot \mathcal{W}^p_{i-1/2}}. \tag{9.76}$$

This is slightly more expensive to implement, but since we compute the left eigenvectors at each cell interface in any case in order to compute the $\mathcal{W}^p_{i-1/2}$, we can also use them to compute $\widehat{\mathcal{W}}^p_{I-1/2}$ at the same time if the computations are organized efficiently.

For most problems the simpler expression (9.74) seems to work well, but for some problems with rapid variation in the eigenvectors the Lax–Liu limiter is superior. For example, in solving the acoustics equations in a rapidly varying periodic medium an instability attributed to nonlinear resonance was observed in [139] using the standard CLAWPACK limiter. A *transmission-based* limiter, closely related to the Lax–Liu limiter, was introduced to address this problem. This is similar to (9.76) but decomposes only the wave $\mathcal{W}^p_{I-1/2}$ into eigenvectors $r^p_{i-1/2}$ rather than starting with the full jump $\Delta Q_{I-1/2}$:

$$\widehat{\mathcal{W}}^p_{I-1/2} = \hat{\alpha}^p_{I-1/2} r^p_{i-1/2} = \left(\ell^p_{i-1/2} \mathcal{W}^p_{I-1/2}\right) r^p_{i-1/2}. \tag{9.77}$$

The formula (9.76) is then used. For acoustics this has the interpretation of taking the wave from the neighboring Riemann problem that is approaching the interface $x_{i-1/2}$ and using only the portion of this wave that would be transmitted through the interface in defining the limiter. See [139] for more discussion of this and several examples where these methods are used in periodic or random media.

9.14 Homogenization of Rapidly Varying Coefficients

If the physical parameters vary substantially within a single grid cell, then it will be necessary to apply some sort of *homogenization theory* to determine appropriate averaged values to use for ρ_i and K_i. This might be the case in a seismology problem, for example, where each grid cell represents a sizable block of earth that is typically not uniform.

Since ρ is the density (mass per unit length, in one dimension), the average density of the ith cell is properly computed as the total mass in the cell divided by the length of the cell, i.e., as the *arithmetic average* of ρ within the cell,

$$\rho_i = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \rho(x)\, dx. \tag{9.78}$$

For the bulk modulus K it turns out that instead of an arithmetic average, it is necessary to use the *harmonic average*

$$K_i = \left(\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{1}{K(x)}\, dx \right)^{-1}. \tag{9.79}$$

Alternatively, we can view this as using the arithmetic average of the parameter $1/K$, which is called the *coefficient of compressibility* (see, e.g., [255]),

$$\frac{1}{K_i} = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{1}{K(x)}\, dx. \tag{9.80}$$

This is most easily motivated by considering the linear elasticity interpretation of acoustics from Section 2.12, with Lamé parameters $\lambda = K$ and $\mu = 0$ and the stress–strain relation

$$\sigma = K\epsilon \tag{9.81}$$

resulting from (2.89) (dropping superscripts on σ^{11} and ϵ^{11}). Rewriting this as

$$\epsilon = \frac{1}{K}\sigma \tag{9.82}$$

shows that this has the form of Hooke's law with spring constant $1/K$. Recalling that $\epsilon = X_x - 1$, we see that for a grid cell of homogeneous material ($K_i = $ constant) with rest length $x_{i-1/2} - x_{i+1/2} = \Delta x$, applying a force σ results in a strain

$$\frac{\big[X\big(x_{i+1/2}\big) - X\big(x_{i-1/2}\big)\big] - \Delta x}{\Delta x} = \left(\frac{1}{K_i}\right)\sigma. \tag{9.83}$$

If the material in the cell is heterogeneous (with σ still constant), then instead we have

$$\begin{aligned} \frac{\big[X\big(x_{i+1/2}\big) - X\big(x_{i-1/2}\big)\big] - \Delta x}{\Delta x} &= \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} [X_x(x) - 1]\, dx \\ &= \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \epsilon(x)\, dx \\ &= \left(\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{1}{K(x)}\, dx \right) \sigma. \end{aligned} \tag{9.84}$$

Comparing this with (9.83) motivates the averaging (9.80).

