Chapter 2

Linear and Nonlinear Waves

To begin exploring the vast mathematical continent that is partial differential equations, our first task is to understand simple 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. Most of our discussion will focus on dynamical models in a single space dimension, bearing in mind that most of the methods can be readily extended to higher dimensional situations. First order partial differential equations and systems model a wide variety of wave phenomena, including transport of solvents in fluids, flood waves, acoustics, gas dynamics, glacier motion, chromatography, traffic flow, and also a variety of biological and ecological systems. As always in mathematical analysis, one must be able to handle relatively tame linear examples before venturing into the nonlinear wilderness.

A basic solution technique relies on an inspired change of variables, rewriting the equation in a moving coordinate frame. In this manner, we are naturally led to the fundamental concept of a characteristic curve. Their physical relevance comes from the fact that signals and disturbances propagate along the characteristic curves in space-time. Indeed, their multi-dimensional counterparts are the light cones of special relativity, [**87**, **91**]. The characteristic method solves a first order *linear* partial differential equation by reducing it to a first order *nonlinear* ordinary differential equation!

In 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 propagate along the characteristics, but in the nonlinear case, characteristic curves may cross each other, precipitating the onset of a shock discontinuity. The characterization of the ensuing shock dynamics is *not* specified by the partial differential equation alone, but relies on additional physical information, in the form of a conservation law and entropy condition.

Having attained a basic understanding of first order wave dynamics, we then focus our attention on the second order wave equation in a single space dimension, used to model waves and vibrations in a violin string, a column of air in a clarinet, or an elastic bar. The wave equation is one of the fundamental partial differential equations that must be mastered in an introductory course. Its multi-dimensional counterparts serve to model vibrations of membranes, solid bodies, water waves, electromagnetic waves, including light, radio and micro-waves, acoustic waves, and many other physical systems. The one-dimensional wave equation is one of a small handful of physically relevant partial differential equations to be favored with an explicit solution formula, originally discovered by the eighteenth century French mathematician (and encyclopedist) Jean d'Alembert. His solution arises from "factoring" the second order wave equation into a pair of first order partial differential equations, of a type solved in the first half of this chapter. Here, we investigate the consequences of d'Alembert's solution formula for the initial value problem on the entire real line. Solutions on bounded intervals, governed by initial-boundary value problems, will be discussed in Chapter 4. Unfortunately, d'Alembert's method is of rather limited applicability, and does not extend beyond the one-dimensional case, nor to equations modeling vibrations of non-uniform media. The analysis of the wave equation in more than one space dimension will be deferred until later in the text, specifically Chapters 11 and 12.

2.1. Stationary Waves.

When beginning a new mathematical subject — in our case partial differential equations — it is incumbent on us to 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, to more and more complicated systems.

The very 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. The proof of this 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 also integrate both sides of the equation from 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

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

Thus, the solution 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 valid classical solution.

The solution (2.2) represents a *stationary wave* meaning that it does not change in time. The initial profile remains frozen in space, and the system is always in equilibrium. Figure 2.1 plots a representative solution as a function of x at three successive times.

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

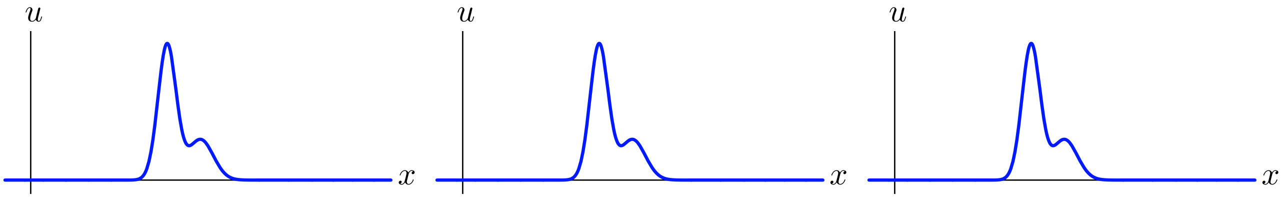

Figure 2.1. Stationary Wave.

The preceding analysis seems very straightforward and perhaps even a little boring. But, to be completely rigorous, we do 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 only defined 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 non-constant 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 only constant 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 only depend on x if we move from one connected component of D to another. Even this 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, $u(1,x) = x^2 \neq u(-1,x) = -x^2$.

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 ■, you are asked to justify these statements.

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

‡ *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. Later developments will reinforce the wisdom of this choice.

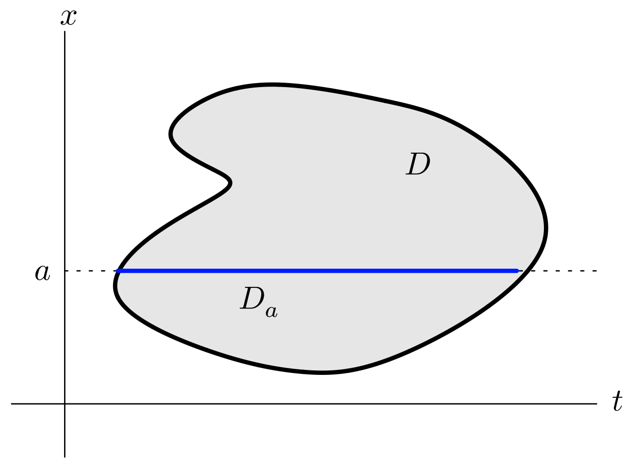

Figure 2.2. Domain for Stationary Wave Solution.

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

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 $u(t,x)$. 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, called the *wave speed*. 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 more challenging, but, as we will see, eminently doable.

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 $f \in \mathrm{C}^1$ is continuously differentiable, the initial conditions serve to specify a unique classical solution, at least for times t sufficiently near t_0. For simplicity, the initial time will be fixed at $t_0 = 0$. Indeed, by replacing the time variable t by $t - t_0$, we can, without loss of generality, always set $t_0 = 0$.

Uniform Transport

Let us begin by assuming that the wave speed c is constant. When confronted with a new equation, one possible solution procedure is to try to convert it into an equation you already know how to solve. In this case, we will introduce a simple change of variables that re-expresses the equation in a moving coordinate system, inspired by the fact that c represents the overall wave speed.

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

$$\xi = x - c\,t \tag{2.6}$$

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, non-relativistic 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 one of the first to champion such "relativistic" moving coordinate systems.

Thus, we will see what happens when we re-express the solution to the transport equation in terms of the moving coordinate frame. We rewrite

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

in terms of the so-called *characteristic variable* $\xi = x - c\,t$, 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**], to express the derivatives of u in terms of those of v:

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

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.

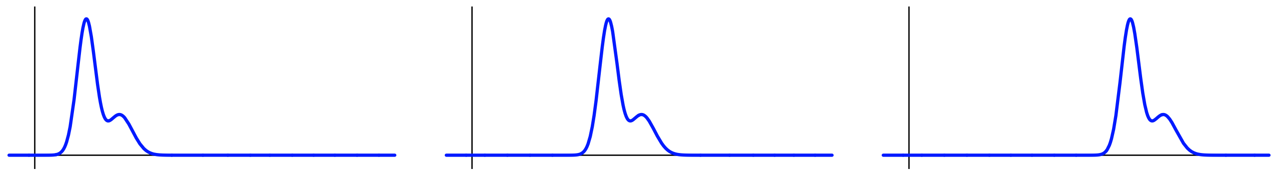

Figure 2.3. Traveling Wave.

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

$$u = v(\xi) = v(x - c\,t)$$

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

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

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

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

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

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

.

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, $(x - c\,t)^2 + 1$, or $\cos(x - c\,t)$, or $e^{x - c\,t}$, to the transport equation with constant wave speed c. And, in accordance with the counting principle, the general solution to this first order partial differential equation in two independent variables depends on one arbitrary function of a single variable.

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

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

and so provides the (unique) solution to the initial value problem (2.4, 12). 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 it only depends on the characteristic variable, every solution to the transport equation is constant on the *characteristic lines* of slope† c, namely

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

† 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.

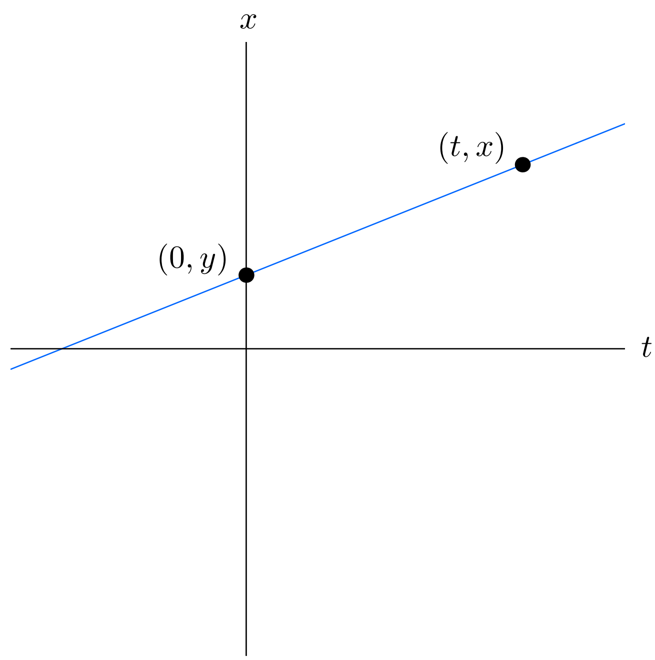

Figure 2.4. Characteristic Line.

where k is an arbitrary constant. At any given time t, the value of the solution at position x only depends on its previous value along the characteristic line passing through (t, x). 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) along the characteristic line $x = c\,t + y$ issuing from the initial point, as illustrated in Figure 2.4.

