Chapter 22

Nonlinear Partial Differential Equations

The last topic to be touched on in this book is the vast and active contemporary research area of nonlinear partial differential equations. Leaving aside quantum mechanics, which remains a purely linear theory, most real-world physical systems, including gas dynamics, fluid mechanics, elasticity, relativity, biology, thermodynamics, and so on, are modeled by nonlinear partial differential equations. Attempts to survey even a tiny fraction of such a all-encompassing range of phenomena, methods, results, and mathematical developments, are necessarily doomed to failure. So we will concentrate on a handful of prototypical, but very important examples, arising in the study of nonlinear waves and heat conduction. Specific topics include shock waves, blow up, similarity solutions, and solitons. We will only be able to consider nonlinear partial differential equations modeling dynamical behavior in one (space) dimension. The much more complicated nonlinear systems that govern our three-dimensional dynamical universe remain on the cutting edge of contemporary research activity.

Historically, i.e., before the advent of high powered computers, relatively little was known about the extraordinary range of behavior exhibited by nonlinear partial differential equations. Most of the most basic phenomena that now drive modern-day research, such as solitons, chaos, stability, blow-up, singularities, asymptotics, etc., remained undetected, or only dimly outlined. The last fifty years has witnessed a remarkable blossoming of our understanding, due in large part to the advent of large scale computing and significant advances in numerical methods for integrating nonlinear systems. Numerical experimentation suddenly exposed many unexpected phenomena, including chaos and solitons, to the light of day. New analytical methods, new mathematical theories, and new computational algorithms have precipitated this revolution in our understanding and study of nonlinear systems, an activity that continues to grow in intensity and breadth. Each leap in computing power and theoretical advances has led to yet deeper understanding of nonlinear phenomena, but also points out how far we have yet to go. To make sense of this bewildering variety of methods, equations, and results, it is essential build upon a firm foundation of, first of all, linear systems theory, and secondly, nonlinear algebraic and ordinary differential equations.

We arrange our presentation according to the order of the underlying differential equation. First order nonlinear partial differential equations govern nonlinear waves and vibrations. Such nonlinear wave motions arise in gas dynamics, water waves, elastodynamics, chemical reactions, flood waves in rivers, chromatography, traffic flow, and a range of biological and ecological systems. One of the most important nonlinear phenomena, with no linear counterpart, is the break down of solutions in finite time, resulting in the forma-
tion of discontinuous shock waves. A striking example is the supersonic boom produced by an airplane that breaks the sound barrier. As in the linear wave equation, the signals propagate along the characteristics, but in the nonlinear case the characteristics can cross each other, indicating the onset of a shock wave.

Second order partial differential equations govern nonlinear diffusion processes, including heat flow and population dynamics. The simplest and most important equation, known as Burgers’ equation, can, surprisingly, be linearized by transforming it to the heat equation. This accident provides an essential glimpse into the world of nonlinear diffusion processes. As we discover, as the diffusion or viscosity tends to zero, the solutions to burg- ers’ equation tend to the shock waves solutions to the first order dispersionless limiting equation.

Third order partial differential equations arise in the study of dispersive wave motion, including water waves, plasma waves and others. We first treat the linear dispersive model, contrasting it with the hyperbolic models we encountered earlier in this book. The distinction between group and wave velocity — seen when waves propagate over water — is exposed. Finally, we introduce the remarkable Korteweg–deVries equation, which serves as a model for nonlinear water waves. Despite being nonlinear, it supports stable localized traveling wave solutions, known as solitons, that even maintain their shape under collisions. The Korteweg–deVries equation is an example of an integrable system, since it can be solved by an associated linear problem.

22.1. Nonlinear Waves and Shocks.

Before attempting to tackle any nonlinear partial differential equations, we should carefully review the solution to the simplest linear first order partial differential equation — the one-way or unidirectional wave equation

\[ u_t + cu_x = 0. \]  

(22.1)

First, assume that the wave velocity \( c \) is constant. According to Proposition 14.7, a solution \( u(t, x) \) to this partial differential equation is constant along the characteristic lines of slope

\[ \frac{dx}{dt} = c, \]

namely

\[ x - ct = \text{constant} \]  

(22.2)

As a consequence, the solutions are all of the form

\[ u = p(x - ct) \]

where \( p(\xi) \) is an arbitrary function of the characteristic variable \( \xi = x - ct \). To a stationary observer, the solution is a wave of unchanging form moving at velocity \( c \). The case \( c > 0 \) corresponds to a wave that translates to the right, as illustrated in Figure 22.1.

Slightly more complicated, but still linear, is the wave equation

\[ u_t + c(x)u_x = 0, \]  

(22.3)

where the variable wave velocity \( c(x) \) depends upon the position of the wave. This equation models unidirectional waves propagating through a non-uniform, but static medium. Generalizing the constant coefficient construction (22.2), we define the characteristic curves
for the wave equation (22.3) to be the solutions to the autonomous ordinary differential equation

\[ \frac{dx}{dt} = c(x). \]  

(22.4)

Thus, unlike the constant velocity version, the characteristics are no longer straight lines. Nevertheless, the preceding observation retains its validity.

**Proposition 22.1.** The solutions to the linear wave equation (22.3) are constant on the characteristic curves.

*Proof:* Let \( x(t) \) be a characteristic curve, i.e., a solution to (22.4), parametrized by the time \( t \). The value of a solution \( u(t, x) \) of the wave equation at the point \( (t, x(t)) \) on the given characteristic curve is \( h(t) = u(t, x(t)) \). Our goal is to prove that \( h(t) \) is a constant function of \( t \), and, as usual, this is done by proving that its derivative is identically zero. To differentiate \( h(t) \), we invoke the chain rule:

\[
\frac{dh}{dt} = \frac{d}{dt} u(t, x(t)) = \frac{\partial u}{\partial t}(t, x(t)) + \frac{dx}{dt} \frac{\partial u}{\partial x}(t, x(t)) = \frac{\partial u}{\partial t}(t, x(t)) + c(x(t)) \frac{\partial u}{\partial x}(t, x(t)) = 0.
\]

We replaced \( dx/dt \) by \( c(x) \) since we are assuming that \( x(t) \) is a characteristic curve, and hence satisfies (22.4). The final combination of derivatives is zero whenever \( u \) solves the wave equation (22.1). Therefore, \( h(t) = u(t, x(t)) \) is constant. Q.E.D.

