3 General First Order Equations

3.1 The general quasilinear first-order equation

Before considering dispersive systems and general first order equations, let us see how far we can push the ideas about characteristics that we developed in the context of traffic flow. We saw that, for an equation of the form

$$\rho_t + c(\rho)\,\rho_x = 0\,,$$

we could determine the solution by following characteristic lines, solving the system of ordinary differential equations

$$\begin{aligned} \frac{dx}{dt} &= c(\rho) \\ \frac{d\rho}{dt} &= 0\,. \end{aligned}$$

This idea can be easily generalized to scalar equations in more dimensions, with more general coefficients and nonzero forcing.

The general *quasi-linear* first order equation for a scalar function $\phi(x)$, $x \in R^n$, has the form

$$\sum_j a_j(\phi, x)\,\frac{\partial \phi}{\partial x_j} = b(\phi, x)\,, \tag{45}$$

where the vector of independent variables x may include time as its first component, and the coefficient a_j's and the forcing b are given by prescribed functions of the independent variables and possibly of the solution, but not of its derivatives. If we introduce *characteristic* lines $x(s)$ in R^n by the system of equations

$$\dot{x}_j(s) = a_j(\phi, x)\,, \tag{46}$$

where a dot denotes derivative with respect to the parameter s, then equation (45) implies that

$$\dot{\phi}(s) = b(\phi, x)\,. \tag{47}$$

The equations in (46, 47) form a closed system of ordinary differential equations, determining lines in the $n+1$ dimensional space (x, ϕ). Given initial data $\phi(x)$ along a surface $S(x) = 0$, the characteristics may be used to find the solution in a neighborhood of the initial surface. For this, however, it is necessary that the characteristics be transversal to the initial surface; i.e. that

$$\nabla S \cdot a \neq 0\,, \tag{48}$$

where a is the vector with components $a_j(\phi, x)$. When the condition (48) is not satisfied, the characteristics are tangent to the surface $S(x) = 0$. Then not only they do not provide information on the solution away from the initial surface, but also they may contradict the initial data if the characteristic equation for ϕ is not satisfied along the surface $S(x) = 0$.

In two dimensions, such as the $x = (x, t)$ of traffic flow, the surface $S(x) = 0$ is in fact a curve, and the non-degeneracy condition (48) simply states that this curve should nowhere be tangent to a characteristic. By extension, surfaces in any dimension not satisfying (48) are called *characteristic surfaces*. Two properties characterize these surfaces: that the data provided on them cannot be arbitrary, and that the equations do not fully specify the variation of the solution in the direction transversal to the surfaces. As a consequence of the latter property, the derivative of ϕ normal to the surface may jump. This provides a further characterization of characteristic surfaces, as those across which the solution admits weak singularities. Later in the class we shall extend this characterization to more general systems of PDEs.

3.2 Dispersive waves. Fourier Synthesis. Phase and group velocities

The study of general first order equations, which we undertake in the next subsection, has many commonalities with the study of dispersive waves. We shall consider first the latter, since these will provide intuition behind the meaning of the general characteristic equations.

Why do waves in nature so often adopt a sinusoidal form? Linear systems of differential equations with constant coefficients typically have exponential solutions, which become sines and cosines for complex arguments. For instance, the wave equation

$$u_{tt} - c^2 \Delta u = 0,$$

studied in the next section, has solutions of the form

$$u = \cos(k \cdot x - \omega t) \quad \text{and} \quad u = \sin(k \cdot x - \omega t),$$

which we may represent succinctly as the complex exponential

$$u = e^{i(k \cdot x - \omega t)}, \tag{49}$$

where the frequency ω and the wavevector k satisfy the *dispersion relation*

$$\omega^2 = c^2 \|k\|^2 . \tag{50}$$

Small-amplitude waves in the surface of lakes and oceans also have the the complex exponentials in (49) as solutions; in this case, the dispersion relation reads

$$\omega^2 = g \, \|k\| \, \tanh(H \|k\|) \, , \tag{51}$$

where $g \approx 10 \, m/s^2$ is the gravity constant, and H is the mean depth of the unperturbed lake. Notice that, for small values of $\|k\|$ (i.e., long waves), this converges to the dispersion relation (50) for the wave equation, with the shallow water wave speed $c^2 = g \, H$. We will derive the more general dispersion relation (51) as an application of Laplace's equation later in class.

