Delta "Functions"

$\delta(x-z)$ (also written $\delta(x,z)$, $\delta_z(x)$, or $\delta_0(x-z)$) is a make-believe function with these properties:

1. $\delta(x-z) = 0$ for all $x \neq z$, and

$$\int_{-\infty}^{\infty} \delta(x-z)\,dx = 1.$$

2. **The key property:** For all continuous functions f,

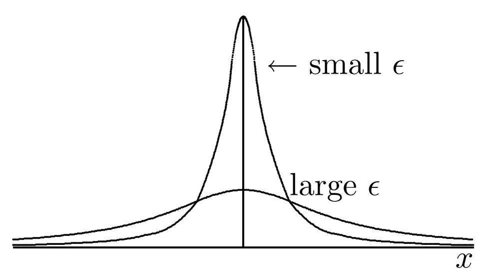
$$\int_{-\infty}^{\infty} \delta(x-z)\,f(x)\,dx = f(z).$$

Also,

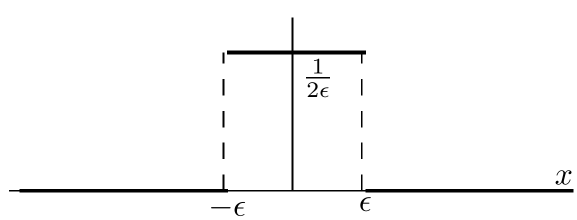
$$\int_a^b \delta(x-z)\,f(x)\,dx = \begin{cases} f(z) & \text{if } z \in (a,b), \\ 0 & \text{if } z \notin [a,b]. \end{cases}$$

3. $\delta(x)$ is the limit of a family of increasingly peaked functions, each with integral 1:

$$\delta(x) = \lim_{\epsilon \downarrow 0} \frac{1}{\pi} \frac{\epsilon}{x^2+\epsilon^2}$$
$$\text{or } \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon\sqrt{\pi}} e^{-x^2/\epsilon^2}$$
$$\text{or } \lim_{\epsilon \downarrow 0} d_\epsilon(x),$$

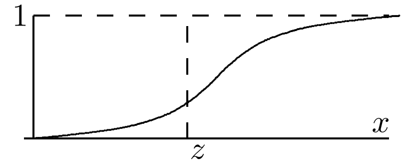


where d_ϵ is a step function of the type drawn here:

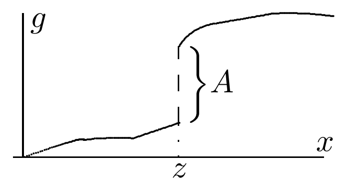


4. $\delta(x-z) = \dfrac{d}{dx} h(x-z)$, where $h(w)$ is the unit step function, or Heaviside function (equal to 1 for $w > 0$ and to 0 for $w < 0$). Note that $h(t-z)$ is the

limit as $\epsilon \downarrow 0$ of a family of functions of this type:



Generalization of 4: If $g(x)$ has a jump discontinuity of size A at $x = z$, then its "derivative" contains a term $A\,\delta(x-z)$. (A may be negative.)

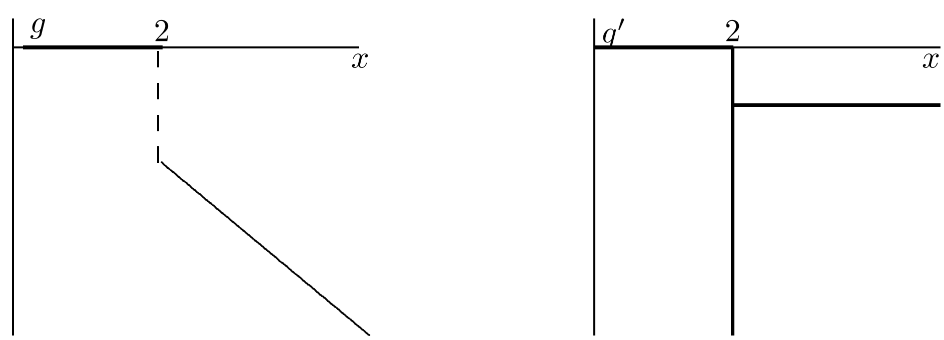


Example:

$$g(x) = \left\{ \begin{array}{rl} 0 & \text{for } x < 2, \\ -x & \text{for } x \geq 2 \end{array} \right\} = -x\,h(x-2).$$

Then

$$\begin{aligned} g'(x) &= -h(x-2) - x\,h'(x-2) \\ &= -h(x-2) - 2\,\delta(x-2). \end{aligned}$$



Interpretation of differential equations involving δ

Example: $$y'' + p(x)y' + q(x)y = A\,\delta(x-z).$$

We expect the solution of this equation to be the limit of the solution of an equation whose source term is a finite but very narrow and hard "kick" at $x = z$. The δ equation is easier to solve than one with a finite peak.

The equation is taken to mean:

$$y'' + py' + qy = 0 \quad \text{for } x < z. \tag{1}$$

$$y'' + py' + qy = 0 \quad \text{for } x > z. \tag{2}$$

$$y \text{ is continuous at } z\text{:} \quad \lim_{x \downarrow z} y(x) = \lim_{x \uparrow z} y(x). \tag{3}$$

[*Notational remarks:* $\lim_{x\downarrow z}$ means the same as $\lim_{x\to z^+}$; $\lim_{x\uparrow z}$ means $\lim_{x\to z^-}$. Also, $\lim_{x\downarrow z} y(x)$ is sometimes written $y(z^+)$, and so on.]

$$\lim_{x \downarrow z} y'(x) = \lim_{x \uparrow z} y'(x) + A. \tag{4}$$

Conditions (3) and (4) tell us how to match solutions of (1) and (2) across the joint. Here is the reasoning behind them:

Assume (3) for the moment. Integrate the ODE from $x = z - \epsilon$ to $x = z + \epsilon$ (where ϵ is very small):

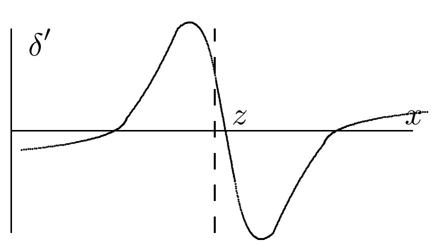
$$\int_{z-\epsilon}^{z+\epsilon} y''\,dx + \int_{z-\epsilon}^{z+\epsilon} (py' + qy)\,dx = A \int_{z-\epsilon}^{z+\epsilon} \delta(x-z)\,dx$$

That is,

$$y'(z+\epsilon) - y'(z-\epsilon) + \text{small term } (\to 0 \text{ as } \epsilon \downarrow 0) = A.$$

In the limit $\epsilon \to 0$, (4) follows.