Transport with Decay

Let $a > 0$ be a positive constant, and c an arbitrary constant. The linear, homogeneous, 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 linear, homogeneous, first order ordinary differential equation, in which the characteristic variable ξ only enters parametrically. The standard solution

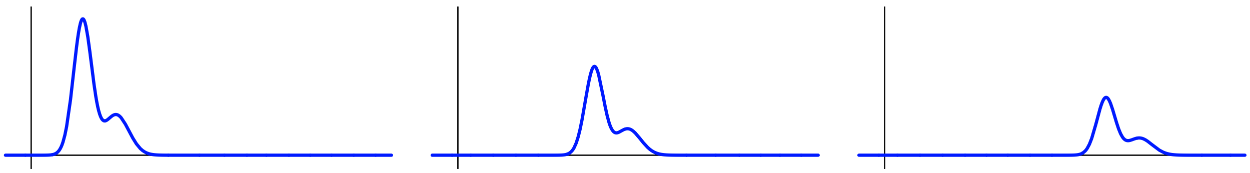

Figure 2.5. Decaying Traveling Wave.

technique learned in elementary ordinary differential equations, [**17**, **23**, **40**], tells us to multiply the equation by the exponential *integrating factor* $e^{a\,t}$, leading to

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

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

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

The resulting function 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.

In this case, the solution (2.15) is no longer constant along the characteristic lines. Nevertheless, signals continue to propagate along characteristics, since the value of the solution at an initial point $(0,y)$ will only affect its (decaying) values along the corresponding characteristic line $x = c\,t + y$.

Non–Uniform Transport

Slightly more complicated, but still linear, is the *non-uniform 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 non-constant, we can no longer expect them to be straight lines. To find the appropriate modification, let's look at how the solution varies along a given curve in the $t\,x$–plane. Assume the curve is identified with the graph of a function $x = x(t)$, and let

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

be the value of the solution along it. We compute the rate of change in the solution along the curve by differentiating h with respect to t. Invoking the multi-variable chain rule, [**8**]:

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

In particular, if $x(t)$ satisfies

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

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

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 our construction in the uniform case.

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

The characteristic curve equation (2.18) is an autonomous first order ordinary differential equation. As such, it can be immediately solved by separating variables, [**23**, **40**]. Assuming $c(x) \neq 0$, we divide both sides of the equation by $c(x)$, and then integrate the resulting equation:

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

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

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

where β^{-1} denotes 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 then the horizontal line $x \equiv x_\star$ is a stationary characteristic curve.

We've shown that the solution $u(t, x)$ is constant along the characteristic curves. Therefore, it must be a function of the *characteristic variable*

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

and hence of the form

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

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 function $v(\xi)$. Keep in mind that the algebraic solution formula (2.21) may fail to be valid at points where the wave speed vanishes: $c(x_\star) = 0$.

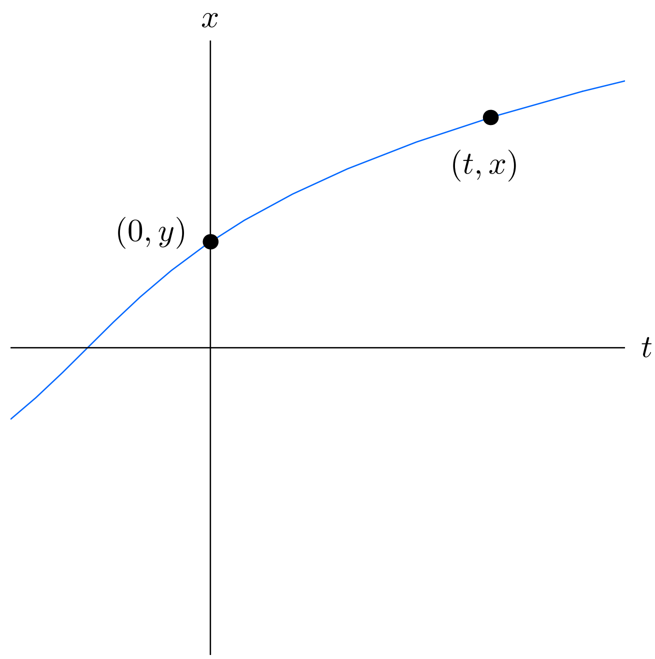

Figure 2.6. Characteristic Curve.

Warning: The definition of characteristic variable used here is a slightly different than what we used in the constant wave speed case, which, according to (2.20), would be $\xi = x/c - t = (x - c\,t)/c$. Clearly, rescaling the characteristic variable by $1/c$ is an unimportant modification of our original definition.

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

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

is not particularly enlightening, but does have a simple graphical interpretation: To find the value of the solution $u(t,x)$ at a given point, we look at the characteristic curve passing through it. If this curve intersects the x axis at the point $(0,y)$, shown in Figure 2.6, then $u(t,x) = u(0,y) = f(y)$, since the solution is constant along the curve. Incidentally, 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 particular non-uniform transport equation

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

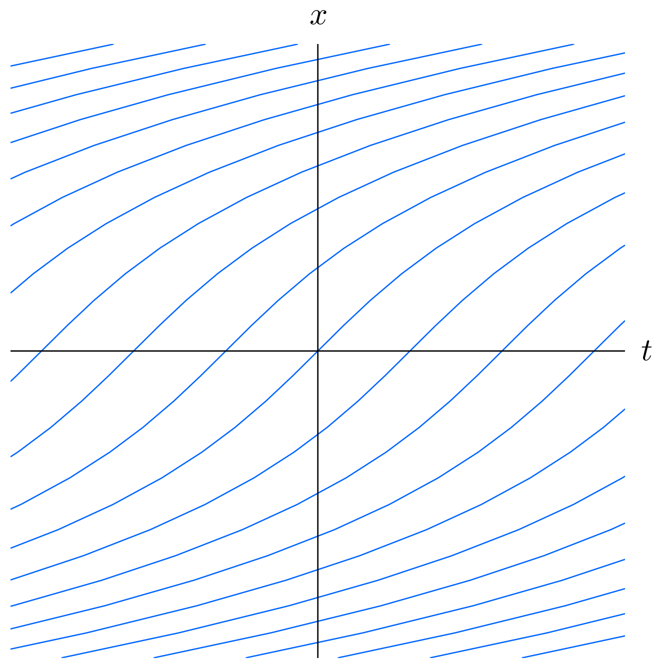

Figure 2.7. Characteristic Curves for $u_t + (x^2+1)^{-1} u_x = 0$.

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 find

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

where k is the integration constant. Representative curves are plotted in Figure 2.7. (In this case, inverting the function β is not pleasant.)

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

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

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

$$u(0,x) = \frac{1}{1+(x+2.75)^2}\,,$$

is plotted at times $t = 0, 2, 5, 10, 25$ and 50 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 non-uniform medium. In this example, since $c(x) > 0$ everywhere, the wave always moves from left to right; each part accelerates as it approaches the origin from the left, and then slows back down once it passes by. 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 $+\infty$.

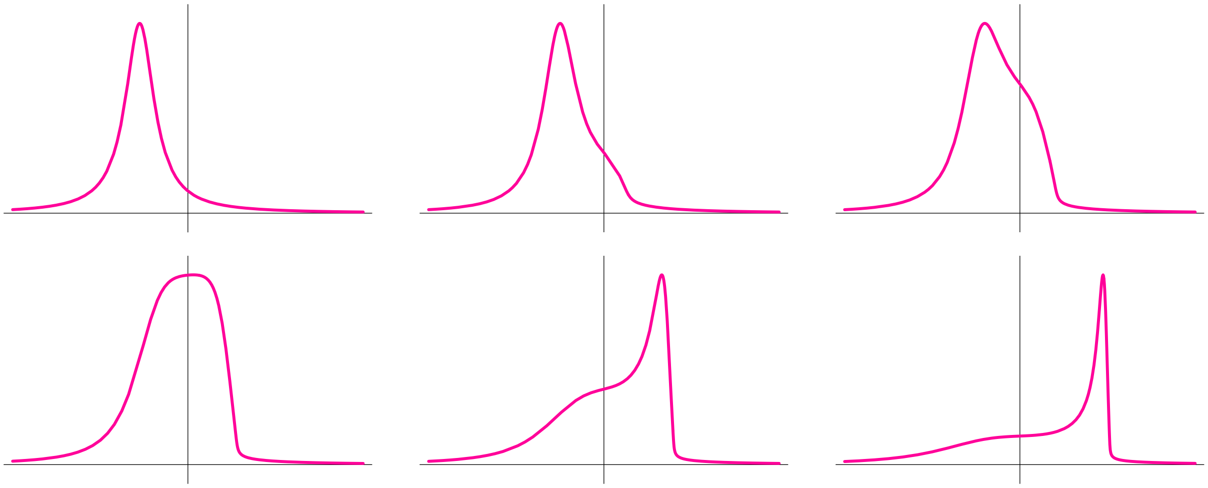

Figure 2.8. Solution to $u_t + \dfrac{1}{x^2+1}\, u_x = 0$.

