Chapter 2
Linear and Nonlinear Waves

Our initial foray into the vast mathematical continent that comprises partial differential equations will begin with some basic first-order equations. In applications, first-order partial differential equations are most commonly used to describe dynamical processes, and so time, t, is one of the independent variables. Our discussion will focus on dynamical models in a single space dimension, bearing in mind that most of the methods we introduce can be extended to higher-dimensional situations. First-order partial differential equations and systems model a wide variety of wave phenomena, including transport of pollutants in fluids, flood waves, acoustics, gas dynamics, glacier motion, chromatography, traffic flow, and various biological and ecological systems.

A basic solution technique relies on an inspired change of variables, which comes from rewriting the equation in a moving coordinate frame. This naturally leads to the fundamental concept of characteristic curve, along which signals and physical disturbances propagate. The resulting method of characteristics is able to solve a first-order *linear* partial differential equation by reducing it to one or more first-order *nonlinear* ordinary differential equations.

Proceeding to the nonlinear regime, the most important new phenomenon is the possible breakdown of solutions in finite time, resulting in the formation of discontinuous shock waves. A familiar example is the supersonic boom produced by an airplane that breaks the sound barrier. Signals continue to propagate along characteristic curves, but now the curves may cross each other, precipitating the onset of a shock discontinuity. The ensuing shock dynamics is *not* uniquely specified by the partial differential equation, but relies on additional physical properties, to be specified by an appropriate conservation law along with a causality condition. A full-fledged analysis of shock dynamics becomes quite challenging, and only the basics will be developed here.

Having attained a basic understanding of first-order wave dynamics, we then focus our attention on the first of three paradigmatic second-order partial differential equations, known as the wave equation, which is used to model waves and vibrations in an elastic bar, a violin string, or a column of air in a wind instrument. Its multi-dimensional versions serve to model vibrations of membranes, solid bodies, water waves, electromagnetic waves, including light, radio waves, microwaves, acoustic waves, and many other physical phenomena. The one-dimensional wave equation is one of a small handful of physically relevant partial differential equations that has an explicit solution formula, originally discovered by the eighteenth-century French mathematician (and encyclopedist) Jean d'Alembert. His solution is the result of being able to "factorize" the second-order wave equation into a pair of first-order partial differential equations, of a type solved in the first part of this

P.J. Olver, *Introduction to Partial Differential Equations*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-02099-0_2, © Springer International Publishing Switzerland 2014

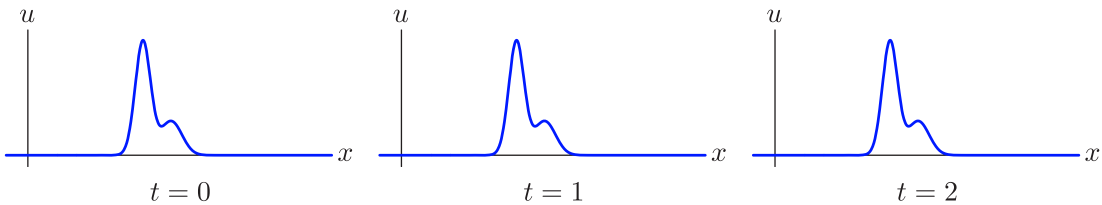

Figure 2.1. Stationary wave.

chapter. We investigate the consequences of d'Alembert's solution formula for the initial value problem on the entire real line; solutions on bounded intervals will be deferred until Chapter 4. Unfortunately, d'Alembert's method is of rather limited scope, and does not extend beyond the one-dimensional case, nor to equations modeling vibrations of nonuniform media. The analysis of the wave equation in more than one space dimension can be found in Chapters 11 and 12.

2.1 Stationary Waves

When entering a new mathematical subject — in our case, partial differential equations — one should first analyze and fully understand the very simplest examples. Indeed, mathematics is, at its core, a bootstrapping enterprise, in which one builds on one's knowledge of and experience with elementary topics — in the present case, ordinary differential equations — to make progress, first with the simpler types of partial differential equations, and then, by developing and applying each newly gained insight and technique, to more and more complicated situations.

The simplest partial differential equation, for a function $u(t,x)$ of two variables, is

$$\frac{\partial u}{\partial t} = 0. \tag{2.1}$$

It is a first-order, homogeneous, linear equation. If (2.1) were an ordinary differential equation[†] for a function $u(t)$ of t alone, the solution would be obvious: $u(t) = c$ must be constant. A proof of this basic fact proceeds by integrating both sides with respect to t and then appealing to the Fundamental Theorem of Calculus. To solve (2.1) as a partial differential equation for $u(t,x)$, let us similarly integrate both sides of the equation from, say, 0 to t, producing

$$0 = \int_0^t \frac{\partial u}{\partial t}(s,x)\,ds = u(t,x) - u(0,x).$$

Therefore, the solution takes the form

$$u(t,x) = f(x), \qquad \text{where} \qquad f(x) = u(0,x), \tag{2.2}$$

and hence is a function of the space variable x alone. The only requirement is that $f(x)$ be continuously differentiable, so $f \in \mathrm{C}^1$, in order that $u(t,x)$ be a bona fide classical

[†] Of course, in this situation, we would write the equation as $du/dt = 0$.

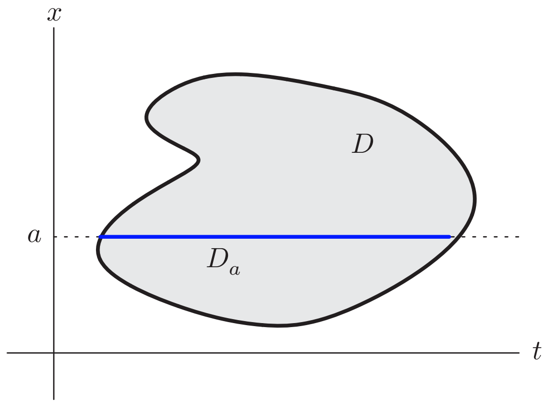

Figure 2.2. Domain for stationary-wave solution.

solution of the first-order partial differential equation (2.1). The solution (2.2) represents a *stationary wave*, meaning that it does not change in time. The initial profile stays frozen in place, and the system remains in equilibrium. Figure 2.1 plots a representative solution as a function of x at three successive times.

The preceding analysis seems very straightforward and perhaps even a little boring. But, to be completely rigorous, we need to take a bit more care. In our derivation, we implicitly assumed that the solution $u(t,x)$ was defined everywhere on $\mathbb{R}^2$. And, in fact, the solution formula (2.2) is *not* completely valid as stated if the solution $u(t,x)$ is defined only on a subdomain $D \subset \mathbb{R}^2$.

Indeed, a solution $u(t)$ to the corresponding ordinary differential equation $du/dt = 0$ is constant, *provided it is defined on a connected subinterval* $I \subset \mathbb{R}$. A solution that is defined on a disconnected subset $D \subset \mathbb{R}$ need only be constant on each connected subinterval $I \subset D$. For instance, the nonconstant function

$$u(t) = \begin{cases} 1, & t > 0, \\ -1, & t < 0, \end{cases} \qquad \text{satisfies} \qquad \frac{du}{dt} = 0$$

everywhere on its domain of definition, that is, $D = \{\, t \neq 0 \,\}$, but is constant only on the connected positive and negative half-lines.

Similar counterexamples can be constructed in the case of the partial differential equation (2.1). If the domain of definition is disconnected, then we do not expect $u(t,x)$ to depend only on x if we move from one connected component of D to another. Even that is not the full story. For example, the function

$$u(t,x) = \begin{cases} 0, & x > 0, \\ x^2, & x \leq 0, \quad t > 0, \\ -x^2, & x \leq 0, \quad t < 0, \end{cases} \tag{2.3}$$

is continuously differentiable† on its domain of definition, namely $D = \mathbb{R}^2 \setminus \{\, (0,x) \mid x \leq 0 \,\}$, satisfies $\partial u/\partial t = 0$ everywhere in D, but, nevertheless, is not a function of x alone, because, for example, $u(1,x) = x^2 \neq u(-1,x) = -x^2$.

† You are asked to rigorously prove differentiability in Exercise 2.1.10.

A completely correct formulation can be stated as follows: If $u(t,x)$ is a classical solution to (2.1), defined on a domain $D \subset \mathbb{R}^2$ whose intersection with any horizontal‡ line, namely $D_a = D \cap \{ (t,a) \mid t \in \mathbb{R} \}$, for each fixed $a \in \mathbb{R}$, is either empty or a connected interval, then $u(t,x) = f(x)$ is a function of x alone. An example of such a domain is sketched in Figure 2.2. In Exercise 2.1.9, you are asked to justify these statements.

We are thus slightly chastened in our dismissal of (2.1) as a complete triviality. The lesson is that, in future, one must *always* be careful when interpreting such "general" solution formulas — since they often rely on unstated assumptions on their underlying domain of definition.

Exercises

2.1.1. Solve the partial differential equation $\dfrac{\partial u}{\partial t} = x$ for $u(t,x)$.

2.1.2. Solve the partial differential equation $\dfrac{\partial^2 u}{\partial t^2} = 0$ for $u(t,x)$.

2.1.3. Find the general solution $u(t,x)$ to the following partial differential equations: (*a*) $u_x = 0$, (*b*) $u_t = 1$, (*c*) $u_t = x - t$, (*d*) $u_t + 3u = 0$, (*e*) $u_x + t\,u = 0$, (*f*) $u_{tt} + 4u = 1$.

2.1.4. Suppose $u(t,x)$ is defined for all $(t,x) \in \mathbb{R}^2$ and solves $\partial u/\partial t + 2u = 0$. Prove that $\lim_{t \to \infty} u(t,x) = 0$ for all x.

2.1.5. Write down the general solution to the partial differential equation $\partial u/\partial t = 0$ for a function of three variables $u(t,x,y)$. What assumptions should be made on the domain of definition for your solution formula to be valid?

2.1.6. Solve the partial differential equation $\dfrac{\partial^2 u}{\partial x\,\partial y} = 0$ for $u(x,y)$.

2.1.7. Answer Exercise 2.1.6 when $u(x,y,z)$ depends on the three independent variables x, y, z.

♡ 2.1.8. Let $u(t,x)$ solve the initial value problem $\dfrac{\partial u}{\partial t} + u^2 = 0$, $u(0,x) = f(x)$, where $f(x)$ is a bounded C^1 function of $x \in \mathbb{R}$. (a) Show that if $f(x) \geq 0$ for all x, then $u(t,x)$ is defined for all $t > 0$, and $\lim_{t \to \infty} u(t,x) = 0$. (*b*) On the other hand, if $f(x) < 0$, then the solution $u(t,x)$ is not defined for all $t > 0$, but in fact, $\lim_{t \to \tau^-} u(t,x) = -\infty$ for some $0 < \tau < \infty$. Given x, what is the corresponding value of τ? (*c*) Given $f(x)$ as in part (*b*), what is the longest time interval $0 < t < t_\star$ on which $u(t,x)$ is defined for all $x \in \mathbb{R}$?

♢ 2.1.9. Justify the claim in the text that if $u(t,x)$ is a solution of $\partial u/\partial t = 0$ that is defined on a domain $D \subset \mathbb{R}^2$ with the property that $D_a = D \cap \{ (t,a) \mid t \in \mathbb{R} \}$ is either empty or a connected interval, then $u(t,x) = v(x)$ depends only on $x \in D$.

♢ 2.1.10. Prove that the function in (2.3) is continuously differentiable at all points (t,x) in its domain of definition.

‡ *Important*: We will adopt the (slightly unusual) convention of displaying the (t,x)–plane with time t along the horizontal axis and space x along the vertical axis — which also conforms with our convention of writing t before x in expressions like $u(t,x)$. Later developments will amply vindicate our adoption of this convention.

2.2 Transport and Traveling Waves

In many respects, the stationary-wave equation (2.1) does not quite qualify as a partial differential equation. Indeed, the spatial variable x enters only parametrically in the solution to what is, in essence (ignoring technical difficulties with domains), a very simple ordinary differential equation.

Let us then turn to a more "genuine" example. Consider the linear, homogeneous first-order partial differential equation

$$\frac{\partial u}{\partial t} + c\,\frac{\partial u}{\partial x} = 0, \tag{2.4}$$

for a function $u(t,x)$, in which c is a fixed, nonzero constant, known as the *wave speed* for reasons that will soon become apparent. We will refer to (2.4) as the *transport equation*, because it models the transport of a substance, e.g., a pollutant, in a uniform fluid flow that is moving with velocity c. In this model, the solution $u(t,x)$ represents the concentration of the pollutant at time t and spatial position x. Other common names for (2.4) are the *first-order* or *unidirectional wave equation*. But for brevity, as well as to avoid any confusion with the second-order, bidirectional wave equation discussed extensively later on, we will stick with the designation "transport equation" here. Solving the transport equation is slightly more challenging, but, as we will see, not difficult.

Since the transport equation involves time, its solutions are distinguished by their initial values. As a first-order equation, we need only specify the value of the solution at an initial time t_0, leading to the initial value problem

$$u(t_0,x) = f(x) \qquad \text{for all} \qquad x \in \mathbb{R}. \tag{2.5}$$

As we will show, as long as $f \in \mathrm{C}^1$, i.e., is continuously differentiable, the initial conditions serve to specify a unique classical solution. Also, by replacing the time variable t by $t - t_0$, we can, without loss of generality, set $t_0 = 0$.

Uniform Transport

Let us begin by assuming that the wave speed c is constant. In general, when one is confronted with a new equation, one solution strategy is to try to convert it into an equation that you already know how to solve. In this case, we will introduce a simple change of variables that effectively rewrites the equation in a moving coordinate system, inspired by the interpretation of c as the overall transport speed.

If x represents the position of an object in a fixed coordinate frame, then

$$\xi = x - ct \tag{2.6}$$

represents the object's position relative to an observer who is uniformly moving with velocity c. Think of a passenger in a moving train to whom stationary objects appear to be moving *backwards* at the train's speed c. To formulate a physical process in the reference frame of the passenger, we replace the stationary space-time coordinates (t,x) by the moving coordinates (t,ξ).

Remark: These are the same changes of reference frame that underlie Einstein's special theory of relativity. However, unlike Einstein, we are working in a purely classical,

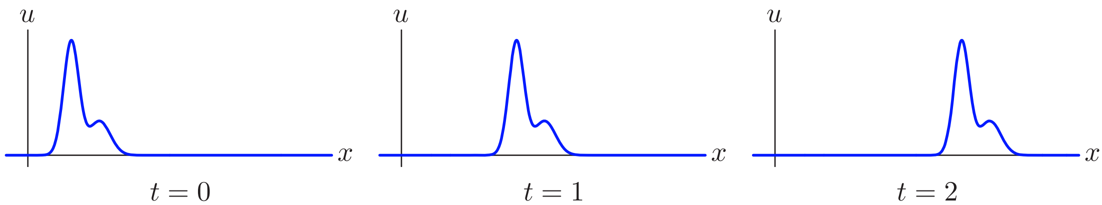

Figure 2.3. Traveling wave with $c > 0$.

nonrelativistic universe here. Such changes to moving coordinates are, in fact, of a much older vintage, and named *Galilean boosts* in honor of Galileo Galilei, who was the first to champion such "relativistic" moving coordinate systems.

Let us see what happens when we re-express the transport equation in terms of the moving coordinate frame. We rewrite

$$u(t,x) = v(t, x - ct) = v(t,\xi) \tag{2.7}$$

in terms of the *characteristic variable* $\xi = x - ct$, along with the time t. To write out the differential equation satisfied by $v(t,\xi)$, we apply the chain rule from multivariable calculus, [**8**, **108**], to express the derivatives of u in terms of those of v:

$$\frac{\partial u}{\partial t} = \frac{\partial v}{\partial t} - c\,\frac{\partial v}{\partial \xi}\,, \qquad\qquad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial \xi}\,.$$

Therefore,

$$\frac{\partial u}{\partial t} + c\,\frac{\partial u}{\partial x} = \frac{\partial v}{\partial t} - c\,\frac{\partial v}{\partial \xi} + c\,\frac{\partial v}{\partial \xi} = \frac{\partial v}{\partial t}\,. \tag{2.8}$$

We deduce that $u(t,x)$ solves the transport equation (2.4) if and only if $v(t,\xi)$ solves the stationary-wave equation

$$\frac{\partial v}{\partial t} = 0. \tag{2.9}$$

Thus, the effect of using a moving coordinate system is to convert a wave moving with velocity c into a stationary wave. Think again of the passenger in the train — a second train moving at the same speed appears as if it were stationary.

According to our earlier discussion, the solution $v = v(\xi)$ to the stationary-wave equation (2.9) is a function of the characteristic variable alone. (For simplicity, we assume that $v(t,\xi)$ has an appropriate domain of definition, e.g., it is defined everywhere on $\mathbb{R}^2$.) Recalling (2.7), we conclude that the solution

$$u = v(\xi) = v(x - ct)$$

to the transport equation must be a function of the characteristic variable only. We have therefore proved the following result:

Proposition 2.1. *If $u(t,x)$ is a solution to the partial differential equation*

$$u_t + c\,u_x = 0, \tag{2.10}$$

which is defined on all of $\mathbb{R}^2$, then

$$u(t,x) = v(x - ct), \tag{2.11}$$

where $v(\xi)$ is a C^1 function of the characteristic variable $\xi = x - ct$.

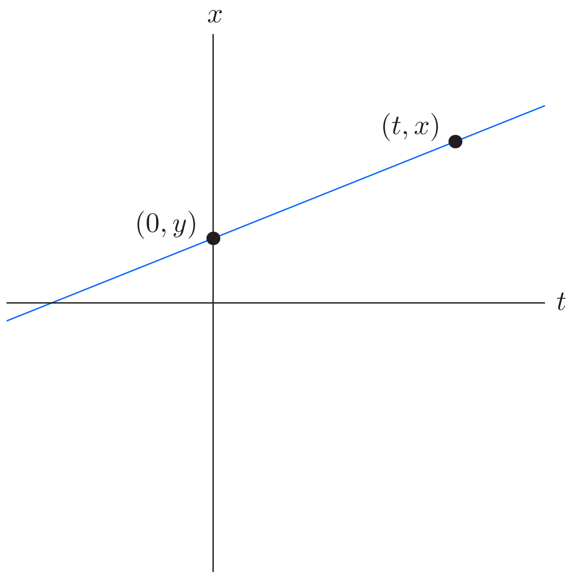

Figure 2.4. Characteristic line.

In other words, *any* (reasonable) function of the characteristic variable, e.g., $\xi^2 + 1$, or $\cos \xi$, or e^{ξ}, will produce a corresponding solution, $(x - ct)^2 + 1$, or $\cos(x - ct)$, or e^{x-ct}, to the transport equation with constant wave speed c. And, in accordance with the counting principle formulated on page 6, the general solution to this first-order partial differential equation in two independent variables depends on one arbitrary function of a single variable.

To a stationary observer, the solution (2.11) appears as a *traveling wave* of unchanging form moving at constant velocity c. When $c > 0$, the wave translates to the right, as illustrated in Figure 2.3. When $c < 0$, the wave translates to the left, while $c = 0$ corresponds to a stationary wave form that remains fixed at its original location, as in Figure 2.1.