Now if y *itself* had a jump at z, then y' would contain $\delta(x-z)$, so y'' would contain $\delta'(x-z)$, which is a singularity "worse" than δ. (It is a limit of functions like the one in the graph shown here.) Therefore, (3) is necessary.



We can solve such an equation by finding the general solution on the interval to the left of z and the general solution to the right of z, and then matching the function and its derivative at z by rules (3) and (4) to determine the undetermined coefficients.

Consider the example

$$y'' = \delta(x-1), \qquad y(0) = 0, \quad y'(0) = 0.$$

For $x < 1$, we must solve the equation $y'' = 0$. The general solution is $y = Ax + B$, and the initial conditions imply then that

$$y = 0 \quad \text{for } x < 1.$$

For $x > 1$, we again must have $y'' = 0$ and hence $y = Cx + D$ (different constants this time). On this interval we have $y' = C$. To find C and D we have to apply rules (3) and (4):

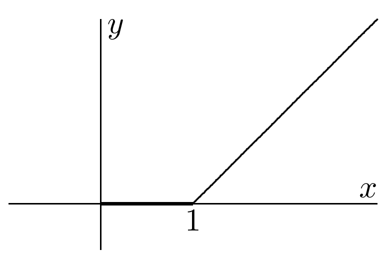
$$0 = y(1^-) = y(1^+) = C + D,$$
$$0 + 1 = y'(1^-) + 1 = y'(1^+) = C.$$

That is,

$$\begin{aligned} C + D &= 0, \\ C \phantom{{}+D} &= 1. \end{aligned}$$

Therefore, $C = 1$ and $D = -1$. Thus $y(x) = x - 1$ for $x > 1$. The complete solution is therefore

$$y(x) = (x-1)\, h(x-1).$$



Delta functions and Green functions

A type of PDE problem which for some reason is not much discussed in courses of this sort, although it certainly might well be, is the *nonhomogeneous* partial differential equation. So far our nonhomogeneities have been initial or boundary data, not terms in the PDE itself. But problems like

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

and

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = j(x,y),$$

where ρ and j are given functions, certainly do arise in practice. Often transform techniques or separation of variables can be used to reduce such PDEs to nonhomogeneous *ordinary* differential equations (a single ODE in situations of extreme symmetry, but more often an infinite family of ODEs).

Here I will show how the delta function and the concept of a Green function can be used to solve nonhomogeneous ODEs.

Example 1: *The Green function for the one-dimensional Dirichlet problem.* Let's start with an equation containing our favorite linear differential operator:

$$\frac{d^2 y}{dx^2} + \omega^2 y = f(x). \qquad (*)$$

We require that

$$y(0) = 0, \qquad y(\pi) = 0.$$

Here ω is positive, and f is a "known" but arbitrary function. Thus our solution will be a formula for y in terms of f. In fact, it will be given by a Green-function integral:

$$y(x) = \int_0^\pi G_\omega(x,z)\, f(z)\, dz,$$

where G is independent of f — but, of course, depends on the left-hand side of the differential equation $(*)$ and on the boundary conditions.

Much as we did earlier (in a homework assignment) when discussing Green functions for *boundary* data in the role of f, we can solve the problem for general f by studying the equation

$$\frac{d^2 y}{dx^2} + \omega^2 y = \delta(x-z) \qquad (*_z)$$

(with the same boundary conditions). We will give the solution of $(*_z)$ the name $G_\omega(x,z)$. Since

$$f(x) = \int_0^\pi \delta(x-z)\, f(z)\, dz$$

(for x in the interval $(0, \pi)$) and since the operator on the left-hand side of $(*)$ is linear, we expect that

$$y(x) \equiv \int_0^\pi G_\omega(x - z)\, f(z)\, dz$$

will be the solution to our problem! That is, since the operator is linear, it can be moved inside the integral (which is a limit of a sum) to act directly on the Green function:

$$\begin{aligned} \frac{d^2 y}{dx^2} + \omega^2 y &= \int_0^\pi \left(\frac{d^2}{dx^2} + \omega^2 \right) G_\omega(x - z)\, f(z)\, dz \\ &= \int_0^\pi \delta(x - z)\, f(z)\, dz \\ &= f(x), \end{aligned}$$

as desired. Furthermore, since G vanishes when $x = 0$ or π, so does the integral defining y; so y satisfies the right boundary conditions.

Therefore, the only task remaining is to solve $(*_z)$. We go about this with the usual understanding that

$$\delta(x - z) = 0 \quad \text{whenever } x \neq z.$$

Thus $(*_z)$ implies

$$\frac{d^2 G_\omega(x, z)}{dx^2} + \omega^2 G_\omega(x, z) = 0 \quad \text{if } x \neq z.$$

Therefore, for some constants A and B,

$$G_\omega(x, z) = A \cos \omega x + B \sin \omega x \quad \text{for } x < z,$$

and, for some constants C and D,

$$G_\omega(x, z) = C \cos \omega x + D \sin \omega x \quad \text{for } x > z.$$

We do *not* necessarily have $A = C$ and $B = D$, because the homogeneous equation for G is not satisfied when $x = z$; that point separates the interval into two disjoint subintervals, and we have a *different* solution of the homogeneous equation on each. Note, finally, that the four unknown "constants" are actually functions of z: there is no reason to expect them to turn out the same for all z's.