Homogenization theory is often used to derive simpler systems of partial differential equations to model the effective behavior of a medium with rapidly varying heterogeneties. As a simple example, consider a layered medium composed of thin layers of width L_1 of material characterized by parameters (ρ_1, K_1), which alternate with thin layers of width L_2 where the parameters are (ρ_2, K_2). Numerically we could solve this problem using the high-resolution methods developed above, provided we can take Δx small enough to resolve the layers. We do this in Example 9.4 below. Analytically, however, we might like to predict the behavior of waves propagating in this layered medium by solving some simplified equation in place of the variable-coefficient acoustics equation. If we consider

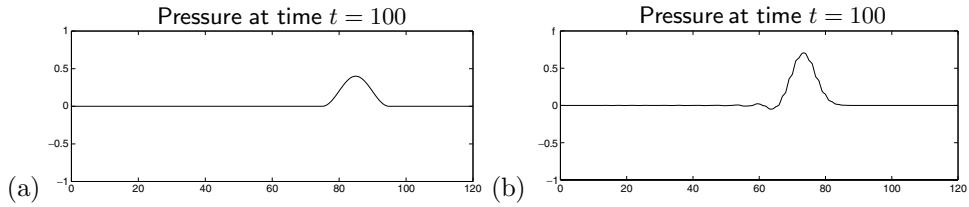

Fig. 9.8. (a) Pressure pulse in a homogeneous medium. (b) Pressure pulse in a layered medium.

waves whose wavelength is considerably longer than L_1 and L_2, then it turns out that the wave propagation can be reasonably well modeled by solving a homogenized system of equations that is simply the constant-coefficient acoustics equation with appropriate averages of ρ and $1/K$,

$$\bar{\rho} = \frac{1}{L_1 + L_2}(L_1\rho_1 + L_2\rho_2) \quad \text{and} \quad \hat{K}^{-1} = \frac{1}{L_1 + L_2}\left(\frac{L_1}{K_1} + \frac{L_2}{K_2}\right). \tag{9.85}$$

This is illustrated in Example 9.4. More accurate homogenized equations can also be derived, which for this problem requires adding higher-order dispersive terms to the constant-coefficient acoustics equation, as shown by Santosa & Symes [393].

In practical problems the material parameters often vary randomly, but with some slowly varying mean values $\rho(x)$ and $K^{-1}(x)$. This might be the case in a seismology problem, for example, where the basic type of rock or soil varies slowly (except for certain sharp interfaces) but is full of random heterogeneous structures at smaller scales. Similar problems arise in studying ultrasound waves in biological tissue, electromagnetic waves in a hazy atmosphere, and many other applications.

Example 9.4. Figure 9.8 shows the results from two acoustics problems with the same initial data and boundary conditions but different material properties. In each case the initial data is $p \equiv u \equiv 0$, and at the left boundary

$$u(0, t) = \begin{cases} 0.2\left[1 + \cos\left(\frac{\pi(t-15)}{10}\right)\right] & \text{if } |t - 15| < 10, \\ 0 & \text{otherwise.} \end{cases} \tag{9.86}$$

This in-and-out wall motion creates a pressure pulse that propagates to the right. In Figure 9.8(a) the material is uniform with $\rho \equiv 1$ and $K \equiv 1$, and so the sound speed is $c \equiv 1$. The peak pressure is generated at time $t = 15$, and at $t = 100$ is visible at $x = 85$, as should be expected.

Figure 9.8(b) shows a case where the material varies periodically with

$$\rho = K = \begin{cases} 3 & \text{for } 2i < x < 2i + 1, \\ 1 & \text{for } 2i + 1 < x < 2i + 2. \end{cases} \tag{9.87}$$

The layers have width $L_1 = L_2 = 1$ as seen in the zoomed view of Figure 9.9(a), where $\rho = K = 3$ in the dark layers and $\rho = K = 1$ in the light layers. For the choice (9.87), the sound speed is $c \equiv 1$ in all layers. However, the pressure pulse does not propagate at

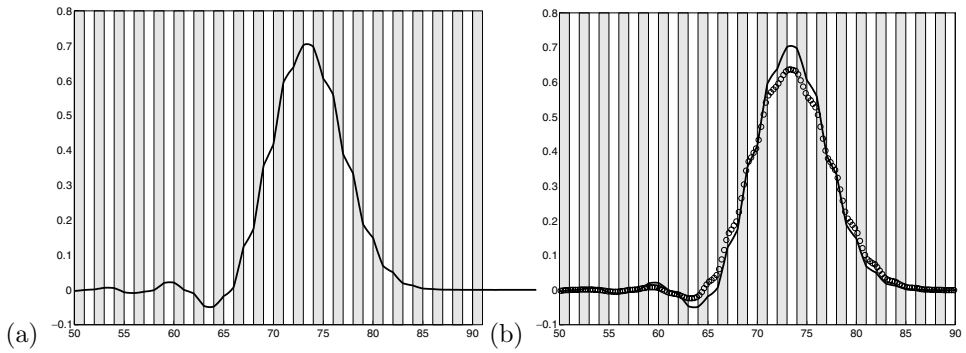

Fig. 9.9. (a) Zoom view of Figure 9.8(b) with the layers indicated. (b) The solid line is the "exact" solution. The symbols indicate the numerical solution computed with only four grid cells per layer. `[claw/book/chap9/acoustics/layered]`

speed 1. Instead, it propagates at the speed predicted by the homogenized acoustic equation as described above. The effective speed is

$$\bar{c} = \sqrt{\hat{K}/\bar{\rho}} \approx 0.866, \tag{9.88}$$

based on the arithmetic average $\bar{\rho} = \frac{1}{2}(3+1) = 2$ and the harmonic average $\hat{K} = (\frac{1}{2}(\frac{1}{3} + 1))^{-1} = 1.5$. The peak pressure generated at $t = 15$ is now observed at $x = (85)(0.866) \approx 73.6$, as is clearly seen in Figure 9.9(a).