At $t = 0$, the wave has the initial profile

$$u(0, x) = v(x), \tag{2.12}$$

and so (2.11) provides the (unique) solution to the initial value problem (2.4, 12). For example, the solution to the particular initial value problem

$$u_t + 2\,u_x = 0, \qquad u(0, x) = \frac{1}{1 + x^2}\,, \qquad \text{is} \qquad u(t, x) = \frac{1}{1 + (x - 2\,t)^2}\,.$$

Since it depends only on the characteristic variable $\xi = x - ct$, every solution to the transport equation is constant on the *characteristic lines* of slope† c, namely

$$x = ct + k, \tag{2.13}$$

where k is an arbitrary constant. At any given time t, the value of the solution at position x depends only on its original value on the characteristic line passing through (t, x).

† This makes use of our convention that the t–axis is horizontal and the x–axis is vertical. Reversing the axes will replace the slope by its reciprocal.

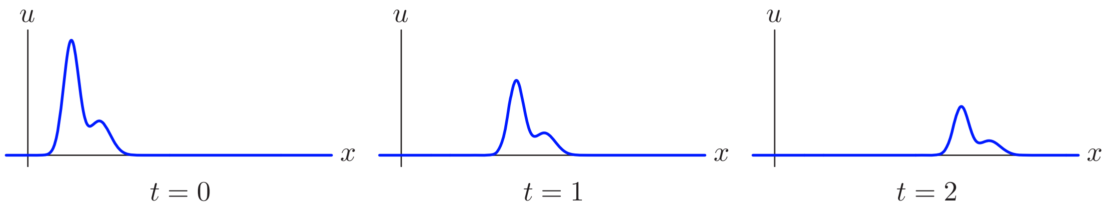

Figure 2.5. Decaying traveling wave.

This is indicative of a general fact concerning such wave models: *Signals propagate along characteristics*. Indeed, a disturbance at an initial point $(0, y)$ only affects the value of the solution at points (t, x) that lie on the characteristic line $x = c\,t + y$ emanating therefrom, as illustrated in Figure 2.4.

Transport with Decay

Let $a > 0$ be a positive constant, and c an arbitrary constant. The homogeneous linear first-order partial differential equation

$$\frac{\partial u}{\partial t} + c\,\frac{\partial u}{\partial x} + a\,u = 0 \tag{2.14}$$

models the transport of, say, a radioactively decaying solute in a uniform fluid flow with wave speed c. The coefficient a governs the rate of decay. We can solve this variant of the transport equation by the self-same change of variables to a uniformly moving coordinate system.

Rewriting $u(t, x)$ in terms of the characteristic variable, as in (2.7), and then recalling our chain rule calculation (2.8), we find that $v(t, \xi) = u(t, \xi + c\,t)$ satisfies the partial differential equation

$$\frac{\partial v}{\partial t} + a\,v = 0.$$

The result is, effectively, a homogeneous linear first-order ordinary differential equation, in which the characteristic variable ξ enters only parametrically. The standard solution technique learned in elementary ordinary differential equations, [**20**, **23**], tells us to multiply the equation by the exponential *integrating factor* $e^{a\,t}$, leading to

$$e^{a\,t}\left(\frac{\partial v}{\partial t} + a\,v\right) = \frac{\partial}{\partial t}\,(e^{a\,t} v) = 0.$$

We conclude that $w = e^{a\,t} v$ solves the stationary-wave equation (2.1). Thus,

$$w = e^{a\,t} v = f(\xi), \qquad \text{and hence} \qquad v(t, \xi) = f(\xi)\,e^{-a\,t},$$

where $f(\xi)$ is an arbitrary function of the characteristic variable. Reverting to physical coordinates, we produce the solution formula

$$u(t, x) = f(x - c\,t)\,e^{-a\,t}, \tag{2.15}$$

which solves the initial value problem $u(0, x) = f(x)$. It represents a wave that is moving along with fixed velocity c while simultaneously decaying at an exponential rate as prescribed by the coefficient $a > 0$. A typical solution, for $c > 0$, is plotted at three successive times in Figure 2.5. While the solution (2.15) is no longer constant on the characteristics, signals continue to propagate along them, since a solution's initial value at a point $(0, y)$ will only affect its subsequent (decaying) values on the associated characteristic line $x = ct + y$.

Exercises

2.2.1. Find the solution to the initial value problem $u_t + u_x = 0$, $u(1, x) = x/(1 + x^2)$.

2.2.2. Solve the following initial value problems and graph the solutions at times $t = 1, 2$, and 3:
(*a*) $u_t - 3u_x = 0$, $u(0, x) = e^{-x^2}$; (*b*) $u_t + 2u_x = 0$, $u(-1, x) = x/(1 + x^2)$;
(*c*) $u_t + u_x + \frac{1}{2}u = 0$, $u(0, x) = \tan^{-1} x$; (*d*) $u_t - 4u_x + u = 0$, $u(0, x) = 1/(1 + x^2)$.

2.2.3. Graph some of the characteristic lines for the following equations, and write down a formula for the general solution:
(*a*) $u_t - 3u_x = 0$, (*b*) $u_t + 5u_x = 0$, (*c*) $u_t + u_x + 3u = 0$, (*d*) $u_t - 4u_x + u = 0$.

2.2.4. Solve the initial value problem $u_t + 2u_x = 1$, $u(0, x) = e^{-x^2}$.

Hint: Use characteristic coordinates.

2.2.5. Answer Exercise 2.2.4 for the initial value problem $u_t + 2u_x = \sin x$, $u(0, x) = \sin x$.

♢ 2.2.6. Let c be constant. Suppose that $u(t, x)$ solves the initial value problem $u_t + cu_x = 0$, $u(0, x) = f(x)$. Prove that $v(t, x) = u(t - t_0, x)$ solves the initial value problem $v_t + cv_x = 0$, $v(t_0, x) = f(x)$.

2.2.7. Is Exercise 2.2.6 valid when the transport equation is replaced by the damped transport equation (2.14)?

2.2.8. Let $c \neq 0$. Prove that if the initial data satisfies $u(0, x) = v(x) \to 0$ as $x \to \pm\infty$, then, for each fixed x, the solution to the transport equation (2.4) satisfies $u(t, x) \to 0$ as $t \to \infty$.

2.2.9. (*a*) Prove that if the initial data is bounded, $|\, f(x)\,| \leq M$ for all $x \in \mathbb{R}$, then the solution to the damped transport equation (2.14) with $a > 0$ satisfies $u(t, x) \to 0$ as $t \to \infty$. (*b*) Find a solution to (2.14) that is defined for all (t, x) but does not satisfy $u(t, x) \to 0$ as $t \to \infty$.

2.2.10. Let $F(t, x)$ be a C^1 function of $(t, x) \in \mathbb{R}^2$. (*a*) Write down a formula for the general solution $u(t, x)$ to the inhomogeneous partial differential equation $u_t = F(t, x)$.
(*b*) Solve the inhomogeneous transport equation $u_t + c\,u_x = F(t, x)$.

♡ 2.2.11. (*a*) Write down a formula for the general solution to the nonlinear partial differential equation $u_t + u_x + u^2 = 0$. (*b*) Show that if the initial data is positive and bounded, $0 \leq u(0, x) = f(x) \leq M$, then the solution exists for all $t > 0$, and $u(t, x) \to 0$ as $t \to \infty$. (*c*) On the other hand, if the initial data is negative somewhere, so $f(x) < 0$ at some $x \in \mathbb{R}$, then the solution *blows up* in finite time: $\lim_{t \to \tau^-} u(t, y) = -\infty$ for some $\tau > 0$ and some $y \in \mathbb{R}$. (*d*) Find a formula for the earliest blow-up time $\tau_\star > 0$.

2.2.12. A sensor situated at position $x = 1$ monitors the concentration of a pollutant $u(t, 1)$ as a function of t for $t \geq 0$. Assuming that the pollutant is transported with wave speed $c = 3$, at what locations x can you determine the initial concentration $u(0, x)$?

2.2.13. Write down a solution to the transport equation $u_t + 2u_x = 0$ that is defined on a connected domain $D \subset \mathbb{R}^2$ and that is *not* a function of the characteristic variable alone.

2.2.14. Let $c > 0$. Consider the uniform transport equation $u_t + c u_x = 0$ restricted to the quarter-plane $Q = \{ x > 0, \ t > 0 \}$ and subject to initial conditions $u(0,x) = f(x)$ for $x \geq 0$, along with boundary conditions $u(t,0) = g(t)$ for $t \geq 0$. (*a*) For which initial and boundary conditions does a classical solution to this initial-boundary value problem exist? Write down a formula for the solution. (*b*) On which regions are the effects of the initial conditions felt? What about the boundary conditions? Is there any interaction between the two?

2.2.15. Answer Exercise 2.2.14 when $c < 0$.

Nonuniform Transport

Slightly more complicated, but still linear, is the *nonuniform transport equation*

$$\frac{\partial u}{\partial t} + c(x)\,\frac{\partial u}{\partial x} = 0, \tag{2.16}$$

where the wave speed $c(x)$ is now allowed to depend on the spatial position. Characteristics continue to guide the behavior of solutions, but when the wave speed is not constant, we can no longer expect them to be straight lines. To adapt the method of characteristics, let us look at how the solution varies along a prescribed curve in the (t,x)–plane. Assume that the curve is identified with the graph of a function $x = x(t)$, and let

$$h(t) = u\big(t, x(t)\big)$$

be the value of the solution on it. We compute the rate of change in the solution along the curve by differentiating h with respect to t. Invoking the multivariable chain rule, we obtain

$$\frac{dh}{dt} = \frac{d}{dt}\,u\big(t,x(t)\big) = \frac{\partial u}{\partial t}\big(t,x(t)\big) + \frac{\partial u}{\partial x}\big(t,x(t)\big)\,\frac{dx}{dt}\,. \tag{2.17}$$

In particular, if $x(t)$ satisfies

$$\frac{dx}{dt} = c\big(x(t)\big), \qquad \text{then} \qquad \frac{dh}{dt} = \frac{\partial u}{\partial t}\big(t,x(t)\big) + c\big(x(t)\big)\,\frac{\partial u}{\partial x}\big(t,x(t)\big) = 0,$$

since we are assuming that $u(t,x)$ solves the transport equation (2.16) for all values of (t,x), including those points $\big(t,x(t)\big)$ on the curve. Since its derivative is zero, $h(t)$ must be a constant, which motivates the following definition.

Definition 2.2. The graph of a solution $x(t)$ to the autonomous ordinary differential equation

$$\frac{dx}{dt} = c(x) \tag{2.18}$$

is called a *characteristic curve* for the transport equation with wave speed $c(x)$.

In other words, at each point (t,x), the slope of the characteristic curve equals the wave speed $c(x)$ there. In particular, if c is constant, the characteristic curves are straight lines of slope c, in accordance with our earlier construction.

Proposition 2.3. *Solutions to the linear transport equation* (2.16) *are constant along characteristic curves.*

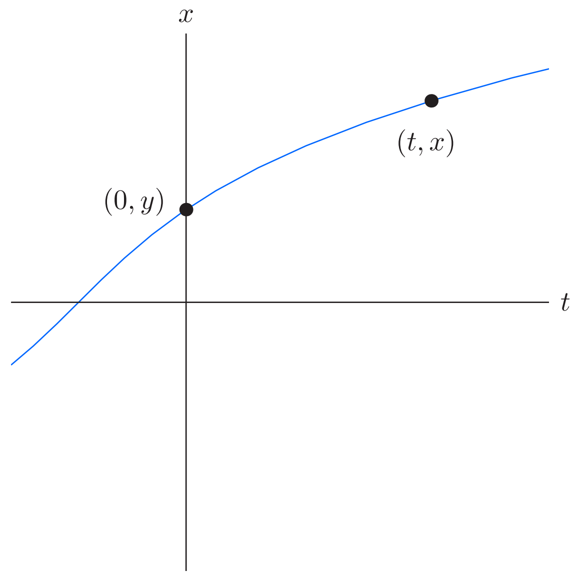

Figure 2.6. Characteristic curve.

The characteristic curve equation (2.18) is an autonomous first-order ordinary differential equation. As such, it can be immediately solved by separating variables, [**20**, **23**]. Assuming $c(x) \neq 0$, we divide both sides of the equation by $c(x)$, and then integrate the resulting equation:

$$\frac{dx}{c(x)} = dt, \qquad \text{whereby} \qquad \beta(x) := \int \frac{dx}{c(x)} = t + k, \tag{2.19}$$

with k denoting the integration constant. For each fixed value of k, (2.19) serves to implicitly define a characteristic curve, namely,

$$x(t) = \beta^{-1}(t + k),$$

with β^{-1} denoting the inverse function. On the other hand, if $c(x_\star) = 0$, then $x_\star$ is a *fixed point* for the ordinary differential equation (2.18), and the horizontal line $x \equiv x_\star$ is a stationary characteristic curve.

Since the solution $u(t, x)$ is constant along the characteristic curves, it must therefore be a function of the *characteristic variable*

$$\xi = \beta(x) - t \tag{2.20}$$

alone, and hence of the form

$$u(t, x) = v\big(\beta(x) - t\big), \tag{2.21}$$

where $v(\xi)$ is an arbitrary C^1 function. Indeed, it is easy to check directly that, provided $\beta(x)$ is defined by (2.19), $u(t, x)$ solves the partial differential equation (2.16) for *any* choice of C^1 function $v(\xi)$. (But keep in mind that the algebraic solution formula (2.21) may fail to be valid at points where the wave speed vanishes: $c(x_\star) = 0$.)

Warning: The definition of characteristic variable used here is slightly different from that in the constant wave speed case, which, by (2.20), would be $\xi = x/c - t = (x - ct)/c$. Clearly, rescaling the characteristic variable by $1/c$ is an inessential modification of our original definition.

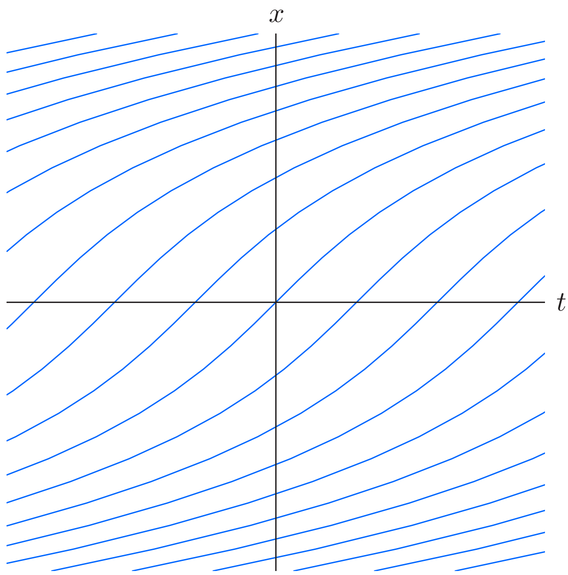

Figure 2.7. Characteristic curves for $u_t + (x^2+1)^{-1} u_x = 0$.

To find the solution that satisfies the prescribed initial conditions

$$u(0,x) = f(x), \tag{2.22}$$

we merely substitute the general solution formula (2.21). This leads to the implicit equation $v(\beta(x)) = f(x)$ for the function $v(\xi) = f \circ \beta^{-1}(\xi)$. The resulting solution formula

$$u(t,x) = f \circ \beta^{-1}\big(\beta(x) - t\big) \tag{2.23}$$

is not particularly enlightening, but it does have a simple graphical interpretation: To find the value of the solution $u(t,x)$, we look at the characteristic curve passing through the point (t,x). If this curve intersects the x–axis at the point $(0,y)$, as in Figure 2.6, then $u(t,x) = u(0,y) = f(y)$, since the solution must be constant along the curve. On the other hand, if the characteristic curve through (t,x) doesn't intersect the x–axis, the solution value $u(t,x)$ is *not* prescribed by the initial data.

Example 2.4. Let us solve the nonuniform transport equation

$$\frac{\partial u}{\partial t} + \frac{1}{x^2+1}\,\frac{\partial u}{\partial x} = 0 \tag{2.24}$$

by the method of characteristics. According to (2.18), the characteristic curves are the graphs of solutions to the first-order ordinary differential equation

$$\frac{dx}{dt} = \frac{1}{x^2+1}\,.$$

Separating variables and integrating, we obtain

$$\beta(x) = \int (x^2+1)\,dx = \tfrac{1}{3}x^3 + x = t + k, \tag{2.25}$$

where k is the integration constant. Representative curves are plotted in Figure 2.7. (In this case, inverting the function β, i.e., solving (2.25) for x as a function of t, is not particularly enlightening.)

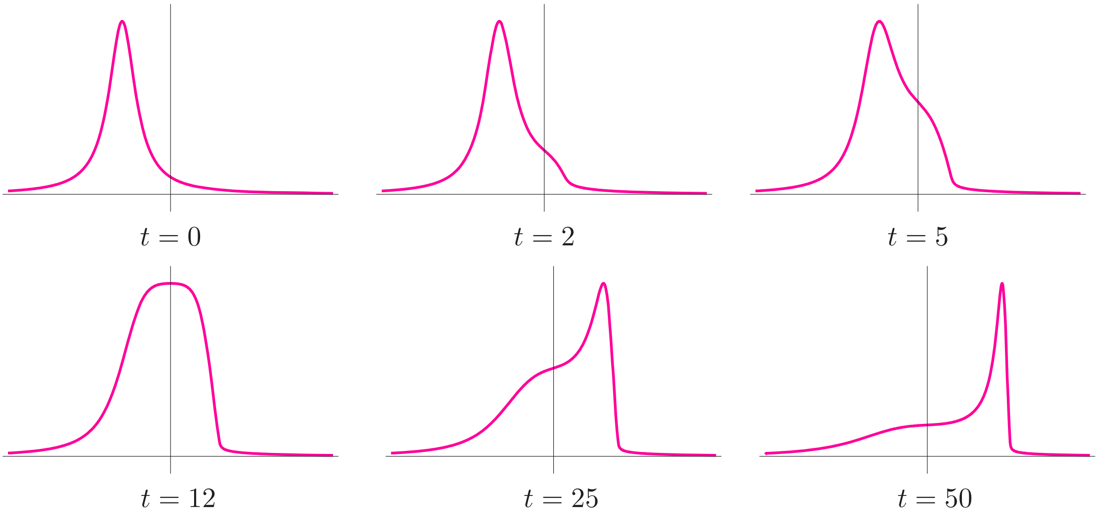

Figure 2.8. Solution to $u_t + \dfrac{1}{x^2+1}\, u_x = 0.$ ⊎

According to (2.20), the characteristic variable is $\xi = \frac{1}{3}x^3 + x - t$, and hence the general solution to the equation takes the form

$$u = v\left(\tfrac{1}{3}x^3 + x - t\right), \tag{2.26}$$

where $v(\xi)$ is an arbitrary C^1 function. A typical solution, corresponding to initial data

$$u(0,x) = \frac{1}{1+(x+3)^2}\,, \tag{2.27}$$

is plotted[†] at the indicated times in Figure 2.8. Although the solution remains constant along each individual curve, a stationary observer will witness a dynamically changing profile as the wave moves through the nonuniform medium. In this example, since $c(x) > 0$ everywhere, the wave always moves from left to right; its speed as it passes through a point x determined by the magnitude of $c(x) = (x^2+1)^{-1}$, with the consequence that each part accelerates as it approaches the origin from the left, and then slows back down once it passes by and $c(x)$ decreases in magnitude. To a stationary observer, the wave spreads out as it speeds through the origin, and then becomes progressively narrower and slower as it gradually moves off to $+\infty$.