We need four equations to determine these four unknowns. Two of them are the boundary conditions:

$$0 = G_\omega(0, z) = A, \qquad 0 = G_\omega(\pi, z) = C \cos \omega\pi + D \sin \omega\pi.$$

The third is that G is continuous at z:

$$A\cos\omega z + B\sin\omega z = G_\omega(z,z) = C\cos\omega z + D\sin\omega z.$$

The final condition is the one we get by integrating $(*_z)$ over a small interval around z:

$$\frac{\partial}{\partial x}G_\omega(z^+,z) - \frac{\partial}{\partial x}G_\omega(z^-,z) = 1.$$

(Notice that although there is no variable "x" left in this equation, the partial derivative with respect to x is still meaningful: it means to differentiate with respect to the first argument of G (*before* letting that argument become equal to the second one).) This last condition is

$$-\omega C\sin\omega z + \omega D\cos\omega z + \omega A\sin\omega z - \omega B\cos\omega z = 1.$$

One of the equations just says that $A=0$. The others can be rewritten

$$\begin{aligned} C\cos\omega\pi + \quad D\sin\omega\pi &= 0, \\ B\sin\omega z - C\cos\omega z - D\sin\omega z &= 0, \\ -\omega B\cos\omega z - \omega C\sin\omega z + \omega D\cos\omega z &= 1. \end{aligned}$$

This system can be solved by Cramer's rule: After a grubby calculation, too long to type, I find that the determinant is

$$\begin{vmatrix} 0 & \cos\omega\pi & \sin\omega\pi \\ \sin\omega z & -\cos\omega z & -\sin\omega z \\ -\omega\cos\omega z & -\omega\sin\omega z & \omega\cos\omega z \end{vmatrix} = -\omega\sin\omega\pi.$$

If ω is *not* an integer, this is nonzero, and so we can go on through additional grubby calculations to the answers,

$$B(z) = \frac{\sin\omega(z-\pi)}{\omega\sin\omega\pi}\,,$$

$$C(z) = -\,\frac{\sin\omega z}{\omega}\,,$$

$$D(z) = \frac{\cos\omega\pi\sin\omega z}{\omega\sin\omega\pi}\,.$$

Thus

$$G_\omega(x,z) = \frac{\sin\omega x\sin\omega(z-\pi)}{\omega\sin\omega\pi} \quad \text{for } x<z,$$

$$G_\omega(x,z) = \frac{\sin\omega z\sin\omega(x-\pi)}{\omega\sin\omega\pi} \quad \text{for } x>z.$$

(Reaching the last of these requires a bit more algebra and a trig identity.)

So we have found the Green function! Notice that it can be expressed in the unified form

$$G_\omega(x,z) = \frac{\sin\omega x_< \sin\omega(x_> - \pi)}{\omega\sin\omega\pi},$$

where

$$x_< \equiv \min(x,z), \qquad x_> \equiv \max(x,z).$$

This symmetrical structure is very common in such problems.

Finally, if ω *is* an integer, it is easy to see that the system of three equations in three unknowns has no solutions. It is no accident that these are precisely the values of ω for which $(*)$'s corresponding *homogeneous* equation,

$$\frac{d^2y}{dx^2} + \omega^2 y = 0,$$

has solutions satisfying the boundary conditions. If the homogeneous problem has solutions (other than the zero function), then the solution of the nonhomogeneous problem (if it exists) must be nonunique, and we have no right to expect to find a formula for it! In fact, the *existence* of solutions to the nonhomogeneous problem also depends upon whether ω is an integer (and also upon f), but we don't have time to discuss the details here. depends heavily

Remark: The algebra in this example could have been reduced by writing the solution for $x > z$ as

$$G_\omega(x,z) = E\sin\omega(x-\pi).$$

(That is, we build the boundary condition at π into the formula by a clever choice of basis solutions.) Then we would have to solve merely two equations in two unknowns (B and E) instead of a 3×3 system.

Example 2: *Variation of parameters in terms of delta and Green functions.* Let's go back to the general second-order linear ODE,

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

and construct the solution satisfying

$$y(0) = 0, \qquad y'(0) = 0.$$

As before, we will solve

$$\frac{\partial^2}{\partial x^2}G(x,z) + p(x)\,\frac{\partial}{\partial x}G(x,z) + q(x)G(x,z) = \delta(x-z)$$

with those initial conditions, and then expect to find y in the form

$$y(x) = \int G(x,z)\,f(z)\,dz.$$

It is not immediately obvious what the limits of integration should be, since there is no obvious "interval" in this problem.

Assume that two linearly independent solutions of the homogeneous equation

$$y'' + p(x)y' + q(x)y = 0$$

are known; call them $y_1(x)$ and $y_2(x)$. Of course, until we are told what p and q are, we can't write down exact formulas for y_1 and y_2; nevertheless, we can solve the problem in the general case — getting an expression for G in terms of y_1 and y_2, whatever they may be.

Since G satisfies the homogeneous equation for $x \neq z$, we have

$$G(x,z) = \begin{cases} A(z)y_1(x) + B(z)y_2(x) & \text{for } x < z, \\ C(z)y_1(x) + D(z)y_2(x) & \text{for } x > z. \end{cases}$$

As before we will get four equations in the four unknowns, two from initial data and two from the continuity of G and the prescribed jump in its derivative at z. Let us consider only the case $z > 0$. Then the initial conditions

$$G(0,z) = 0, \qquad \frac{\partial}{\partial x}G(0,z) = 0$$

force $A = 0 = B$. The continuity condition, therefore, says that $G(z,z) = 0$, or

$$C(z)y_1(z) + D(z)y_2(z) = 0. \tag{1}$$

The jump condition

$$\frac{\partial}{\partial x}G(z^+,z) - \frac{\partial}{\partial x}G(z^-,z) = 1$$

now becomes

$$C(z)y_1'(z) + D(z)y_2'(z) = 1. \tag{2}$$

Solve (1) and (2): The determinant is the Wronskian

$$\begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix} = y_1y_2' - y_2y_1' \equiv W(z).$$

Then

$$C = -\frac{y_2}{W}, \qquad D = \frac{y_1}{W}.$$

Thus our conclusion is that (for $z > 0$)

$$G(x,z) = \begin{cases} 0 & \text{for } x < z, \\ \dfrac{1}{W(z)}\big(y_1(z)y_2(x) - y_2(z)y_1(x)\big) & \text{for } x > z. \end{cases}$$