Since the characteristic curve differential equation (22.4) is autonomous, it can be immediately solved:

\[ h(x) = \int \frac{dx}{c(x)} = t + \delta, \]

(22.5)

where \( \delta \) is the constant of integration. Therefore, the characteristic curves are defined by the formula \( x = g(t + \delta) \), where \( g = h^{-1} \) is the inverse function.

Any function which is constant along the curves defined by (22.5) must be a function of the characteristic variable \( \xi = h(x) - t \). As a consequence, Proposition 22.1 implies that we can write the solution to the wave equation in the form

\[ u(t, x) = p(h(x) - t), \]

(22.6)

\[ \dagger \] The present definition of characteristic variable has changed slightly from the constant velocity case.
where $p(\xi)$ is an arbitrary function of the characteristic variable. It is easy to verify directly that (22.6) that, provided $h(x)$ is defined by (22.5), $u(t, x)$ solves the partial differential equation (22.3) for any choice of function $p(\xi)$.

To find the solution that satisfies the given initial conditions

$$u(0, x) = f(x)$$

we merely substitute the solution (22.6), leading to an implicit equation

$$p(h(x)) = f(x), \quad \text{and hence} \quad p(\xi) = f \circ h^{-1}(\xi) = f[g(\xi)].$$

Graphically, the solution must be constant along each characteristic curve. Therefore, to find the value of the solution $u(t, x)$ at a given point, we look at the characteristic curve passing through $(t, x)$. If this curve intersects the $x$ axis at the point $(0, y)$, then $u(t, x) = u(0, y) = f(y)$. The construction is illustrated in Figure ccx.[1]

**Example 22.2.** Consider the equation

$$\frac{\partial u}{\partial t} + \frac{1}{x^2 + 1} \frac{\partial u}{\partial x} = 0. \tag{22.8}$$

According to (22.4), the characteristic curves are the solutions to the first order ordinary differential equation

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

Integrating, we find

$$\int (x^2 + 1) \, dx = \frac{1}{3} x^3 + x = t + \delta,$$

and the resulting characteristic curves are plotted in Figure wcxx.[1]

The general solution to the equation takes the form

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

where $p(\xi)$ is an arbitrary function of the characteristic variable $\xi = \frac{1}{3} x^3 + x - t$. A typical solution, corresponding to initial data $u(t, 0) = \bullet$ is plotted in Figure wcxx.[1]. The fact that the characteristic curves are not straight means that, although the wave remains constant along each individual curve, a stationary observer will witness a dynamically changing profile as the wave moves along. Waves speed up as they arrive at the origin, and then slow down once they pass by. As a result, we observe the wave spreading out as it approaches the origin, and then contracting as it moves off to the right.

**Example 22.3.** Consider the equation

$$u_t - x u_x = 0. \tag{22.9}$$

The characteristic curves are the solutions to

$$\frac{dx}{dt} = -x, \quad \text{and so} \quad x e^t = c. \tag{22.10}$$
where $c$ is the constant of integration. The solution takes the form

$$u = p(xe^t),$$

(22.11)

where $p(ξ)$ is an arbitrary function. Therefore, for initial data

$$u(0, x) = f(x) \quad \text{the solution is} \quad u = f(xe^t).$$

For example, the solution

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

corresponding to initial data $u(t, 0) = f(x) = (x^2 + 1)^{-1}$ is plotted in Figure wavecx. Note that since the characteristic curves all converge on the $t$ axis, the solution becomes more and more concentrated at the origin.

A Nonlinear Wave Equation

One of the simplest possible nonlinear partial differential equations is the nonlinear wave equation

$$u_t + uu_x = 0.$$ (22.12)

first systematically studied by Riemann. Since it appears in so many applications, this equation goes under a variety of names in the literature, including the Riemann, inviscid Burgers’, dispersionless Korteweg-deVries, and so on.

The equation (22.12) has the form of a unidirectional wave equation $u_t + cu_x = 0$ in which the wave velocity $c = u$ depends, not on the position $x$, but rather on the magnitude of the wave. Larger waves move faster, and overtake smaller waves. Waves of depression, where $u < 0$, move in the reverse direction.

Fortunately, the method of characteristics that was developed for linear wave equations also works in the present nonlinear situation and leads to a complete solution to the equation. Mimicking our previous construction, (22.4), let us define the characteristic curves of the nonlinear wave equation (22.12) by the formula

$$\frac{dx}{dt} = u(t, x).$$ (22.13)

In this case, the characteristics depend upon the solution, and so it appears that we will be not able to specify the characteristics until we know the solution $u(t, x)$. Be that as it may, the solution $u(t, x)$ remains constant along its characteristic curves, and this observation will allow us to pin both down.

† In addition to his contributions to complex analysis, partial differential equations and number theory, Riemann also was the inventor of Riemannian geometry, which proved absolutely essential for Einstein’s theory of general relativity some 70 years later!
First, to prove this claim, assume that \( x = x(t) \) parametrizes a characteristic curve. We need to show that \( h(t) = u(t, x(t)) \) is constant along the curve. As before, we differentiate. Using the chain rule and (22.13), we deduce that
\[
\frac{dh}{dt} = \frac{d}{dt} u(t, x(t)) = \frac{\partial u}{\partial t} (t, x(t)) + \frac{dx}{dt} \frac{\partial u}{\partial x} (t, x(t)) = \frac{\partial u}{\partial t} (t, x(t)) + u(t, x(t)) \frac{\partial u}{\partial x} (t, x(t)) = 0.
\]
The final expression vanishes because \( u \) is assumed to solve the wave equation (22.12) at all values of \((t, x)\). Since the derivative of \( h(t) = u(t, x(t)) \) is zero, this quantity must be a constant, as stated.