Still another example is provided by the Schrödinger equation of Quantum Mechanics for a free particle,

$$-i\psi_t = \Delta\psi, \tag{52}$$

with dispersion relation

$$\omega = \|k\|^2. \tag{53}$$

Examples abound (see the book by Whitham for many more examples, as well as for a far more thorough treatment that we can afford here) of *dispersive systems*, which have much of their behavior encapsulated in a dispersion relation of the form

$$\omega = \Omega(k)\,, \tag{54}$$

valid for exponential solutions of the form (49).

Introducing the *phase* of the wave, $\theta(x,t) = k \cdot x - \omega t$, we note that this (and hence the solution) is constant along the line

$$x = x_0 + \frac{\omega}{\|k\|}\frac{k}{\|k\|}t$$

(Here the factor $\frac{k}{\|k\|}$, absent in one dimension, is a unit vector in the direction of propagation of the wave.) The speed

$$C_{ph} = \frac{\Omega(k)}{\|k\|}\frac{k}{\|k\|} \tag{55}$$

at which the fronts move normal to themselves is called the *phase velocity* of the waves. For sinusoidal waves, such as those given by the real or imaginary components of (49), this is the speed at which crests and troughs propagate. It appears, at first sight, that this is the most fundamental speed associated with a wave. Hence we might be surprised when further analysis leads us to a quite different conclusion, downplaying the role of the phase velocity, and highlighting instead the *group velocity*

$$C_g = \nabla_k \Omega(k)\,, \tag{56}$$

where ∇_k denotes the gradient in wave vector space.

Before exploring this issue further, notice that a knowledge of the dispersion relation (54) allows us to solve the initial value problem for linear waves in homogeneous media rather concisely. For we can superimpose many solutions of the form (49), and write

$$u(x,t) = \int \hat{u_0}(k)\, e^{i(k\cdot x - \Omega(k)t)}\, dk\,, \tag{57}$$

where $\hat{u_0}(k)$ is the Fourier Transform of the initial data $u(x,0) = u_0(x)$:

$$\hat{u_0}(k) = \left(\frac{1}{2\pi}\right)^d \int u_0(x)\, e^{-i\,k\cdot x}\, dx\,, \tag{58}$$

and d is the dimension of the space. When more than one function is required as initial condition (for instance both $u(x,0)$ and $u_t(x,0)$ are required for the wave equation), this typically results in the dispersion relation $\Omega(k)$ being multivalued, as is the case of (50). Then the integral (57) becomes a sum of integrals involving the various branches of $\Omega(k)$, and with amplitudes that can be related to the Fourier Transforms of the various functions provided as initial data.

We can generalize the globally uniform exponential solution (49), and consider *modulated waves* of the form

$$u(x,t) = \rho(x,t)\, e^{i\,\theta(x,t)}\,, \tag{59}$$

where we now *define* the wave vector and frequency to be

$$k = k(x,t) = \nabla_x \theta\,, \quad \omega = \omega(x,t) = -\theta_t\,. \tag{60}$$

(Notice that these agree with the former definitions when $\theta = k\cdot x - \omega t$.) The idea behind writing a solution in this form is to describe approximately sinusoidal waves that may either be moving through a variable medium, or have been generated by a variable source. An example of a wave moving through a variable medium is that of ocean waves approaching a shore, and hence moving through waters of variable depth. Technological applications of variable sources include telephone, radio, TV and internet signals sent through space via modulated electromagnetic waves.

Behind the ansatz of a modulated wave lies the assumption of *scale separation*: the amplitude $\rho(x,t)$ and the wave vector and frequency $k(x,t)$, $\omega(x,t)$ need vary little within one period of the wave. Plugging the ansatz (59) into a linear system under this assumption yields a dispersion dispersion as in (54), with the caveat that now the frequency Ω may depend not just on the wave vector k, but also on space and time:

$$\omega = \Omega(k,x,t)\,. \tag{61}$$

On the other hand, because of the definition of k and ω as derivatives of the phase, their cross–derivatives need to match:

$$k_t + \nabla_x \omega = 0\,. \tag{62}$$

Equation (62) can be thought of as a conservation law for the number of waves, where the wavenumber k represents the spatial density of waves, and the frequency ω represent the flux of waves per unit time (You may want to play with this idea first in one spatial dimension, where the intuition behind it is easier.) Replacing (61) in (62) yields the equation

$$k_t + (C_g \cdot \nabla_x)\, k = -\nabla_x \Omega. \tag{63}$$

(Try deriving this equation: you will need the fact that $\frac{\partial k_i}{\partial x_j} = \frac{\partial^2 \theta}{\partial x_i \partial x_j} = \frac{\partial k_j}{\partial x_i}$.)