Now recall that the solution of the original ODE,

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

was supposed to be

$$y(x) = \int G(x,z)\, f(z)\, dz.$$

Assume that $f(z) \neq 0$ only for $z > 0$, where our result for G applies. Then the integrand is 0 for $z < 0$ (because $f = 0$ there) and also for $z > x$ (because $G = 0$ there). Thus

$$\begin{aligned} y(x) &= \int_0^x G(x,z)\, f(z)\, dz \\ &= \int_0^x \frac{y_1(z) f(z)}{W(z)}\, dz\, y_2(x) - \int_0^x \frac{y_2(z) f(z)}{W(z)}\, dz\, y_1(x). \end{aligned}$$

This is exactly the same solution that is found in differential equations textbooks by making the ansatz

$$y(x) = u_1(x) y_1(x) + u_2(x) y_2(x)$$

and deriving a system of first-order differential equations for u_1 and u_2. That method is called "variation of parameters". Writing the variation-of-parameters solution in terms of the Green function G shows in a precise and clear way how the solution y depends on the nonhomogeneous term f as f is varied. That formula is a useful starting point for many further investigations — for instance, a proof that a WKB approximation to y really is a good approximation to the exact y satisfying the same boundary conditions.

Delta functions and Fourier transforms

Formally, the Fourier transform of a delta function is a complex exponential function, since

$$\int_{-\infty}^{\infty} \delta(x-z)\, e^{-i\omega x}\, dx = e^{-i\omega z}.$$

According to the Fourier inversion formula, therefore, we should have

$$\begin{aligned} \delta(x-z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega z}\, e^{i\omega x}\, d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(x-z)}\, d\omega. \end{aligned}$$

This is a very useful formula! Here is another way of seeing what it means and why it is true:

Recall that

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega)\, e^{i\omega x}\, d\omega,$$

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(z)\, e^{-i\omega z}\, dz.$$

Let us substitute the second formula into the first:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\omega(x-z)}\, f(z)\, dz\, d\omega.$$

Of course, this equation is useless for computing $f(x)$, since it just goes in a circle; its significance lies elsewhere. If we're willing to play fast and loose with the order of the integrations, we can write it

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{i\omega(x-z)}\, d\omega \right] f(z)\, dz,$$

which says precisely that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(x-z)}\, d\omega$$

satisfies the defining property of $\delta(x-z)$. Our punishment for playing fast and loose is that this integral does not converge (in the usual sense), and there is no function δ with the desired property. Nevertheless, both the integral and the object δ itself can be given a rigorous meaning in the modern theory of *distributions*; crudely speaking, they both make perfect sense as long as you keep them inside other integrals (multiplied by continuous functions) and do not try to evaluate them at a point to get a number.

What would happen if we tried this same trick with the Fourier *series* formulas? Let's consider the sine series,

$$f(x) = \sum_{n=1}^{\infty} b_n \sin nx,$$

$$b_n = \frac{2}{\pi} \int_0^{\pi} f(z)\, \sin nz\, dz.$$

This gives

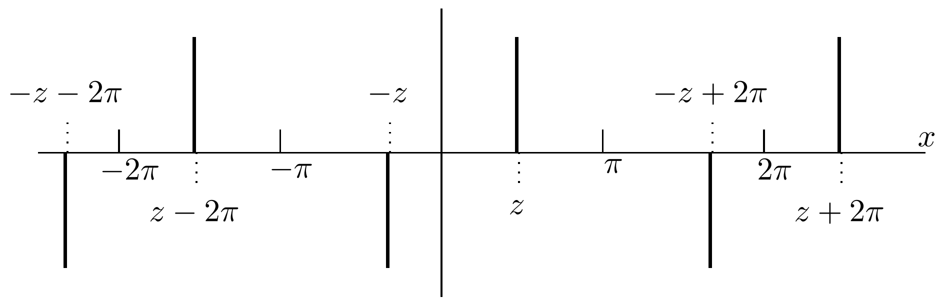
$$f(x) = \frac{2}{\pi} \int_0^{\pi} \left[\sum_{n=1}^{\infty} \sin nx\, \sin nz \right] f(z)\, dz. \qquad (\dagger)$$

Does this entitle us to say that

$$\delta(x-z) = \frac{2}{\pi}\sum_{n=1}^{\infty} \sin nx \, \sin nz \, ? \tag{‡}$$

Yes and no. In (†) the variables x and z are confined to the interval $[0,\pi]$. (‡) *is a valid representation of the delta function when applied to functions whose domain is* $[0,\pi]$. If we applied it to a function on a larger domain, it would act like the odd, periodic extension of $\delta(x-z)$, as is always the case with Fourier sine series:

$$\frac{2}{\pi}\sum_{n=1}^{\infty} \sin nx \, \sin nz = \sum_{M=-\infty}^{\infty} [\delta(x-z+2\pi M) - \delta(x+z+2\pi M)].$$



Orthogonal Functions and Sturm–Liouville Problems

So far all of our example PDEs have led to separated equations of the form $X'' + \omega^2 X = 0$, with standard Dirichlet or Neumann boundary conditions. Not surprisingly, more complicated equations often come up in practical problems. For example, if the medium in a heat or wave problem is inhomogeneous, the relevant equation may look like

$$X'' - V(x)X = -\omega^2 X$$

for some function V, or even

$$a(x)X'' + b(x)X' + c(x)X = -\omega^2 X.$$

Also, if the boundary in a problem is a circle, cylinder, or sphere, the solution of the problem is simplified by converting to polar, cylindrical, or spherical coordinates, so that the boundary is a surface of constant radial coordinate. This simplification of the boundary conditions is bought at the cost of complicating the differential equation itself: we again have to deal with ODEs with nonconstant coefficients, such as

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} - \frac{n^2}{r^2}R = -\omega^2 R.$$

The good news is that many of the properties of Fourier series carry over to these more general situations. As before, we can consider the *eigenvalue problem* defined by such an equation together with appropriate boundary conditions: Find all functions that satisfy the ODE (for *any* value of ω) and also satisfy the boundary conditions. And it is still true (under certain conditions) that the set of all eigenfunctions is *complete*: *Any* reasonably well-behaved function can be expanded as an infinite series where each term is proportional to one of the eigenfunctions. This is what allows arbitrary data functions in the original PDE to be matched to a sum of separated solutions! Also, the eigenfunctions are *orthogonal* to each other; this leads to a simple formula for the coefficients in the eigenfunction expansion, and also to a Parseval formula relating the norm of the function to the sum of the squares of the coefficients, and to a representation of the delta function as a series in the eigenfunctions.