Now, since the solution \( u(t, x(t)) \) is constant on the characteristic curve, the right hand side of its defining equation (22.13) is a constant. Therefore, the derivative \( dx/dt \) is constant, and the characteristic curve is a \textit{straight line}! Consequently, each characteristic curve
\[
x = ut + \delta,
\]
is a straight line of slope \( u \), which we call the \textit{characteristic slope} of the line. The value of the solution on each characteristic line is its characteristic slope. The larger \( u \) is, the steeper the characteristic line, and the faster the wave moves.

The characteristic variable \( \xi = x - tu \) depends upon the solution, which can now be written in implicit form
\[
u = f(x - tu), \tag{22.14}
\]
where \( f(\xi) \) is an arbitrary function of the characteristic variable. For example, if \( f(\xi) = \alpha \xi + \beta \) is an affine function, then
\[
u = \alpha(x - tu) + \beta, \quad \text{and hence} \quad u(t, x) = \frac{\alpha x + \beta}{1 + \alpha t}. \tag{22.15}
\]
If \( \alpha > 0 \), this represents a straight line solution that gradually flattens out as \( t \to \infty \). On the other hand, if \( \alpha < 0 \), the line rapidly steepens to vertical as \( t \to t_* = -1/\alpha \) when the solution blows up.

To construct a solution \( u(t, x) \) to the initial value problem
\[
u(0, x) = f(x), \tag{22.16}
\]
we note that, at \( t = 0 \), the implicit solution formula formula (22.14) reduces to \( u(0, x) = f(x) \), and hence the function \( f \) coincides with the initial data! However, because (22.14) defines \( u(t, x) \) implicitly, it is not clear
\( (a) \) whether it can be solved to give a well-defined value for the solution \( u(t, x) \), and,
\( (b) \) if so, what are the solution’s qualitative features and dynamical behavior.

A more instructive and revealing strategy is based on the following geometrical construction. Through each point \((y, 0)\) on the \( x \) axis, draw the characteristic line \( x = tf(y) + y \) whose slope \( f(y) \) equals the value of the initial data at that point. According to the preceding argument, the solution will have the same value, \( u = f(y) \), on the entire characteristic line \( x = tf(y) + y \). For example, if \( f(y) = y \), then \( u(t, x) = y \) whenever \( x = ty + y \). Eliminating \( y \), we recover the previous solution \( u(t, x) = x/(t + 1) \).
Now, the problem with this construction is immediately apparent from the illustrative Figure Rsol. Any characteristic lines which are not parallel must cross each other. The value of the solution is supposed to be equal to the slope of the characteristic line, and so at the point of crossing, the solution is supposed to have two different values, one corresponding to each line. Something is clearly amiss, and we need to understand the construction and the resulting solution in more depth.

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

The next simplest case occurs when the initial data \( f(x) \) is everywhere increasing, so \( f(x) \leq f(y) \) whenever \( x \leq y \), which is assured if the derivative \( f'(x) \geq 0 \) is never negative. In this case, as in Figure chli, the characteristic lines emanating from the \( x \) axis fan out into the right half plane, and so never cross each other at any \( t \geq 0 \). Each point \( (t, x) \) for \( t \geq 0 \) lies on a unique characteristic line, and the value of the solution at \( (t, x) \) is equal to the slope of the line. Consequently, the solution is well-defined at all future times. Physically, such a solution represents a wave of rarefaction, which gradually spreads out as time progresses. A typical example is plotted in Figure 22.2, corresponding to initial data \( u(0, x) = \tan^{-1} x + \frac{\pi}{2} \).

The more interesting case is when \( f'(x) < 0 \). Now the characteristic lines starting at \( t = 0 \) cross at some future time. If a point \( (t, x) \) lies on two or more characteristic lines of different slopes, the value of the solution \( u(t, x) \), which should equal the characteristic slope, is no longer uniquely determined. Although one might be tempted to deal with such multiply-valued solutions in a purely mathematical framework, from a physical standpoint this is unacceptable. The solution \( u \) is supposed to represent a physical quantity, e.g., density, velocity, pressure, etc., and must therefore have a unique value at each point. The mathematical model has broken down, and fails to agree with the physical process.

Before confronting this difficulty, let us first, theoretically, understand what happens when we continue the solution as a multiply-valued function. To be specific, consider the initial data
\[
  u(0, x) = \frac{\pi}{2} - \tan^{-1} x, \quad (22.17)
\]
plotted in the first figure in Figure shu. In the companion picture we plot the characteristic lines for this particular initial data. Initially, the characteristic lines do not cross, and the solution is a well-defined single-valued functions. However, there is a critical time \( t = t_* > 0 \) when the first two lines cross each other at a particular point \( (t_*, x_*) \). After the
critical time, the \((t, x)\) plane contains a wedge-shaped region, each point of which lies on the intersection of three different characteristic lines with different slopes; at such points, the solution will achieve three different values. Outside the wedge, the points only belong to a single characteristic line, and the solution remains single valued there. (The boundary of the wedge is special, consisting of points where only two characteristic lines cross.)

To understand what is going on, look now at the sequence of pictures of the solution at successive times in Figure 22.3. Since the initial data is positive, \(f(x) > 0\), all the characteristic slopes are positive. As a consequence, all the points on the solution curve will move to the right, at a speed equal to their height. Since the initial data is decreasing, points to the left will move faster than those to the right, and eventually overtake them. Thus, as time progresses, the solution gradually steepens. At the critical time \(t_*\) when the first two characteristic lines cross, the tangent to the solution curve \(u(t_*, x)\) has become vertical at the point \(x_*, tar\), and so \(u_x(t, x_*) \to \infty\) as \(t \to t_*\). After the critical time, the solution graph \(u(t, x)\) for fixed \(t > t_*\) remains a continuous curve in the \((x, u)\) plane, but no longer represents the graph of a single-valued function. The overlapping lobes correspond to points \((t, x)\) lying in the aforementioned wedge.

