

HANDOUT #9: INTRODUCTION TO WAVES

1 What is a wave?

In everyday life we are familiar with lots of examples of **waves**: ripples on the surface of a pond; a traveling wave along a rope; the oscillations of a violin string or an organ pipe; sound waves in the air; electromagnetic waves (i.e., light, radio, etc.); and many other examples. But it is surprisingly difficult to give a general definition of what one means by a wave!

The main idea to keep in mind is that a wave is a *pattern*, not a thing; usually it is a pattern that moves or changes in some way.¹ Of course, this pattern is usually formed (or “carried”) by the motion of some “medium”: for instance, the ripples on the surface of a pond are formed by the motion of water molecules; a sound wave is formed by the motion of molecules in the air; and so forth. But the motion of the wave can be very different — both in direction and in magnitude — from the motion of the medium that “carries” the wave:

- When we drop a stone into a placid pond, a wave propagates concentrically outwards along the surface of the pond (in principle to infinity, and in practice for many metres). But the water molecules move *up and down* (and only a few millimetres).
- When a person speaks, a sound wave propagates concentrically outwards in three-dimensional space (sometimes to hundreds of metres). But the molecules in the air oscillate *back and forth*, and only very small distances (comparable to the typical separation between molecules in the air).

Here are two even more bizarre examples:

- Imagine a queue to buy tickets at the cinema. When the person at the front of the queue finishes buying his ticket, he departs from the queue, leaving an empty space at the front. Then the second person moves into this empty space, leaving in turn an empty space in position #2; then the third person moves into position #2, leaving an empty space in position #3; and so forth. The pattern here is the empty space: it moves *backwards* (quite possibly for a long distance). The medium that carries this pattern is the group of people: they each move *forwards* (but only one step).
- In a Mexican wave in a stadium, successive rows of spectators stand, raise their arms, and then sit again. This gives rise to a pattern (standing spectators) that moves *horizontally* through the stadium (quite possibly for a long distance). But the medium that carries this pattern, namely the spectators, simply move up and down (one metre

¹For this point I am indebted to Ralph Baierlein, *Newton to Einstein: The Trail of Light* (Cambridge University Press, Cambridge–New York, 1992), p. 60.

or so) without leaving their seats. [This is of course completely analogous to a ripple on a pond.]

Waves can be classified as **longitudinal** or **transverse**, according as the motion of the medium is parallel or perpendicular to the direction of propagation of the wave. For instance, sound waves are longitudinal: the molecules in the air oscillate back and forth along the direction of propagation of the wave. By contrast, waves on the surface of water, or traveling waves along a rope, are transverse: the molecules in the medium oscillate up and down while the wave propagates horizontally. Electromagnetic waves are also transverse: the electric and magnetic field vectors are perpendicular to the direction of propagation (and also perpendicular to each other). Waves on a slinky can be either longitudinal or transverse (that is one of the reasons that a slinky is so much fun to play with!).

Waves in free space are typically **propagating waves**: a pattern propagates in some way (perhaps in one direction, perhaps concentrically outwards); the pattern may or may not change in shape as it propagates.² On the other hand, when waves are constrained by *boundary conditions* — for instance, a violin string tied down at both ends, or an organ pipe closed at one or both ends — it is possible to set up **standing waves**: these are fixed shapes whose amplitude oscillates in time, so that the wave $f(x, t)$ is of the simple product form $g(x)h(t)$ [where $h(t)$ is usually a simple oscillation $A \cos \omega t$, while $g(x)$ is the **normal mode**].

2 Waves on an elastic string

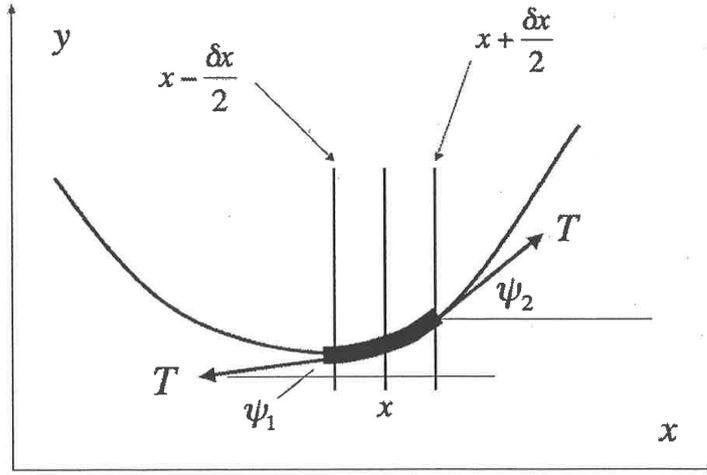
Let us now derive a differential equation for the transverse motion of a string, in the linearized (i.e. small-amplitude) approximation. The equation will describe the vertical height $y(x, t)$ of the string at point x along the string and time t . Since y depends on both x and t , we will have a *partial differential equation*.