Orthonormal bases

Consider an interval $[a,b]$ and the real-valued (or complex-valued) functions defined on it. A sequence of functions $\{\phi_n(x)\}$ is called *orthogonal* if

$$\int_a^b \phi_n(x)^* \, \phi_m(x)\, dx = 0 \quad \text{whenever } m \neq n.$$

It is called *orthonormal* if, *in addition*,

$$\int_a^b |\phi_n(x)|^2\,dx = 1.$$

This normalization condition is merely a convenience; the important thing is the orthogonality. (If we are lucky enough to have an orthogonal set, we can always convert it to an orthonormal set by dividing each function by the square root of its normalization integral:

$$\psi_n(x) \equiv \frac{\phi_n(x)}{\sqrt{\int_a^b |\phi_n(z)|^2\,dz}} \Rightarrow \int_a^b |\psi_n(x)|^2\,dx = 1.$$

However, in certain cases this may make the formula for ψ_n more complicated, so that the redefinition is hardly worth the effort. A prime example is the eigenfunctions in the Fourier sine series:

$$\phi_n(x) \equiv \sin nx \;\Rightarrow\; \int_0^\pi |\phi_n(x)|^2\,dx = \frac{\pi}{2}\,;$$

therefore,

$$\psi_n(x) \equiv \sqrt{\frac{2}{\pi}}\,\sin nx$$

are the elements of the orthonormal basis. The old dispute about whether to put the entire $2/\pi$ in one place or to put half of it in the Fourier series and half in the cofficient formula is just a disagreement over whether normalizing the eigenfunctions is more of a nuisance than a help.)

Now let $f(x)$ be an arbitrary (nice) function on $[a,b]$. *If* f has an expansion as a linear combination of the ϕ's,

$$f(x) = \sum_{n=1}^{\infty} c_n \phi_n(x),$$

then

$$\int_a^b \phi_m(x)^*\,f(x)\,dx = \sum_{n=1}^{\infty} c_n \int_a^b \phi_m(x)^*\,\phi_n(x)\,dx = c_m \int_a^b |\phi_m(x)|^2\,dx$$

by orthogonality. If the set is orthonormal, this just says

$$\boxed{c_m = \int_a^b \phi_m(x)^*\,f(x)\,dx.} \qquad (*)$$

(In the rest of this discussion, I shall assume that the orthogonal set is orthonormal. This greatly simplifies the formulas of the general theory, even while possibly complicating the expressions for the eigenfunctions in any particular case.)

It can be shown (see the Schaum's book, pp. 57–58) that

$$\boxed{\int_a^b |f(x)|^2\,dx = \sum_{n=1}^{\infty} |c_n|^2.}$$

This is the *Parseval equation* associated to this orthonormal set. Furthermore, if f is *not* of the form $\sum_{n=1}^{\infty} c_n\phi_n(x)$, then

$$\text{(1)} \qquad \sum_{n=1}^{\infty} |c_n|^2 < \int_a^b |f(x)|^2\,dx$$

(called *Bessel's inequality*), and (2) the *best approximation* to $f(x)$ of the form $\sum c_n\phi_n(x)$ is the one where the coefficients are computed by formula $(*)$. These last two statements remain true when $\{\phi_n\}$ is a *finite* set — in which case, obviously, the probability that a given f will not be exactly a linear combination of the ϕ's is greatly increased. (The precise meaning of (2) is that the choice $(*)$ of the c_n minimizes the integral

$$\int_a^b \left| f(x) - \sum_{n=1}^{\infty} c_n\phi_n(x)\right|^2 dx.$$

That is, we are talking about *least squares approximation*. It is understood in this discussion that f itself is square-integrable on $[a,b]$.)

Now suppose that *every* square-integrable f is the limit of a series $\sum_{n=1}^{\infty} c_n\phi_n$. (This series is supposed to converge "in the mean" — that is, the least-squares integral

$$\int_a^b \left| f(x) - \sum_{n=1}^{M} c_n\phi_n(x)\right|^2 dx$$

for a partial sum approaches 0 as $M\to\infty$.) Then $\{\phi_n\}$ is called a *complete set* or an *orthonormal basis*. This is the analogue of the mean convergence theorem for Fourier series. Under certain conditions there may also be pointwise or uniform convergence theorems, but these depend more on the special properties of the particular functions ϕ being considered.

So far this is just a definition, not a theorem. To guarantee that our orthonormal functions form a basis, we have to know where they came from. The miracle of the subject is that the eigenfunctions that arise from variable-separation problems *do* form orthonormal bases:

Theorem: Suppose that the ODE that arises from some separation of variables is

$$\mathcal{L}[X] = -\omega^2 r(x) X \quad \text{on } (0, L),$$

where $\mathcal{L}$ is an abbreviation for a second-order *linear differential operator*

$$\mathcal{L}[X] \equiv a(x)X'' + b(x)X' + c(x)X,$$

a, b, c, and r are continuous on $[0, L]$, and $a(x) > 0$ and $r(x) > 0$ on $[0, L]$. Suppose further that

$$\int_0^L \big(\mathcal{L}[u](x)\big)^* \, v(x)\, dx = \int_0^L u(x)^* \big(\mathcal{L}[v](x)\big)\, dx \qquad (\dagger)$$

for all functions u and v satisfying the boundary conditions of the problem. Then:

(1) All the eigenvalues ω^2 are real (but possibly negative).

(2) The eigenfunctions corresponding to different ω's are orthogonal with respect to the *weight function* $r(x)$:

$$\int_0^L \phi_n(x)^* \, \phi_m(x)\, r(x)\, dx = 0 \quad \text{if } n \neq m.$$

(Everything said previously about orthonormality can be generalized to the case of a nontrivial positive weight function.)

(3) The eigenfunctions are complete. (This implies that the corresponding PDE can be solved for arbitrary boundary data, in precise analogy to Fourier series problems!)