What's going on? Even though the sound speed $c \equiv 1$ is constant everywhere, the impedance $Z = \rho c = \rho$ is not, and so the eigenvectors of the coefficient matrix are different in each layer. Recall from Section 9.8 that the wave will be partially reflected at each interface in this case. The wave observed in Figure 9.8(a) is not a simple wave in the 2-characteristic family translating to the right as in the constant-coefficient case. Instead it is composed of waves moving in both directions that are constantly bouncing back and forth between the interfaces, so that the energy moves more slowly to the right than we would expect from the sound speed. It is the group velocity that we are observing rather than the phase velocity.

Notice from Figure 9.9(a) that the waveform is not completely smooth, but appears to have jump discontinuities in p_x at each interface. This is expected, since

$$u_t + \frac{1}{\rho} p_x = 0.$$

The velocity u must be continuous at all times, from which it follows that u_t and hence p_x/ρ must also be continuous. This means that p_x will be discontinuous at points where ρ is discontinuous.

Figure 9.9(a) also shows dispersive oscillations beginning to form behind the propagating pressure pulse in the periodic medium. This is *not* due to numerical dispersion, but rather is a correct feature of the exact solution, as predicted by the dispersive homogenized equation derived in [393].

In fact, the curve shown in Figure 9.8(b) and Figure 9.9(a) is essentially the exact solution. It was calculated on a fine grid with 2400 cells, and hence 20 grid cells in each layer. More importantly, the time step $\Delta t = \Delta x$ was used. Since $c \equiv 1$ everywhere, this means

that each wave propagates through exactly one grid cell each time step, and there are no numerical errors introduced by averaging onto the grid. In essence we are using the method of characteristics, with the Riemann solver providing the correct characteristic decomposition at each cell interface.

We might now ask how accurate the high-resolution finite volume method would be on this problem, since for practical problems it will not always be possible to choose the time step so that $c\,\Delta t = \Delta x$ everywhere. This can be tested by choosing $\Delta t < \Delta x$, so that the method no longer reduces to the method of characteristics. Now the piecewise linear reconstruction with limiter functions is being used together with averaging onto grid cells.

Using $\Delta t = 0.8\,\Delta x$ on the same grid with 2400 cells gives numerical results that are nearly indistinguishable from the exact solution shown in Figure 9.9(a). As a much more extreme test, the calculation was performed on a grid with only four grid cells in each layer. The results are still reasonable, as shown by the circles in Figure 9.9(b). Some accuracy is lost at this resolution, but the basic structure of the solution is clearly visible, with the wave propagating at the correct effective velocity.

Exercises

9.1. Compute the local truncation error for the method (9.18) in the case where no limiter is applied. Show that when $u(x)$ is not constant the method fails to be formally second-order accurate.

9.2. Use the Lax–Wendroff approach described in Section 6.1, expanding in Taylor series, to derive a method that is formally second-order accurate for the variable-coefficient color equation (9.12). Compare this method with (9.18) using (9.17).

9.3. Show that if we add second-order correction terms (with no limiters) to (9.41) based on (9.35), then we obtain a second-order accurate method for the conservative advection equation.

9.4. Explain why (9.32) is consistent with the notion of density as "cars per unit length."

9.5. In the discussion of variable-coefficient advection we have assumed $u(x)$ has the same sign everywhere. Consider the conservative equation $q_t + (u(x)q)_x = 0$ with $q(x, 0) \equiv 1$ and

$$u(x) = \begin{cases} -1 & \text{if } x < 0, \\ +1 & \text{if } x > 0, \end{cases} \quad \text{or} \quad u(x) = \begin{cases} +1 & \text{if } x < 0, \\ -1 & \text{if } x > 0. \end{cases}$$

What is the solution to each? (You might want to consider the conveyer-belt interpretation of Section 9.4.1.) How well will the methods presented here work for these problems? (See also Exercise 13.11.)

9.6. Show that the method developed in Section 9.5.2 with the correction terms (9.18) added is formally second-order accurate if no limiter is applied.