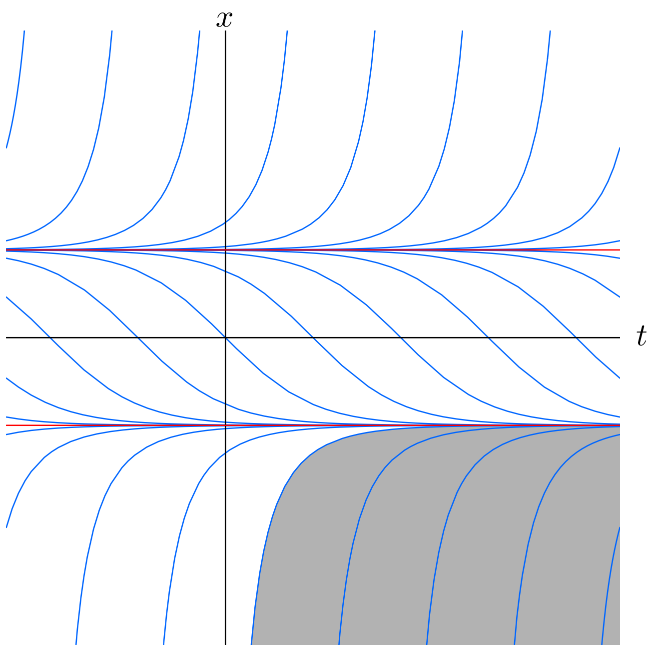

Figure 2.9. Characteristic Curves for $u_t + (x^2 - 1) u_x = 0$.

Example 2.5. Consider the non-uniform transport equation

$$u_t + (x^2 - 1) u_x = 0. \tag{2.25}$$

In this case, the characteristic curves are the solutions to

$$\frac{dx}{dt} = x^2 - 1, \tag{2.26}$$

and so

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

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, as $t \to \infty$, those curves starting below $x_+ = 1$ converge to the stable fixed point $x_- = -1$, 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. However, 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 is *not* prescribed by the initial data; these are the points in the shaded region in Figure 2.9. 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

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

As t increases the 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 (non-uniformly) to a step function:

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

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, [**17**, **23**, **62**], and assuming $c(x)$ is continuously differentiable†:

(*a*) There is a unique characteristic curve passing through each point $(t, x) \in \mathbb{R}^2$.

(*b*) Characteristic curves cannot cross each other.

(*c*) If $t = \beta(x)$ is a characteristic curve, then so are all its horizontal translates: $t = \beta(x) + k$ for any k.

(*d*) 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. A wave can never reverse its direction.

† For those who know about such things, this assumption can be weakened to just Lipschitz continuity.

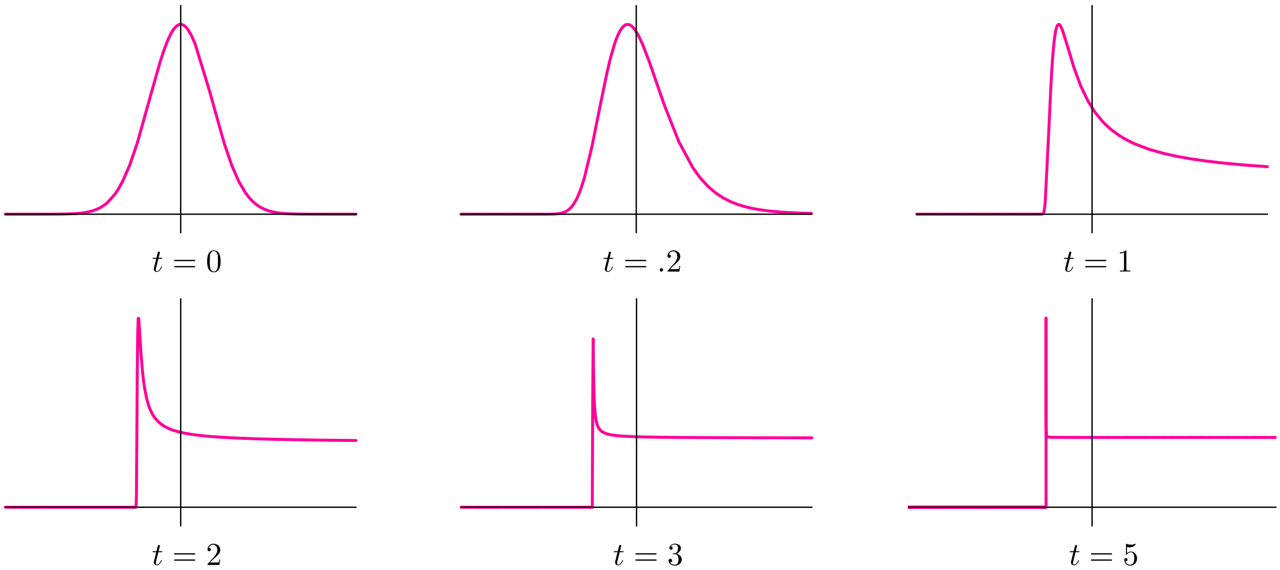

Figure 2.10. Solution to $u_t + (x^2 - 1)u_x = 0$.

(*e*) As t increases, the characteristic curve either asymptotes to a fixed point $x_\star$, with $c(x_\star) = 0$, as $t \to \infty$, or goes off to $\pm\infty$ in either finite or infinite time.

Proofs of these statements are assigned to the reader in the exercises.

2.3. Nonlinear Transport and Shocks.

Perhaps the simplest nonlinear partial differential equation is

$$u_t + u\,u_x = 0. \tag{2.27}$$

It 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 was first systematically studied by the famous mathematicians Siméon–Denis Poisson and Bernhard Riemann in the early nineteenth century. 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 many other areas. As a nonlinear partial differential equation, although we will be able to write down a solution formula, its full analysis is far from trivial, and will require us to confront the possibility of discontinuous shock waves and weak solutions. Motivated readers are referred to Whitham's book, [**132**], for further developments and details on applications.

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 (2.27) to be a solution to the ordinary differential equation

$$\frac{dx}{dt} = u(t,x). \tag{2.28}$$

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 for the given solution $u(t,x)$. As before, our task is to show that $h(t) = u(t,x(t))$, which is obtained by evaluating the solution along the curve, is constant. As usual, constancy is proved by checking that its derivative is identically zero. Repeating our chain rule computation (2.17), and using (2.28), we find

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

since u is assumed to solve the nonlinear transport equation (2.27) 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 (2.28) 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*,

$$x = u\,t + k, \tag{2.29}$$

whose slope equals the value assumed by the solution u on it. The larger u is, the steeper the characteristic line, and the faster that part of the wave travels.

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

$$\xi = x - t\,u \tag{2.30}$$

alone, and so

$$u = f(x - t\,u), \tag{2.31}$$

where $f(\xi)$ is an arbitrary C^1 function. Formula (2.31) is an algebraic equation that implicitly defines the solution $u(t,x)$ as a function of t and x.

Example 2.6. If

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

with α, β constant, then (2.31) becomes

$$u = \alpha(x - t\,u) + \beta, \qquad \text{and hence} \qquad u(t,x) = \frac{\alpha\,x + \beta}{1 + \alpha\,t} \tag{2.32}$$

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: $u(t,x) \to 0$

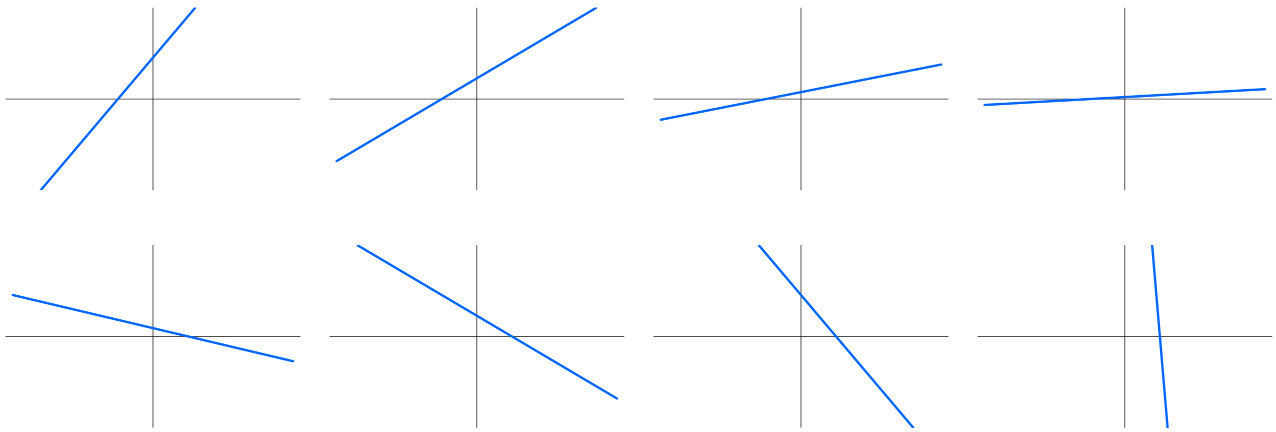

Figure 2.11. Two Solutions to $u_t + u\, u_x = 0$.

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 $\alpha = 1$, $\beta = .5$, 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 $t \to 5$, at which time it ceases to exist.

Remark: Although (2.32) is a valid solution formula after the blow-up time, $t > 5$, this is *not* to be considered as a part of the original solution. Solutions cease to exist after the appearance of such a singularity.