Example 2.5. Consider the nonuniform transport equation

$$u_t + (x^2 - 1)\, u_x = 0. \tag{2.28}$$

[†] The required function $v(\xi)$ in (2.26) is implicitly given by the equation $v\left(\frac{1}{3}x^3 + x\right) = u(0,x)$, and so the explicit formula for $u(t,x)$ is not very instructive or useful. Indeed, to make the plots, we instead sampled the initial data (2.27) at a collection of uniformly spaced points $y_1 < y_2 < \cdots < y_n$. Since the solution is constant along the characteristic curve (2.25) passing through each sample point $(0, y_i)$, we can find nonuniformly spaced sample values for $u(t, x_i)$ at any later time. The smooth solution curve $u(t,x)$ is then approximated using spline interpolation, [**89**; §11.4], on these sample values.

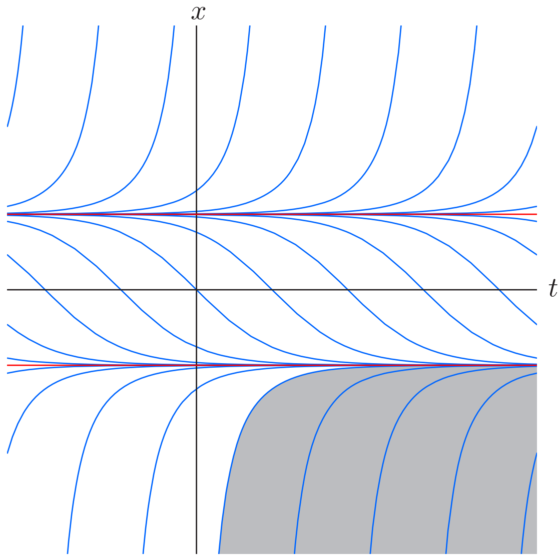

Figure 2.9. Characteristic curves for $u_t + (x^2 - 1)u_x = 0$.

In this case, the characteristic curves are the solutions to

$$\frac{dx}{dt} = x^2 - 1,$$

and so

$$\beta(x) = \int \frac{dx}{x^2 - 1} = \frac{1}{2}\,\log\left|\frac{x-1}{x+1}\right| = t + k. \tag{2.29}$$

One must also include the horizontal lines $x = x_\pm = \pm 1$ corresponding to the roots of $c(x) = x^2 - 1$. The curves are graphed in Figure 2.9. Note that those curves starting below $x_+ = 1$ converge to $x_- = -1$ as $t \to \infty$, while those starting above $x_+ = 1$ veer off to ∞ in finite time. Owing to the sign of $c(x) = x^2 - 1$, points on the graph of $u(0, x)$ lying over $|\,x\,| < 1$ will move to the left, while those over $|\,x\,| > 1$ will move to the right.

In Figure 2.10, we graph several snapshots of the solution whose initial value is a bell-shaped Gaussian profile

$$u(0, x) = e^{-x^2}.$$

The initial conditions uniquely prescribe the value of the solution along the characteristic curves that intersect the x–axis. On the other hand, if

$$x \le \frac{1 + e^{2t}}{1 - e^{2t}} \qquad \text{for} \qquad t > 0,$$

the characteristic curve through (t, x) does not intersect the x–axis, and hence the value of the solution at such points, lying in the shaded region in Figure 2.9, is *not* prescribed by the initial data. Let us arbitrarily assign the solution to be $u(t, x) = 0$ at such points. At other values of (t, x) with $t \ge 0$, the solution (2.23) is

$$u(t, x) = \exp\left[-\left(\frac{x + 1 + (x-1)\,e^{-2t}}{x + 1 - (x-1)\,e^{-2t}}\right)^2\right]. \tag{2.30}$$

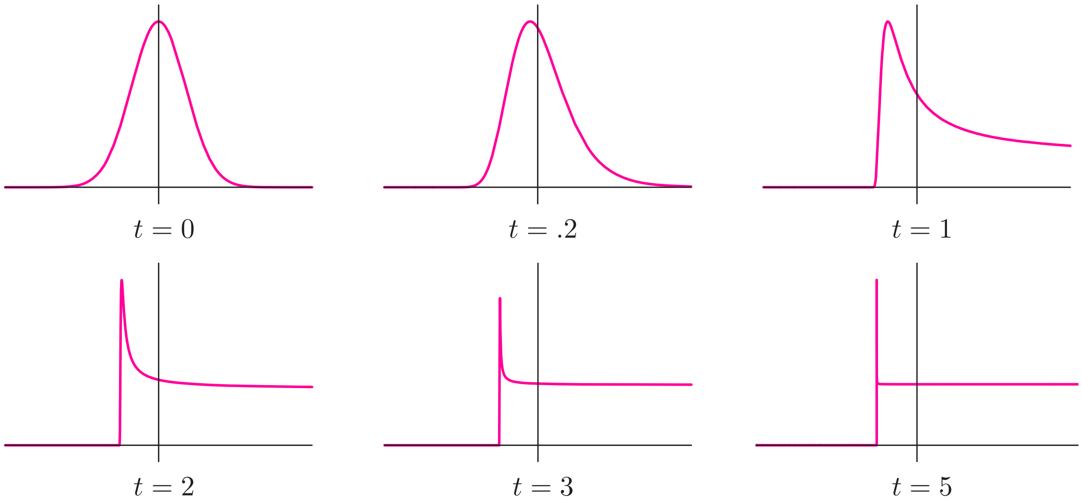

Figure 2.10. Solution to $u_t + (x^2 - 1)u_x = 0$.

(The derivation of this solution formula is left as Exercise 2.2.23.) As t increases, the solution's peak becomes more and more concentrated near $x_- = -1$, while the section of the wave above $x > x_+ = 1$ rapidly spreads out to ∞. In the long term, the solution converges (albeit nonuniformly) to a step function of height $1/e$:

$$u(t,x) \longrightarrow s(x) = \begin{cases} 1/e \approx .367879, & x \geq -1, \\ 0, & x < -1, \end{cases} \qquad \text{as} \qquad t \longrightarrow \infty.$$

Let us finish by making a few general observations concerning the characteristic curves of transport equations whose wave speed $c(x)$ depends only on the position x. Using the basic existence and uniqueness theory for such autonomous ordinary differential equations, [**20**, **23**, **52**], and assuming that $c(x)$ is continuously differentiable:†

- There is a unique characteristic curve passing through each point $(t,x) \in \mathbb{R}^2$.
- Characteristic curves cannot cross each other.
- If $t = \beta(x)$ is a characteristic curve, then so are all its horizontal translates: $t = \beta(x) + k$ for any k.
- Each non-horizontal characteristic curve is the graph of a strictly monotone function. Thus, each point on a wave always moves in the same direction, and can never reverse its direction of propagation.
- As t increases, the characteristic curve either tends to a fixed point, $x(t) \to x_\star$ as $t \to \infty$, with $c(x_\star) = 0$, or goes off to $\pm\infty$ in either finite or infinite time.

Proofs of these statements are assigned to the reader in Exercise 2.2.25.

† For those who know about such things, [**18**, **52**], this assumption can be weakened to just Lipschitz continuity.

Exercises

2.2.16. (a) Find the general solution to the first-order equation $u_t + \frac{3}{2}u_x = 0$.
(b) Find a solution satisfying the initial condition $u(1,x) = \sin x$. Is your solution unique?

2.2.17. (a) Solve the initial value problem $u_t - x\,u_x = 0$, $u(0,x) = (x^2+1)^{-1}$.
(b) Graph the solution at times $t = 0,1,2,3$. (c) What is $\lim_{t\to\infty} u(t,x)$?

2.2.18. Suppose the initial data $u(0,x) = f(x)$ of the nonuniform transport equation (2.28) is continuous and satisfies $f(x) \to 0$ as $|\,x\,| \to \infty$. What is the limiting solution profile $u(t,x)$ as (a) $t \to \infty$? (b) $t \to -\infty$?

♡ 2.2.19. (a) Find and graph the characteristic curves for the equation $u_t + (\sin x)\,u_x = 0$. Suppose you are given initial data (i) $u(0,x) = \left|\cos\frac{1}{2}x\right|$, (ii) $u(0,x) = \cos\frac{1}{2}\pi x$.
(b) Write down a formula for the solution. (c) Graph your solution at times $t = 0,1,2,3,5$, and 10. (d) What is the limiting solution profile as $t \to \infty$?

2.2.20. Consider the linear transport equation $u_t + (1+x^2)\,u_x = 0$. (a) Find and sketch the characteristic curves. (b) Write down a formula for the general solution. (c) Find the solution to the initial value problem $u(0,x) = f(x)$ and discuss its behavior as t increases.

2.2.21. Prove that, for $t \gg 0$, the speed of the wave in Example 2.4 is asymptotically proportional to $t^{-2/3}$.

2.2.22. Verify directly that formula (2.21) defines a solution to the differential equation (2.16).

♢ 2.2.23. Explain how to derive the solution formula (2.30). Justify that it defines a solution to equation (2.28).

2.2.24. Let $c(x)$ be a bounded C^1 function, so $|\,c(x)\,| \le c_\star < \infty$ for all x. Let $f(x)$ be any C^1 function. Prove that the solution $u(t,x)$ to the initial value problem $u_t + c(x)\,u_x = 0$, $u(0,x) = f(x)$, is uniquely defined for all $(t,x) \in \mathbb{R}^2$.

♡ 2.2.25. Suppose that $c(x) \in \mathrm{C}^1$ is continuously differentiable for all $x \in \mathbb{R}$. (a) Prove that the characteristic curves of the transport equation (2.16) cannot cross each other. (b) A point where $c(x_\star) = 0$ is known as a *fixed point* for the characteristic equation $dx/dt = c(x)$. Explain why the characteristic curve passing through a fixed point $(t, x_\star)$ is a horizontal straight line. (c) Prove that if $x = g(t)$ is a characteristic curve, then so are all the horizontally translated curves $x = g(t+\delta)$ for any δ. (d) *True or false*: Every characteristic curve has the form $x = g(t+\delta)$, for some fixed function $g(t)$. (e) Prove that each non-horizontal characteristic curve is the graph $x = g(t)$ of a strictly monotone function. (f) Explain why a wave cannot reverse its direction. (g) Show that a non-horizontal characteristic curve starts, in the distant past, $t \to -\infty$, at either a fixed point or at $-\infty$ and ends, as $t \to +\infty$, at either the next-larger fixed point or at $+\infty$.

♡ 2.2.26. Consider the transport equation $\dfrac{\partial u}{\partial t} + c(t,x)\,\dfrac{\partial u}{\partial x} = 0$ with time-varying wave speed. Define the corresponding characteristic ordinary differential equation to be $\dfrac{dx}{dt} = c(t,x)$, the graphs of whose solutions $x(t)$ are the *characteristic curves*. (a) Prove that any solution $u(t,x)$ to the partial differential equation is constant on each characteristic curve. (b) Suppose that the general solution to the characteristic equation is written in the form $\xi(t,x) = k$, where k is an arbitrary constant. Prove that $\xi(t,x)$ defines a *characteristic variable*, meaning that $u(t,x) = f(\xi(t,x))$ is a solution to the time-varying transport equation for any continuously differentiable scalar function $f \in \mathrm{C}^1$.

2.2.27. (a) Apply the method in Exercise 2.2.26 to find the characteristic curves for the equation $u_t + t^2\,u_x = 0$. (b) Find the solution to the initial value problem $u(0,x) = e^{-x^2}$, and discuss its dynamic behavior.

2.2.28. Solve Exercise 2.2.27 for the equation $u_t + (x - t)\,u_x = 0$.

♡ 2.2.29. Consider the first-order partial differential equation $u_t + (1 - 2t)\,u_x = 0$. Use Exercise 2.2.26 to: (*a*) Find and sketch the characteristic curves. (*b*) Write down the general solution. (*c*) Solve the initial value problem with $u(0,x) = \dfrac{1}{1+x^2}$. (*d*) Describe the behavior of your solution $u(t,x)$ from part (*c*) as $t \to \infty$. What about $t \to -\infty$?

2.2.30. Discuss which of the conclusions of Exercise 2.2.25 are valid for the characteristic curves of the transport equation with time-varying wave speed, as analyzed in Exercise 2.2.26.

♢ 2.2.31. Consider the two-dimensional transport equation $\dfrac{\partial u}{\partial t} + c(x,y)\,\dfrac{\partial u}{\partial x} + d(x,y)\,\dfrac{\partial u}{\partial y} = 0$, whose solution $u(t,x,y)$ depends on time t and space variables x, y. (*a*) Define a characteristic curve, and prove that the solution is constant along it. (*b*) Apply the method of characteristics to solve the initial value problem $u_t + y\,u_x - x\,u_y$, $u(0,x,y) = e^{-(x-1)^2-(y-1)^2}$. (*c*) Describe the behavior of your solution.

2.3 Nonlinear Transport and Shocks

The first-order nonlinear partial differential equation

$$u_t + u\,u_x = 0 \tag{2.31}$$

has the form of a transport equation (2.4), but the wave speed $c = u$ now depends, not on the position x, but rather on the size of the disturbance u. Larger waves will move faster, and overtake smaller, slower-moving waves. Waves of elevation, where $u > 0$, move to the right, while waves of depression, where $u < 0$, move to the left. This equation is considerably more challenging than the linear transport models analyzed above, and was first systematically studied in the early nineteenth century by the influential French mathematician Siméon–Denis Poisson and the great German mathematician Bernhard Riemann.† It and its multi-dimensional and multi-component generalizations play a crucial role in the modeling of gas dynamics, acoustics, shock waves in pipes, flood waves in rivers, chromatography, chemical reactions, traffic flow, and so on. Although we will be able to write down a solution formula, the complete analysis is far from trivial, and will require us to confront the possibility of discontinuous shock waves. Motivated readers are referred to Whitham's book, [**122**], for further details.

Fortunately, the method of characteristics that was developed for linear transport equations also works in the present context and leads to a complete mathematical solution. Mimicking our previous construction, (2.18), but now with wave speed $c = u$, let us define a *characteristic curve* of the nonlinear wave equation (2.31) to be the graph of a solution $x(t)$ to the ordinary differential equation

$$\frac{dx}{dt} = u(t,x). \tag{2.32}$$

† In addition to his fundamental contributions to partial differential equations, complex analysis, and number theory, Riemann also was the inventor of Riemannian geometry, which turned out to be absolutely essential for Einstein's theory of general relativity some 70 years later!

As such, the characteristics depend upon the solution u, which, in turn, is to be specified by its characteristics. We appear to be trapped in a circular argument.

The resolution of the conundrum is to argue that, as in the linear case, the solution $u(t,x)$ remains constant along its characteristics, and this fact will allow us to simultaneously specify both. To prove this claim, suppose that $x = x(t)$ parametrizes a characteristic curve associated with the given solution $u(t,x)$. Our task is to show that $h(t) = u\big(t, x(t)\big)$, which is obtained by evaluating the solution along the curve, is constant, which, as usual, is proved by checking that its derivative is identically zero. Repeating our chain rule computation (2.17), and using (2.32), we deduce that

$$\frac{dh}{dt} = \frac{d}{dt}\,u\big(t,x(t)\big) = \frac{\partial u}{\partial t}\big(t,x(t)\big) + \frac{dx}{dt}\,\frac{\partial u}{\partial x}\big(t,x(t)\big) = \frac{\partial u}{\partial t}\big(t,x(t)\big) + u\big(t,x(t)\big)\,\frac{\partial u}{\partial x}\big(t,x(t)\big) = 0,$$

since u is assumed to solve the nonlinear transport equation (2.31) at all values of (t,x), including those on the characteristic curve. We conclude that $h(t)$ is constant, and hence u is indeed constant on the characteristic curve.

Now comes the clincher. We know that the right-hand side of the characteristic ordinary differential equation (2.32) is a constant whenever $x = x(t)$ defines a characteristic curve. This means that the derivative dx/dt is a constant — namely the fixed value of u on the curve. Therefore, the characteristic curve must be a *straight line*,

$$x = u\,t + k, \tag{2.33}$$

whose slope equals the value assumed by the solution u on it.

And, as before, since the solution is constant along each characteristic line, it must be a function of the *characteristic variable*

$$\xi = x - t\,u \tag{2.34}$$

alone, and so

$$u = f(x - t\,u), \tag{2.35}$$

where $f(\xi)$ is an arbitrary C^1 function. Formula (2.35) should be viewed as an algebraic equation that implicitly defines the solution $u(t,x)$ as a function of t and x. Verification that the resulting function is indeed a solution to (2.31) is the subject of Exercise 2.3.14.

Example 2.6. Suppose that

$$f(\xi) = \alpha\,\xi + \beta,$$

with α, β constant. Then (2.35) becomes

$$u = \alpha(x - t\,u) + \beta, \qquad \text{and hence} \qquad u(t,x) = \frac{\alpha\,x + \beta}{1 + \alpha\,t} \tag{2.36}$$

is the corresponding solution to the nonlinear transport equation. At each fixed t, the graph of the solution is a straight line. If $\alpha > 0$, the solution flattens out: $u(t,x) \to 0$ as $t \to \infty$. On the other hand, if $\alpha < 0$, the straight line rapidly steepens to vertical as t approaches the critical time $t_\star = -1/\alpha$, at which point the solution ceases to exist. Figure 2.11 graphs two representative solutions. The top row shows the solution with $\alpha = 1,\ \beta = .5$, plotted at times $t = 0, 1, 5$, and 20; the bottom row takes $\alpha = -.2,\ \beta = .1$, and plots the solution at times $t = 0, 3, 4$, and 4.9. In the second case, the solution *blows up* by becoming vertical as $t \to 5$.

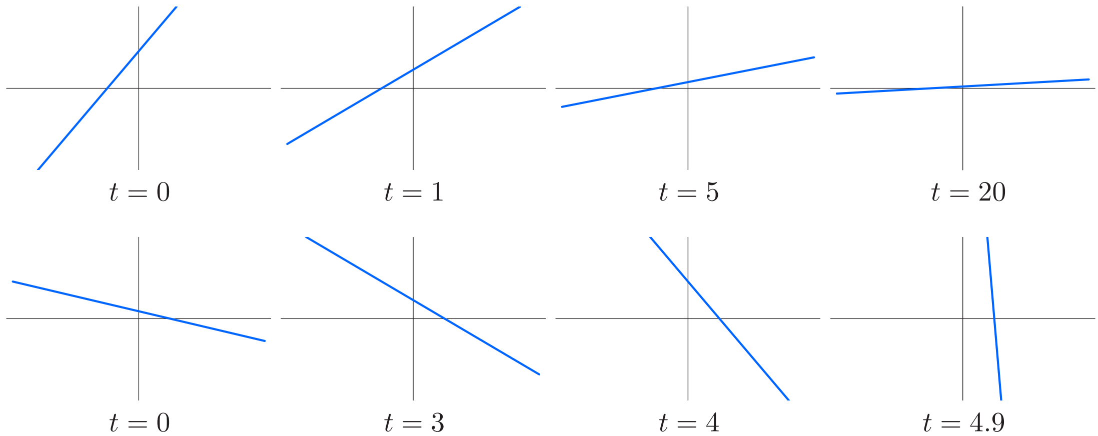

Figure 2.11. Two solutions to $u_t + u\,u_x = 0$.