We work in the absence of gravity; therefore, “horizontal” is simply a synonym for “longitudinal” (i.e. along the direction of the string), and “vertical” is simply a synonym for “transverse” (i.e. perpendicular to the direction of the string).

2.1 Direct derivation

We assume that the string has a constant linear density (= mass per unit length) ρ and is held at a tension T . Because of our small-amplitude approximation, we may assume that the tension T does not change as the string oscillates. Let us now consider a small segment of the string, of width δx , centered at x , as shown in the figure below:

²We will see that the one-dimensional wave equation has solutions that propagate to the left or right without changing their shape. On the other hand, for the wave equation in two or more spatial dimensions, solutions do usually change their shape as they propagate.



The mass of this segment of string is $\rho \delta x$. Consider now the forces acting on this segment of string. The string to the right of this segment pulls on it with a horizontal force $T \cos \psi_2$ and a vertical force $T \sin \psi_2$, where ψ_2 is the angle that the tangent to the string makes with the horizontal axis at the rightmost end of our segment. Likewise, the string to the left of this segment pulls on it with a horizontal force $-T \cos \psi_1$ and a vertical force $-T \sin \psi_1$, where ψ_1 is the angle that the tangent to the string makes with the horizontal axis at the leftmost end of our segment. The net vertical force on this segment is therefore

$$F_{\text{vert}} = T \sin \psi_2 - T \sin \psi_1 . \quad (1)$$

Our small-amplitude approximation means that we assume the angles ψ_1, ψ_2 are small (they are exaggerated in the figure for clarity); therefore $\sin \psi_1 \approx \psi_1$ and $\sin \psi_2 \approx \psi_2$, and hence in this approximation we have

$$F_{\text{vert}} = T(\psi_2 - \psi_1) . \quad (2)$$

On the other hand,

$$\frac{\partial y}{\partial x}(x + \delta x/2, t) = \tan \psi_2 \approx \psi_2 \quad (3a)$$

$$\frac{\partial y}{\partial x}(x - \delta x/2, t) = \tan \psi_1 \approx \psi_1 \quad (3b)$$

Now³

$$\frac{\partial y}{\partial x}(x + \delta x/2, t) - \frac{\partial y}{\partial x}(x - \delta x/2, t) = \frac{\partial^2 y}{\partial x^2}(x, t) \delta x + O((\delta x)^2) , \quad (4)$$

so the net vertical force on this segment of string is

$$F_{\text{vert}} = T \frac{\partial^2 y}{\partial x^2}(x, t) \delta x + O((\delta x)^2) . \quad (5)$$

³Here we have just used the Taylor expansion $f(x + \epsilon) = f(x) + \epsilon f'(x) + O(\epsilon^2)$ for the function $f = \partial y / \partial x$ and $\epsilon = \pm \delta x / 2$; the variable t just goes for the ride.

By Newton's second law, the net vertical force on this segment should equal its mass $\rho \delta x$ multiplied by its vertical acceleration of its center of mass. But the vertical acceleration of every point in this small segment is essentially the same as the vertical acceleration of its center point, which is $\frac{\partial^2 y}{\partial t^2}(x, t)$. (More precisely, different points within this segment will have vertical accelerations that differ from this by an amount of order δx .) We therefore have

$$(\rho \delta x) \left[\frac{\partial^2 y}{\partial t^2}(x, t) + O(\delta x) \right] = T \frac{\partial^2 y}{\partial x^2}(x, t) \delta x + O((\delta x)^2). \quad (6)$$

Dividing through by δx and letting $\delta x \rightarrow 0$, we obtain the partial differential equation

$$\rho \frac{\partial^2 y}{\partial t^2} = T \frac{\partial^2 y}{\partial x^2}. \quad (7)$$

It is convenient to define $c = \sqrt{T/\rho}$ (we will see later that c is the speed of propagation of waves along this string); we therefore have

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}. \quad (8)$$

Remark. The net *horizontal* force on this segment is

$$F_{\text{horiz}} = T \cos \psi_2 - T \cos \psi_1, \quad (9)$$

which is *zero* in the small-amplitude approximation (i.e., expanding everything through *linear* order in the angles ψ) because to this order we have $\cos \psi_1 \approx 1$ and $\cos \psi_2 \approx 1$. So there is no horizontal motion.

Let us henceforth rename the dependent variable y as f , in order to stress that the “thing that oscillates” need not be the height of a string; it could be the pressure in a sound wave, the voltage in an electrical transmission line, or many other things. The partial differential equation

$$\frac{\partial^2 f}{\partial t^2} = c^2 \frac{\partial^2 f}{\partial x^2} \quad (10)$$

for the function $f(x, t)$ is called the **wave equation in one spatial dimension** — or the **one-dimensional wave equation** for short. Sometimes we prefer to divide this equation by c^2 and write instead

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 f}{\partial t^2}. \quad (11)$$

And sometimes we write this equation in the form

$$\left(\frac{\partial^2}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) f = 0 \quad (12)$$

in order to stress the role played by the **one-dimensional wave operator** $\frac{\partial^2}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}$.