To solve the general initial value problem

$$u(0,x) = f(x), \tag{2.33}$$

we note that, at $t = 0$, the implicit solution formula (2.31) reduces to (2.33), and hence the function f coincides with the initial data. However, because our solution formula (2.31) 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 strategy is based on the following geometrical construction. Through each point $(0,y)$ on the x axis, draw the characteristic line

$$x = t\, f(y) + y \tag{2.34}$$

whose slope, namely $f(y) = u(0,y)$, equals the value of the initial data (2.33) at that point. According to the preceding discussion, the solution will have the same value on the entire characteristic line (2.34), and so

$$u(t, t\, f(y) + y) = f(y) \qquad \text{for all} \quad t. \tag{2.35}$$

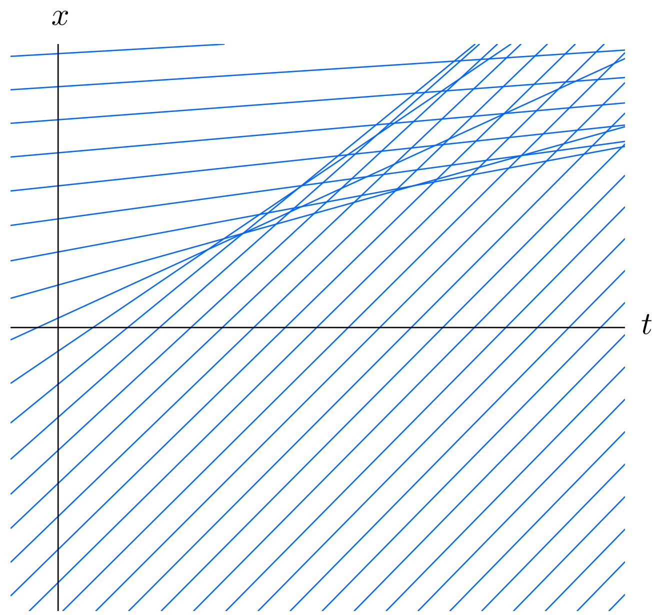

Figure 2.12. Characteristics for $u(0,x) = \frac{1}{2}\pi - \tan^{-1} x$.

For example, if $f(y) = y$, then $u(t,x) = y$ whenever $x = ty + y$; eliminating y, we recover $u(t,x) = x/(t+1)$, which agrees with one of our straight line solutions (2.32).

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

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

Two characteristic lines that are not parallel must cross each other somewhere. The value of the solution at a point is supposed equal to the slope of the characteristic line passing through the point. Hence, at such 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 analyze the equation in greater depth in order to resolve this paradox.

It turns out that there are three basic scenarios. The first, trivial situation is 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 trivial constant solution.

The next simplest case occurs when the initial data is everywhere non-decreasing, 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 $t \geq 0$. 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

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

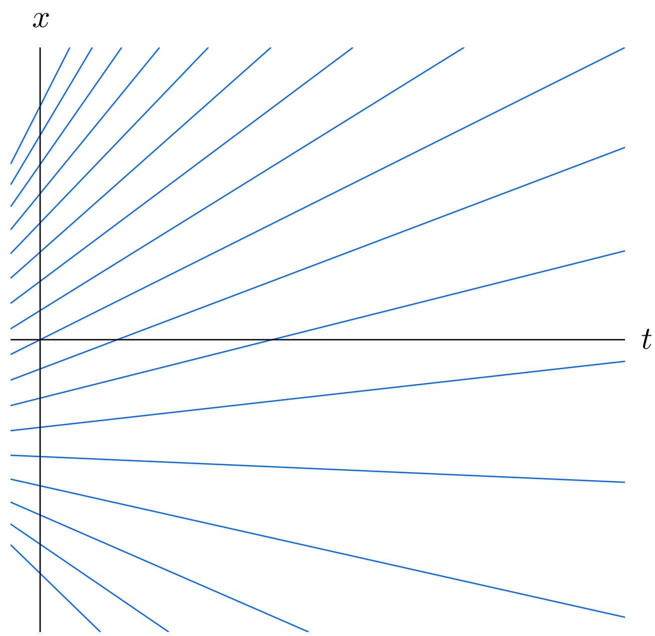

Figure 2.13. Characteristic Lines for a Rarefaction Wave.

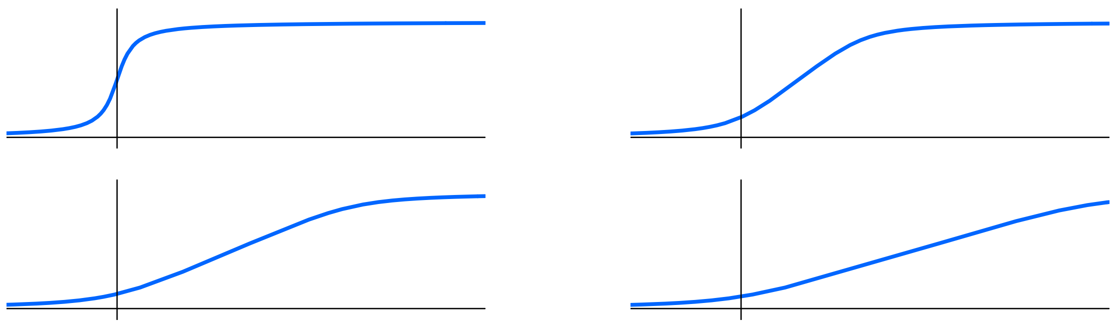

Figure 2.14. Rarefaction Wave.

has its characteristic lines plotted in Figure 2.13. Figure 2.14 graphs the solution at the successive times $t = 0, 1, 2$, and 3. Note how the rarefaction wave gradually spreads out as time increases.

The more interesting case is 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, velocity, pressure, etc., and must therefore assume a unique value at each point. The mathematical model has broken down, and no longer conforms with physical reality.

Before confronting this difficulty, let us first, from a purely theoretical standpoint, try to understand what happens if we were to mathematically continue the solution as a

Figure 2.15. Multiply–Valued Solution.

multiply-valued function. To be specific, consider the initial data

$$u(0,x) = \tfrac{1}{6}\,\pi - \tfrac{1}{3}\tan^{-1} x, \tag{2.36}$$

appearing in the first graph in Figure 2.15. The corresponding characteristic lines can be seen 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 which lie on the intersection of three 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 on 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 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 aforementioned wedge.

The critical time $t_\star$ can be determined from the implicit solution formula (2.31). Indeed, if we differentiate with respect to x, we find

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

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,

$$t_\star = \min\left\{ -\frac{1}{f'(x)} \;\middle|\; f'(x) < 0 \right\}. \tag{2.37}$$

For instance, for the particular initial configuration (2.36) 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

$$t_\star = \min\{\,1 + x^2\,\} = 1, \qquad \text{with} \qquad x_\star = f(0)\,t_\star = \tfrac{1}{6}\pi$$

giving the position of the vertical point on the graph.

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 are we trying to model.

To be specific, suppose the transport equation is viewed 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 down the end of a long pipe then the gas will move ahead of the piston and thereby be compressed. However, if the piston moves too rapidly, 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 formula (2.37) will give us the time of the onset of the shock wave. Our goal now is to predict its subsequent behavior.

Shock Dynamics

The way to resolve our mathematical dilemma is to appeal to a suitable physical conservation law. In many applications, the solution $u(t, x)$ represents density — density of gas in a pipe or traffic density on a highway. In such contexts, one expects mass to be conserved, even through a shock discontinuity — since even there atoms and cars can neither be created nor destroyed. (The traffic model does not allow for accidents!)

Before investigating the implications of conservation of mass for the motion of shocks, let's 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 at position x and time t, then the total mass in a interval $a \leq x \leq b$ at time t is calculated by integrating the density:

$$M_{a,b}(t) = \int_a^b u(t,x)\,dx, \tag{2.38}$$

or, equivalently, by the area under its graph. Assuming that u is a classical solution to the nonlinear transport equation (2.27), we can determine the rate of change of mass on this interval by differentiation:

$$\begin{aligned} \frac{dM_{a,b}}{dt} &= \frac{d}{dt}\int_a^b u(t,x)\,dx = \int_a^b \frac{\partial u}{\partial t}(t,x)\,dx \\ &= -\int_a^b u(t,x)\,\frac{\partial u}{\partial x}(t,x)\,dx = -\int_a^b \frac{\partial}{\partial x}\left[\,\tfrac{1}{2}\,u(t,x)^2\,\right] dx \\ &= -\,\tfrac{1}{2}\,u(t,x)^2\,\Big|_{x=a}^{b} = \tfrac{1}{2}\,u(t,a)^2 - \tfrac{1}{2}\,u(t,b)^2. \end{aligned} \tag{2.39}$$

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 the flux of mass through the endpoints; within the interval, mass can neither be created nor be destroyed. This is the precise content of a conservation law in continuum mechanics. In particular, if there is zero net mass flux, then the total mass is constant — mass is conserved. For example, if the initial data (2.33) has finite total mass,

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

which requires that $f(x) \to 0$ sufficiently 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:

$$\int_{-\infty}^{\infty} u(t,x)\,dx = \int_{-\infty}^{\infty} u(0,x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx. \tag{2.40}$$

Similarly, if $u(t,x)$ represents the traffic density on a highway at time t at position x, then the integrated conservation law (2.39) tells us that the rate of change in the number of vehicles on a stretch of road from a to b equals the number of vehicles entering at point a minus the number leaving at point b — which assumes there are no other exits or entrances on this part of the highway. Thus, in the traffic model, the conservation law (2.42) represents 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

$$\frac{\partial T}{\partial t} + \frac{\partial X}{\partial x} = 0. \tag{2.41}$$

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 upon derivatives of u, have only relatively recently become important, for so-called integrable partial differential equations, [**43**, **103**].) For example, the nonlinear transport equation (2.27) is a conservation law, since it can be written in the form

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\tfrac{1}{2}u^2\right) = 0, \tag{2.42}$$

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 (2.39) work. The general result, proved by an analogous computation, justifies calling (2.41) a conservation law.