Remark: Although (2.36) remains a valid solution formula after the blow-up time, $t > 5$, this is *not* to be viewed as a part of the original solution. With the appearance of such a singularity, the physical solution has broken down, and we stop tracking it.

To solve the general initial value problem

$$u(0, x) = f(x), \tag{2.37}$$

we note that, at $t = 0$, the implicit solution formula (2.35) reduces to (2.37), and hence the function f coincides with the initial data. However, because our solution formula (2.35) is an implicit equation, it is not immediately evident

(a) whether it can be solved to give a well-defined function $u(t, x)$, and,

(b) even granted this, how to describe the resulting solution's qualitative features and dynamical behavior.

A more instructive approach is founded on the following geometrical construction. Through each point $(0, y)$ on the x–axis, draw the characteristic line

$$x = t\,f(y) + y \tag{2.38}$$

whose slope, namely $f(y) = u(0, y)$, equals the value of the initial data (2.37) at that point. According to the preceding discussion, the solution will have the same value on the entire characteristic line (2.38), and so

$$u(t, t\,f(y) + y) = f(y) \qquad \text{for all} \;\; t. \tag{2.39}$$

For example, if $f(y) = y$, then $u(t, x) = y$ whenever $x = t\,y + y$; eliminating y, we find $u(t, x) = x/(t + 1)$, which agrees with one of our straight line solutions (2.36).

Now, the problem with this construction is immediately apparent from Figure 2.12, which plots the characteristic lines associated with the initial data

$$u(0, x) = \tfrac{1}{2}\pi - \tan^{-1} x.$$

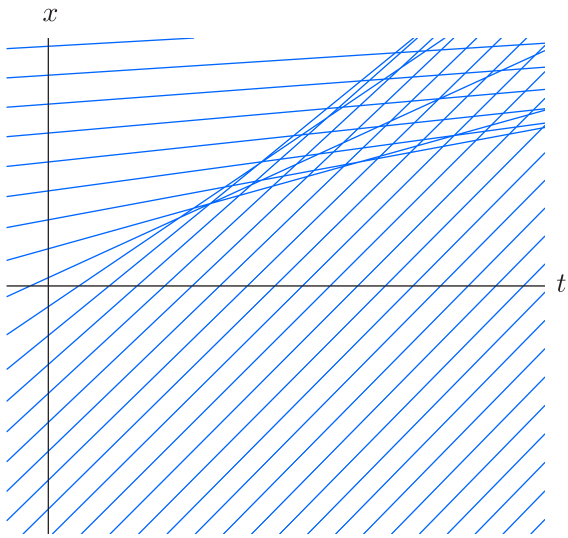

Figure 2.12. Characteristics lines for $u(0,x) = \frac{1}{2}\pi - \tan^{-1} x$.

Two characteristic lines that are not parallel must cross each other somewhere. The value of the solution is supposed to equal the slope of the characteristic line passing through the point. Hence, at a crossing point, the solution is required to assume two *different* values, one corresponding to each line. Something is clearly amiss, and we need to resolve this apparent paradox.

There are three principal scenarios. The first, trivial, situation occurs when all the characteristic lines are parallel, and so the difficulty does not arise. In this case, they all have the same slope, say c, which means that the solution has the same value on each one. Therefore, $u(t,x) \equiv c$ is a constant solution.

The next-simplest case occurs when the initial data is everywhere *nondecreasing*, so $f(x) \leq f(y)$ whenever $x \leq y$, which is assured if its derivative is never negative: $f'(x) \geq 0$. In this case, as sketched in Figure 2.13, the characteristic lines emanating from the x axis fan out into the right half-plane, and so never cross each other at any future time $t > 0$. Each point (t,x) with $t \geq 0$ lies on a unique characteristic line, and the value of the solution at (t,x) is equal to the slope of the line. We conclude that the solution $u(t,x)$ is well defined at all future times $t \geq 0$. Physically, such solutions represent *rarefaction waves*, which spread out as time progresses. A typical example, corresponding to initial data

$$u(0,x) = \tfrac{1}{2}\pi + \tan^{-1}(3x),$$

has its characteristic lines plotted in Figure 2.13, while Figure 2.14 graphs some representative solution profiles.

The more interesting case occurs when the initial data is a decreasing function, and so $f'(x) < 0$. Now, as in Figure 2.12, some of the characteristic lines starting at $t = 0$ will cross at some point in the future. If a point (t,x) lies on two or more distinct characteristic lines, the value of the solution $u(t,x)$, which should equal the characteristic slope, is no longer uniquely determined. Although, in a purely mathematical context, one might be tempted to allow such multiply valued solutions, from a physical standpoint this is unacceptable. The solution $u(t,x)$ is supposed to represent a measurable quantity, e.g., concentration,

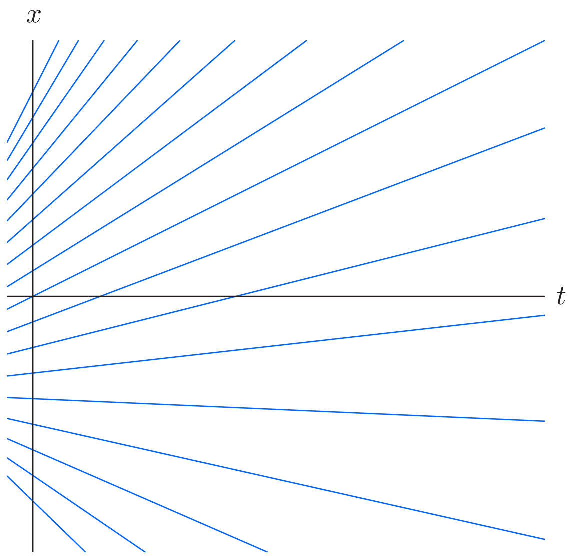

Figure 2.13. Characteristic lines for a rarefaction wave.

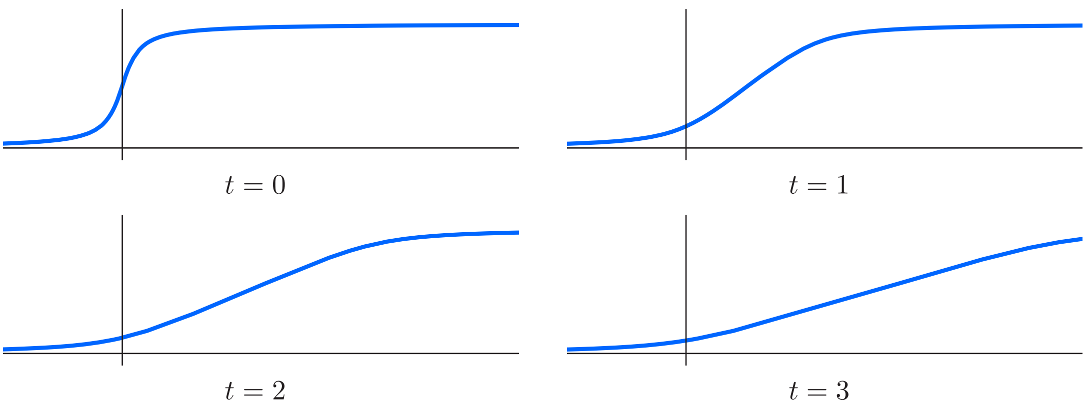

Figure 2.14. Rarefaction wave.

velocity, pressure, and must therefore assume a unique value at each point. In effect, the mathematical model has broken down and no longer conforms to physical reality.

However, before confronting this difficulty, let us first, from a purely theoretical standpoint, try to understand what happens if we mathematically continue the solution as a multiply valued function. For specificity, consider the initial data

$$u(0,x) = \tfrac{1}{2}\pi - \tan^{-1} x, \tag{2.40}$$

appearing in the first graph in Figure 2.15. The corresponding characteristic lines are displayed in Figure 2.12. Initially, they do not cross, and the solution remains a well-defined, single-valued function. However, after a while one reaches a critical time, $t_\star > 0$, when the first two characteristic lines cross each other. Subsequently, a wedge-shaped region appears in the (t,x)–plane, consisting of points that lie on the intersection of three

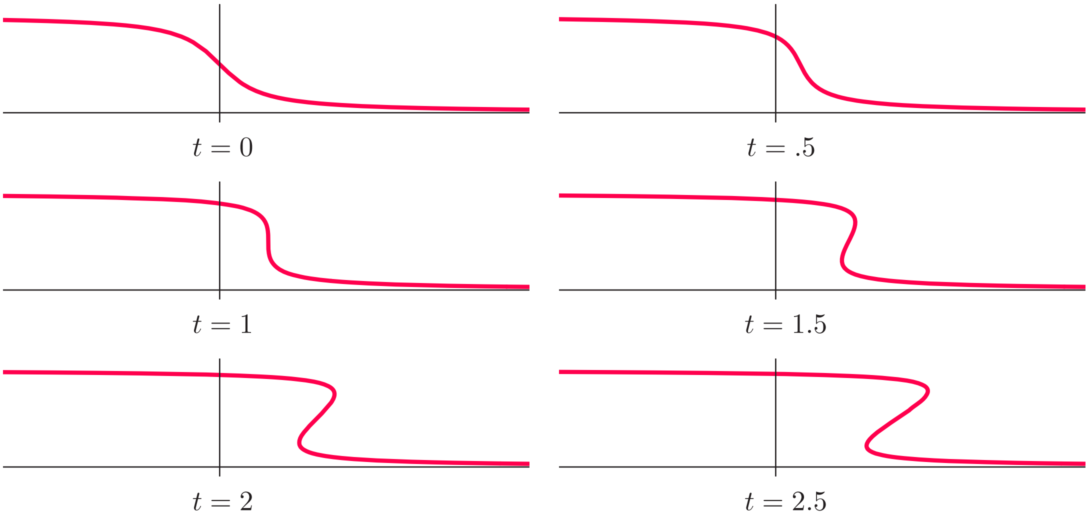

Figure 2.15. Multiply valued compression wave.

distinct characteristic lines with different slopes; at such points, the mathematical solution achieves three distinct values. Points outside the wedge lie on a single characteristic line, and the solution remains single-valued there. The boundary of the wedge consists of points where precisely two characteristic lines cross.

To fully appreciate what is going on, look now at the sequence of pictures of the multiply valued solution in Figure 2.15, plotted at six successive times. Since the initial data is positive, $f(x) > 0$, all the characteristic slopes are positive. As a consequence, every point on the solution curve moves to the right, at a speed equal to its height. Since the initial data is a decreasing function, points on the graph lying to the left will move faster than those to the right and eventually overtake them. At first, the solution merely steepens into a *compression wave*. At the critical time $t_\star$ when the first two characteristic lines cross, say at position $x_\star$, so that $(t_\star, x_\star)$ is the tip of the aforementioned wedge, the solution graph has become vertical:

$$\frac{\partial u}{\partial x}(t, x_\star) \longrightarrow \infty \qquad \text{as} \qquad t \longrightarrow t_\star,$$

and $u(t,x)$ is no longer a classical solution. Once this occurs, the solution graph ceases to be a single-valued function, and its overlapping lobes lie over the points (t,x) belonging to the wedge.

The critical time $t_\star$ can, in fact, be determined from the implicit solution formula (2.35). Indeed, if we differentiate with respect to x, we obtain

$$\frac{\partial u}{\partial x} = \frac{\partial}{\partial x} f(\xi) = f'(\xi)\,\frac{\partial \xi}{\partial x} = f'(\xi)\left(1 - t\,\frac{\partial u}{\partial x}\right), \qquad \text{where} \qquad \xi = x - t\,u.$$

Solving for

$$\frac{\partial u}{\partial x} = \frac{f'(\xi)}{1 + t\, f'(\xi)},$$

we see that the slope blows up:

$$\frac{\partial u}{\partial x} \longrightarrow \infty \qquad \text{as} \qquad t \longrightarrow -\frac{1}{f'(\xi)}\,.$$

In other words, if the initial data has negative slope at position x, so $f'(x) < 0$, then the solution along the characteristic line emanating from the point $(0, x)$ will fail to be smooth at the time $-1/f'(x)$. The earliest critical time is, thus,

$$t_\star := \min\left\{ -\frac{1}{f'(x)} \;\middle|\; f'(x) < 0 \right\}. \tag{2.41}$$

If x_0 is the value of x that produces the minimum $t_\star$, then the slope of the solution profile will first become infinite at the location where the characteristic starting at x_0 is at time $t_\star$, namely

$$x_\star = x_0 + f(x_0)\, t_\star. \tag{2.42}$$

For instance, for the particular initial configuration (2.40) represented in Figure 2.15,

$$f(x) = \frac{\pi}{2} - \tan^{-1} x, \qquad\qquad f'(x) = -\frac{1}{1+x^2}\,,$$

and so the critical time is

$$t_\star = \min\,\{\,1 + x^2\,\} = 1, \qquad \text{with} \qquad x_\star = f(0)\, t_\star = \tfrac{1}{2}\,\pi,$$

since the minimum value occurs at $x_0 = 0$.

Now, while mathematically plausible, such a multiply valued solution is physically untenable. So what really happens after the critical time $t_\star$? One needs to decide which (if any) of the possible solution values is physically appropriate. The mathematical model, in and of itself, is incapable of resolving this quandary. We must therefore revisit the underlying physics, and ask what sort of phenomenon we are trying to model.

Shock Dynamics

To be specific, let us regard the transport equation (2.31) as a model of compressible fluid flow in a single space variable, e.g., the motion of gas in a long pipe. If we push a piston into the pipe, then the gas will move ahead of it and thereby be compressed. However, if the piston moves too rapidly, then the gas piles up on top of itself, and a shock wave forms and propagates down the pipe. Mathematically, the shock is represented by a discontinuity where the solution abruptly changes value. The formulas (2.41) and (2.42) determine the time and position for the onset of the shock-wave discontinuity. Our goal now is to predict its subsequent behavior, and this will be based on use of a suitable physical conservation law. Indeed, one expects mass to be conserved – even through a shock discontinuity — since gas atoms can neither be created nor destroyed. And, as we will see, conservation of mass (almost) suffices to prescribe the subsequent motion of the shock wave.

Before investigating the implications of conservation of mass, let us first convince ourselves of its validity for the nonlinear transport model. (Just because a mathematical equation models a physical system does not automatically imply that it inherits any of its physical conservation laws.) If $u(t,x)$ represents density, then, at time t, the total mass lying in an interval $a \le x \le b$ is calculated by integration:

$$M_{a,b}(t) = \int_a^b u(t,x)\,dx. \tag{2.43}$$

Assuming that $u(t,x)$ is a classical solution to the nonlinear transport equation (2.31), we can determine the rate of change of mass on this interval by differentiation:

$$\begin{aligned} \frac{dM_{a,b}}{dt} &= \frac{d}{dt}\int_a^b u(t,x)\,dx = \int_a^b \frac{\partial u}{\partial t}(t,x)\,dx = -\int_a^b u(t,x)\,\frac{\partial u}{\partial x}(t,x)\,dx \\ &= -\int_a^b \frac{\partial}{\partial x}\left[\,\tfrac{1}{2}u(t,x)^2\,\right]dx = -\,\tfrac{1}{2}u(t,x)^2\,\Big|_{x=a}^{b} = \tfrac{1}{2}u(t,a)^2 - \tfrac{1}{2}u(t,b)^2. \end{aligned} \tag{2.44}$$

The final expression represents the net *mass flux* through the endpoints of the interval. Thus, the only way in which the mass on the interval $[a,b]$ changes is through its endpoints; inside, mass can be neither created nor destroyed, which is the precise meaning of the mass conservation law in continuum mechanics. In particular, if there is zero net mass flux, then the total mass is constant, and hence conserved. For example, if the initial data (2.37) has finite total mass,

$$\left|\int_{-\infty}^{\infty} f(x)\,dx\right| < \infty, \tag{2.45}$$

which requires that $f(x) \to 0$ reasonably rapidly as $|\,x\,| \to \infty$, then the total mass of the solution — at least up to the formation of a shock discontinuity — remains constant and equal to its initial value:

$$\int_{-\infty}^{\infty} u(t,x)\,dx = \int_{-\infty}^{\infty} u(0,x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx. \tag{2.46}$$

Similarly, if $u(t,x)$ represents the traffic density on a highway at time t and position x, then the integrated conservation law (2.44) tells us that the rate of change in the number of vehicles on the stretch of road between a and b equals the number of vehicles entering at point a minus the number leaving at point b — which assumes that there are no other exits or entrances on this part of the highway. Thus, in the traffic model, (2.44) represents the conservation of vehicles.

The preceding calculation relied on the fact that the integrand can be written as an x derivative. This is a common feature of physical conservation laws in continuum mechanics, and motivates the following general definition.

Definition 2.7. A *conservation law*, in one space dimension, is an equation of the form

$$\frac{\partial T}{\partial t} + \frac{\partial X}{\partial x} = 0. \tag{2.47}$$

The function T is known as the *conserved density*, while X is the associated *flux*.

In the simplest situations, the conserved density $T(t,x,u)$ and flux $X(t,x,u)$ depend on the time t, the position x, and the solution $u(t,x)$ to the physical system. (Higher-order conservation laws, which also depend on derivatives of u, arise in the analysis of integrable partial differential equations; see Section 8.5 and [**36**, **87**].) For example, the nonlinear transport equation (2.31) is itself a conservation law, since it can be written in the form

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\tfrac{1}{2}u^2\right) = 0, \tag{2.48}$$

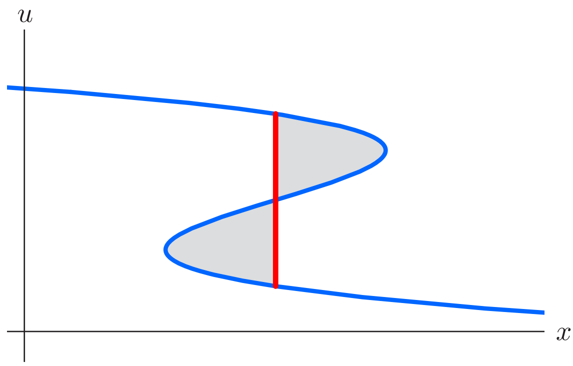

Figure 2.16. Equal Area Rule.

and so the conserved density is $T = u$ and the flux is $X = \frac{1}{2}u^2$. And indeed, it was this identity that made our computation (2.44) work. The general result, proved by an analogous computation, justifies calling (2.47) a conservation law.