More generally, the **wave equation in n spatial dimensions** is

$$\Delta f = \frac{1}{c^2} \frac{\partial^2 f}{\partial t^2} \quad (13)$$

where

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_n^2} \quad (14)$$

is the **Laplacian** in n spatial dimensions (also denoted ∇^2). It turns out that the wave equation in $n \geq 2$ spatial dimensions is much more interesting and subtle than the case $n = 1$ that we are studying here! But understanding the solution of the one-dimensional problem is a good warm-up for studying the more difficult multidimensional problem.

2.2 Alternate derivation as limit of discrete system

The equation (7) for the motion of the constant-density string can also be derived in a slightly different way, by starting from the equations of motion of the massless string loaded with discrete masses (derived in Section 5 of Handout #8) and taking the “continuum limit”. Recall the situation we treated there: A massless string of length $L = (n + 1)d$ is tied down at the two ends $x = 0$ and $x = L$ and held at tension T ; on this string we attach n particles, each of mass m , at the locations $x = d, 2d, 3d, \dots, nd$. So the j th particle ($j = 1, 2, \dots, n$) has longitudinal location $x = jd$; we denote its transverse displacement by y_j .

We now consider the continuous string of linear density ρ to be a limit of the discrete loaded string in which we take $d \rightarrow 0$ and $m \rightarrow 0$ with the ratio $d/m = \rho$ held fixed.⁴ By equation (44) of Handout #8 together with Newton’s second law, the equation of motion for the discrete loaded string (in the linearized approximation) is

$$m\ddot{y}_j = \frac{T}{d}(y_{j-1} - 2y_j + y_{j+1}). \quad (15)$$

We now assume that the y_j are the values of a smooth function $y(x)$ evaluated at the points $x = jd$. Applying Taylor’s formula

$$f(x + \epsilon) = f(x) + \epsilon f'(x) + \frac{1}{2}\epsilon^2 f''(x) + O(\epsilon^3) \quad (16)$$

with $f = y$, $x = jd$ and $\epsilon = d$, we obtain

$$y_{j-1} - 2y_j + y_{j+1} = d^2 \frac{\partial^2 y}{\partial x^2}(x, t) + O(d^3) \quad (17)$$

— that is, the “second difference” $y_{j-1} - 2y_j + y_{j+1}$ is a discrete approximation to the second derivative $\partial^2 y / \partial x^2$ (up to a factor d^2). On the other hand, \ddot{y}_j is simply $\partial^2 y / \partial t^2$. Since $m = \rho d$, we have

$$\rho d \frac{\partial^2 y}{\partial t^2} = T d \frac{\partial^2 y}{\partial x^2} + O(d^2). \quad (18)$$

Dividing by d and taking the limit $d \rightarrow 0$, we obtain the one-dimensional wave equation (7).

⁴If the length L of the string is fixed, this means that $n \rightarrow \infty$.

Note that in this derivation we nowhere used the length L of the string or the boundary conditions at the ends. Indeed, in this section we are simply deriving the differential equation (7) obtained by applying Newton’s laws of motion to the *interior* elements of the string. The same differential equation (7) holds irrespective of whether the string has boundaries or not, and irrespective of what goes on at those boundaries if they do exist. Rather, if the string does have boundaries, then the differential equation has to be supplemented by **boundary conditions** — we will discuss this aspect of the problem in Section 4 below.

The connection seen here between the discrete loaded string and the continuous massive string is prototypical of many similar situations arising in both pure and applied mathematics. A continuous problem can frequently be understood as a “continuum limit” of a discrete problem when some sort of “lattice spacing” tends to zero: for instance, the derivative $f'(x)$ is the limit of a difference quotient $[f(x + \epsilon) - f(x)]/\epsilon$ when $\epsilon \rightarrow 0$; the Riemann integral is the limit of Riemann sums; and so forth. This connection between a continuous problem and a sequence of discrete problems can be exploited in various directions:

- Sometimes we use the connection for theoretical purposes. For instance, we might prove upper or lower bounds on a sum by relating it to an integral; this can be useful if the sum is difficult to perform but the integral is easy.⁵ Or we might find that the best way to make sense of an ill-defined continuum expression is to write it as a limit of well-defined discrete expressions (this situation arises, for instance, in quantum field theory).
- Sometimes it is the continuous problem that is of interest to us, but we study the discrete problem (for n finite but large) as an approximation. For instance, if we want to solve a differential equation (ordinary or partial) numerically on the computer, we *must* discretize it, replacing derivatives by finite differences. (It is then prudent to perform the numerical solution for at least two different values of the “lattice spacing” ϵ , in order to test numerically whether the solutions seem to be almost converged.)
- Sometimes it is the discrete problem that is of interest to us, but we start by studying a closely related continuous problem because it is easier to analyze. For instance, differential equations are often simpler to analyze than difference equations, especially if they are nonlinear.⁶

You will see many examples of these techniques in your future study of mathematics.