Hence, along the characteristic lines

$$\dot{x} = C_g = \nabla_k \Omega\,, \tag{64}$$

the wavenumber satisfies

$$\dot{k} = -\nabla_x \Omega. \tag{65}$$

Notice that the frequency $\omega = \Omega(k, x)$ also satisfies a similar equation:

$$\dot{\omega} = \Omega_t + \dot{x} \cdot \nabla_x \Omega + \dot{k} \cdot \nabla_k \Omega = \Omega_t\,. \tag{66}$$

In particular, if the dispersion relation does not depend on x and t (i.e., the medium is homogeneous and stationary), then both the wavenumber and the frequency are constant along characteristics. Hence the group velocity C_g can be thought of as the speed at which the wavenumber and the frequency propagate. These quantities are far more fundamental for the characterization of the wave than its phase θ, moving at the phase speed C_{ph}. Here go two examples that illustrate the significance of this distinction:

- Wave packets: Looking at the Hudson or at a pond in Central Park in a breezy day, you will notice that small capillary surface waves, a centimeter or two in wavelength, move in *packets*, broader wave envelopes that propagate at a speed different from their individual wave components. The latter appear to be born continuously at one end of the packet, and vanish at the other. The individual waves move at the phase velocity C_{ph}, the packet as a whole at the group velocity C_g.

- Waves and particles: One of the peculiar phenomena that gave rise to quantum mechanics is the observation that small particles, such as electrons, sometimes diffract as if they were instead waves. It turns out that these particles can be modeled as wave packet solutions to a dispersive system, the *Schrödinger equation (52)*, moving as particles at the group velocity, but with phases moving at the phase speed.

3.3 Characteristics of nonlinear equations

A dispersion relation (61) relates the spatial and temporal derivatives of the phase of the solution to a linear system. In general, a first order partial differential equation relates the first order derivatives of a function, the function itself, and the position in space and time. It adopts the general form

$$F(x, \phi, p) = 0\,, \tag{67}$$

where $x = (x_j)$ is the vector of independent variables (possibly including time), $\phi(x)$ is the scalar function that we would like to determine from (67), and $p = (p_j) = \left(\frac{\partial \phi}{\partial x_j}\right)$ is a vector containing the partial derivatives of ϕ.

In the particular case of quasilinear equations, studied in subsection (3.1), F is linear in the p's; i.e. it has the form

$$F = \sum_j a_j(x, \phi)\, p_j - b(x, \phi)\,.$$

In this case, we found that special *characteristic* lines exist, $x = x(s)$, $\phi = \phi(s)$, along which the partial differential equation (67) yields a system of ordinary differential equations for the x's and ϕ. Now we will show that this construction extends to the solution of equation (67) even when F depends nonlinearly on p. In fact, this was shown already in subsection 3.2 for functions F that do not involve ϕ (the phase θ) explicitly, and had been solved for ϕ_t in terms of the other variables, to yield the dispersion relation (61).

In this subsection, we build the characteristics or *rays* of (67) fast and mechanically. Time allowing, we shall explore later the geometrical and physical meaning of this powerful, seemingly magical technique. Assume that special lines exist, and write them in terms of a parameter s in the form

$$\begin{aligned} x &= x(s) \\ \phi &= \phi(s) \\ p &= p(s) \,. \end{aligned}$$

Denoting by dots the derivatives with respect to s, the components of p satisfy the equation

$$\dot{p}_j = \dot{\frac{\partial \phi}{\partial x_j}} = \sum_i \frac{\partial^2 \phi}{\partial x_j \partial x_i} \, \dot{x}_i \,. \tag{68}$$

In order to get closed equations involving only x, ϕ and p, we need to eliminate the second derivatives of ϕ from (68). To do this, differentiate (67) with respect to x_i, where F is considered as a function of the x's alone, through the –yet unknown– dependence of ϕ and p on x:

$$0 = \frac{dF}{dx_j} = \frac{\partial F}{\partial x_j} + p_j \, \frac{\partial F}{\partial \phi} + \sum_i \frac{\partial^2 \phi}{\partial x_j \partial x_i} \, \frac{\partial F}{\partial p_i} \,. \tag{69}$$