Proposition 2.8. *Given a conservation law* (2.41), *on any closed interval* $a \le x \le b$,

$$\frac{d}{dt}\int_a^b T\,dx = -\,X\,\Big|_{x=a}^{b}. \tag{2.43}$$

Proof: The proof of is an immediate consequence of the Fundamental Theorem:

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

We will refer to (2.43) as the *integrated form* of the conservation law (2.41). 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 the same form, or its multi-dimensional generalization, [**103**].

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. This will turn out to (almost) uniquely prescribe the shock motion.

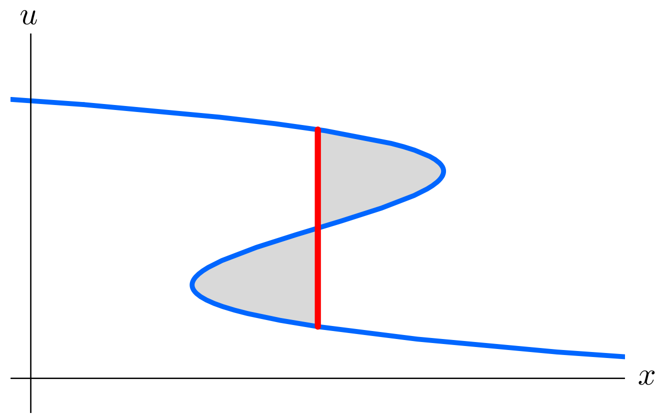

Figure 2.16. Equal Area Rule.

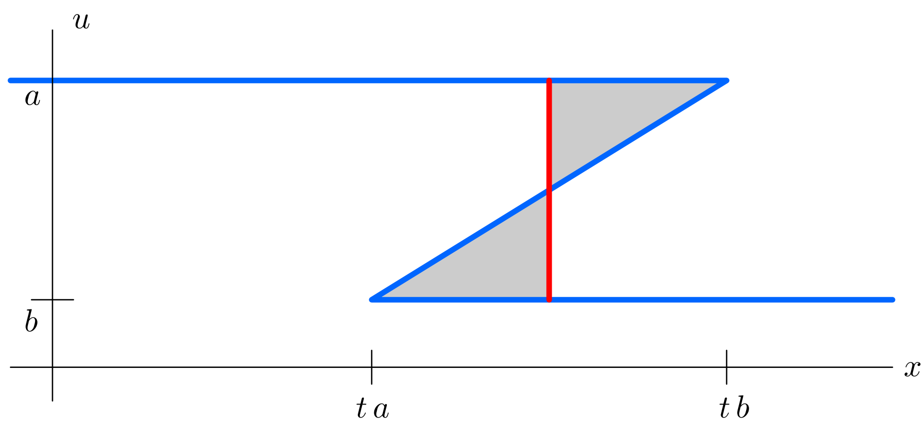

Figure 2.17. Multiply–Valued Step Wave.

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 uses a line integral $\displaystyle\int_C u\,dx$, 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 merely draws a vertical shock line where the areas of the two shaded lobes on either side are equal, as in Figure 2.16. This *Equal Area Rule* 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 is when the initial data has the form of a *step function* with a single discontinuity at the origin:

$$u(0,x) = \begin{cases} a, & x < 0, \\ b, & x > 0. \end{cases} \tag{2.44}$$

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-

† We are implicitly assuming that the mass if finite, as in (2.40), although the construction does not rely on this restriction.

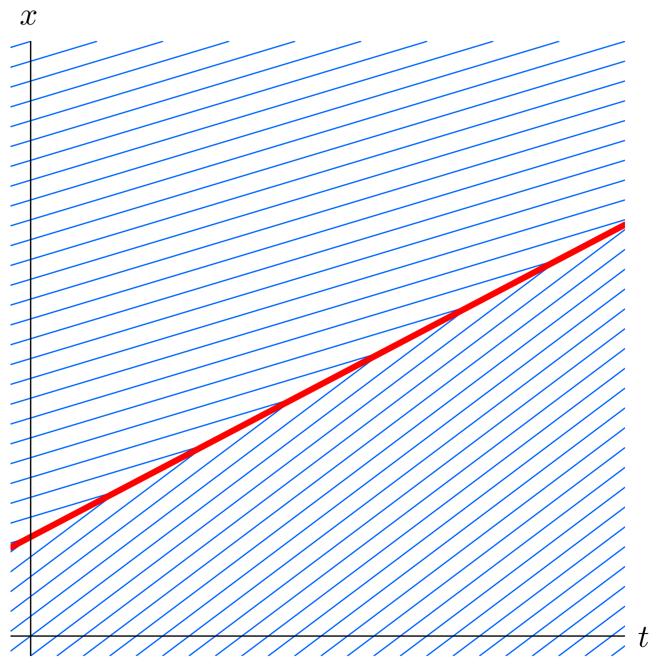

Figure 2.18. Characteristic Lines for the Step Wave Shock.

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 its characteristic a distance of $t u$. 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. This resulting shock wave solution is

$$u(t,x) = \begin{cases} a, & x < ct, \\ b, & x > ct, \end{cases} \qquad \text{where} \qquad c = \frac{a+b}{2}. \tag{2.45}$$

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 slopes.

The fact that the shock speed equals the *average* of the solution values on either side is, in fact, of general validity. This conclusion is known the *Rankine–Hugoniot condition*, named after two of the nineteenth century pioneers in the study of gas dynamics: the Scottish physicist/engineer William Rankine and the French engineer Pierre Hugoniot.

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 right and left hand limits*

$$u_-(t) = u(t,\sigma(t)^-) = \lim_{x \to \sigma(t)^-} u(t,x), \qquad u_+(t) = u(t,\sigma(t)^+) = \lim_{x \to \sigma(t)^+} u(t,x), \tag{2.46}$$

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:

$$\frac{d\sigma}{dt} = \frac{u_-(t) + u_+(t)}{2}. \tag{2.47}$$

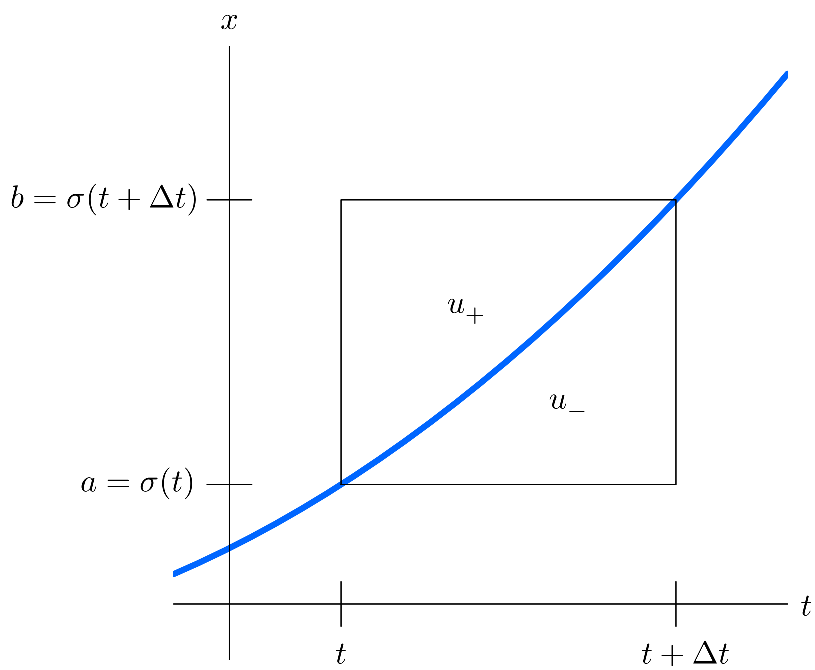

Figure 2.19. Conservation of Mass Near a Shock.

Proof: Referring to Figure 2.19, consider a small time interval, from t to $t + \Delta t$. 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

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

where we assume $\Delta t \ll 1$ is very small, and so the integrand is well approximated by its limiting value (2.46). Similarly, after the shock has passed, the total mass remaining in the interval is

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

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

$$\begin{aligned} \frac{dM}{dt} &= \lim_{\Delta t \to 0} \frac{M(t + \Delta t) - M(t)}{\Delta t} = \big[\,u_-(t) - u_+(t)\,\big] \lim_{\Delta t \to 0} \frac{\sigma(t + \Delta t) - \sigma(t)}{\Delta t} \\ &= \big[\,u_-(t) - u_+(t)\,\big]\,\frac{d\sigma}{dt}\;. \end{aligned}$$

On the other hand, at any $t < \tau < t + \Delta t$, the mass flux into the interval $[a, b]$ through the endpoints is, according to the right hand side of (2.39),

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

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

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

Solving for $d\sigma/dt$ establishes (2.47). *Q.E.D.*

Example 2.11. By way of contrast, consider the case when the initial data is a step function (2.44), 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 (2.32) produces the function

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

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. The resulting solution is the piecewise affine *rarefaction wave*

$$u(t,x) = \begin{cases} a, & x \le a\,t, \\ x/t, & a\,t \le x \le b\,t, \\ b, & x \ge b\,t, \end{cases} \tag{2.48}$$

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 (2.45). Which of the two competing solutions should we use? The first, (2.48), 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, cf. Exercise ■. Moreover, the discontinuous solution (2.45) 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 be done in a variety of ways. The discontinuous solution is found to violate *causality*, meaning that the solution profile at any given time uniquely prescribes its subsequent motion. Causality requires that characteristics are only allowed to terminate at a shock discontinuity; they cannot begin there, because their individual slopes cannot be uniquely prescribed by the shock profile. Causality requires that 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*

$$u_-(t) > \frac{d\sigma}{dt} = \frac{u_-(t) + u_+(t)}{2} > u_+(t). \tag{2.49}$$

It can be shown, [**68**], that, under the entropy condition (2.49), the rarefaction wave (2.48) is the unique solution to the initial value problem.