The critical time can be determined from the implicit solution formula (22.14). Indeed, if we differentiate with respect to \(x\), we find

\[
\frac{\partial u}{\partial x} = \frac{\partial}{\partial x} f(x - tu) = f'(\xi) \left( 1 - t \frac{\partial u}{\partial x} \right),
\]

where \(\xi = x - tu\) is the characteristic variable, which is constant along the characteristic lines. Solving,

\[
\frac{\partial u}{\partial x} = \frac{f'(\xi)}{1 + t f'(\xi)}.
\]

Therefore, the slope

\[
\frac{\partial u}{\partial x} \to \infty \quad \text{as} \quad t \to -\frac{1}{f'(\xi)}.
\]

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 \((x, 0)\) will break down at the time \(-1/f'(x)\). As a consequence, the first of these critical times is at

\[
t_* = \min \left\{-\frac{1}{f'(x)} \bigg| f'(x) < 0 \right\}.
\]

(22.18)
For instance, for the particular initial configuration (22.17) represented by the pictures,

\[ f(x) = \frac{\pi}{2} - \tan^{-1} x, \quad f'(x) = -\frac{1}{1 + x^2}, \]

and so the critical time is

\[ t_* = \min(1 + x^2) = 1. \]

As noted above, the triply-valued matheamtical solution is physically untenable. So what happens after the critical time \( t_* \)? One needs to choose which of the three possible values should be used at each given point in the triply-valued wedge.

The mathematics is incapable of providing us with the answer, and we must reconsider the physical system that we are modeling.

The mathematics by itself incapable of telling us how to continue with this solution past the critical time at which the characteristics begin to cross. We therefore need to return to the physics underlying the partial differential equation, and ask what sort of phenomenon we are trying to model. The most instructive is to view the equation as a simple model of compressible fluid flow in a single space variable, e.g., gas in a pipe. If we push a piston down the end of a long pipe then the gas will move ahead of the piston and compress. If we push the piston too fast, the gas will compress near the piston. However, if the piston moves too rapidly the gas piles up on top of itself and a shock wave forms.

The physical assumption that underlies the specification of where the shock wave appears is known as an entropy condition. The simplest version, which applies to many physical systems, is an equal area rule. Draw the vertical shock line where the areas of the two lobes in the multiply valued solution are equal, as in Figure ea.

Note that this implies irreversibility of the solutions to the nonlinear wave equation. One cannot simply run time backwards and expect the shock to disappear. However, this is a different issue than the irreversibility of the heat equation, which was due to its ill-posedness in backwards time. One can run the nonlinear wave equation backwards, but this would result, typically, in the formation of a different collection of shocks.

**Example 22.4.** An interesting case is when the initial data has the form of a step function with a single jump discontinuity:

\[ u(0, x) = f(x) = a + b \sigma(x) = \begin{cases} a, & x < 0, \\ b, & x > 0. \end{cases} \] (22.19)

If \( a > b > 0 \), then the initial data is in the form of a shock. If we use the mathematical solution by continuing along the characteristic lines, the solution at time \( t \) is multiply-valued in the region \( bt < x < at \) where it assumes both values \( a \) and \( b \) as illustrated in Figure sws. If we use the equal area rule, we draw the shock line halfway along, at \( x = \frac{1}{2} (a + b) t \). Therefore, the shock moves with speed \( \frac{1}{2} (a + b) \) equal to one half the magnitude of the jump (and the value of the step function at the jump according to the Fourier convention). Behind the shock the solution has value \( a \) and in front the smaller value \( b \). A graph of the characteristic lines appears in Figure swsch.

By way of contrast, if \( 0 < a < b \), then the characteristic lines diverge from the shock point, and the mathematical solution is not well-defined in the wedge-shaped region.
at < x < bt. We must decide how to connect the two regions where the solution is defined. Physical reasoning points to using an affine or straight line to connect the two parts of the solution. Indeed, a simple modification of the solution (22.15) yields the function

\[ u(t, x) = \frac{x}{t}, \]

which not only solves the differential equation, but also has the required values \( u(t, at) = a, \) and \( u(t, bt) = b \) at the edge of the wedge. Therefore, the desired solution is the rarefaction wave

\[ u(t, x) = \begin{cases} a, & x \leq at, \\ \frac{x}{t}, & at \leq x \leq bt, \\ b, & x \geq bt \end{cases} \]

which is graphed in Figure swsr.

These two simple solutions epitomize the basic phenomenon modeled by our nonlinear wave equation — rarefaction wave where the solution is spreading out, that correspond to regions where \( f'(x) > 0, \) and waves of compression, where \( f'(x) < 0, \) where the solution contracts and eventually lead to shocks. Anyone caught in a traffic jam recognized the compression waves, where the traffic is bunched together and almost stationary, while the interspersed rarefaction waves represent freely moving traffic. (An intelligent drive will recognize the rarefaction waves moving through the jam and use them to switch lanes!) The observed, and frustrating traffic jam phenomenon is a direct result of the nonlinear wave model for traffic flow.

An entropy condition such as the equal area rule allows us to progress beyond the formation of a simple shock. As other characteristic lines cross, additional shocks form. The shocks themselves continue propagate, often at different velocities. When a fast-moving shock catches up with a slow moving shock, one must decide how to merge the shocks together to retain a physically consistent solution. At this point, the mathematical details have become too complicated for us to pursue in any more detail, and we refer the reader to [147] for a detailed discussion, along with applications to equations of gas dynamics, flood waves in rivers, motion of glaciers, chromatography, traffic flow and many other physical systems.

22.2. Nonlinear Diffusion.

First order partial differential equations, beginning with the simple scalar equation (22.12), and progressing through the equations of gas dynamics on to the full-blown Euler equations of fluid mechanics, model conservative wave motion. Such models fail to account for frictional and viscous effects.

When a shock wave forms, there is a breakdown in the mathematical solution to the equation. But the physical processes continue. This indicates that our assumptions governing the physical situation modeled by the partial differential equation are not complete, and are neglecting certain significant physical effects. In the case of gas dynamics, the nonlinear wave equation (22.12) does not build in any damping effects due to viscosity in the fluid. Dissipative or frictional or viscous effects are, as we know, governed by second
order differential operators. The simplest is the linear heat equation which models a broad range of dissipative phenomena, but fails to take into account nonlinear physical effects.