To show that a given $\mathcal{L}$ satisfies $(\dagger)$ (or doesn't satisfy it, as the case may be), just integrate by parts. It turns out that $(\dagger)$ will be satisfied if $\mathcal{L}$ has the form

$$\frac{d}{dx} p(x) \frac{d}{dx} + q(x)$$

(with p and q real-valued and well-behaved) and the boundary conditions are of the type

$$\alpha X'(0) + \beta X(0) = 0, \qquad \gamma X'(L) + \delta X(L) = 0$$

with α, etc., real. Such an eigenvalue problem is called a *regular Sturm–Liouville problem.*

The proof of the conclusions (1) and (2) of the theorem is quite simple and is a generalization of the proof of the corresponding theorem for eigenvalues and eigenvectors of a *symmetric matrix.* The latter theorem is part of many physics courses, not to mention linear-algebra courses. See Schaum, pp. 59–60. Part (3) is harder to prove, like the convergence theorems for Fourier series (which are a special case of it).

The simplest nontrivial example of a Sturm–Liouville problem ("nontrivial" in the sense that it gives something other than a Fourier series) is the usual spatially homogeneous heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2},$$

$$0 < x < L, \qquad 0 < t < \infty,$$

with boundary conditions such as

$$u(t,0) = 0, \qquad \frac{\partial u}{\partial x}(t,L) + \beta u(t,L) = 0$$

and initial data

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

In a realistic problem, the zeros in the BC would be replaced by constants; as usual, we would take care of that complication by subtracting off a steady-state solution. Physically, the constant value of $\frac{\partial u}{\partial x}(t,L) + \beta u(t,L)$ is proportional to the temperature of the air (or other fluid medium) to which the right-hand endpoint of the bar is exposed; heat is lost through that end by convection, according to "Newton's law of cooling".

The separation of variables proceeds just as in the more standard heat problems, up to the point

$$T(t) = e^{-\omega^2 t}, \qquad X(x) = \sin \omega x.$$

To get the sine I used the boundary condition $X(0) = 0$. The other BC is

$$X'(L) + \beta X(L) = 0,$$

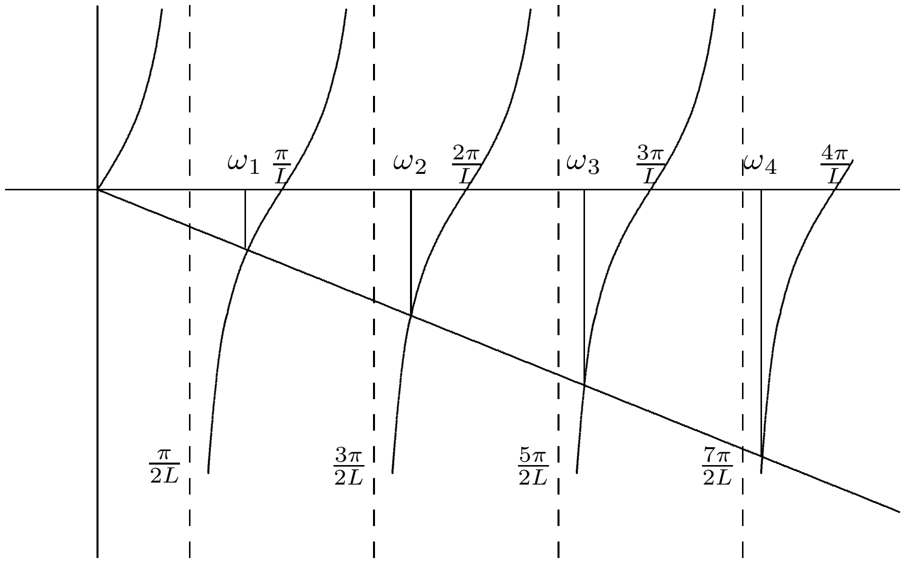
or

$$\frac{\omega}{\beta} \cos \omega L + \sin \omega L = 0, \tag{‡}$$

or

$$\tan \omega L = -\frac{1}{\beta}\, \omega. \tag{$*$}$$

It is easy to find the *approximate* locations of the eigenvalues, ω^2, by graphing the two sides of ($*$) (as functions of ω) and picking out the points of intersection.



The nth root, ω_n, is somewhere between $\left(n-\frac{1}{2}\right)\frac{\pi}{L}$ and $\frac{n\pi}{L}$; as $n \to \infty$, ω_n becomes arbitrarily close to $\left(n-\frac{1}{2}\right)\frac{\pi}{L}$, the vertical asymptote of the tangent function. For smaller n one could guess ω_n by eye and then improve the guess by, for example, Newton's method. (Because of the violent behavior of tan near the asymptotes, Newton's method does not work well when applied to ($*$); it is more fruitful to work with (‡) instead.)

To complete the solution, we write a linear combination of the separated solutions,

$$u(t,x) = \sum_{n=1}^{\infty} b_n \sin \omega_n x \, e^{-\omega_n{}^2 t},$$

and seek to determine the coefficients from the initial condition,

$$f(x) = u(0,x) = \sum_{n=1}^{\infty} b_n \sin \omega_n x.$$

This problem satisfies the conditions of the Sturm–Liouville theorem, so the eigenfunctions

$$\phi_n \equiv \sin \omega_n x$$

are guaranteed to be orthogonal. This can be verified by direct computation (making use of the fact that ω_n satisfies (‡)). Thus

$$\int_0^L f(x) \sin \omega_m x \, dx = b_m \int_0^L \sin^2 \omega_m x \, dx.$$

However, the ϕ_n have not been normalized, so we have to calculate

$$\int_0^L \sin^2 \omega_m x \, dx \equiv \|\phi_m\|^2$$

and divide by it. (This number is *not* just $\frac{1}{2}L$. See Schaum, p. 61.) Alternatively, we could construct *orthonormal* basis functions by dividing by the square root of this quantity:

$$\psi_n \equiv \frac{\phi_n}{\|\phi_n\|}\,.$$

Then the coefficient formula is simply

$$B_m = \int_0^L f(x)\,\psi_m(x)\,dx$$

(where $f(x) = \sum_m B_m \psi_m$, so $B_m = \|\phi_m\| b_m$).