⁵In analysis (MATH0003) you may have seen the example of the *harmonic sum*

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n},$$

which is not given by any simple explicit formula but which can be bounded above and below in terms of the corresponding definite integral $\int_1^n \frac{1}{t} dt$, which *is* easily computable (of course it equals $\log n$).

⁶For instance, the **logistic differential equation**

$$\frac{dy}{dt} = y(1 - y)$$

3 Solutions of the one-dimensional wave equation in infinite space

In this section we shall find the *general* solution of the one-dimensional wave equation (11) when the spatial variable x ranges over the entire real line — that is, for an infinitely long string, without boundaries.

3.1 Guessing the general solution

We begin by observing that the one-dimensional wave equation admits traveling-wave solutions of arbitrary shape, which move either to the right or to the left at speed c . To see this, let F be *any* twice-continuously-differentiable function of a single real variable, and let us consider the function

$$f(x, t) = F(x - ct) . \tag{19}$$

Physically, this represents a pulse that has shape $F(x)$ at time 0 [since $f(x, 0) = F(x)$] and that moves to the right at speed c without changing its shape. (You should make sure you understand why this pulse moves *to the right* as time goes on, i.e. moves towards larger positive x .)⁷ Now compute the first and second partial derivatives of this $f(x, t)$ with respect to x :

$$\frac{\partial f}{\partial x} = F'(x - ct) \tag{20a}$$

$$\frac{\partial^2 f}{\partial x^2} = F''(x - ct) \tag{20b}$$

(why?). By similar reasoning we may differentiate with respect to t :

$$\frac{\partial f}{\partial t} = -cF'(x - ct) \tag{21a}$$

$$\frac{\partial^2 f}{\partial t^2} = c^2F''(x - ct) \tag{21b}$$

has applications in population biology and many other fields; and it is very easy to solve:

$$y(t) = \frac{1}{1 + e^{-(t-t_0)}}$$

where t_0 is the time at which $y = \frac{1}{2}$. (You should derive this solution: note that the equation is separable.) On the other hand, the discrete-time **logistic map**

$$y_{n+1} = ry_n(1 - y_n) ,$$

where $r \in [0, 4]$ is a parameter — which also has applications in population biology — has incredibly complicated behavior for many values of the parameter r . This example shows that in some cases the simplified continuous model is often a *bad* approximation to the discrete model; it all depends on what aspects of the behavior one is interested in.

⁷To keep at the same point along the pulse, $x - ct$ must be kept constant. Therefore, as t increases, the value of x must also increase in order to stay at the same value of $x - ct$. Indeed, if t increases by Δt , then x must increase by $c\Delta t$ in order to stay at the same value of $x - ct$; so the pulse moves to the right at speed c .

(why?). It follows that $f(x, t) = F(x - ct)$ is a solution of the one-dimensional wave equation (11), and this *no matter what* the function F is, provided only that it is smooth (i.e. twice continuously differentiable).

By an analogous argument, the function

$$f(x, t) = G(x + ct) \tag{22}$$

is also a solution (whenever G is twice continuously differentiable); it corresponds physically to a pulse that moves to the *left* at speed c .

And by linearity, the sum of two solutions is also a solution: therefore,

$$f(x, t) = F(x - ct) + G(x + ct) \tag{23}$$

is a solution of the one-dimensional wave equation (11). Physically, it corresponds to a pair of pulses — one of shape F moving to the right, the other of shape G moving to the left — that propagate without interacting with each other (i.e., the solution is just the arithmetic sum of the two).⁸

Remark. Suppose that we had tried the more general Ansatz

$$f(x, t) = F(x - vt) \tag{24}$$

where the velocity v is unspecified. Then we would have

$$\frac{\partial^2 f}{\partial x^2} = F''(x - vt) \tag{25}$$

and

$$\frac{\partial^2 f}{\partial t^2} = v^2 F''(x - vt), \tag{26}$$

so that this $f(x, t)$ solves one-dimensional wave equation (11) if and only if $v^2 = c^2$, i.e. $v = \pm c$. (And this no matter what the function F is, provided only that it is twice differentiable.) So this more general Ansatz would lead us back to the two solutions $v = c$ and $v = -c$ that we found by clever guessing.⁹

⁸The fact that the two pulses propagate without interacting with each other is, of course, a consequence of the fact that we are dealing here with a *linear* differential equation, so that any linear combination of two solutions is also a solution. The behavior of *nonlinear* partial differential equations is in general very different!