Proposition 2.8. *Given a conservation law* (2.47)*, then, on any closed interval* $a \le x \le b$,

$$\frac{d}{dt}\int_a^b T\,dx = -\,X\,\Big|_{x=a}^{b}. \tag{2.49}$$

Proof: The proof is an immediate consequence of the Fundamental Theorem of Calculus — assuming sufficient smoothness that allows one to bring the derivative inside the integral sign:

$$\frac{d}{dt}\int_a^b T\,dx = \int_a^b \frac{\partial T}{\partial t}\,dx = -\int_a^b \frac{\partial X}{\partial x}\,dx = -\,X\,\Big|_{x=a}^{b}. \qquad Q.E.D.$$

We will refer to (2.49) as the *integrated form* of the conservation law (2.47). It states that the rate of change of the total density, integrated over an interval, is equal to the amount of flux through its two endpoints. In particular, if there is no net flux into or out of the interval, then the integrated density is *conserved*, meaning that it remains constant over time. All physical conservation laws — mass, momentum, energy, and so on — for systems governed by partial differential equations are of this form or its multi-dimensional extensions, [**87**].

With this in hand, let us return to the physical context of the nonlinear transport equation. By definition, a *shock* is a discontinuity in the solution $u(t,x)$. We will make the physically plausible assumption that mass (or vehicle) conservation continues to hold even within the shock. Recall that the total mass, which at time t is the area† under the curve $u(t,x)$, must be conserved. This continues to hold even when the mathematical solution becomes multiply valued, in which case one employs a line integral $\int_C u\,dx$, where C represents the graph of the solution, to compute the mass/area. Thus, to construct a discontinuous shock solution with the *same* mass, one replaces part of the multiply valued

† We are implicitly assuming that the mass is finite, as in (2.45), although the overall construction does not rely on this restriction.

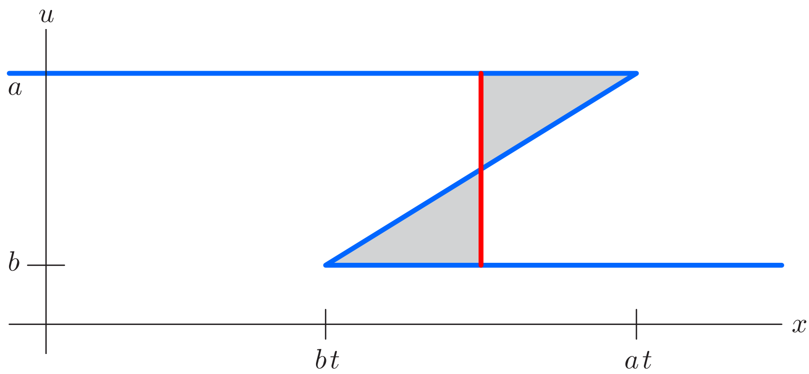

Figure 2.17. Multiply valued step wave.

graph by a vertical shock line in such a way that the resulting function is single-valued and has the same area under its graph. Referring to Figure 2.16, observe that the region under the shock graph is obtained from that under the multi-valued solution graph by deleting the upper shaded lobe and appending the lower shaded lobe. Thus the resulting area will be the same, provided the shock line is drawn so that the areas of the two shaded lobes are equal. This construction is known as the *Equal Area Rule*; it ensures that the total mass of the shock solution matches that of the multiply valued solution, which in turn is equal to the initial mass, as required by the physical conservation law.

Example 2.9. An illuminating special case occurs when the initial data has the form of a *step function* with a single discontinuity at the origin:

$$u(0,x) = \begin{cases} a, & x < 0, \\ b, & x > 0. \end{cases} \tag{2.50}$$

If $a > b$, then the initial data is already in the form of a shock wave. For $t > 0$, the mathematical solution constructed by continuing along the characteristic lines is multiply valued in the region $bt < x < at$, where it assumes both values a and b; see Figure 2.17. Moreover, the initial vertical line of discontinuity has become a tilted line, because each point $(0, u)$ on it has moved along the associated characteristic a distance ut. The Equal Area Rule tells us to draw the shock line halfway along, at $x = \frac{1}{2}(a+b)\,t$, in order that the two triangles have the same area. We deduce that the shock moves with speed $c = \frac{1}{2}(a+b)$, equal to the average of the two speeds at the jump. The resulting shock-wave solution is

$$u(t,x) = \begin{cases} a, & x < ct, \\ b, & x > ct, \end{cases} \qquad \text{where} \qquad c = \frac{a+b}{2}\,. \tag{2.51}$$

A plot of its characteristic lines appears in Figure 2.18. Observe that colliding pairs of characteristic lines terminate at the shock line, whose slope is the average of their individual slopes.

The fact that the shock speed equals the *average* of the solution values on either side is, in fact, of general validity, and is known as the *Rankine–Hugoniot condition*, named after the nineteenth-century Scottish physicist William Rankine and French engineer Pierre Hugoniot, although historically these conditions first appeared in a 1849 paper by George Stokes, [**109**]. However, intimidated by criticism by his contemporary applied mathematicians Lords Kelvin and Rayleigh, Stokes thought he was mistaken, and even ended up

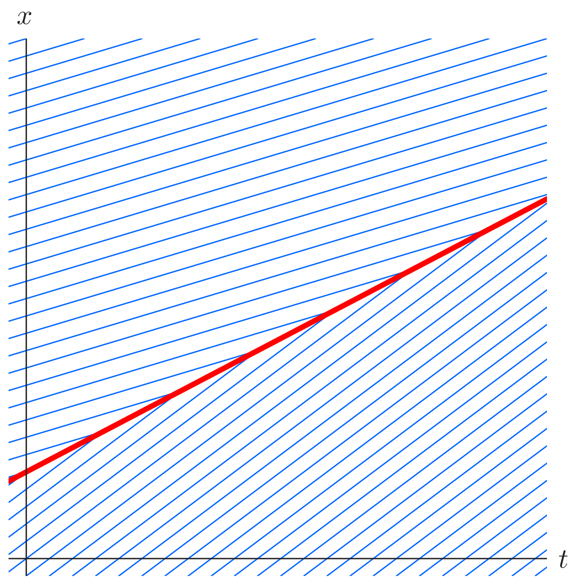

Figure 2.18. Characteristic lines for the step wave shock.

deleting the relevant part when his collected works were published in 1883, [**110**]. The missing section was restored in the 1966 reissue, [**111**].

Proposition 2.10. *Let $u(t,x)$ be a solution to the nonlinear transport equation that has a discontinuity at position $x=\sigma(t)$, with finite, unequal left- and right-hand limits*

$$u^-(t) = u\bigl(t,\sigma(t)^-\bigr) = \lim_{x\,\to\,\sigma(t)^-} u(t,x), \qquad u^+(t) = u\bigl(t,\sigma(t)^+\bigr) = \lim_{x\,\to\,\sigma(t)^+} u(t,x), \tag{2.52}$$

on either side of the shock discontinuity. Then, to maintain conservation of mass, the speed of the shock must equal the average of the solution values on either side:

$$\frac{d\sigma}{dt} = \frac{u^-(t)+u^+(t)}{2}\,. \tag{2.53}$$

Proof: Referring to Figure 2.19, consider a small time interval, from t to $t+\Delta t$, with $\Delta t > 0$. During this time, the shock moves from position $a=\sigma(t)$ to position $b=\sigma(t+\Delta t)$. The total mass contained in the interval $[a,b]$ at time t, before the shock has passed through, is

$$M(t) = \int_a^b u(t,x)\,dx \approx u^+(t)\,(b-a) = u^+(t)\,\bigl[\,\sigma(t+\Delta t)-\sigma(t)\,\bigr],$$

where we assume that $\Delta t \ll 1$ is very small, and so the integrand is well approximated by its limiting value (2.52). Similarly, after the shock has passed, the total mass remaining in the interval is

$$M(t+\Delta t) = \int_a^b u(t+\Delta t,x)\,dx \approx u^-(t+\Delta t)\,(b-a) = u^-(t+\Delta t)\,\bigl[\,\sigma(t+\Delta t)-\sigma(t)\,\bigr].$$

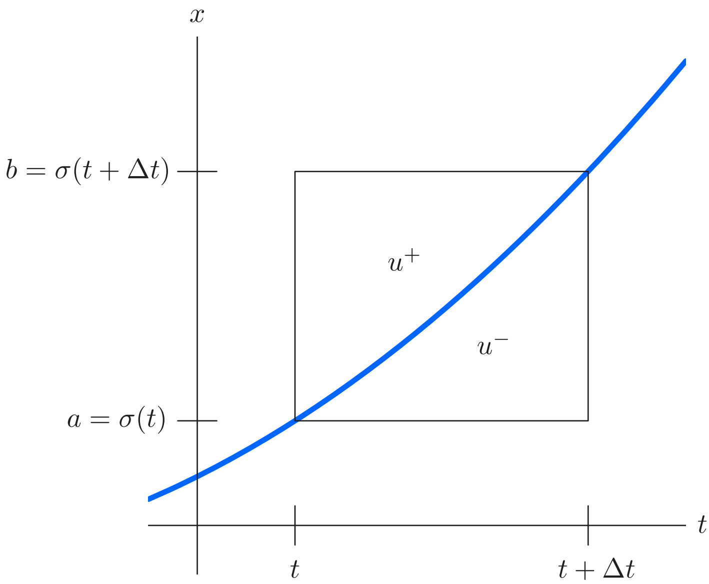

Figure 2.19. Conservation of mass near a shock.

Thus, the rate of change in mass across the shock at time t is given by

$$\frac{dM}{dt} = \lim_{\Delta t \to 0} \frac{M(t+\Delta t) - M(t)}{\Delta t}$$
$$= \lim_{\Delta t \to 0} \left[\, u^-(t+\Delta t) - u^+(t) \,\right] \frac{\sigma(t+\Delta t) - \sigma(t)}{\Delta t} = \left[\, u^-(t) - u^+(t) \,\right] \frac{d\sigma}{dt} \,.$$

On the other hand, at any $t < \tau < t + \Delta t$, the mass flux into the interval $[\,a,b\,]$ through the endpoints is given by the right-hand side of (2.44):

$$\tfrac{1}{2}\left[\, u(\tau,a)^2 - u(\tau,b)^2 \,\right] \longrightarrow \tfrac{1}{2}\left[\, u^-(t)^2 - u^+(t)^2 \,\right], \qquad \text{since} \quad \tau \to t \quad \text{as} \quad \Delta t \to 0.$$

Conservation of mass requires that the rate of change in mass be equal to the mass flux:

$$\frac{dM}{dt} = \left[\, u^-(t) - u^+(t) \,\right] \frac{d\sigma}{dt} = \tfrac{1}{2}\left[\, u^-(t)^2 - u^+(t)^2 \,\right].$$

Solving for $d\sigma/dt$ establishes (2.53). *Q.E.D.*

Example 2.11. By way of contrast, let us investigate the case when the initial data is a step function (2.50), but with $a < b$, so the jump goes upwards. In this case, the characteristic lines diverge from the initial discontinuity, and the mathematical solution is not specified at all in the wedge-shaped region $a\,t < x < b\,t$. Our task is to decide how to "fill in" the solution values between the two regions where the solution is well defined and constant.

One possible connection is by a straight line. Indeed, a simple modification of the rational solution (2.36) produces the *similarity solution*†

$$u(t,x) = \frac{x}{t}\,,$$

† See Section 8.2 for general techniques for constructing similarity (scale-invariant) solutions to partial differential equations.

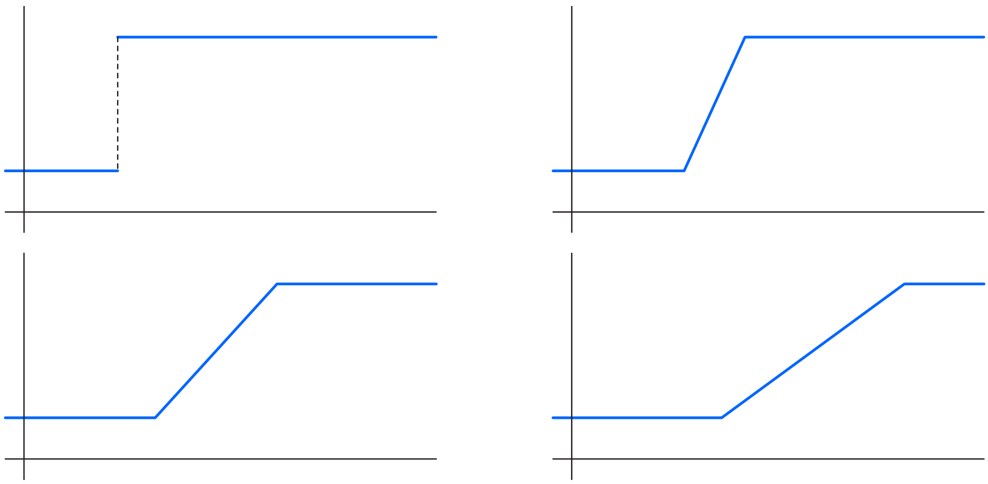

Figure 2.20. Rarefaction wave.

which not only solves the differential equation, but also has the required values $u(t, a\,t) = a$ and $u(t, b\,t) = b$ at the two edges of the wedge. This can be used to construct the piecewise affine *rarefaction wave*

$$u(t,x) = \begin{cases} a, & x \le a\,t, \\ x/t, & a\,t \le x \le b\,t, \\ b, & x \ge b\,t, \end{cases} \tag{2.54}$$

which is graphed at four representative times in Figure 2.20.

A second possibility would be to continue the discontinuity as a shock wave, whose speed is governed by the Rankine-Hugoniot condition, leading to a discontinuous solution having the same formula as (2.51). Which of the two competing solutions should we use? The first, (2.54), makes better physical sense; indeed, if we were to smooth out the discontinuity, then the resulting solutions would converge to the rarefaction wave and not the reverse shock wave; see Exercise 2.3.13. Moreover, the discontinuous solution (2.51) has characteristic lines emanating from the discontinuity, which means that the shock is creating new values for the solution as it moves along, and this can, in fact, be done in a variety of ways. In other words, the discontinuous solution violates *causality*, meaning that the solution profile at any given time uniquely prescribes its subsequent motion. Causality requires that, while characteristics may terminate at a shock discontinuity, they cannot begin there, because their slopes will not be uniquely prescribed by the shock profile, and hence the characteristics to the left of the shock must have larger slope (or speed), while those to the right must have smaller slope. Since the shock speed is the average of the two characteristic slopes, this requires the *Entropy Condition*

$$u^-(t) > \frac{d\sigma}{dt} = \frac{u^-(t) + u^+(t)}{2} > u^+(t). \tag{2.55}$$

With further analysis, it can be shown, [**57**], that the rarefaction wave (2.54) is the unique solution† to the initial value problem satisfying the entropy condition (2.55).

† Albeit not a classical solution, but rather a weak solution, as per Section 10.4.

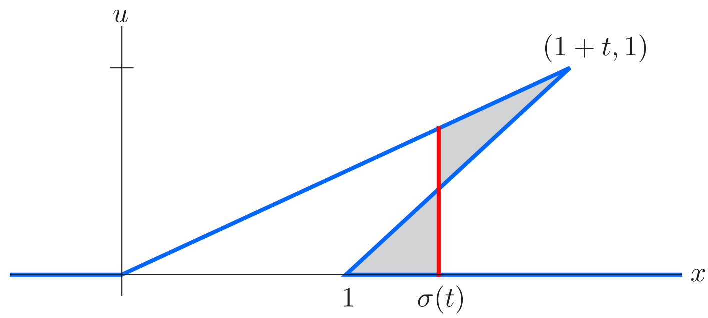

Figure 2.21. Equal Area Rule for the triangular wave.

These prototypical solutions epitomize the basic phenomena modeled by the nonlinear transport equation: *rarefaction waves*, which emanate from regions where the initial data satisfies $f'(x) > 0$, causing the solution to spread out as time progresses, and *compression waves*, emanting from regions where $f'(x) < 0$, causing the solution to progressively steepen and eventually break into a shock discontinuity. Anyone caught in a traffic jam recognizes the compression waves, where the vehicles are bunched together and almost stationary, while the interspersed rarefaction waves correspond to freely moving traffic. (An intelligent driver will take advantage of the rarefaction waves moving backwards through the jam to switch lanes!) The familiar, frustrating traffic jam phenomenon, even on accident- or construction-free stretches of highway, is, thus, an intrinsic effect of the nonlinear transport models that govern traffic flow, [**122**].

Example 2.12. *Triangular wave*: Suppose the initial data has the triangular profile

$$u(0,x) = f(x) = \begin{cases} x, & 0 \le x \le 1, \\ 0, & \text{otherwise}, \end{cases}$$

as in the first graph in Figure 2.22. The initial discontinuity at $x = 1$ will propagate as a shock wave, while the slanted line behaves as a rarefaction wave. To find the profile at time t, we first graph the multi-valued solution obtained by moving each point on the graph of f to the right an amount equal to t times its height. As noted above, this motion preserves straight lines. Thus, points on the x–axis remain fixed, and the diagonal line now goes from $(0,0)$ to $(1+t,1)$, which is where the uppermost point $(1,1)$ on the graph of f has moved to, and hence has slope $(1+t)^{-1}$, while the initial vertical shock line has become tilted, going from $(1,0)$ to $(0,1+t)$. We now need to find the position $\sigma(t)$ of the shock line in order to satisfy the Equal Area Rule, namely so that the areas of the two shaded regions in Figure 2.21 are identical. The reader is invited to determine this geometrically; instead, we invoke the Rankine–Hugoniot condition (2.53). At the shock line, $x = \sigma(t)$, the left- and right-hand limiting values are, respectively,

$$u^-(t) = u\big(t,\sigma(t)^-\big) = \frac{\sigma(t)}{1+t}, \qquad u^+(t) = u\big(t,\sigma(t)^+\big) = 0,$$

and hence (2.53) prescribes the shock speed to be

$$\frac{d\sigma}{dt} = \frac{1}{2}\left(\frac{\sigma(t)}{1+t} + 0\right) = \frac{\sigma(t)}{2\,(1+t)}\,.$$

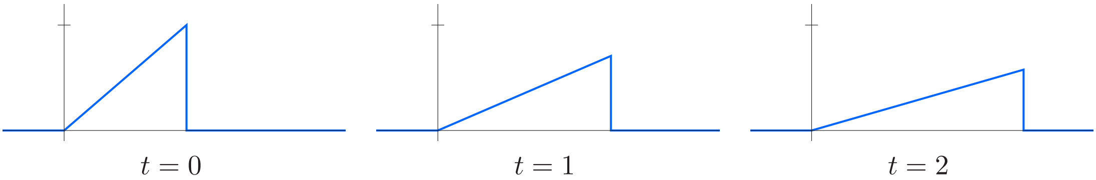

Figure 2.22. Triangular-wave solution.