The theorem also guarantees that the eigenfunctions are complete, so this solution is valid for any reasonable f.

Eigenfunctions, delta functions, and Green functions

Let's return to the general case and assume that the eigenfunctions have been chosen orthonormal. We have an expansion formula

$$f(x) = \sum_{n=1}^{\infty} c_n \phi_n(x) \tag{$\sharp$}$$

and a coefficient formula

$$c_m = \int_a^b \phi_m(z)^* \, f(z)\, dz. \tag{$\flat$}$$

Substituting ($\flat$) into ($\sharp$) and interchanging the order of summation and integration yields

$$f(x) = \int_a^b dz\, f(z) \left[\sum_{n=1}^{\infty} \phi_n(x)\phi_n(z)^*\right].$$

In other words, when acting on functions with domain (a, b),

$$\boxed{\delta(x-z) = \sum_{n=1}^{\infty} \phi_n(x)\phi_n(z)^*.}$$

This is called the *completeness relation* for the eigenfunctions $\{\phi_n\}$, since it expresses the fact that the whole function f can be built up from the pieces $c_n\phi_n$. In the special case of the Fourier sine series, we looked at this formula earlier.

We can also substitute (♯) into (♭), getting

$$c_m = \sum_{n=1}^{\infty} c_n \left[\int_a^b \phi_m(x)^* \, \phi_n(x) \, dx \right].$$

This equation is equivalent to

$$\boxed{\int_a^b \phi_m(x)^* \, \phi_n(x) \, dx = \delta_{mn} \,,}$$

where

$$\delta_{mn} \equiv \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n. \end{cases}$$

(This is called the *Kronecker delta symbol*; it is the discrete analogue of the Dirac delta function — or, rather, Dirac's delta function is a continuum generalization of it!) This *orthogonality relation* summarizes the fact that the ϕ's form an orthonormal basis.

Note that the completeness and orthogonality relations are very similar in structure. Basically, they differ only in that the variables x and n interchange roles (along with their alter egos, z and m). The different natures of these variables causes a sum to appear in one case, an integral in the other.

Finally, consider the result of substituting (♭) into the solution of an initial-value problem involving the functions ϕ_n. For example, for a certain heat-equation problem we would get

$$u(t,x) = \sum_{n=1}^{\infty} c_n \, \phi_n(x) \, e^{-\omega_n{}^2 t}.$$

This becomes

$$u(t,x) = \int_a^b dz \, f(z) \left[\sum_{n=1}^{\infty} \phi_n(x) \, \phi_n(z)^* \, e^{-\omega_n{}^2 t} \right].$$

Therefore, the Green function for that problem is

$$\boxed{G(x,z;t) = \sum_{n=1}^{\infty} \phi_n(x) \, \phi_n(z)^* \, e^{-\omega_n{}^2 t}.}$$

When $t = 0$ this reduces to the completeness relation, since

$$\lim_{t \downarrow 0} G(x,z;t) = \delta(x - z).$$

It may be easier to solve for the Green function directly than to sum the series in this formula. In fact, such formulas are often used in the reverse direction, to obtain information about the eigenfunctions and eigenvalues from independently obtained information about the Green function.

Bessel Functions

Reference: J. D. Jackson, *Mathematics for Quantum Mechanics* (Benjamin, 1962), Secs. 2.2 and 3.4 and Appendix A.)

Now we will look at a Sturm–Liouville equation with *nonconstant coefficients*, which arises naturally from the geometry of a physical problem. Consider the heat equation in a disc,

$$0 \le x^2 + y^2 < 1, \qquad 0 < t < \infty,$$

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\,,$$

$$u(t,x,y) = 0 \quad \text{for } (x,y) \text{ on the boundary } (x^2+y^2=1),$$

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

The circular boundary condition is difficult to handle in Cartesian coordinates, so we transform to polar coordinates:

$$0 \le r < 1, \qquad 0 \le \theta \le 2\pi,$$

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}\,,$$

$$u(t,1,\theta) = 0,$$

$$u(t,r,0) = u(t,r,2\pi), \qquad \frac{\partial u}{\partial \theta}(t,r,0) = \frac{\partial u}{\partial \theta}(t,r,2\pi),$$

$$u(0,r,\theta) = g(r,\theta).$$

(The next-to-last condition expresses the periodicity in θ — the facts that 0 and 2π label the same ray, and u must be just as smooth there as it is everywhere else. The last condition simply means that the initial temperature distribution is some arbitrary function g. In terms of the Cartesian formulation,

$$g(r,\theta) = f(r\cos\theta, r\sin\theta).$$

Conceptually, g and f are the same thing, and they may be denoted informally by the same letter when there seems to be little chance of confusion.)

Let's separate variables:

$$u = T(t)R(r)\Theta(\theta).$$

$$T'R\Theta = TR''\Theta + \frac{1}{r}TR'\Theta + \frac{1}{r^2}TR\Theta''.$$

$$-\omega^2 = \frac{T'}{T} = \frac{R''}{R} + \frac{1}{r}\,\frac{R'}{R} + \frac{1}{r^2}\,\frac{\Theta''}{\Theta}\,.$$

Thus $T(t) = e^{-\omega^2 t}$ as usual. To solve the (r,θ) equation we need to separate variables again, and to do so we must multiply the equation by r^2 first:

$$r^2\,\frac{R''}{R} + r\,\frac{R'}{R} + \frac{\Theta''}{\Theta} + r^2\omega^2 = 0.$$

Set $\frac{\Theta''}{\Theta} = -\nu^2$. Then the periodic boundary condition will be satisfied if ν is an integer, and we have a familiar set of eigenfunctions:

$$e^{i\nu\theta} \qquad (-\infty < \nu < \infty).$$

(Of course, one could use sines and cosines instead of complex exponentials.)

To save writing, I'll define

$$n \equiv |\nu|.$$

Thus n is a *nonnegative* integer, and $\nu^2 = n^2$.