⁹There is actually another solution that we have missed here: Can you see what it is? The equation

$$v^2 F''(x - vt) = c^2 F''(x - vt)$$

is solved by an arbitrary function F when $v^2 = c^2$; but it is *also* solved for an *arbitrary* value of v whenever $F''(x - vt)$ is identically zero, i.e. when $F(x) = a + bx$. But this is a rather bizarre solution from a physical point of view, since the string's displacement gets huge as $x \rightarrow \pm\infty$.

3.2 Proof that it is the general solution

In fact, it turns out that (23) is the *general* solution to the one-dimensional wave equation, i.e. *every* solution is of this form. To show this, let us make the change of variables from (x, t) to new variables (ξ, η) defined by

$$\xi = x - ct \tag{27a}$$

$$\eta = x + ct \tag{27b}$$

And let us re-express the operators of partial differentiation with respect to x and/or t as operators of partial differentiation with respect to ξ and/or η .

Warning! When computing partial derivatives arising from changes of variable in functions of two or more variables, one must be very careful! The reason is this: when we have a function $f(x, y)$ and we write $\frac{\partial f}{\partial x}$, we mean the rate of change of f as x varies *with y held fixed*, but our notation fails to show this latter aspect; to be more precise we ought to write $\left. \frac{\partial f}{\partial x} \right|_y$ in order to make clear not only which variable is being varied but also which variable is being held fixed. Ordinarily this does not matter, because we know implicitly which variable is being held fixed; but when we make changes of variable this can become ambiguous. For instance, suppose that we change from y to the new variable $y' = y + x^2$. We can write f either as a function of x and y or as a function of x and y' . But $\left. \frac{\partial f}{\partial x} \right|_y \neq \left. \frac{\partial f}{\partial x} \right|_{y'}$! This fact causes endless complication (and confusion) in some applied subjects, such as thermodynamics.

In the case at hand, there will be no ambiguity: whenever we write partial derivatives with respect to x and/or t , we assume that everything is being written as a function of x and t ; so, for instance, $\frac{\partial}{\partial x}$ means $\left. \frac{\partial}{\partial x} \right|_t$. Likewise, whenever we write partial derivatives with respect to ξ and/or η , we assume that everything is being written as a function of ξ and η , so that $\frac{\partial}{\partial \xi}$ means $\left. \frac{\partial}{\partial \xi} \right|_\eta$. We never mix the two sets of variables.

If we increase x by ϵ with t being held fixed, this means that both ξ and η increase by ϵ ; therefore

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta} . \tag{28}$$

Likewise, if we increase t by ϵ with x being held fixed, this causes η to increase by $c\epsilon$ and ξ to decrease by $c\epsilon$; hence

$$\frac{\partial}{\partial t} = c \frac{\partial}{\partial \eta} - c \frac{\partial}{\partial \xi} . \tag{29}$$

The wave operator $\frac{\partial^2}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}$ can therefore be re-expressed in terms of ξ and η as

$$\frac{\partial^2}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} = \left(\frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta} \right)^2 - \frac{1}{c^2} \left(c \frac{\partial}{\partial \eta} - c \frac{\partial}{\partial \xi} \right)^2 \tag{30a}$$

$$= \left(\frac{\partial^2}{\partial \xi^2} + 2 \frac{\partial^2}{\partial \xi \partial \eta} + \frac{\partial^2}{\partial \eta^2} \right) - \left(\frac{\partial^2}{\partial \eta^2} - 2 \frac{\partial^2}{\partial \xi \partial \eta} + \frac{\partial^2}{\partial \xi^2} \right) \quad (30b)$$

$$= 4 \frac{\partial^2}{\partial \xi \partial \eta} . \quad (30c)$$

So the one-dimensional wave equation can be rewritten in the variables ξ and η as

$$\frac{\partial^2 f}{\partial \xi \partial \eta} = 0 . \quad (31)$$

Now, what is the general solution to the partial differential equation (31)? That is, which functions $f(\xi, \eta)$ have the property that their mixed partial derivative $\partial^2 f / \partial \xi \partial \eta$ vanishes identically? It is easy to see that any function of the form $f(\xi, \eta) = F(\xi) + G(\eta)$ has this property; and with a bit more thought one can see that *only* functions of this form have this property.

Here is a proof of the latter statement. Let us rewrite the equation $\partial^2 f / \partial \xi \partial \eta = 0$ as

$$\frac{\partial}{\partial \xi} \left(\frac{\partial f}{\partial \eta} \right) = 0 . \quad (32)$$

This says that the function $g = \partial f / \partial \eta$ does not depend on ξ ; therefore it is a function solely of η , i.e.

$$\frac{\partial f}{\partial \eta} = H(\eta) \quad (33)$$

for some function H . We now solve this latter equation: f is the antiderivative of $H(\eta)$ [let us call this $G(\eta)$] plus a “constant of integration” that does not depend on η but may depend on ξ [let us call this $F(\xi)$]. Therefore $f(\xi, \eta) = G(\eta) + F(\xi)$.