As in the linear heat equation, dissipative effects such as friction and viscosity are governed by second order elliptic differential operators, and hence introduce second order terms into the wave model. In this section, we study the very simplest model that includes both nonlinear wave motion and dissipation, known as Burgers’ equation.

**Burgers’ Equation**

The simplest nonlinear diffusion equation is known as\(^1\) Burgers’ equation, and takes the form

\[
\frac{\partial u}{\partial t} + \beta u \frac{\partial u}{\partial x} = \gamma \frac{\partial^2 u}{\partial x^2},
\]

(22.20)

The term \(\gamma \frac{\partial^2 u}{\partial x^2}\) represents linear diffusion, as in the heat equation. The diffusion coefficient \(\gamma > 0\) must be positive in order that the equation be well-posed. The second, nonlinear term represents a simple advection. In the inviscid limit, as the diffusion goes to zero, \(\gamma \to 0\), Burgers’ equation reduces to the nonlinear wave equation (22.12), which, as a result, is often referred to as the inviscid Burgers’ equation in the literature. One can also interpret the linear term as modeling viscosity, and so Burgers’ equation represents a very simplified version of the equations of viscous fluid mechanics. The higher the viscosity \(\gamma\), the faster the diffusion.

We will only consider solutions \(u(t, x)\) which are globally defined for all \(-\infty < x < \infty\) and for times \(t > t_0\) after an initial time which, for simplicity, we take as \(t_0 = 0\). As with both the heat equation and the nonlinear wave equation (22.12), the solution will be specified by its initial values

\[
\frac{\partial u}{\partial t}(0, x) = f(x), \quad -\infty < x < \infty.
\]

(22.21)

The initial data \(f(x)\) is assumed to be reasonably smooth, say \(C^1\), and bounded as \(|x| \to \infty\).

Small solutions of Burgers’ equation, \(|u(t, x)| \ll 1\), will tend to act like solutions to the heat equation, since the nonlinear terms will be negligible. On the other hand, for large solutions \(|u(t, x)| \gg 1\), the nonlinear terms will dominate the equation, and we expect the solution to behave like the nonlinear waves we analyzed in Section 22.1. Thus, the question naturally arises: do the solutions of Burgers’ equation experience shocks, or does the diffusion have a sufficient effect to smooth out any potential discontinuities. As we will see, the latter scenario is correct. Assuming \(\gamma > 0\), it can be proved, [147], that the initial value problem (22.20), (22.21) for Burgers’ equation has a unique solution \(u(t, x)\) that is smooth and exists for all positive time \(t > 0\). The diffusion, no matter how small, is sufficient to prevent the formation of any shocks or other discontinuities in the solution.

A typical simple solution is plotted in Figure Burgers. The initial data is the same as in the shock wave solution for the inviscid version that we plotted in Figure swa. We

\(^1\) Note that the apostrophe goes after the “s” since the equation is named after the applied mathematician J. Burgers, [Burgers]. It was first studied in this context by Bateman, [16], although it does appear in older pure mathematical texts.
take the diffusion coefficient to be small: $\gamma = .01$, and the nonlinearity $\beta = 1$. As you can see, the wave initially steepens just like its inviscid counterpart. However, at a certain point, the diffusion prevents the wave from becoming vertical and then moving into the shock regime. Instead, once the initial steepening is finished, the wave takes the form of a very sharp, but nevertheless smooth, transition, looking like a slightly smoothed-out form of the equal area shock wave solution that we found in Section 22.1.

Indeed, the profound fact is that, as the diffusion $\gamma \to 0$ becomes very small, the solutions to Burgers’ equation (22.20) converge, in the inviscid limit, to the shock wave solution to (22.12) constructed by the equal area rule. This observation is in accordance with our physical intuition, that all physical systems retain a very small dissipative component, that serves to smooth out discontinuities that appear in the theoretical model that fails to take the dissipation/viscosity/damping/etc. into account. It also has very important theoretical consequences. The way to characterize the discontinuous solutions to the inviscid nonlinear wave equation is as the limit, as the viscosity goes to zero, of classical solutions to the Burgers’ equation. This is known as viscosity solution method. If the viscous terms are as in Burgers’ equation, the resulting viscosity solutions are consistent with the equal area rule for drawing the shocks. More generally, this method allows one to continue the solutions into the regimes where multiple shocks merge and interact.

*The Hopf–Cole Transformation*

While the Burgers’ equation is a fully nonlinear partial differential equation, there is a remarkable nonlinear transformation that converts it into the linear heat equation. This result first appears in an exercise in the nineteenth century differential equations textbook by Forsyth, [59; vol. 6, p. 102]. Its modern rediscovery by Eberhard Hopf, [83], and Julian Cole, [35], was a milestone in the study of nonlinear partial differential equations.

One simple-minded way to convert a linear equation into a more complicated nonlinear equation is to make a nonlinear change of variables. The resulting nonlinear equation is said to be linearizable since it can be linearized by inverting the change of variables. Recognizing when a nonlinear equation can, in fact, be linearized by a suitable change of variables is a challenging problem, and the subject of much contemporary research, [E]. In practice, “nonlinearizing” a linear equation by a randomly chosen changes of variables rarely leads to a nonlinear equation of any interest. However, sometimes there is a luck accident, and the linearizing change of variables can make a profound impact on our understanding of the nonlinear version.

Our starting point is the linear heat equation

$$v_t = \gamma v_{xx}. \quad (22.22)$$

Among all the possible nonlinear changes of dependent variable, one of the simplest that might spring to mind is an exponential function. Consider the nonlinear change of variables

$$v = e^{\sigma \varphi}, \quad (22.23)$$

\[\dagger\] Nonlinear changes of the independent variables $t, x$ alone will only lead to a linear partial differential equation, albeit with nonconstant coefficients; see Exercise \[●\].
where $\varphi(t, x)$ is a new function. The change of variables is valid provided $v(t, x) > 0$ is a positive solution to the heat equation. Fortunately, this is not hard to arrange: if the initial data $v(0, x) > 0$ is strictly positive, then the solution is positive for all $t > 0$. Physically, if the temperature in a fully insulated bar starts out everywhere above freezing, in the absence of external heat sources, it can never dip below freezing at any later time.