These prototypical solutions epitomize the basic phenomena modeled by the nonlinear transport equation: *rarefaction waves*, that emanate from regions where the initial data satisfies $f'(x) > 0$, where the solution spreads out as time progresses, and *compression waves*, emanting from regions where $f'(x) < 0$, that 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

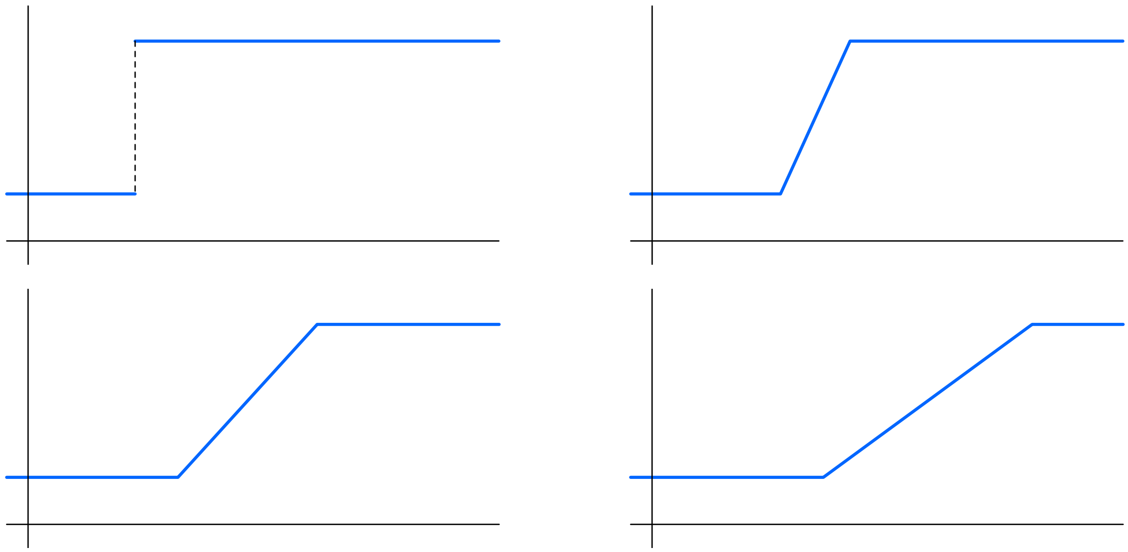

Figure 2.20. Rarefaction Wave.

interspersed rarefaction waves correspond to freely moving traffic. (An intelligent driver will take advantage of the rarefaction waves moving through the jam to switch lanes!) The familiar, frustrating traffic jam phenomenon, even on accident- or construction-free stretches of highway, is an intrinsic effect of the nonlinear transport model that governs the traffic flow, [**132**].

Continuing on past the initial shock formation, as other characteristic lines start to cross, additional shocks appear. 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 together to retain a physically consistent solution. However, at this point, the mathematical details have become too complicated for us to pursue in any more detail, and we refer the interested reader to Whitham's book, [**132**], which includes a wide range of applications to equations of gas dynamics, flood waves in rivers, motion of glaciers, chromotography, traffic flow, and many other physical systems. See also [**68**] for a proof of the following existence theorem for shock wave solutions to the nonlinear transport equation.

Theorem 2.12. *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 solution to the nonlinear transport equation* (2.27) *that also satisfies the Rankine–Hugoniot condition* (2.47) *and the entropy condition* (2.49).

Remark: Our derivation of the Rankine–Hugoniot shock speed condition (2.47) 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 (2.27) by u allows us write it in the alternative conservative

[†] See Section 3.2 for a precise definition.

form

$$u\,\frac{\partial u}{\partial t} + u^2\,\frac{\partial u}{\partial x} = \frac{\partial}{\partial t}\left(\tfrac{1}{2}\,u^2\right) + \frac{\partial}{\partial x}\left(\tfrac{1}{3}\,u^3\right) = 0. \tag{2.50}$$

In this formulation, the conserved density is $T = \frac{1}{2}\,u^2$, and the associated flux $X = \frac{1}{3}\,u^3$. The integrated form (2.43) of the conservation law (2.50) is

$$\frac{d}{dt}\int_a^b \tfrac{1}{2}\,u(t,x)^2\,dx = \tfrac{1}{3}\left[\,u(t,a)^3 - u(t,b)^3\,\right]. \tag{2.51}$$

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 only enter the interval as a flux through its ends. If we assume that energy is conserved at a shock, then, repeating our previous argument, we are led to an alternative equation

$$\frac{d\sigma}{dt} = \frac{\tfrac{1}{3}\left[\,u_-(t)^3 - u_+(t)^3\,\right]}{\tfrac{1}{2}\left[\,u_-(t)^2 - u_+(t)^2\,\right]} = \frac{2}{3}\;\frac{u_-(t)^2 + u_-(t)\,u_+(t) + u_+(t)^2}{u_-(t) + u_+(t)} \tag{2.52}$$

for the shock speed. Thus, a shock that conserves energy moves at a different speed than one that conserves mass! The evolution of a shock 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

$$u_t + c(u)\,u_x = 0, \tag{2.53}$$

whose wave speed is a more general function of the disturbance u. (Further extensions, allowing c to also depend on t and x are discussed in the exercises.) Most of the development is directly parallel to the special case (2.27) 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

$$\frac{dx}{dt} = c(u(t,x)). \tag{2.54}$$

As usual, 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,

$$u(0,x) = f(x), \tag{2.55}$$

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

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\,C(u) = 0, \qquad \text{where} \qquad C(u) = \int c(u)\,du \tag{2.56}$$

is any convenient anti-derivative of the wave speed. Thus, following the same computation as in (2.39), we discover that conservation of mass now takes the integrated form

$$\frac{d}{dt}\int_a^b u(t,x)\,dx = C(u(t,a)) - C(u(t,b)). \tag{2.57}$$

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

$$\frac{d\sigma}{dt} = \frac{C(u_-(t)) - C(u_+(t))}{u_-(t) - u_+(t)}\,. \tag{2.58}$$

Note that if

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

and so (2.58) reduces to our earlier rule (2.47). Moreover, as the shock magnitude $u_-(t) - u_+(t) \to 0$, the right hand side of (2.58) converges to the derivative $C'(u) = c(u)$ and hence recovers the wave speed, as it should.

2.4. The Wave Equation — d'Alembert's Solution.

Newton's Law, that force equals mass times acceleration, is absolutely fundamental to 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 model governing small vibrations is the second order partial differential equation

$$\rho(x)\,\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left(\kappa(x)\,\frac{\partial u}{\partial x}\right), \tag{2.59}$$

first written down by d'Alembert. Here $u(t,x)$ represents the displacement of the string (or bar) at position x and time t, while $\rho(x) > 0$ indicates its density and $\kappa(x) > 0$ denotes its stiffness or tension, which are here assumed to not 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 times acceleration. A detailed derivation of the model from first principles can be found in [**131**].

The general vibration equation (2.59) is too complicated at this early stage in our study of partial differential equations. We will thus concentrate on the simplest situation, in which the medium is homogeneous, and so both its density and stiffness are constant. The general one-dimensional vibration equation (2.59) reduces to the *wave equation*

$$\frac{\partial^2 u}{\partial t^2} = c^2\,\frac{\partial^2 u}{\partial x^2}\,, \qquad \text{where the constant} \qquad c = \sqrt{\frac{\kappa}{\rho}}\ >\ 0 \tag{2.60}$$

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

To uniquely specify the solution to any dynamical system arising from Newton's Law, one must fix both its initial position and initial velocity. Thus, the initial conditions for the one-dimensional wave equation (2.60) (or for the more general vibration equation (2.59)) take the form

$$u(0,x) = f(x), \qquad \frac{\partial u}{\partial t}(0,x) = g(x), \tag{2.61}$$

where, for simplicity, we set the initial time $t_0 = 0$. (See also Exercise ■.) The *initial value problem* seeks the corresponding C^2 function $u(t,x)$ that solves the wave equation (2.60) and has the required initial values (2.61). In this section, we will learn how to solve the initial value problem (2.60–61) 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 higher dimensional version of the wave equation appears in Chapters 11 and 12.

d'Alembert's Solution

Let us now derive the explicit solution formula for the basic wave equation (2.60) first found by d'Alembert. The starting point is to write the equation in the suggestive form

$$\Box\, u = (\partial_t^2 - c^2\,\partial_x^2)\, u = u_{tt} - c^2\, u_{xx} = 0. \tag{2.62}$$

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†:

$$\Box = \partial_t^2 - c^2\,\partial_x^2 = (\partial_t - c\,\partial_x)\,(\partial_t + c\,\partial_x). \tag{2.63}$$

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

$$(\partial_t + c\,\partial_x)\, u = u_t + c\, u_x = 0, \tag{2.64}$$

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 (2.64) as the uniform transport equation, cf. (2.4), with constant wave speed c. Proposition 2.1 tells us that its solutions are traveling waves with wave speed c:

$$u(t,x) = p(\xi) = p(x - c\,t), \tag{2.65}$$

† The cross terms cancel thanks to the equality of mixed partial derivatives: $\partial_t\partial_x u = \partial_x\partial_t u$.

where p is an arbitrary function of the characteristic variable $\xi = x - c\,t$. As long as $p \in \mathrm{C}^2$ (i.e., twice continuously differentiable), the resulting function $u(t,x)$ is a classical solution to the wave equation (2.60), as you can easily check.