Returning to the variables x and t , it follows that the general solution to the one-dimensional wave equation (11) is

$$f(x, t) = F(x - ct) + G(x + ct) . \quad (34)$$

3.3 Interpretation as the initial-value problem

We can understand the general solution (34) in another way, in connection with the **initial-value problem** for the one-dimensional wave equation. Recall first that an n th-order *ordinary* differential equation for a function $f(t)$ has a general solution that involves n arbitrary “constants of integration”; and these n numbers can be determined, for any particular solution, by specifying n *initial conditions* (usually the value of f and its first $n - 1$ derivatives at some initial time). Likewise, a partial differential equation that is of n th order in time, for a function $f(t, \mathbf{x})$ [here \mathbf{x} denotes one or more spatial variables], will have a general solution that involves n arbitrary “functions of integration” (that is, “constants of integration” that are constant as concerns t but can be arbitrary functions of \mathbf{x}); and these n functions can be determined, for any particular solution, by specifying n *initial functions* (usually the value of f and its first $n - 1$ partial derivatives with respect to time, throughout space, at some initial time).

Since the wave equation (11) [or (13)] is of *second* order in time, we will need to specify *two* initial functions: for instance, the values of f and $\partial f/\partial t$ throughout space at time 0. Let us denote these by φ and ψ :

$$\varphi(x) \stackrel{\text{def}}{=} f(x, 0) \tag{35a}$$

$$\psi(x) \stackrel{\text{def}}{=} \frac{\partial f}{\partial t}(x, 0) \tag{35b}$$

In our general solution $f(x, t) = F(x - ct) + G(x + ct)$, we have

$$\varphi(x) = F(x) + G(x) \tag{36a}$$

$$\psi(x) = c[G'(x) - F'(x)] \tag{36b}$$

Letting $\Psi(x)$ be an antiderivative of $\psi(x)$, it follows that

$$G(x) - F(x) = \frac{\Psi(x)}{c} + \kappa \tag{37}$$

where κ is some constant. We then have

$$F(x) = \frac{1}{2} \left[\varphi(x) - \frac{\Psi(x)}{c} - \kappa \right] \tag{38a}$$

$$G(x) = \frac{1}{2} \left[\varphi(x) + \frac{\Psi(x)}{c} + \kappa \right] \tag{38b}$$

and hence

$$f(x, t) = F(x - ct) + G(x + ct) \tag{39a}$$

$$= \frac{\varphi(x - ct) + \varphi(x + ct)}{2} + \frac{\Psi(x + ct) - \Psi(x - ct)}{2c} \tag{39b}$$

$$= \frac{\varphi(x - ct) + \varphi(x + ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} \psi(s) ds. \tag{39c}$$

This explicit expression for the solution of the one-dimensional wave equation in terms of the initial data is called **d'Alembert's formula**.¹⁰

3.4 Normal modes via separation of variables

Let us now consider the **normal modes** of an infinite string: that is, we are interested in the solutions of the one-dimensional wave equation (with x ranging over the entire real line) that are of the special form

$$f(x, t) = g(x) e^{i\omega t}. \tag{40}$$

¹⁰Jean-Baptiste le Rond d'Alembert (1717–1783) was a French mathematician and physicist who made many important contributions (some of which we will see later in this course). He studied the one-dimensional wave equation, and found this general formula for its solution, in a 1747 paper.

But it is instructive to see how the “oscillatory Ansatz” (40) arises as the outcome of a more general technique called **separation of variables**: here we search for solutions of the more general form

$$f(x, t) = g(x) h(t) \quad (41)$$

without imposing any *a priori* requirement on what the function h (or g) is.

So let us plug the “separation-of-variables Ansatz” (41) into the one-dimensional wave equation (11): we obtain

$$g''(x) h(t) = \frac{1}{c^2} g(x) h''(t) \quad (42)$$

or, rearranging,

$$c^2 \frac{g''(x)}{g(x)} = \frac{h''(t)}{h(t)}. \quad (43)$$

Note now that the left-hand side of (43) depends only on x , while the right-hand side depends only on t . Therefore, if they are equal to each other, they must both be equal to the same *constant*, i.e. independent of both x and t ! Let us call this constant κ . We then have the pair of equations