To find the differential equation satisfied by the new function $\varphi$, we compute the relations among their derivatives using the chain rule:

$$v_t = \sigma \varphi_t e^{\sigma \varphi}, \quad v_x = \sigma \varphi_x e^{\sigma \varphi}, \quad v_{xx} = \left( \sigma \varphi_{xx} + \sigma^2 \varphi_x^2 \right) e^{\sigma \varphi}.$$  

We substitute the first and last formulae into the heat equation (22.22) and canceling a common exponential factor. We conclude that $\varphi(t, x)$ satisfies the nonlinear partial differential equation

$$\varphi_t = \gamma \varphi_{xx} + \gamma \sigma \varphi_x^2; \quad (22.24)$$

known as the potential Burgers' equation, for reasons that will soon become apparent.

The second step in the procedure is to differentiate the potential Burgers' equation with respect to $x$; the result is

$$\varphi_{tx} = \gamma \varphi_{xxx} + 2\gamma \sigma \varphi_x \varphi_{xx}. \quad (22.25)$$

If we now set

$$\frac{\partial \varphi}{\partial x} = u, \quad (22.26)$$

then the resulting partial differential equation is a form of Burgers’ equation (22.20):

$$u_t = \gamma u_{xx} + 2\gamma \sigma uu_x.$$  

Indeed, if we define $\beta = -2\gamma \sigma$, then the two equations coincide. Let us summarize the resulting Hopf–Cole transformation.

**Theorem 22.5.** If $v(t, x) > 0$ is any positive solution to the linear heat equation $v_t = \gamma v_{xx}$, then

$$u(t, x) = \frac{\partial}{\partial x} \left( -\frac{\gamma}{2\beta} \log v(t, x) \right) = -\frac{\gamma}{2\beta} \frac{v_x}{v}, \quad (22.27)$$

solves Burgers' equation $u_t + \beta uu_x = \gamma u_{xx}$.

Do all solutions to Burgers’ equation arise in this way? In order to decide, we run the derivation of the transformation in reverse. The intermediate function $\varphi(t, x)$ in (22.26) can be interpreted as a potential function for the solution $u(t, x)$, which can be physically interpreted as a fluid velocity given as the one-dimensional gradient of its potential. The potential is only determined up to a constant, or, more accurately, up to a function of $t$, and so $\tilde{\varphi}(t, x) = \varphi(t, x) + h(t)$ is an equally valid potential. Substituting the potential relation (22.26) into Burgers’ equation (22.20) leads to (22.25)

$$\varphi_{tx} + \beta \varphi_x \varphi_{xx} = \gamma \varphi_{xxx}.$$  

$2/25/04$ 961  

© 2004 Peter J. Olver
Integrating both sides with respect to $x$ produces

$$
\varphi_t = \gamma \varphi_{xx} + \gamma \sigma \varphi_x^2 + h(t),
$$

for some integration “constant” $h(t)$. Replacing $\varphi = \bar{\varphi} + H(t)$ where $H' = h$, we find that the alternative potential $\bar{\varphi}$ does satisfy the potential Burgers’ equation, and thus comes from a positive solution to the heat equation by the exponential changes of variables. Thus, the answer is yes: every solution to Burgers’ equation comes from a positive solution to the heat equation via the Hopf–Cole transformation.

**Example 22.6.** For example, the separable solution

$$
v(t, x) = \alpha + \beta e^{-\omega^2 t} \cos \omega x
$$

to the heat equation leads to the solution

$$
u(t, x) = \frac{\gamma}{2\beta} \frac{\beta \omega e^{-\omega^2 t} \sin \omega x}{\alpha + \beta e^{-\omega^2 t} \cos \omega x}
$$
to Burgers’ equation; a typical example is plotted in Figure Bcos. We must require that $\alpha > |\beta|$ in order that $v(t, x) > 0$ be a positive solution to the heat equation for $t \geq 0$; otherwise the solution to Burgers’ equation would have singularities at the roots of $u$. This particular solution primarily feels the effects of the diffusivity, and goes to zero exponentially fast.

To solve the initial value problem (22.20), (22.21) we note that the initial conditions transform, via (22.27), to

$$
v(0, x) = h(x) = \exp \varphi(0, x) = \exp \int_0^x dy f(x), \quad (22.28)
$$

where 0 can be replaced by any convenient starting point for the integral, e.g., $-\infty$. According to the solution formula (14.59) (adapted to general diffusivity; see Exercise), the solution to the initial value problem (22.22), (22.28) for the heat equation can be written as a convolution with the fundamental solution:

$$
v(t, x) = \frac{1}{2\sqrt{\pi \gamma t}} \int_{-\infty}^\infty e^{-(x-y)^2/(4\gamma t)} h(y) \, dy.
$$

Therefore, the solution to the Burgers’ initial value problem (22.20), (22.21) is

$$
u(t, x) = \frac{\int_{-\infty}^\infty \frac{x-y}{t} e^{-G(t,x,y)/2\gamma} \, dy}{\int_{-\infty}^\infty e^{-G(t,x,y)/2\gamma} \, dy} \quad \text{where} \quad G(t, x, y) = \int_0^y f(z) \, dz + \frac{(x-y)^2}{2\gamma}.
$$

(22.29)
Example 22.7. To demonstrate the smoothing effect of the diffusion terms, let us take the initial data
\[ u(0, x) = \begin{cases} a, & x < 0, \\ b, & x > 0, \end{cases} \] (22.30)
in the form of a step function. We assume that \( a > b \), which would correspond to a shock wave in the inviscid limit. (The reader is asked to analyze the case \( a < b \) which corresponds to a rarefaction wave.) The solution takes the form
\[ u(t, x) = b + \frac{a - b}{1 + w(t, x) \exp \left( \frac{a - b}{2 \gamma} (x - ct) \right)} \]
where
\[ c = \frac{a + b}{2}, \quad w(t, x) = \begin{cases} \frac{\sqrt{\pi}}{2} - \text{erf} \left( \frac{bt - x}{\sqrt{4 \pi \gamma t}} \right), & y < 0, \\ \frac{\sqrt{\pi}}{2} - \text{erf} \left( \frac{x - at}{\sqrt{4 \pi \gamma t}} \right), & y > 0. \end{cases} \]
where \( \text{erf} z \) denotes the error function (14.61). The solution is plotted in Figure Bshock. Note that the sharp transition region for the shock has been smoothed out. The larger the diffusion coefficient in relation to the initial solution heights \( a, b \), the more significant the smoothing effect.