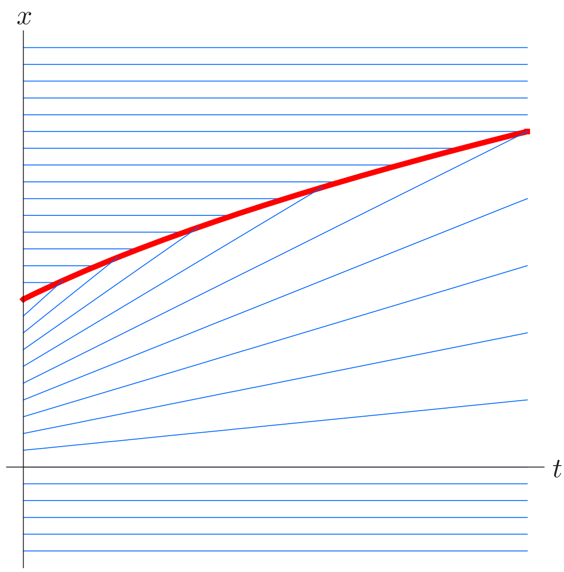

Figure 2.23. Characteristic lines for the triangular-wave shock.

The solution to the resulting separable ordinary differential equation is easily found. Since the shock starts out at $\sigma(0) = 1$, we deduce that

$$\sigma(t) = \sqrt{1+t}\,, \qquad \text{with} \qquad \frac{d\sigma}{dt} = \frac{1}{2\sqrt{1+t}}\,.$$

Further, the strength of the shock, namely its height, is

$$u^-(t) = \frac{\sigma(t)}{1+t} = \frac{1}{\sqrt{1+t}}\,.$$

We conclude that, as t increases, the solution remains a triangular wave, of steadily decreasing slope, while the shock moves off to $x = +\infty$ at a progressively slower speed and smaller height. Its position follows a parabolic trajectory in the (t, x)–plane. See Figure 2.22 for representative plots of the triangular wave solution, while Figure 2.23 illustrates the characteristic lines and shock wave trajectory.

In more general situations, continuing on after the initial shock formation, other characteristic lines may start to cross, thereby producing new shocks. The shocks themselves continue to propagate, often at different velocities. When a fast-moving shock catches up with a slow-moving shock, one must then decide how to merge the shocks so as to retain a physically meaningful solution. The Rankine–Hugoniot (Equal Area) and Entropy Conditions continue to uniquely specify the dynamics. However, at this point, the mathematical details have become too intricate for us to pursue any further, and we refer the interested reader to Whitham's book, [**122**]. See also [**57**] for a proof of the following existence theorem for shock-wave solutions to the nonlinear transport equation.

Theorem 2.13. *If the initial data* $u(0,x) = f(x)$ *is piecewise*† C^1 *with finitely many jump discontinuities, then, for* $t > 0$, *there exists a unique* (*weak*) *solution to the nonlinear transport equation* (2.31) *that also satisfies the Rankine–Hugoniot condition* (2.53) *and the entropy condition* (2.55).

Remark: Our derivation of the Rankine–Hugoniot shock speed condition (2.53) relied on the fact that we can write the original partial differential equation in the form of a conservation law. But there are, in fact, other ways to do this. For instance, multiplying the nonlinear transport equation (2.31) by u allows us write it in the alternative conservative form

$$u\,\frac{\partial u}{\partial t} + u^2\,\frac{\partial u}{\partial x} = \frac{\partial}{\partial t}\left(\tfrac{1}{2}u^2\right) + \frac{\partial}{\partial x}\left(\tfrac{1}{3}u^3\right) = 0. \tag{2.56}$$

In this formulation, the conserved density is $T = \frac{1}{2}u^2$, and the associated flux is $X = \frac{1}{3}u^3$. The integrated form (2.49) of the conservation law (2.56) is

$$\frac{d}{dt}\int_a^b \tfrac{1}{2}u(t,x)^2\,dx = \tfrac{1}{3}\left[\,u(t,a)^3 - u(t,b)^3\,\right]. \tag{2.57}$$

In some physical models, the integral on the left-hand side represents the energy within the interval $[a,b]$, and the conservation law tells us that energy can enter the interval as a flux only through its ends. If we assume that energy is conserved at a shock, then, repeating our previous argument, we are led to the alternative equation

$$\frac{d\sigma}{dt} = \frac{\frac{1}{3}\left[\,u^-(t)^3 - u^+(t)^3\,\right]}{\frac{1}{2}\left[\,u^-(t)^2 - u^+(t)^2\,\right]} = \frac{2}{3}\;\frac{u^-(t)^2 + u^-(t)\,u^+(t) + u^+(t)^2}{u^-(t) + u^+(t)} \tag{2.58}$$

for the shock speed. Thus, a shock that conserves energy moves at a different speed from one that conserves mass! The evolution of a shock wave depends not just on the underlying differential equation, but also on the physical assumptions governing the selection of a suitable conservation law.

More General Wave Speeds

Let us finish this section by considering a nonlinear transport equation

$$u_t + c(u)\,u_x = 0, \tag{2.59}$$

whose wave speed is a more general function of the disturbance u. (Further extensions, allowing c to depend also on t and x, are discussed in Exercise 2.3.20.) Most of the

† Meaning continuous everywhere, and continuously differentiable except at a discrete set of points; see Definition 3.7 below for the precise definition.

development is directly parallel to the special case (2.31) discussed above, and so the details are left for the reader to fill in, although the shock dynamics does require some care.

In this case, the *characteristic curve* equation is

$$\frac{dx}{dt} = c\big(u(t,x)\big). \tag{2.60}$$

As before, the solution u is constant on characteristics, and hence the characteristics are straight lines, now with slope $c(u)$. Thus, to solve the initial value problem

$$u(0,x) = f(x), \tag{2.61}$$

through each point $(0,y)$ on the x–axis, one draws the characteristic line of slope $c(u(0,y)) = c(f(y))$. Until the onset of a shock discontinuity, the solution maintains its initial value $u(0,y) = f(y)$ along the characteristic line.

A shock forms whenever two characteristic lines cross. As before, the mathematical equation no longer uniquely specifies the subsequent dynamics, and we need to appeal to an appropriate conservation law. We write the transport equation in the form

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} C(u) = 0, \qquad \text{where} \qquad C(u) = \int c(u)\, du \tag{2.62}$$

is any convenient anti-derivative of the wave speed. Thus, following the same computation as in (2.44), we discover that conservation of mass now takes the integrated form

$$\frac{d}{dt}\int_a^b u(t,x)\,dx = C(u(t,a)) - C(u(t,b)), \tag{2.63}$$

with $C(u)$ playing the role of the mass flux. Requiring the conservation of mass, i.e., of the area under the graph of the solution, means that the Equal Area Rule remains valid. However, the Rankine–Hugoniot shock-speed condition must be modified in accordance with the new dynamics. Mimicking the preceding argument, but with the modified mass flux, we find that the shock speed is now given by

$$\frac{d\sigma}{dt} = \frac{C(u^-(t)) - C(u^+(t))}{u^-(t) - u^+(t)}\,. \tag{2.64}$$

Note that if

$$c(u) = u, \qquad \text{then} \qquad C(u) = \int u\, du = \tfrac{1}{2}\, u^2,$$

and so (2.64) reduces to our earlier formula (2.53). Moreover, in the limit as the shock magnitude approaches zero, $u^-(t) - u^+(t) \to 0$, the right-hand side of (2.64) converges to the derivative $C'(u) = c(u)$ and hence recovers the wave speed, as it should.

Exercises

2.3.1. Discuss the behavior of the solution to the nonlinear transport equation (2.31) for the following initial data:

(a) $u(0,x) = \begin{cases} 2, & x < -1, \\ 1, & x > -1; \end{cases}$ (b) $u(0,x) = \begin{cases} -2, & x < -1, \\ 1, & x > -1; \end{cases}$ (c) $u(0,x) = \begin{cases} 1, & x < 1, \\ -2, & x > 1. \end{cases}$

2.3.2. Solve the following initial value problems: (a) $u_t + 3uu_x = 0,\ u(0,x) = \begin{cases} 2, & x < 1, \\ 0, & x > 1; \end{cases}$

(b) $u_t - uu_x = 0,\ u(1,x) = \begin{cases} -1, & x < 0, \\ 3, & x > 0; \end{cases}$ (c) $u_t - 2uu_x = 0,\ u(0,x) = \begin{cases} 1, & x < 1, \\ 0, & x > 1. \end{cases}$

2.3.3. Let $u(0,x) = (x^2+1)^{-1}$. Does the resulting solution to the nonlinear transport equation (2.31) produce a shock wave? If so, find the time of onset of the shock, and sketch a graph of the solution just before and soon after the shock wave. If not, explain what happens to the solution as t increases.

2.3.4. Solve Exercise 2.3.3 when $u(0,x) =$ (a) $-(x^2+1)^{-1}$, (b) $x(x^2+1)^{-1}$.

2.3.5. Consider the initial value problem $u_t - 2uu_x = 0,\ u(0,x) = e^{-x^2}$. Does the resulting solution produce a shock wave? If so, find the time of onset of the shock and the position at which it first forms. If not, explain what happens to the solution as t increases.

2.3.6. (a) For what values of $\alpha, \beta, \gamma, \delta$ is $u(t,x) = \dfrac{\alpha x + \beta}{\gamma t + \delta}$ a solution to (2.31)?

(b) For what values of $\alpha, \beta, \gamma, \delta, \lambda, \mu$ is $u(t,x) = \dfrac{\lambda t + \alpha x + \beta}{\gamma t + \mu x + \delta}$ a solution to (2.31)?

2.3.7. A *triangular wave* is a shock-wave solution to the initial value problem for (2.31) that has initial data $u(0,x) = \begin{cases} mx, & 0 \le x \le \ell, \\ 0, & \text{otherwise.} \end{cases}$ Assuming $m > 0$, write down a formula for the triangular-wave solution at times $t > 0$. Discuss what happens to the triangular wave as time progresses.

2.3.8. Solve Exercise 2.3.7 when $m < 0$.

2.3.9. Solve (2.31) for $t > 0$ subject to the following initial conditions, and graph your solution at some representative times. In what sense does your solution conserve mass?

(a) $u(0,x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{otherwise,} \end{cases}$ (b) $u(0,x) = \begin{cases} x, & -1 < x < 1, \\ 0, & \text{otherwise,} \end{cases}$

(c) $u(0,x) = \begin{cases} -x, & -1 < x < 1, \\ 0, & \text{otherwise,} \end{cases}$ (d) $u(0,x) = \begin{cases} 1 - |\,x\,|, & -1 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$

2.3.10. An N*–wave* is a solution to the nonlinear transport equation (2.31) that has initial conditions $u(0,x) = \begin{cases} mx, & -\ell \le x \le \ell, \\ 0, & \text{otherwise,} \end{cases}$ where $m > 0$. (a) Write down a formula for the N–wave solution at times $t > 0$. (b) What about when $m < 0$?

◇ 2.3.11. Suppose $u(t,x)$ and $\widetilde{u}(t,x)$ are two solutions to the nonlinear transport equation (2.31) such that, for some $t_\star > 0$, they agree: $u(t_\star,x) = \widetilde{u}(t_\star,x)$ for all x. Do the solutions necessarily have the same initial conditions: $u(0,x) = \widetilde{u}(0,x)$? Use your answer to discuss the uniqueness of solutions to the nonlinear transport equation.

2.3.12. Suppose that $x_1 < x_2$ are such that the characteristic lines of (2.31) through $(0,x_1)$ and $(0,x_2)$ cross at a shock at $(t,\sigma(t))$ and, moreover, the left- and right-hand shock values (2.52) are $f(x_1) = u^-(t)$, $f(x_1) = u^+(t)$. Explain why the signed area of the region between the graph of $f(x)$ and the secant line connecting $(x_1, f(x_1))$ to $(x_2, f(x_2))$ is zero.

◇ 2.3.13. Consider the initial value problem $u^\varepsilon(0,x) = 2 + \tan^{-1}(x/\varepsilon)$ for the nonlinear transport equation (2.31). (a) Show that, as $\varepsilon \to 0^+$, the initial condition converges to a step function (2.51). What are the values of a, b? (b) Show that, moreover, the resulting solution $u^\varepsilon(0,x)$ to the nonlinear transport equation converges to the corresponding rarefaction wave (2.54) resulting from the limiting initial condition.

♢ 2.3.14. (a) Under what conditions can equation (2.35) be solved for a single-valued function $u(t,x)$? *Hint*: Use the Implicit Function Theorem. (b) Use implicit differentiation to prove that the resulting function $u(t,x)$ is a solution to the nonlinear transport equation.

2.3.15. For what values of $\alpha, \beta, \gamma, \delta, k$ is $u(t,x) = \left(\dfrac{\alpha x + \beta}{\gamma t + \delta}\right)^k$ a solution to the transport equation $u_t + u^2 u_x = 0$?

2.3.16. (a) Solve the initial value problem $u_t + u^2 u_x = 0$, $u(0,x) = f(x)$, by the method of characteristics. (b) Discuss the behavior of solutions and compare/contrast with (2.31).

2.3.17. (a) Determine the Rankine–Hugoniot condition, based on conservation of mass, for the speed of a shock for the equation $u_t + u^2 u_x = 0$. (b) Solve the initial value problem $u(0,x) = \begin{cases} a, & x < 0, \\ b, & x > 0, \end{cases}$ when (i) $|a| > |b|$, (ii) $|a| < |b|$. *Hint*: Use Exercise 2.3.15 to determine the shape of a rarefaction wave.

2.3.18. Solve Exercise 2.3.17 when the wave speed $c(u) =$ (i) $1 - 2u$, (ii) u^3, (iii) $\sin u$.

♢ 2.3.19. Justify the shock-speed formula (2.58).

♢ 2.3.20. Consider the general quasilinear first-order partial differential equation

$$\frac{\partial u}{\partial t} + c(t,x,u)\,\frac{\partial u}{\partial x} = h(t,x,u).$$

Let us define a *lifted characteristic curve* to be a solution $(t, x(t), u(t))$ to the system of ordinary differential equations $\dfrac{dx}{dt} = c(t,x,u)$, $\dfrac{du}{dt} = h(t,x,u)$. The corresponding *characteristic curve* $\big(t, x(t)\big)$ is obtained by projecting to the (t,x)–plane. Prove that if $u(t,x)$ is a solution to the partial differential equation, and $u(t_0, x_0) = u_0$, then the lifted characteristic curve passing through (t_0, x_0, u_0) lies on the graph of $u(t,x)$. Conclude that the graph of the solution to the initial value problem $u(t_0, x) = f(x)$ is the union of all lifted characteristic curves passing through the initial data points $\big(t_0, x_0, f(x_0)\big)$.

2.3.21. Let $a > 0$. (a) Apply the method of Exercise 2.3.20 to solve the initial value problem for the *damped transport equation*: $u_t + u\,u_x + a\,u = 0$, $u(0,x) = f(x)$. (b) Does the damping eliminate shocks?

2.3.22. Apply the method of Exercise 2.3.20 to solve the initial value problem

$$u_t + t\,u_x = u^2, \qquad u(0,x) = \frac{1}{1+x^2}\,.$$

2.4 The Wave Equation: d'Alembert's Formula

Newton's Second Law states that force equals mass times acceleration. It forms the bedrock underlying the derivation of mathematical models describing all of classical dynamics. When applied to a one-dimensional medium, such as the transverse displacements of a violin string or the longitudinal motions of an elastic bar, the resulting model governing small vibrations is the second-order partial differential equation

$$\rho(x)\,\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left(\kappa(x)\,\frac{\partial u}{\partial x}\right). \tag{2.65}$$

Here $u(t,x)$ represents the displacement of the string or bar at time t and position x, while $\rho(x) > 0$ denotes its density and $\kappa(x) > 0$ its stiffness or tension, both of which are assumed not to vary with t. The right-hand side of the equation represents the restoring force due to a (small) displacement of the medium from its equilibrium, whereas the left-hand side is the product of mass per unit length and acceleration. A correct derivation of the model from first principles would require a significant detour, and we refer the reader to [**120**, **124**] for the details.

We will simplify the general model by assuming that the underlying medium is *uniform*, and so both its density ρ and stiffness κ are constant. Then (2.65) reduces to the one-dimensional *wave equation*

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \qquad \text{where the constant} \qquad c = \sqrt{\frac{\kappa}{\rho}} > 0 \tag{2.66}$$

is known as the *wave speed*, for reasons that will soon become apparent.

In general, to uniquely specify the solution to any dynamical system arising from Newton's Second Law, including the wave equation (2.66) and the more general vibration equation (2.65), one must fix both its initial position and initial velocity. Thus, the initial conditions take the form

$$u(0,x) = f(x), \qquad \frac{\partial u}{\partial t}(0,x) = g(x), \tag{2.67}$$

where, for simplicity, we set the initial time $t_0 = 0$. (See also Exercise 2.4.6.) The *initial value problem* seeks the corresponding C^2 function $u(t,x)$ that solves the wave equation (2.66) and has the required initial values (2.67). In this section, we will learn how to solve the initial value problem on the entire line $-\infty < x < \infty$. The analysis of the wave equation on bounded intervals will be deferred until Chapters 4 and 7. The two- and three-dimensional versions of the wave equation are treated in Chapters 11 and 12, respectively.

d'Alembert's Solution

Let us now derive the explicit solution formula for the second-order wave equation (2.66) first found by d'Alembert. The starting point is to write the partial differential equation in the suggestive form

$$\Box\, u = (\partial_t^2 - c^2\,\partial_x^2)\, u = u_{tt} - c^2\, u_{xx} = 0. \tag{2.68}$$

Here

$$\Box = \partial_t^2 - c^2\,\partial_x^2$$

is a common mathematical notation for the *wave operator*, which is a linear second-order partial differential operator. In analogy with the elementary polynomial factorization

$$t^2 - c^2\,x^2 = (t - c\,x)(t + c\,x),$$

we can factor the wave operator into a product of two first-order partial differential operators:†

$$\Box = \partial_t^2 - c^2\,\partial_x^2 = (\partial_t - c\,\partial_x)\,(\partial_t + c\,\partial_x). \tag{2.69}$$

† The cross terms cancel, thanks to the equality of mixed partial derivatives: $\partial_t\partial_x u = \partial_x\partial_t u$. Constancy of the wave speed c is essential here.

Now, if the second factor annihilates the function $u(t,x)$, meaning

$$(\partial_t + c\,\partial_x)\,u = u_t + c\,u_x = 0, \tag{2.70}$$

then u is automatically a solution to the wave equation, since

$$\Box\, u = (\partial_t - c\,\partial_x)\,(\partial_t + c\,\partial_x)\,u = (\partial_t - c\,\partial_x)\,0 = 0.$$

We recognize (2.70) as the first-order transport equation (2.4) with constant wave speed c. Proposition 2.1 tells us that its solutions are traveling waves with wave speed c:

$$u(t,x) = p(\xi) = p(x - c\,t), \tag{2.71}$$

where p is an arbitrary function of the characteristic variable $\xi = x - c\,t$. As long as $p \in \mathrm{C}^2$ (i.e., is twice continuously differentiable), the resulting function $u(t,x)$ is a classical solution to the wave equation (2.66), as you can easily check.