$$\frac{g''(x)}{g(x)} = \frac{\kappa}{c^2} \quad (44a)$$

$$\frac{h''(t)}{h(t)} = \kappa \quad (44b)$$

or equivalently

$$\frac{d^2g}{dx^2} - \frac{\kappa}{c^2}g = 0 \quad (45a)$$

$$\frac{d^2h}{dt^2} - \kappa h = 0 \quad (45b)$$

The general solutions are therefore

$$g(x) = Ae^{(\sqrt{\kappa}/c)x} + Be^{-(\sqrt{\kappa}/c)x} \quad (46a)$$

$$h(t) = Ce^{\sqrt{\kappa}t} + De^{-\sqrt{\kappa}t} \quad (46b)$$

If $\kappa > 0$ these are exponential functions, which get very large as x (or t) tends to $+\infty$ or $-\infty$ or both, and are therefore not of much physical interest (we have assumed that the vibrations of our string have *small* amplitude, and these solutions violate that assumption). So let us concentrate on the case $\kappa < 0$: for convenience we write $\kappa = -\omega^2$, so that the solutions are

$$g(x) = Ae^{i(\omega/c)x} + Be^{-i(\omega/c)x} \quad (47a)$$

$$h(t) = Ce^{i\omega t} + De^{-i\omega t} \quad (47b)$$

We may then write the solution $f(x, t) = g(x)h(t)$ in a shorthand manner as

$$f(x, t) \sim e^{\pm i(\omega/c)x} e^{\pm i\omega t} \quad (48a)$$

$$\sim e^{\pm i(\omega/c)(x \pm ct)} \quad (48b)$$

where what this means is that the actual $f(x, t)$ arising from our separation-of-variables Ansatz is a linear combination of the four expressions gotten from making all possible choices of + and - signs in (48), namely

$$f(x, t) = A_1 e^{i(\omega/c)(x+ct)} + A_2 e^{i(\omega/c)(x-ct)} + A_3 e^{-i(\omega/c)(x+ct)} + A_4 e^{-i(\omega/c)(x-ct)}. \quad (49)$$

These four contributions, taken for all possible ω , are the **normal modes** for the one-dimensional wave equation in infinite space.

Another way of expressing this is to say that the normal modes are

$$f(x, t) = e^{i(kx \pm \omega t)}, \quad (50)$$

where the **wavenumber** $k \in \mathbb{R}$ is related to the **angular frequency** $\omega \in [0, \infty)$ by

$$\omega = c|k|, \quad (51)$$

where c is the **speed of propagation**. Since the wavelength is $\lambda = 2\pi/k$ and the frequency is $f = \omega/2\pi$, the relation (51) can be rewritten as

$$f\lambda = c, \quad (52)$$

which you probably learned in high school.

For the n -dimensional wave equation (13) the normal modes are

$$f(\mathbf{x}, t) = e^{i(\mathbf{k} \cdot \mathbf{x} \pm \omega t)} \quad (53)$$

where the **wavenumber** $\mathbf{k} \in \mathbb{R}^n$ is related to the **angular frequency** $\omega \in [0, \infty)$ by

$$\omega = c|\mathbf{k}|. \quad (54)$$

In general, formulae expressing the angular frequency ω as a function of the wavenumber \mathbf{k} are called **dispersion relations**. The dispersion relation for the (one-dimensional or n -dimensional) wave equation is *linear*, in the sense that ω is strictly proportional to $|\mathbf{k}|$. By contrast, other linear partial differential equations with constant coefficients can have *nonlinear* dispersion relations.

Note that in infinite space (unlike the case of a finite string tied down at the ends) there are no boundary conditions forcing the frequency ω to lie in some special set of values; *all* frequencies ω give rise to legitimate solutions.

The *general* solution of the one-dimensional wave equation in infinite space is then a linear combination of these normal modes. Since ω ranges over the entire real line (rather than just a discrete set), this “linear combination” may have to be interpreted as an integral rather than as a finite or countably infinite sum. Since the normal modes are just complex exponentials in both space and time, what we are really doing is **Fourier analysis**.

4 Solutions of the one-dimensional wave equation on a finite interval

Let us now take a look at the solutions of the one-dimensional wave equation (11) when the spatial variable x ranges over a finite interval of the real line, say $0 \leq x \leq L$. In this case the partial differential equation (11) has to be supplemented with **boundary conditions** specifying how the solution $f(x, t)$ is constrained to behave at the two endpoints $x = 0$ and $x = L$. The two most commonly arising types of boundary conditions are:

- **Dirichlet boundary conditions.** We impose $f = 0$ at the given endpoint, at all times. If f is the transverse displacement of a vibrating string, this means that the string is tied down at the given endpoint. If f is the air pressure in an organ pipe, this means that the pipe is *open* at the given endpoint.¹¹
- **Neumann boundary conditions.** We impose $\partial f / \partial x = 0$ at the given endpoint, at all times. If f is the transverse displacement of a vibrating string, this means that at the given endpoint we attach a loop to the string and allow it to slide frictionlessly along a vertical pole.¹² If f is the air pressure in an organ pipe, this means that the pipe is *closed* at the given endpoint.