Example 22.8. Consider the case when the initial data \( u(0, x) = a \delta(x) \) is a concentrated delta function impulse at the origin. In the solution formula (22.29), starting the integral for \( G(t, x, y) \) at 0 is problematic, but as noted earlier, any other starting point will lead to a valid formula. Thus, we take
\[ G(t, x, y) = \int_{-\infty}^{y} a \delta(z) dz + \frac{(x - y)^2}{2 \gamma} = \begin{cases} \frac{(x - y)^2}{2 \gamma}, & y < 0, \\ a + \frac{(x - y)^2}{2 \gamma}, & y > 0. \end{cases} \]
Substituting this into (22.29), we can evaluate the upper integral in elementary terms, while the lower integral involves the error function (14.61):
\[ u(t, x) = \sqrt{\frac{\gamma}{c}} \left( e^{\alpha/2 \gamma} - 1 \right) e^{-x^2/(4 \gamma t)} \]
A graph of this solution appears in Figure hump. As you can see, the delta function diffuses out, but, unlike the heat equation, the wave does not remain symmetric owing to the advection terms in the equation. The effect is to steepen in front as it propagates. We note that (22.31) is in the form of a similarity solution
\[ u(t, x) = \sqrt{\frac{\gamma}{c}} \frac{x}{\sqrt{4 \pi \gamma t}} \]
which could perhaps have been predicted from the scaling invariance of the initial data.
If $a \ll 1$ is small, then the nonlinear terms in burgers’ equation are negligible, and the solution is very close to the fundamental source solution to the heat equation. On the other hand, for large $a \gg 1$, one would expect the advection terms to dominate, and the only effect of diffusion being a smoothing at any abrupt discontinuity. Indeed, for large $a$ the leading edge of the solution is in the form of a shock wave. As $a \to \infty$, the solution converges to the similarity solution

$$u(t, x) \rightarrow \begin{cases} \frac{x}{t}, & 0 \leq x \leq \sqrt{2at}, \\ 0, & \text{otherwise}. \end{cases}$$

of the inviscid wave equation (22.12).

### 22.3. Dispersion and Solitons.

Finally, we look at a remarkable third order evolution equation that originally arose in the modeling of surface water waves, that serves to introduce yet further phenomena, both linear and nonlinear. The third order derivative models dispersion, in which waves of different frequencies move at different speeds. Coupled with the same nonlinearity as in the inviscid and viscous Burgers’ (22.12), (22.20), the result is one of the most remarkable equations in all of mathematics, with far-reaching implications, not only in fluid mechanics and applications, but even in complex function theory, physics, etc., etc.

**Linear Dispersion**

The simplest linear partial differential equation of a type that we have not yet considered is the third order equation

$$u_t = u_{xxx} \quad (22.32)$$

It is the third member of the hierarchy of simple evolution equations that starts with the simple ordinary differential equation $u_t = u$, then proceeds to the unidirectional wave equation $u_t = u_x$, and then the heat equation $u_t = u_{xx}$. Each member of the hierarchy has its own range of phenomenology. The third order case is a simple model for linear dispersive waves.

We shall only look at the equation on the entire line, so $x \in \mathbb{R}$, and so can ignore additional complications caused by boundary conditions. The solution to the equation is uniquely specified by initial data

$$u(0, x) = f(x), \quad -\infty < x < \infty.$$ 

See [X] for a proof.

Let us apply the Fourier transform to solve the equation. Using separation of variables Substitute

$$u(t, x) = e^{i\omega t + ikx}$$

where $\omega$ is the frequency and $k$ is called the *wave number*. We find $\omega = k^3$ is the dispersion relation. Therefore, the solution is given by superposition as a Fourier integral

$$u(t, x) = \int_{-\infty}^{\infty} e^{ik^3 t + ikx} \hat{f}(k) \, dk$$
In particular, the solution with a concentrated initial disturbance

\[ u(0, x) = \delta(x) \quad \text{is} \quad u(t, x) = \text{Ai} \left( \frac{x}{t^{1/3}} \right) \]

in terms of the Airy function. See Figure ee3 for a graph.

Fundamental solution and superposition

Although energy is conserved, unlike the heat and diffusion equations, the dispersion of waves means that the solution dies out.

group velocity and wave velocity.

The Korteweg–deVries Equation

The simplest wave equation that combines dispersion with nonlinearity is the celebrated Korteweg–deVries equation

\[ u_t + u_{xxx} + uu_x = 0. \quad (22.33) \]

The equation was first derived by the French applied mathematician Boussinesq, \cite{22; eq. (30), p. 77}, \cite{23; eqs. (283, 291)}, in 1872 as a model for surface water waves. It was rediscovered by the Dutch mathematicians Korteweg and de Vries, \cite{96}, over two decades later. More recently, in the early 1960’s, Kruskal and Zabusky, \cite{155}, rederived it as a continuum limit of a model of nonlinear mass-spring chains studied by Fermi, Pasta and Ulam, \cite{54}. Their numerical experiments on the equation opened the door to the understanding of its many remarkable properties. It has a critical balance between nonlinear effects and dispersion, leading to integrability.