Now, the factorization (2.63) can equally well be written in reverse order:

$$\Box = \partial_t^2 - c^2\,\partial_x^2 = (\partial_t + c\,\partial_x)\,(\partial_t - c\,\partial_x). \tag{2.66}$$

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

$$u_t - c\,u_x = 0, \tag{2.67}$$

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 (2.67) has the form

$$u(t,x) = q(\eta) = q(x + c\,t) \tag{2.68}$$

where q is an arbitrary function of the alternative characteristic variable $\eta = x + c\,t$. The solutions (2.68) represent traveling waves moving to the *left* with constant speed $c > 0$. Provided $q \in \mathrm{C}^2$, the functions (2.68) will provide a second class of solutions to the wave equation.

We conclude that the wave equation (2.62) 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 which are superpositions of left and right traveling waves. The remarkable fact is that *every* solution to the wave equation can be so represented.

Theorem 2.13. *Every solution to the wave equation* (2.60) *can be written as a superposition,*

$$u(t,x) = p(\xi) + q(\eta) = p(x - c\,t) + q(x + c\,t), \tag{2.69}$$

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

$$\xi = x - c\,t, \qquad\qquad \eta = x + c\,t. \tag{2.70}$$

Proof: As when solving the transport equation, we simplify the wave equation through an inspired change of variables. In this case, the new independent variables will be the two characteristic variables ξ, η defined by (2.70). We set

$$u(t,x) = v(x - c\,t, x + c\,t) = v(\xi,\eta), \qquad \text{whereby} \qquad v(\xi,\eta) = u\left(\frac{\eta-\xi}{2\,c}, \frac{\eta+\xi}{2}\right). \tag{2.71}$$

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

$$\frac{\partial u}{\partial t} = c\left(\frac{\partial v}{\partial \xi} - \frac{\partial v}{\partial \eta}\right), \qquad\qquad \frac{\partial u}{\partial x} = \frac{\partial v}{\partial \xi} + \frac{\partial v}{\partial \eta}\,. \tag{2.72}$$

and, further,

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

Therefore
$$\Box\, u = \frac{\partial^2 u}{\partial t^2} - c^2\, \frac{\partial^2 u}{\partial x^2} = -\,4\,c^2\, \frac{\partial^2 v}{\partial \xi\, \partial \eta}\,. \tag{2.73}$$
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 integration "constant". Replacing the characteristic variables by their formulae in terms of t and x completes the proof. *Q.E.D.*

Let us now see how this can be used to solve the initial value problem (2.61). Substituting the solution formula (2.69) into the initial conditions, we find
$$u(0,x) = p(x) + q(x) = f(x), \qquad \frac{\partial u}{\partial t}\,(0,x) = -\,c\,p'(x) + c\,q'(x) = g(x). \tag{2.74}$$
To solve this pair of equations for 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,
$$p(x) = \frac{1}{2}\, f(x) - \frac{1}{2\,c} \int_0^x g(z)\, dz + a,$$
where a is an integration constant. The first equation in (2.74) then yields
$$q(x) = f(x) - p(x) = \frac{1}{2}\, f(x) + \frac{1}{2\,c} \int_0^x g(z)\, dz - a.$$

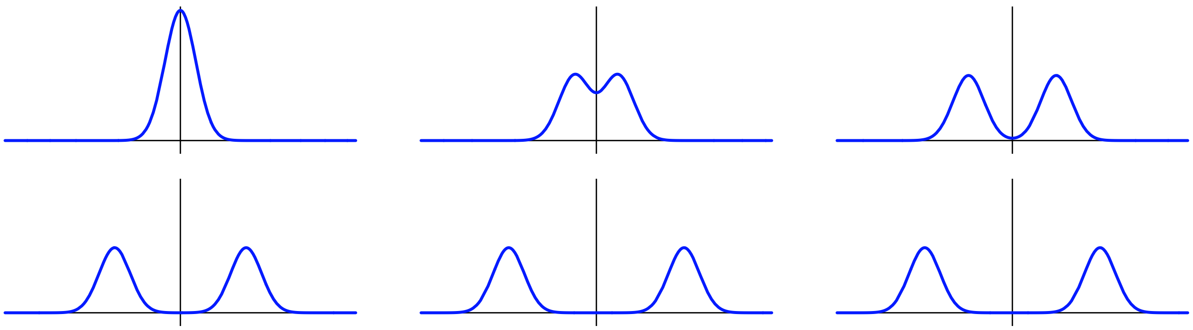

Figure 2.21. Splitting of Waves.

Substituting these two expressions back into our solution formula (2.69), we find

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

where ξ, η are the characteristic variables (2.70). In this manner, we deduce *d'Alembert's solution* to the initial value problem for the wave equation on the entire line $-\infty < x < \infty$.

Theorem 2.14. *The solution to the initial value problem*

$$\frac{\partial^2 u}{\partial t^2} = c^2\,\frac{\partial^2 u}{\partial x^2}, \qquad u(0,x) = f(x), \qquad \frac{\partial u}{\partial t}(0,x) = g(x), \qquad -\infty < x < \infty, \tag{2.75}$$

is given by

$$u(t,x) = \frac{f(x-c\,t)+f(x+c\,t)}{2} + \frac{1}{2\,c}\int_{x-c\,t}^{x+c\,t} g(z)\,dz. \tag{2.76}$$

Remark: In order that (2.76) 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 non-classical, as they do fit under the more general concept of "weak solution", [**73, 106**].

Example 2.15. Suppose there is no initial velocity, so $g(x) \equiv 0$, so the motion is purely the result of the initial displacement $u(0,x) = f(x)$. In this case, (2.76) reduces to

$$u(t,x) = \tfrac{1}{2}\,f(x-c\,t) + \tfrac{1}{2}\,f(x+c\,t). \tag{2.77}$$

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

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

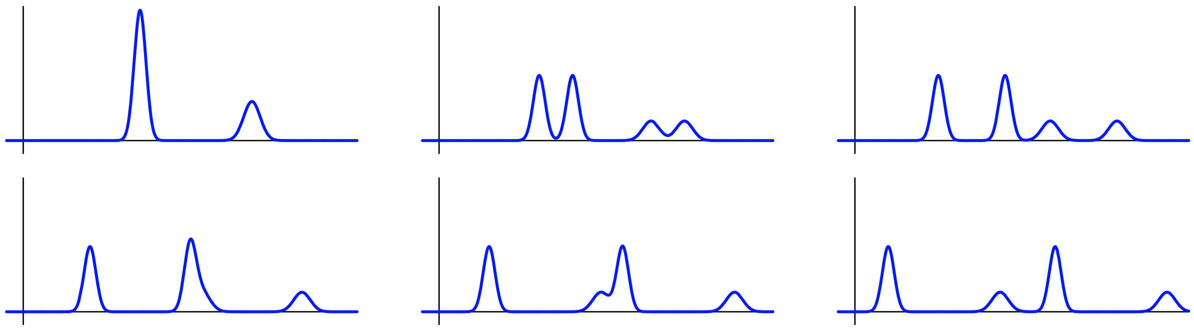

Figure 2.22. Interaction of Waves.

then the solution

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

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 2.21.

If we take two separated pulses, say

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

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

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

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 2.22, 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 go off in tandem in opposing directions, having 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 right and left-moving waves will separate, and the observer will see two half size replicas running away, with speed c, in opposite directions. If the displacement 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

$$\cos c t \;\; \cos x = \tfrac{1}{2} \cos(x - ct) + \tfrac{1}{2} \cos(x + ct). \tag{2.78}$$

In accordance with the left hand expression, an observer will see a cosinusoidal wave $\cos x$, standing still while vibrating with frequency c. However, the d'Alembert form of the

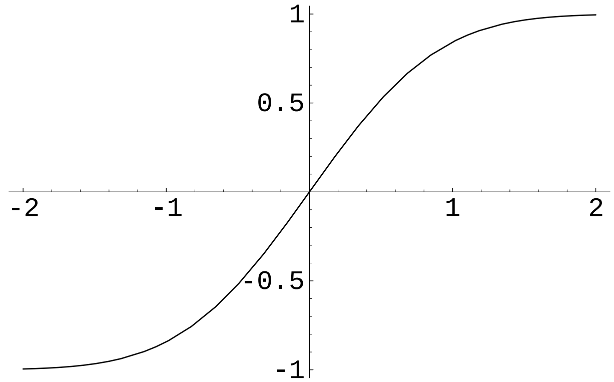

Figure 2.23. The Error Function.

solution on the right hand side says that this is just the sum of right and left traveling cosine waves! The interactions of their peaks and troughs reproduce the standing wave. So, the same solution can be interpreted in two seemingly incompatible ways!