Let us not attempt to find the general solution to the one-dimensional wave equation on a finite interval with boundary conditions, but simply pass to the computation of the normal modes.

4.1 Normal modes via separation of variables

We can copy verbatim the analysis from Section 3.4, up through equation (47): we make the separation-of-variables Ansatz $f(x, t) = h(x)g(t)$ and find the solution

$$g(x) = Ae^{i(\omega/c)x} + Be^{-i(\omega/c)x} \quad (55a)$$

$$h(t) = Ce^{i\omega t} + De^{-i\omega t} \quad (55b)$$

or equivalently

$$g(x) = A' \cos(\omega x/c) + B' \sin(\omega x/c) \quad (56a)$$

$$h(t) = C' \cos(\omega t) + D' \sin(\omega t) \quad (56b)$$

Now we impose the boundary conditions, first at $x = 0$:

¹¹At the end of a pipe open to the air, the pressure cannot oscillate; rather, it is fixed at the ambient pressure p_0 of the surrounding air. Since the constant function p_0 solves the one-dimensional wave equation, and this equation is linear, we can make a change of dependent variable $f \rightarrow f - p_0$, i.e. we can reinterpret f as meaning the *difference* between the observed pressure and the ambient pressure. With this reinterpretation, we have the boundary condition $f = 0$ at an open end of the pipe.

¹²The loop is assumed massless. Therefore the net vertical force on it must be zero, so the tangent to the string must be horizontal at the endpoint, i.e. $\partial f / \partial x = 0$.

- If the boundary condition at $x = 0$ is Dirichlet [i.e., $f(0, t) = 0$ for all t], then we must choose $g(0) = 0$, hence $A' = 0$ — that is, the g solution is pure sine.
- If the boundary condition at $x = 0$ is Neumann [i.e., $(\partial f / \partial x)(0, t) = 0$ for all t], then we must choose $g'(0) = 0$, hence $B' = 0$ — that is, the g solution is pure cosine.

We then impose in the same way the boundary condition at $x = L$, namely $g(L) = 0$ for a Dirichlet boundary condition or $g'(L) = 0$ for a Neumann boundary condition; this will restrict the frequency ω to lie in a suitable discrete set. There are four cases:

- **Dirichlet at both ends.** The Dirichlet condition at $x = 0$ forces $g(x) \sim \sin(\omega x/c)$. The Dirichlet condition at $x = L$ then forces $\omega L/c$ to be an integer multiple of π , i.e.

$$\omega = j \frac{\pi c}{L} \quad \text{for } j = 1, 2, \dots \quad (57)$$

- **Dirichlet at one end and Neumann at the other.** There are two possibilities (D–N and N–D): let us consider the first. The Dirichlet condition at $x = 0$ again forces $g(x) \sim \sin(\omega x/c)$. The Neumann condition at $x = L$ then forces $\omega L/c$ to be an odd-integer multiple of $\pi/2$, i.e.

$$\omega = (2j - 1) \frac{\pi c}{2L} = (j - \frac{1}{2}) \frac{\pi c}{L} \quad \text{for } j = 1, 2, \dots \quad (58)$$

The second possibility (N–D) can be treated analogously. But there is no need: the two possibilities (D–N and N–D) are interchanged by $x \leftrightarrow L - x$ (which leaves the wave equation invariant), so the frequencies are the same for both (and the normal modes are related by $x \leftrightarrow L - x$).

- **Neumann at both ends.** The Neumann condition at $x = 0$ forces $g(x) \sim \cos(\omega x/c)$. The Neumann condition at $x = L$ then forces $\omega L/c$ to be an integer multiple of π , i.e.

$$\omega = j \frac{\pi c}{L} \quad \text{for } j = 1, 2, \dots \quad (59)$$

(In principle $j = 0$ is also allowed, but this corresponds to the trivial solution $f(x, t) = \text{const}$ of the one-dimensional wave equation.) The frequencies are thus the same as in the Dirichlet-at-both-ends case, although the normal modes look very different (multiples of half-cycles of \cos instead of \sin).

Thus, for instance, an organ pipe that is closed at one end and open at the other will have a fundamental frequency ($j = 1$) that is *half* that of a pipe of the same length that is open at both ends, i.e. it will sound one octave lower. Also, a pipe that is closed at one end and open at the other resonates only at *odd* harmonics, while a pipe that is open at both ends resonates at *all* harmonics.

We can use these normal modes to solve the **initial-value problem**: the idea is to determine the coefficient of each normal mode by Fourier-analyzing the initial data $\varphi(x) = f(x, 0)$ and $\psi(x) = \frac{\partial f}{\partial t}(x, 0)$. But we have probably already done enough for an introduction to waves!