Now, the factorization (2.69) can equally well be written in the reverse order:

$$\Box = \partial_t^2 - c^2\,\partial_x^2 = (\partial_t + c\,\partial_x)\,(\partial_t - c\,\partial_x). \tag{2.72}$$

The same argument tells us that any solution to the "backwards" transport equation

$$u_t - c\,u_x = 0, \tag{2.73}$$

with constant wave speed $-\,c$, also provides a solution to the wave equation. Again, by Proposition 2.1, with c replaced by $-\,c$, the general solution to (2.73) has the form

$$u(t,x) = q(\eta) = q(x + c\,t), \tag{2.74}$$

where q is an arbitrary function of the alternative characteristic variable $\eta = x + c\,t$. The solutions (2.74) represent traveling waves moving to the *left* with constant speed $c > 0$. Provided $q \in \mathrm{C}^2$, the functions (2.74) will provide a second family of solutions to the wave equation.

We conclude that, unlike first-order transport equations, the wave equation (2.68) is *bidirectional* in that it admits both left and right traveling-wave solutions. Moreover, by linearity the sum of any two solutions is again a solution, and so we can immediately construct solutions that are superpositions of left and right traveling waves. The remarkable fact is that *every* solution to the wave equation can be so represented.

Theorem 2.14. *Every solution to the wave equation* (2.66) *can be written as a superposition,*

$$u(t,x) = p(\xi) + q(\eta) = p(x - c\,t) + q(x + c\,t), \tag{2.75}$$

of right and left traveling waves. Here $p(\xi)$ and $q(\eta)$ are arbitrary C^2 functions, each depending on its respective characteristic variable

$$\xi = x - c\,t, \qquad\qquad \eta = x + c\,t. \tag{2.76}$$

Proof: As in our treatment of the transport equation, we will simplify the wave equation through an inspired change of variables. In this case, the new independent variables are the characteristic variables ξ, η defined by (2.76). We set

$$u(t,x) = v(x - c\,t, x + c\,t) = v(\xi,\eta), \qquad \text{whereby} \qquad v(\xi,\eta) = u\left(\frac{\eta-\xi}{2\,c}, \frac{\eta+\xi}{2}\right). \tag{2.77}$$

Then, employing the chain rule to compute the partial derivatives,

$$\frac{\partial u}{\partial t} = c\left(-\frac{\partial v}{\partial \xi} + \frac{\partial v}{\partial \eta} \right), \qquad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial \xi} + \frac{\partial v}{\partial \eta}, \tag{2.78}$$

and, further,

$$\frac{\partial^2 u}{\partial t^2} = c^2 \left(\frac{\partial^2 v}{\partial \xi^2} - 2\,\frac{\partial^2 v}{\partial \xi\,\partial \eta} + \frac{\partial^2 v}{\partial \eta^2} \right), \qquad \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial \xi^2} + 2\,\frac{\partial^2 v}{\partial \xi\,\partial \eta} + \frac{\partial^2 v}{\partial \eta^2}.$$

Therefore

$$\Box\, u = \frac{\partial^2 u}{\partial t^2} - c^2\,\frac{\partial^2 u}{\partial x^2} = -\,4\,c^2\,\frac{\partial^2 v}{\partial \xi\,\partial \eta}. \tag{2.79}$$

We conclude that $u(t,x)$ solves the wave equation $\Box\, u = 0$ if and only if $v(\xi,\eta)$ solves the second-order partial differential equation

$$\frac{\partial^2 v}{\partial \xi\,\partial \eta} = 0,$$

which we write in the form

$$\frac{\partial}{\partial \xi}\left(\frac{\partial v}{\partial \eta} \right) = \frac{\partial w}{\partial \xi} = 0, \qquad \text{where} \qquad w = \frac{\partial v}{\partial \eta}.$$

Thus, applying the methods of Section 2.1 (and making the appropriate assumptions on the domain of definition of w), we deduce that

$$w = \frac{\partial v}{\partial \eta} = r(\eta),$$

where r is an arbitrary function of the characteristic variable η. Integrating both sides of the latter partial differential equation with respect to η, we find

$$v(\xi,\eta) = p(\xi) + q(\eta), \qquad \text{where} \qquad q(\eta) = \int r(\eta)\, d\eta,$$

while $p(\xi)$ represents the η integration "constant". Replacing the characteristic variables by their formulas in terms of t and x completes the proof. *Q.E.D.*

Let us see how the solution formula (2.75) can be used to solve the initial value problem (2.67). Substituting into the initial conditions, we deduce that

$$u(0,x) = p(x) + q(x) = f(x), \qquad \frac{\partial u}{\partial t}(0,x) = -\,c\,p'(x) + c\,q'(x) = g(x). \tag{2.80}$$

To solve this pair of equations for the functions p and q, we differentiate the first,

$$p'(x) + q'(x) = f'(x),$$

and then subtract off the second equation divided by c; the result is

$$2\,p'(x) = f'(x) - \frac{1}{c}\,g(x).$$

Therefore,

$$p(x) = \frac{1}{2}\,f(x) - \frac{1}{2\,c}\int_0^x g(z)\,dz + a,$$

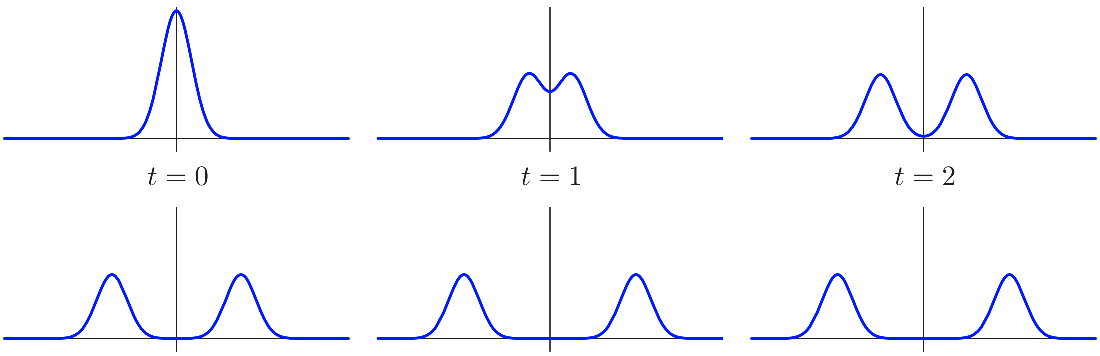

Figure 2.24. Splitting of waves.

where a is an integration constant. The first equation in (2.80) then yields

$$q(x) = f(x) - p(x) = \frac{1}{2}\, f(x) + \frac{1}{2\,c} \int_0^x g(z)\, dz - a.$$

Substituting these two expressions back into our solution formula (2.75), we obtain

$$\begin{aligned} u(t,x) = p(\xi) + q(\eta) &= \frac{f(\xi) + f(\eta)}{2} - \frac{1}{2\,c}\int_0^{\xi} g(z)\,dz + \frac{1}{2\,c}\int_0^{\eta} g(z)\,dz \\ &= \frac{f(\xi) + f(\eta)}{2} + \frac{1}{2\,c}\int_{\xi}^{\eta} g(z)\,dz, \end{aligned}$$

where ξ, η are the characteristic variables (2.76). In this manner, we have arrived at *d'Alembert's solution* to the initial value problem for the wave equation on the real line.

Theorem 2.15. *The solution to the initial value problem*

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \qquad u(0,x) = f(x), \qquad \frac{\partial u}{\partial t}(0,x) = g(x), \qquad -\infty < x < \infty, \tag{2.81}$$

is given by

$$u(t,x) = \frac{f(x - c\,t) + f(x + c\,t)}{2} + \frac{1}{2\,c}\int_{x - c\,t}^{x + c\,t} g(z)\,dz. \tag{2.82}$$

Remark: In order that (2.82) define a classical solution to the wave equation, we need $f \in \mathrm{C}^2$ and $g \in \mathrm{C}^1$. However, the formula itself makes sense for more general initial conditions. We will continue to treat the resulting functions as solutions, albeit nonclassical, since they fit under the more general rubric of "weak solution", to be developed in Section 10.4.

Example 2.16. Suppose there is no initial velocity, so $g(x) \equiv 0$, and hence the motion is purely the result of the initial displacement $u(0,x) = f(x)$. In this case, (2.82) reduces to

$$u(t,x) = \tfrac{1}{2}\, f(x - c\,t) + \tfrac{1}{2}\, f(x + c\,t). \tag{2.83}$$

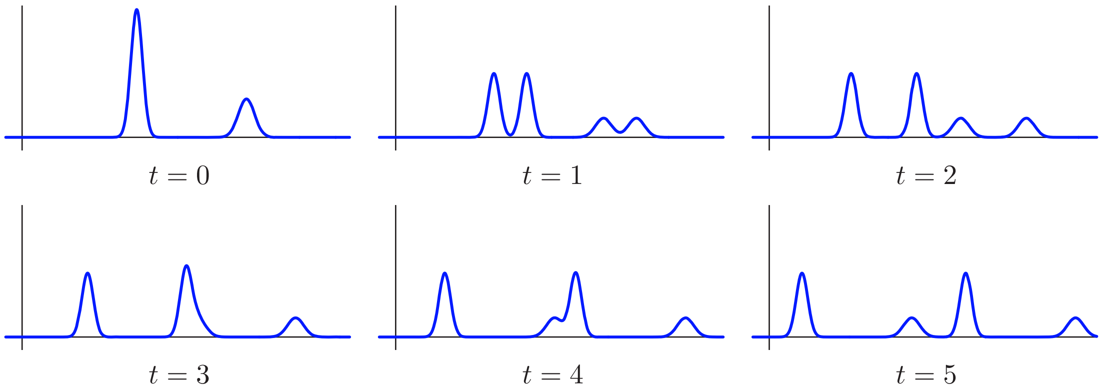

Figure 2.25. Interaction of waves.

The effect is that the initial displacement splits into two waves, one moving to the right and the other moving to the left, each of constant speed c, and each of exactly the same shape as $f(x)$, but only half as tall. For example, if the initial displacement is a localized pulse centered at the origin, say

$$u(0,x) = e^{-x^2}, \qquad \frac{\partial u}{\partial t}(0,x) = 0,$$

then the solution

$$u(t,x) = \tfrac{1}{2}\, e^{-(x-ct)^2} + \tfrac{1}{2}\, e^{-(x+ct)^2}$$

consists of two half size pulses running away from the origin with the same speed c, but in opposite directions. A graph of the solution at several successive times can be seen in Figure 2.24.

If we take two initially separated pulses, say

$$u(0,x) = e^{-x^2} + 2\, e^{-(x-1)^2}, \qquad \frac{\partial u}{\partial t}(0,x) = 0,$$

centered at $x = 0$ and $x = 1$, then the solution

$$u(t,x) = \tfrac{1}{2}\, e^{-(x-ct)^2} + e^{-(x-1-ct)^2} + \tfrac{1}{2}\, e^{-(x+ct)^2} + e^{-(x-1+ct)^2}$$

will consist of four pulses, two moving to the right and two to the left, all with the same speed. An important observation is that when a right-moving pulse collides with a left-moving pulse, they emerge from the collision unchanged, which is a consequence of the inherent linearity of the wave equation. In Figure 2.25, the first picture plots the initial displacement. In the second and third pictures, the two localized bumps have each split into two copies moving in opposite directions. In the fourth and fifth, the larger right-moving bump is in the process of interacting with the smaller left-moving bump. Finally, in the last picture the interaction is complete, and the individual pairs of left- and right-moving waves move off in tandem in opposing directions, experiencing no further collisions.

In general, if the initial displacement is localized, so that $|\, f(x)\, | \ll 1$ for $|\, x\, | \gg 0$, then, after a finite time, the left- and right-moving waves will separate, and the observer will see two half-size replicas running away, with speed c, in opposite directions. If the displacement

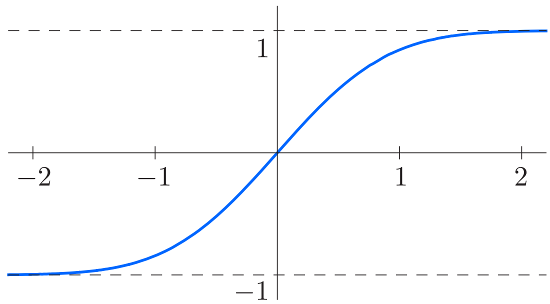

Figure 2.26. The error function erf x.

is not localized, then the left and right traveling waves will never fully disengage, and one might be hard pressed to recognize that a complicated solution pattern is, in reality, just the superposition of two simple traveling waves. For example, consider the elementary trigonometric solution

$$\cos ct \ \cos x = \tfrac{1}{2}\cos(x - ct) + \tfrac{1}{2}\cos(x + ct). \qquad (2.84)$$

In accordance with the left-hand expression, an observer will see a standing cosinusoidal wave that vibrates up and down with frequency c. However, the d'Alembert form of the solution on the right-hand side says that this is just the sum of left- and right-traveling cosine waves! The interactions of their peaks and troughs reproduce the standing wave. Thus, the same solution can be interpreted in two seemingly incompatible ways. And, in fact, this paradox lies at the heart of the perplexing wave-particle duality of quantum physics.

Example 2.17. By way of contrast, suppose there is no initial displacement, so $f(x) \equiv 0$, and the motion is purely the result of the initial velocity $u_t(0,x) = g(x)$. Physically, this models a violin string at rest being struck by a "hammer blow" at the initial time. In this case, the d'Alembert formula (2.82) reduces to

$$u(t,x) = \frac{1}{2\,c}\int_{x-ct}^{x+ct} g(z)\,dz. \qquad (2.85)$$

For example, when $u(0,x) = 0$, $u_t(0,x) = e^{-x^2}$, the resulting solution (2.85) is

$$u(t,x) = \frac{1}{2\,c}\int_{x-ct}^{x+ct} e^{-x^2}\,dz = \frac{\sqrt{\pi}}{4\,c}\big[\operatorname{erf}(x+ct) - \operatorname{erf}(x-ct)\big], \qquad (2.86)$$

where

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}\,dz \qquad (2.87)$$

is known as the *error function* due to its many applications throughout probability and statistics, [**39**]. The error function integral cannot be written in terms of elementary functions; nevertheless, its properties have been well studied and its values tabulated, [**86**]. A graph appears in Figure 2.26. The constant in front of the integral (2.87) has been chosen so that the error function has asymptotic values

$$\lim_{x\to\infty} \operatorname{erf} x = 1, \qquad \lim_{x\to-\infty} \operatorname{erf} x = -1, \qquad (2.88)$$

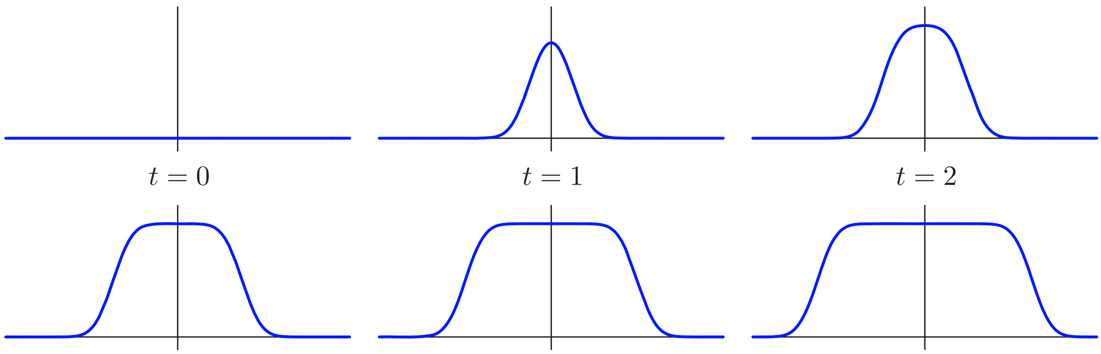

Figure 2.27. Error function solution to the wave equation.

which follow from a well-known integration formula to be derived in Exercise 2.4.21.

A graph of the solution (2.86) at successive times is displayed in Figure 2.27. The first graph shows the zero initial displacement. Gradually, the effect of the initial hammer blow is felt further and further away along the string, as the two wave fronts propagate away from the origin, both with speed c, but in opposite directions. Thus, unlike the case of a nonzero initial displacement in Figure 2.24, where the solution eventually returns to its equilibrium position $u = 0$ after the wave passes by, a nonzero initial velocity leaves the string permanently deformed.

In general, the lines of slope $\pm c$, where the respective characteristic variables are constant,

$$\xi = x - ct = a, \qquad \eta = x + ct = b, \tag{2.89}$$

are known as the *characteristics* of the wave equation. Thus, the second-order wave equation has *two* distinct characteristic lines passing through each point in the (t, x)–plane.

Remark: The characteristic lines are the one-dimensional counterparts of the light cone in Minkowski space-time, which plays a starring role in special relativity, [**70**, **75**]. See Section 12.5 for further details.

In Figure 2.28, we plot the two characteristics going through a point $(0, y)$ on the x axis. The wedge-shaped region $\{\, y - ct \le x \le y + ct,\ t \ge 0 \,\}$ lying between them is known as the *domain of influence* of the point $(0, y)$, since, in general, the value of the initial data at a point will affect the subsequent solution values only in its domain of influence. Indeed, the effect of an initial displacement at the point y propagates along the two characteristic lines, while the effect of an initial velocity there will be felt at every point in the triangular wedge.

External Forcing and Resonance

When a homogeneous vibrating medium is subjected to external forcing, the wave equation acquires an additional, inhomogeneous term:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} + F(t, x), \tag{2.90}$$

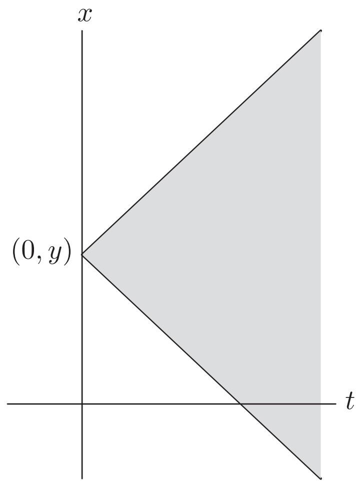

Figure 2.28. Characteristic lines and domain of influence.

in which $F(t, x)$ represents a force imposed at time t and spatial position x. With a bit more work, d'Alembert's solution technique can be readily adapted to incorporate the forcing term.