Example 2.16. 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 represents a violin string at rest being struck by some sort of hammer at the initial time. In this case, the d'Alembert formula (2.76) reduces to

$$u(t,x) = \frac{1}{2\,c} \int_{x-c\,t}^{x+c\,t} g(z)\,dz. \tag{2.79}$$

Note that, at time $t > 0$, the effect of the hammer blow is felt along an entire interval $[x - c\,t, x + c\,t]$. For example, given $u(0,x) = 0$, $u_t(0,x) = e^{-x^2}$, the solution (2.79) is

$$u(t,x) = \frac{1}{2\,c} \int_{x-c\,t}^{x+c\,t} e^{-x^2}\,dz = \frac{\sqrt{\pi}}{4\,c} \bigl[\,\mathrm{erf}(x + c\,t) - \mathrm{erf}(x - c\,t)\,\bigr], \tag{2.80}$$

where

$$\mathrm{erf}\,x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2}\,dz \tag{2.81}$$

is known as the *error function* due to its applications throughout probability and statistics, [**46**]. The error function integral cannot be written in terms of elementary functions. Nevertheless, its importance in a wide range of applications means that its properties have been well studied, and its values tabulated, [**3**]. A graph appears in Figure 2.23. The constant in the front of the integral (2.81) has been chosen so that the error function has asymptotic values

$$\lim_{x \to \infty} \mathrm{erf}\,x = 1, \qquad \lim_{x \to -\infty} \mathrm{erf}\,x = -1, \tag{2.82}$$

which follow from a well known integration formula, cf. Exercise ■.

A graph of the solution (2.80) at successive times is displayed in Figure 2.24. 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 at equal and opposite velocities $\pm c$. Thus, unlike the case of a non-zero

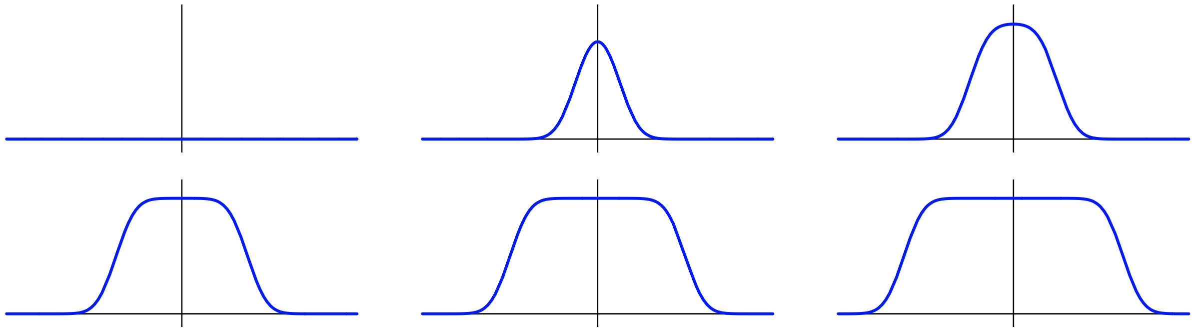

Figure 2.24. Error Function Solution to the Wave Equation.

initial displacement in Figure 2.21, where the solution eventually returns to its equilibrium position $u = 0$ after the wave passes by, for a nonzero initial velocity, the string remains permanently deformed.

In general, the lines of slope $\pm c$ in the (t,x)–plane, where the characteristic variables are constant,

$$\xi = x - c\,t = a, \qquad \eta = x + c\,t = b, \tag{2.83}$$

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 for physicists: The characteristic lines in space-time are the one-dimensional counterparts of the light cone in Minkowski space appearing in special relativity, [**91**]. See Section 12.5 for further details.

In Figure 2.25, we plot the two characteristics going through a point $(0,y)$ on the x axis. The wedge-shaped region $\{\, y - c\,t \leq x \leq y + c\,t,\ t \geq 0 \,\}$ 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 that point will only affect the solution values in its domain of influence. Indeed, the effect of the initial displacement at the point y propagates along the two characteristic lines, while the effect of an initial velocity at the point $(y,0)$ will be felt at every point in the triangular wedge.

External Forcing and Resonance

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

$$\frac{\partial^2 u}{\partial t^2} = c^2\,\frac{\partial^2 u}{\partial x^2} + F(t,x). \tag{2.84}$$

representing the external force 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 equation is supplemented by homogeneous initial conditions,

$$u(0,x) = 0, \qquad u_t(0,x) = 0, \tag{2.85}$$

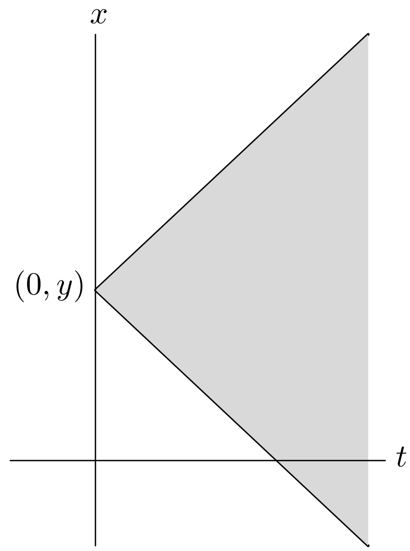

Figure 2.25. Characteristic Lines and Domain of Influence.

and so there is no initial displacement nor velocity. To solve the initial value problem (2.84–85), we switch to the same characteristic coordinates (2.70), setting

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

Applying the chain rule formulae (2.73), we find that the forced equation (2.84) becomes

$$\frac{\partial^2 v}{\partial\xi\,\partial\eta} = -\,\frac{1}{4\,c^2}\,F\left(\frac{\eta-\xi}{2\,c}\,,\frac{\eta+\xi}{2}\right). \tag{2.86}$$

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

$$\frac{\partial v}{\partial\xi}\,(\xi,\eta) - \frac{\partial v}{\partial\xi}\,(\xi,\xi) = -\,\frac{1}{4\,c^2}\int_\xi^\eta F\left(\frac{\tau-\xi}{2\,c}\,,\frac{\tau+\xi}{2}\right)d\tau. \tag{2.87}$$

But, recalling (2.72),

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

and so, in particular,

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

due to our choice of initial conditions (2.85), which imply that both $\dfrac{\partial u}{\partial t}\,(0,x) = 0$ and, by differentiating the first initial condition $u(0,x) = 0$ with respect to x, also $\dfrac{\partial u}{\partial x}\,(0,x) = 0$ for all x, including $x = \xi$. As a result, (2.87) simplifies to

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

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

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

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

$$T(\xi,\eta) = \{\, (\sigma,\tau) \mid \ \xi \le \sigma \le \tau \le \eta \,\}. \tag{2.88}$$

Let us revert to physical integration variables by setting

$$\sigma = y - c\,s, \qquad \tau = y + c\,s.$$

The triangle (2.88) becomes

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

which can be rewritten as

$$D(t,x) = \{\, (s,y) \mid \ x - c\,(t-s) \le y \le x + c\,(t-s),\ 0 \le s \le t \,\}. \tag{2.89}$$

The change of variables formula for double integrals requires that we compute the Jacobian determinant

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

and so $d\sigma\, d\tau = 2\,c\, ds\, dy$. Therefore,

$$u(t,x) = \frac{1}{2\,c}\iint_{D(t,x)} F(s,y)\, ds\, dy = \frac{1}{2\,c}\int_0^t \int_{x-c\,(t-s)}^{x+c\,(t-s)} F(s,y)\, dy.\, ds \tag{2.90}$$

This gives the solution to 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 (2.90) to the forced wave equation subject to homogeneous initial conditions plus the d'Alembert solution (2.76) to the unforced equation subject to inhomogeneous boundary conditions.

Theorem 2.17. *The solution to the initial value problem*

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

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

$$u(t,x) = \frac{f(x-c\,t) + f(x+c\,t)}{2} + \frac{1}{2\,c}\int_{x-c\,t}^{x+c\,t} g(y)\, dy + \frac{1}{2\,c}\int_0^t \int_{x-c\,(t-s)}^{x+c\,(t-s)} F(s,y)\, dy\, ds. \tag{2.91}$$

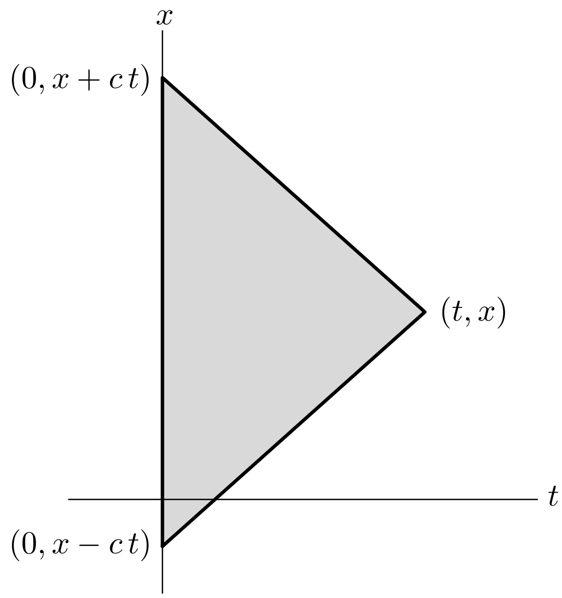

Figure 2.26. Domain of Dependence.

The triangular integration region (2.89), lying between the x axis and the characteristic lines going backwards from the point (t, x) is known as the *domain of dependence* of the point; see Figure 2.26. This is because, for $t > 0$, 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 (2.91) requires only the initial displacement at the corners $(0, x + c\,t)$, $(0, x - c\,t)$; 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 2.18. 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 continually subject to a sinusoidal forcing function whose amplitude varies periodically in time at frequency $\omega > 0$. According to formula (2.90), the solution is

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

Notice that, when $0 < \omega \neq 1$, the solution is bounded, being a combination of two vibrational modes: an externally induced mode at frequency ω and an internal mode, at frequency 1. If $\omega = p/q \neq 1$ is a rational number, then the solution varies periodically in time. On the other hand, if ω is irrational, then the solution is only *quasi-periodic*, 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 partial differential equations in more detail in Chapters 4 and 5.