Comparing (68) and (69), we notice that the choice

$$\dot{x}_j = \frac{\partial F}{\partial p_j} \tag{70}$$

turns (68) into

$$\dot{p}_j = -\frac{\partial F}{\partial x_j} - p_j \, \frac{\partial F}{\partial \phi} \,. \tag{71}$$

To complete the system (70, 71), we need an equation for $\dot{\phi}$; this is given by

$$\dot{\phi} = \sum_j \frac{\partial \phi}{\partial x_j} \, \dot{x}_j = \sum_j p_j \, \frac{\partial F}{\partial p_j} \,. \tag{72}$$

The equations in (70, 71, 72) form a closed system, whose solutions determine a family of lines in the $2n+1$ dimensional *extended phase space* (x, ϕ, p). Initial data are typically given by the values of $\phi = \phi(q)$ over a surface $x = x(q)$, where $q = (q_j)$ is an $n-1$ dimensional parameter. To start the solution along characteristics, though, we need to compute the $p_j(q)$'s as well. Their values follow from the system

$$\frac{\partial \phi}{\partial q_j} = \sum_i p_i \frac{\partial x_i}{\partial q_j} \tag{73}$$

$$0 = F(x(q), \phi(q), p(q))\,. \tag{74}$$

For each value of q, this system for p is not degenerate if the normal to the surface (74) in p space, $n^s = (n_i^s) = \left(\frac{\partial F}{\partial p_i}\right)$, is not in the space generated by the normals to the planes (73), $n^j = \left(n_i^j\right) = \frac{\partial x_i}{\partial q_j}$. In view of (70), this is equivalent to saying that the characteristic passing though the point $(x(q), \phi(q), p(q))$, projected onto the space x_j of independent variables, should not be tangent to the surface $x(q)$ at the point q. When this condition is not satisfied, the initial data and the differential equation may contradict each other and, even if this is not the case, they are not enough to extend the solution away from the initial surface. Such degenerate surfaces are called, again, *characteristics*. A partial justification for such an abuse of notation is that, when there are only two independent variables, the characteristic surfaces and the characteristic lines are the same geometrical objects.

3.4 Hamilton-Jacobi equations

Back in the section on dispersive waves, we wrote a general dispersion relation (61) relating the frequency ω and the wave vector k. Here we rewrite the same equation, just changing notation:

$$S_t + H(x, t, p) = 0\,, \tag{75}$$

where $p = \nabla_x S$. Clearly H, the *Hamiltonian*, plays the role of the dispersion relation Ω, S, the *action*, that of the phase ϕ, p, the *momenta*, replace the wave vector k, and the frequency ω just lost its name, being now simply refered to by its definition, $-S_t$. In this new language, an equation of the general form (75) is called a *Hamilton–Jacobi* equation, and plays a central role in theoretical physics. There is more than a historical reason for this duality of language: the fact that the fundamental notions of waves and mechanics can be formulated in identical terms amazed Hamilton when he discovered it, and lies at the heart of their identification in quantum mechanics.

The characteristic equations for (75) are, using the time t for their parameterization,

$$\dot{x} = \nabla_p H \tag{76}$$

$$\dot{p} = -\nabla_x H \tag{77}$$

$$\dot{H} = H_t\,, \tag{78}$$

the *Hamilton equations* of classical mechanics. Notice that, if we identify $x(t)$ with the trajectory of a particle or group of particles, the vector p with the actual mechanical momenta $m\,\dot{x}$ (where m denotes the particle's mass, taken here for notational simplicity to be the same for all particles), and the Hamiltonian with the system's total energy

$$H = \frac{|p|^2}{2\,m} + V(x)\,, \tag{79}$$

(where V represents the system's potential energy and the quadratic form in the p's its kinetic energy), then the characteristic equations become

$$\dot{x} = \frac{p}{m} \tag{80}$$

$$\dot{p} = -\nabla_x V\,, \tag{81}$$

Newton's equations of motion in a potential field. Behind this fact lie the variational principles of mechanics and optics and the formal grounds for Quantum Mechanics. I hope that you will find the time to explore these fascinating fields, which unfortunately lie outside the scope of this class.