Let us, for simplicity, assume that the differential equation is supplemented by homogeneous initial conditions,

$$u(0, x) = 0, \qquad u_t(0, x) = 0, \tag{2.91}$$

meaning that there is no initial displacement or velocity. To solve the initial value problem (2.90–91), we switch to the same characteristic coordinates (2.76), setting

$$v(\xi, \eta) = u\left(\frac{\eta - \xi}{2\,c}, \frac{\eta + \xi}{2}\right).$$

Invoking the chain rule formulas (2.79), we find that the forced equation (2.90) becomes

$$\frac{\partial^2 v}{\partial \xi\, \partial \eta} = -\,\frac{1}{4\,c^2}\, F\left(\frac{\eta - \xi}{2\,c}, \frac{\eta + \xi}{2}\right). \tag{2.92}$$

Let us integrate both sides of the equation with respect to η, on the interval $\xi \le \zeta \le \eta$:

$$\frac{\partial v}{\partial \xi}(\xi, \eta) - \frac{\partial v}{\partial \xi}(\xi, \xi) = -\,\frac{1}{4\,c^2} \int_\xi^\eta F\left(\frac{\zeta - \xi}{2\,c}, \frac{\zeta + \xi}{2}\right) d\zeta. \tag{2.93}$$

But, recalling (2.78),

$$\frac{\partial v}{\partial \xi}(\xi, \eta) = -\,\frac{1}{2\,c}\,\frac{\partial u}{\partial t}\left(\frac{\eta - \xi}{2\,c}, \frac{\eta + \xi}{2}\right) + \frac{1}{2}\,\frac{\partial u}{\partial x}\left(\frac{\eta - \xi}{2\,c}, \frac{\eta + \xi}{2}\right),$$

and so, in particular,

$$\frac{\partial v}{\partial \xi}(\xi, \xi) = -\,\frac{1}{2\,c}\,\frac{\partial u}{\partial t}(0, \xi) + \frac{1}{2}\,\frac{\partial u}{\partial x}(0, \xi) = 0,$$

which vanishes owing to our choice of homogeneous initial conditions (2.91). Indeed, the initial velocity condition says that $u_t(0, x) = 0$, while differentiating the initial displacement condition $u(0,x)=0$ with respect to x implies that $u_x(0,x)=0$ for all x, including $x=\xi$. As a result, (2.93) simplifies to

$$\frac{\partial v}{\partial \xi}(\xi,\eta) = -\,\frac{1}{4\,c^2}\int_\xi^\eta F\left(\frac{\zeta-\xi}{2\,c},\frac{\zeta+\xi}{2}\right)d\zeta.$$

We now integrate the latter equation with respect to ξ on the interval $\xi \le \chi \le \eta$, producing

$$-\,v(\xi,\eta) = v(\eta,\eta) - v(\xi,\eta) = -\,\frac{1}{4\,c^2}\int_\xi^\eta\int_\chi^\eta F\left(\frac{\zeta-\chi}{2\,c},\frac{\zeta+\chi}{2}\right)d\zeta\,d\chi,$$

since $v(\eta,\eta) = u(0,\eta) = 0$, thanks again to the initial conditions. In this manner, we have produced an explicit formula for the solution to the characteristic variable version of the forced wave equation subject to the homogeneous initial conditions. Reverting to the original physical coordinates, the left-hand side of this equation becomes $-\,u(t,x)$. As for the double integral on the right-hand side, it takes place over the triangular region

$$T(\xi,\eta) = \{\,(\chi,\zeta)\ |\ \ \xi\le\chi\le\zeta\le\eta\,\}. \tag{2.94}$$

Let us introduce "physical" integration variables by setting

$$\chi = y - c\,s, \qquad \zeta = y + c\,s.$$

The defining inequalities of the triangle (2.94) become

$$x - ct \le y - c\,s \le y + c\,s \le x + ct,$$

and so, in the physical coordinates, the triangular integration domain assumes the form

$$D(t,x) = \{\,(s,y)\ |\ \ x - c\,(t-s)\le y\le x + c\,(t-s),\ 0\le s\le t\,\}, \tag{2.95}$$

which is graphed in Figure 2.29. The change of variables formula for double integrals requires that we compute the Jacobian determinant

$$\det\begin{pmatrix}\partial\chi/\partial y & \partial\chi/\partial s\\ \partial\zeta/\partial y & \partial\zeta/\partial s\end{pmatrix} = \det\begin{pmatrix}1 & -c\\ 1 & c\end{pmatrix} = 2\,c,$$

and so $d\chi\,d\zeta = 2\,c\,ds\,dy$. Therefore,

$$u(t,x) = \frac{1}{2\,c}\iint_{D(t,x)} F(s,y)\,ds\,dy = \frac{1}{2\,c}\int_0^t\int_{x-c\,(t-s)}^{x+c\,(t-s)} F(s,y)\,dy\,ds, \tag{2.96}$$

which gives the solution formula for the forced wave equation when subject to homogeneous initial conditions.

To solve the general initial value problem, we appeal to linear superposition, writing its solution as a sum of the solution (2.96) to the forced wave equation subject to homogeneous initial conditions plus the d'Alembert solution (2.82) to the unforced equation subject to inhomogeneous boundary conditions.

Theorem 2.18. *The solution to the general initial value problem*

$$u_{tt} = c^2 u_{xx} + F(t,x), \qquad u(0,x) = f(x), \qquad u_t(0,x) = g(x), \qquad -\infty < x < \infty, \qquad t>0,$$

for the wave equation subject to an external forcing is given by

$$u(t,x) = \frac{f(x-ct)+f(x+ct)}{2} + \frac{1}{2\,c}\int_{x-ct}^{x+ct} g(y)\,dy + \frac{1}{2\,c}\int_0^t\int_{x-c\,(t-s)}^{x+c\,(t-s)} F(s,y)\,dy\,ds. \tag{2.97}$$

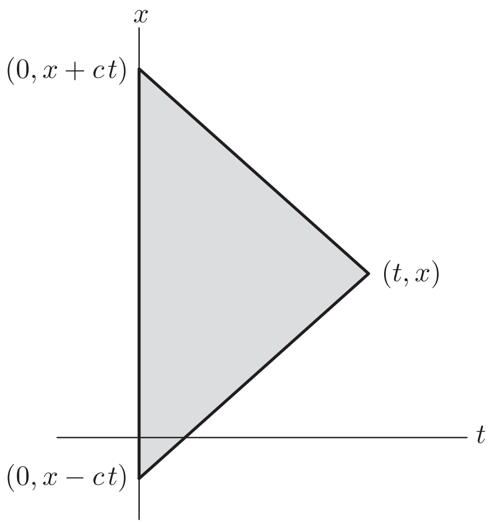

Figure 2.29. Domain of dependence.

Observe that the solution is a linear superposition of the respective effects of the initial displacement, the initial velocity, and the external forcing. The triangular integration region (2.95), lying between the x–axis and the characteristic lines going backwards from (t,x), is known as the *domain of dependence* of the point (t,x). This is because, for any $t>0$, the solution value $u(t,x)$ depends only on the values of the initial data and the forcing function at points lying within the domain of dependence $D(t,x)$. Indeed, the first term in the solution formula (2.97) requires only the initial displacement at the corners $(0,x+ct)$, $(0,x-ct)$; the second term requires only the initial velocity at points on the x–axis lying on the vertical side of $D(t,x)$; while the final term requires the value of the external force on the entire triangular region.

Example 2.19. Let us solve the initial value problem

$$u_{tt}=u_{xx}+\sin\omega t\ \sin x, \qquad u(0,x)=0, \qquad u_t(0,x)=0,$$

for the wave equation with unit wave speed subject to a sinusoidal forcing function whose amplitude varies periodically in time with frequency $\omega>0$. According to formula (2.96), the solution is

$$\begin{aligned} u(t,x) &= \frac{1}{2}\int_0^t \int_{x-t+s}^{x+t-s} \sin\omega s\,\sin y\,dy\,ds \\ &= \frac{1}{2}\int_0^t \sin\omega s\,\big[\cos(x-t+s)-\cos(x+t-s)\big]\,ds \\ &= \begin{cases} \dfrac{\sin\omega t-\omega\sin t}{1-\omega^2}\,\sin x, & 0<\omega\neq 1, \\[2ex] \dfrac{\sin t-t\cos t}{2}\,\sin x, & \omega=1. \end{cases} \end{aligned}$$

Notice that, when $\omega\neq 1$, the solution is bounded, being a combination of two vibrational modes: an externally induced mode at frequency ω along with an internal mode, at frequency 1. If $\omega=p/q\neq 1$ is a rational number, then the solution varies periodically in

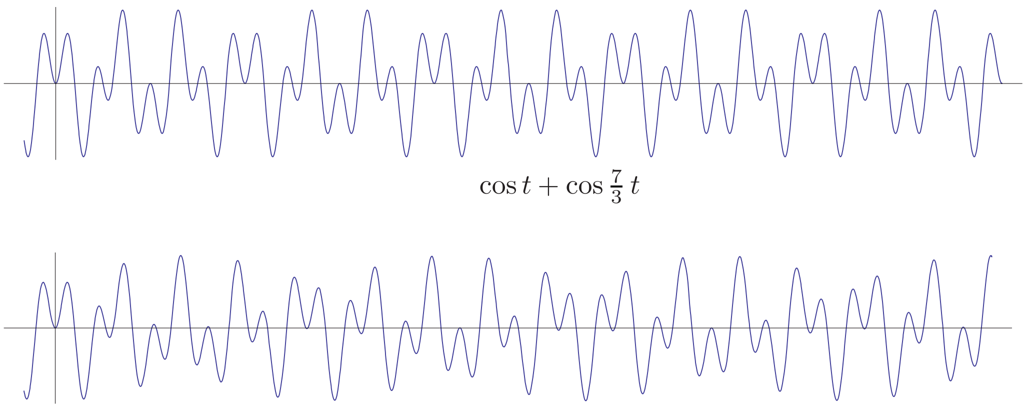

Figure 2.30. Periodic and quasiperiodic functions.

time. On the other hand, if ω is irrational, then the solution is only *quasiperiodic*, and never exactly repeats itself. Finally, if $\omega = 1$, the solution grows without limit as t increases, indicating that this is a *resonant frequency*. We will investigate external forcing and the mechanisms leading to resonance in dynamical partial differential equations in more detail in Chapters 4 and 6.

Example 2.20. To appreciate the difference between periodic and quasiperiodic vibrations, consider the elementary trigonometric function

$$u(t) = \cos t + \cos \omega t,$$

which is a linear combination of two simple periodic vibrations, of frequencies 1 and ω. If $\omega = p/q$ is a rational number, then $u(t)$ is a periodic function of period $2\pi q$, so $u(t+2\pi q) = u(t)$. However, if ω is an irrational number, then $u(t)$ is not periodic, and never repeats. You are encouraged to inspect the graphs in Figure 2.30. The first is periodic — can you spot where it begins to repeat? — whereas the second is only quasiperiodic. The only quasiperiodic functions we will encounter in this text are linear combinations of periodic trigonometric functions whose frequencies are *not* all rational multiples of each other. To the uninitiated, such quasiperiodic motions may appear to be random, even though they are built from a few simple periodic constituents. While ostensibly complicated, quasiperiodic motion is *not* true chaos, which is is an inherently nonlinear phenomenon, [**77**].

Exercises

2.4.1. Solve the initial value problem $u_{tt} = c^2 u_{xx}$, $u(0,x) = e^{-x^2}$, $u_t(0,x) = \sin x$.

2.4.2. (*a*) Solve the wave equation $u_{tt} = u_{xx}$ when the initial displacement is the box function $u(0,x) = \begin{cases} 1, & 1 < x < 2, \\ 0, & \text{otherwise}, \end{cases}$ while the initial velocity is 0.

(*b*) Sketch the resulting solution at several representative times.

2.4.3. Answer Exercise 2.4.2 when the initial velocity is the box function, while the initial displacement is zero.

2.4.4. Write the following solutions to the wave equation $u_{tt} = u_{xx}$ in d'Alembert form (2.82). *Hint*: What is the appropriate initial data?
(a) $\cos x \cos t$, (b) $\cos 2x \sin 2t$, (c) e^{x+t}, (d) $t^2 + x^2$, (e) $t^3 + 3tx^2$.

♡ 2.4.5. (a) Solve the *dam break problem*, that is, the wave equation when the initial displacement is a step function $\sigma(x) = \begin{cases} 1, & x > 0, \\ 0, & x < 0, \end{cases}$ and there is no initial velocity. (b) Analyze the case in which there is no initial displacement, while the initial velocity is a step function. (c) Are your solutions classical solutions? Explain your answer. (d) Prove that the step function is the limit, as $n \to \infty$, of the functions $f_n(x) = \dfrac{1}{\pi} \tan^{-1} nx + \dfrac{1}{2}$. (e) Show that, in both cases, the step function solution can be realized as the limit, as $n \to \infty$, of solutions to the initial value problems with the functions $f_n(x)$ as initial displacement or velocity.

◇ 2.4.6. Suppose $u(t,x)$ solves the initial value problem $u(0,x) = f(x)$, $u_t(0,x) = g(x)$, for the wave equation (2.66). Prove that the solution to the initial value problem $u(t_0,x) = f(x)$, $u_t(t_0,x) = g(x)$, is $u(t - t_0, x)$.

2.4.7. Find all resonant frequencies for the wave equation with wave speed c when subject to the external forcing function $F(t,x) = \sin \omega t \, \sin kx$ for fixed $\omega, k > 0$.

2.4.8. Consider the initial value problem $u_{tt} = 4u_{xx} + F(t,x)$, $u(0,x) = f(x)$, $u_t(0,x) = g(x)$. Determine (a) the domain of influence of the point $(0,2)$; (b) the domain of dependence of the point $(3,-1)$; (c) the domain of influence of the point $(3,-1)$.

2.4.9. (a) A solution to the wave equation $u_{tt} = 2u_{xx}$ is generated by a displacement concentrated at position $x_0 = 1$ and time $t_0 = 0$, but no initial velocity. At what time will an observer at position $x_1 = 5$ feel the effect of this displacement? Will the observer continue to feel an effect in the future? (b) Answer part (a) when there is an initial velocity concentrated at position $x_0 = 1$ and time $t_0 = 0$, but no initial displacement.

2.4.10. Suppose $u(t,x)$ solves the initial value problem $u_{tt} = 4u_{xx} + \sin \omega t \, \cos x$, $u(0,x) = 0$, $u_t(0,x) = 0$. Is $h(t) = u(t,0)$ a periodic function?

♡ 2.4.11. (a) Write down an explicit formula for the solution to the initial value problem

$$\frac{\partial^2 u}{\partial t^2} - 4\frac{\partial^2 u}{\partial x^2} = 0, \qquad u(0,x) = \sin x, \qquad \frac{\partial u}{\partial t}(0,x) = \cos x, \qquad -\infty < x < \infty, \quad t \ge 0.$$

(b) *True or false*: The solution is a periodic function of t.
(c) Now solve the forced initial value problem

$$\frac{\partial^2 u}{\partial t^2} - 4\frac{\partial^2 u}{\partial x^2} = \cos 2t, \qquad u(0,x) = \sin x, \qquad \frac{\partial u}{\partial t}(0,x) = \cos x, \qquad -\infty < x < \infty, \quad t \ge 0.$$

(d) *True or false*: The forced equation exhibits resonance. Explain.
(e) Does the answer to part (d) change if the forcing function is $\sin 2t$?

2.4.12. Given a classical solution $u(t,x)$ of the wave equation, let $E = \frac{1}{2}(u_t^2 + c^2 u_x^2)$ be the associated *energy density* and $P = u_t u_x$ the *momentum density*.
(a) Prove that $\partial P/\partial t = \partial E/\partial x$ and $\partial E/\partial t = c^2 \partial P/\partial x$. Explain why both E and P are conserved densities for the wave equation.
(b) Show that $E(t,x)$ and $P(t,x)$ both satisfy the wave equation.
(c) Suppose that both $u_t(t,x) \to 0$ and $u_x(t,x) \to 0$ as $|x| \to \infty$ sufficiently rapidly in order that the integrals defining the *total momentum* $\mathcal{P}(t) = \int_{-\infty}^{\infty} P(t,x)\,dx$ and the *total energy* $\mathcal{E}(t) = \int_{-\infty}^{\infty} E(t,x)\,dx$ are defined and finite for each $t \in \mathbb{R}$. Show that $\mathcal{P}(t)$ and $\mathcal{E}(t)$ are conserved quantities, i.e., they are constants, independent of the time t.

♢ 2.4.13. Let $u(t,x)$ be a classical solution to the wave equation $u_{tt} = c^2 u_{xx}$. The *total energy*

$$E(t) = \int_{-\infty}^{\infty} \frac{1}{2} \left[\left(\frac{\partial u}{\partial t} \right)^2 + c^2 \left(\frac{\partial u}{\partial x} \right)^2 \right] dx \tag{2.98}$$

represents the sum of kinetic and potential energies of the displacement $u(t,x)$ at time t. Suppose that $\nabla u \to \mathbf{0}$ sufficiently rapidly as $x \to \pm\infty$; more precisely, one can find $\alpha > \frac{1}{2}$ and $C(t) > 0$ such that $|\, u_t(t,x) \,|,\ |\, u_x(t,x) \,| \leq C(t)/|\, x \,|^\alpha$ for each fixed t and all sufficiently large $|\, x \,| \gg 0$. For such solutions, establish the *Law of Conservation of Energy* by showing that $E(t)$ is finite and constant. *Hint*: You do not need the formula for the solution.

♢ 2.4.14. (*a*) Use Exercise 2.4.13 to prove that the only classical solution to the initial-boundary value problem $u_{tt} = c^2 u_{xx}$, $u(0,x) = 0$, $u_t(0,x) = 0$, satisfying the indicated decay assumptions is the trivial solution $u(t,x) \equiv 0$. (*b*) Establish the following *Uniqueness Theorem* for the wave equation: there is at most one such solution to the initial-boundary value problem $u_{tt} = c^2 u_{xx}$, $u(0,x) = f(x)$, $u_t(0,x) = g(x)$.

2.4.15. The *telegrapher's equation* $u_{tt} + a\, u_t = c^2 u_{xx}$, with $a > 0$, models the vibration of a string under frictional damping. (a) Show that, under the decay assumptions of Exercise 2.4.13, the wave energy (2.98) of a classical solution is a nonincreasing function of t. (*b*) Prove uniqueness of such solutions to the initial value problem for the telegrapher's equation.

2.4.16. What happens to the proof of Theorem 2.14 if $c = 0$?

2.4.17. (*a*) Explain why the d'Alembert factorization method doesn't work when the wave speed $c(x)$ depends on the spatial variable x.
(*b*) Does it work when $c(t)$ depends only on the time t?

2.4.18. The *Poisson–Darboux equation* is $\dfrac{\partial^2 u}{\partial t^2} - \dfrac{\partial^2 u}{\partial x^2} - \dfrac{2}{x} \dfrac{\partial u}{\partial x} = 0$. Solve the initial value problem $u(0,x) = 0$, $u_t(0,x) = g(x)$, where $g(x) = g(-x)$ is an even function. *Hint*: Set $w = x\, u$.

♡ 2.4.19. (*a*) Solve the initial value problem $u_{tt} - 2\, u_{tx} - 3\, u_{xx} = 0$, $u(0,x) = x^2$, $u_t(0,x) = e^x$. *Hint*: Factor the associated linear differential operator. (*b*) Determine the domain of influence of a point $(0,x)$. (*c*) Determine the domain of dependence of a point (t,x) with $t > 0$.

♢ 2.4.20. (*a*) Use polar coordinates to prove that, for any $a > 0$,

$$\iint_{\mathbb{R}^2} e^{-a(x^2+y^2)}\, dx\, dy = \frac{\pi}{a}. \tag{2.99}$$

(*b*) Explain why

$$\int_{-\infty}^{\infty} e^{-a x^2}\, dx = \sqrt{\frac{\pi}{a}}. \tag{2.100}$$

♢ 2.4.21. Use Exercise 2.4.20 to prove the error function formulae (2.88).