The most important special solutions to the Korteweg–deVries equation are the traveling waves. We assume that

\[ u = v(\xi) = v(x - ct) \]

to be a wave of permanent form, translating to the right with speed \( c \), that is, a solution to \( u_t + c u_x = 0 \). Note that

\[ \frac{\partial u}{\partial t} = -cv'(\xi), \quad \frac{\partial u}{\partial x} = v'(\xi), \quad \frac{\partial^3 u}{\partial x^3} = v'''(\xi). \]

Therefore, \( v(\xi) \) satisfies the third order nonlinear ordinary differential equation

\[ v''' + vv' - cv = 0. \quad (22.34) \]

Moreover, we impose boundary conditions

\[ \lim_{\xi \to -\infty} v(\xi) = \lim_{\xi \to -\infty} v'(\xi) = \lim_{\xi \to -\infty} v''(\xi) = 0. \quad (22.35) \]

This equation can be integrated. First, note that it can be written as a derivative:

\[ \frac{d}{d\xi} \left[ v'' + \frac{1}{2} v^2 - cv \right] = 0, \quad \text{and hence} \quad v'' + \frac{1}{2} v^2 - cv = a, \]
where \( a \) is a constant of integration. However, the boundary conditions as \( \pm \infty \) imply that \( a = 0 \). Multiplying the latter equation by \( v' \) allows us to integrate a second time

\[
\frac{d}{d\xi} \left[ \frac{1}{2} (v')^2 + \frac{1}{6} v^3 - \frac{1}{2} cv^2 \right] = v' \left[ v'' + \frac{1}{2} v^2 - cv \right] = 0.
\]

Integrating both sides of the equation,

\[
\frac{1}{2} (v')^2 + \frac{1}{6} v^3 - \frac{1}{2} cv^2 = b,
\]

where \( b \) is a second constant of integration, which, again by the boundary conditions (22.35), is also \( b = 0 \).

We also assume that the wave is localized, meaning that \( u \) and its derivatives tend to 0 as \( |x| \to \infty \). Therefore, \( v(\xi) \) satisfies the first order autonomous ordinary differential equation

\[
\frac{dv}{d\xi} = v \sqrt{c - \frac{1}{3} v}.
\]

We integrate by the usual method:

\[
\int \frac{dv}{v \sqrt{c - \frac{1}{3} v}} = \xi + \delta.
\]

The solution has the form

\[
v(\xi) = 3c \text{sech}^2 \left[ \frac{1}{2} \sqrt{c} \xi + \delta \right],
\]

where

\[
\text{sech} y = \frac{1}{\cosh y} = \frac{2}{e^y + e^{-y}},
\]

is the hyperbolic secant function. Hence, the localized traveling wave solutions of the Korteweg–deVries equation equation are of the form

\[
u(t, x) = 3c \text{sech}^2 \left[ \frac{1}{2} \sqrt{c} (x - ct) + \delta \right], \tag{22.36}
\]

where \( c > 0 \) and \( \delta \) are arbitrary constants. The parameter \( c \) measures the velocity of the wave. It also measures its amplitude, since the maximum value of \( u(t, x) \) is \( 3c \) since \( \text{sech} y \) has a maximum value of 1 at \( y = 0 \). Therefore, the taller the wave, the faster it moves. See Figure soliton for a graph.

The solution (22.36) is known as a solitary wave solution since it represents a localized wave that travels unchanged in shape. Such waves were first observed by the British engineer J. Scott Russell, [131], who tells the remarkable incident of chasing such a wave generated by the sudden motion of a barge along an Edinburgh canal on horseback for several miles. The mathematician Airy claimed that such waves could not exist, but he based his analysis upon a linearized theory. Boussinesq’s establishment of the surface wave model demonstrated that such localized disturbances can result from nonlinear effects in the system.
Remark: In the Korteweg–deVries equation model, one can find arbitrarily tall soliton solutions. In physical water waves, if the wave is too tall it will break. Indeed, it can be rigorously proved that the full water wave equations admit solitary wave solutions, but there is a wave of greatest height, beyond which a wave will tend to break. The solitary water waves are not genuine solitons, since there is a small, but measureable, effect when two waves collide.

These nonlinear traveling wave solutions were discovered by Kruskal and Zabusky, [155], to have remarkable properties. For this reason they have been given a special new name — soliton. Ordinarily, combining two solutions to a nonlinear equation can be quite unpredictable, and one might expect any number of scenarios to occur. If you start with initial conditions representing a taller wave to the left of a shorter wave, the solution of the Korteweg–deVries equation runs as follows. The taller wave moves faster, and so catches up the shorter wave. They then have a very complicated nonlinear interaction, as expected. But, remarkably, after a while they emerge from the interaction unscathed. the smaller wave is now in back and the larger one in front. After this, they proceed along their way, with the smaller one lagging behind the high speed tall wave. the only effect of their encounter is a phase shift, meaning a change in the value of the phase parameter \( \delta \) in each wave. See Figure solitons. After the interaction, the position of the soliton if it had traveled unhindered by the other is shown in a dotted line. Thus, they behave like colliding paricles, which is the genesis of the word “soliton”.

A similar phenomenon holds for several such soliton solutions. After some time where the various waves interact, they finally emerge with the largest soliton in front, and then in order to the smallest one in back, all progressing at their own speed, and so gradually drawing apart.

Moreover, starting with an arbitrary initial disturbance

\[
  u(0, x) = f(x)
\]

it can be proved that after some time, the solution disintegrates into a finite number of solitons of different heights, moving off to the right, plus a small dispersive tail moving to the left that rapidly disappears. Proving this remarkable result is beyond the scope of this book. It relies on the method of inverse scattering, that connects the Korteweg–deVries equation with a linear eigenvalue problem of fundamental importance in one-dimensional quantum mechanics. The solitons correspond to the bound states of a quantum potential. We refer the interested reader to the introductory text [50] and the more advanced monograph [1] for details.

Chaos and integrability are the two great themes in modern nonlinear applied mathematics, and the student is well-advised to pursue both.

There is a remarkable transformation, known as the inverse scattering transform, which is a form of nonlinear Fourier transform, that can be used to solve the Korteweg–deVries equation. Its fascinating properties continue to be of great current research interest to this day.

22.4. Conclusion and Bon Voyage.

These are your first wee steps in a vast new realm. We are unable to discuss nonlinear
partial differential equations arising in fluid mechanics, in elasticity, in relativity, in differential geometry, in computer vision, in mathematical biology. We bid the reader adieu and farewell.