Now the r equation becomes

$$R'' + \frac{1}{r}\,R' + \left(\omega^2 - \frac{n^2}{r^2}\right) R = 0.$$

This equation depends on two parameters, ω and n. However, one of them can be scaled away: Let

$$z \equiv \omega r, \qquad R(r) \equiv Z(z) = Z(\omega r).$$

Then all the terms which don't already have a factor ω^2 will acquire one, and it can be cancelled out:

$$\frac{d^2Z}{dz^2} + \frac{1}{z}\,\frac{dZ}{dz} + \left(1 - \frac{n^2}{z^2}\right) Z = 0.$$

This is the famous *Bessel equation of order n*. (Here we have assumed that ω^2 is positive. When ω is 0, R is a power of r. When ω is imaginary, the equation is a *modified Bessel equation*, whose solutions are related to the Bessel functions described here in the same way that hyperbolic functions are related to trigonometric functions.)

The solutions, called *Bessel functions of order n*, have been thoroughly studied, and their principal properties are well summarized in Jackson's Appendix A. (See also Chap. 6 of the Schaum's Outline book.) For fixed n, there are, of course, two linearly independent solutions. By the method of Frobenius (power series solution around the singular point, $z = 0$), one finds that one of these is asymptotically proportional to z^n as $z \to 0$,

$$J_n(z) \sim \frac{1}{n!}\left(\frac{z}{2}\right)^n,$$

and all the others (that is, the solutions linearly independent of J_n) are asymptotically proportional to z^{-n} if $n \neq 0$, to $\ln z$ if $n = 0$. There is a standard choice for the second solution, denoted $Y_n(z)$ in some books and $N_n(z)$ in others. Since we want u to be a smooth solution of the heat equation at the origin ($r = 0 \iff z = 0$), we reject those solutions in the present problem and keep only the J's; this amounts to a new boundary condition. Thus $R(r) = J_n(\omega r)$.

If we were solving a PDE in a ring (annulus), defined by $a < r < b$ with $a > 0$, then the correct solutions of the radial equation would be those that satisfy prescribed homogeneous boundary conditions at a and b (such as $u(t, a, \theta) = 0$), and they would be certain linear combinations of the J's and Y's. So the Y's are not totally useless. If we were solving our PDE in a pie-shaped region (sector) ($0 < r < 1$, $0 < \theta < \frac{\pi}{2}$, say), it is still correct to use only the J's, although the argument about "a smooth solution at the origin" no longer applies. A full justification in that case requires some more advanced concepts; see I. Stakgold, *Green's Functions and Boundary Value Problems*, pp. 448–450. (Note also that in sector problems the allowed values of n will be different.)

At infinity, it can be shown (by a method closely related to the WKB theory studied earlier in this course) that

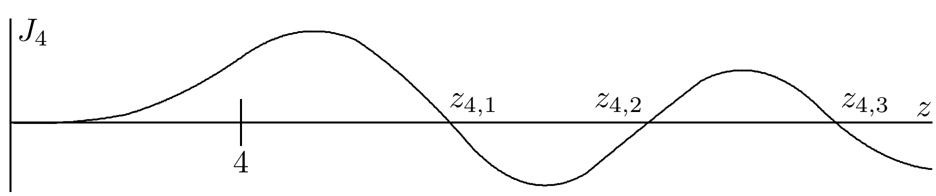
$$J_n(z) \sim \sqrt{\frac{2}{\pi z}} \cos\left(z - \frac{n\pi}{2} - \frac{\pi}{4}\right).$$

The second solution, Y_n, is actually *defined* so that it has matching behavior:

$$Y_n(z) \sim \sqrt{\frac{2}{\pi z}} \sin\left(z - \frac{n\pi}{2} - \frac{\pi}{4}\right).$$

That is, the Bessel functions eventually act much like ordinary trig functions, except for a gentle decay (note the modulating factor $z^{-\frac{1}{2}}$).

In the region around $z = n$ there is a gradual transition from the initial power-law behavior to the final oscillatory behavior. Thus a typical $J_n(z)$ has a graph like this:



In particular, $J_n(z)$ has infinitely many positive real *roots*, z_{ni} (n fixed, $i = 1, 2, \ldots$). We can now handle the remaining homogeneous boundary condition in our original problem, $u(t, 1, \theta) = 0$. This translates to $R(1) = 0$, or $J_n(\omega r)\big|_{r=1} = 0$.

This shows that the allowed values of ω are the roots z_{ni}. (If the radius of the disc, r_0, had been something other than 1, then ω_{ni} would have been z_{ni}/r_0.)

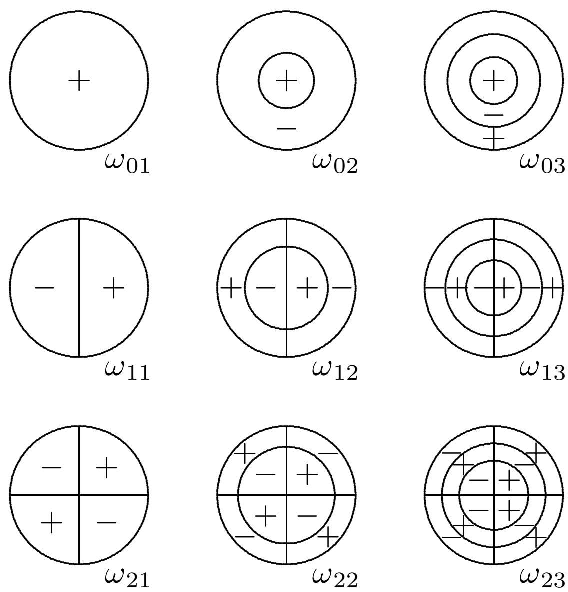
Putting everything together, we have the eigenvalues

$$\omega_{ni}{}^2 = z_{ni}{}^2 \tag{1}$$

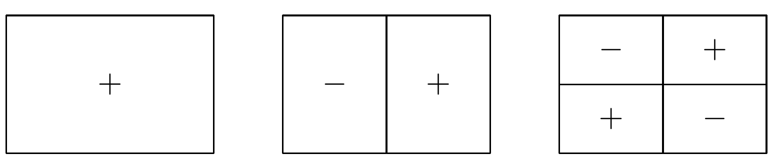
and corresponding eigenfunctions

$$\psi_{\nu i}(r,\theta) = R(r)\Theta(\theta) = J_n(\omega_{ni} r)\, e^{i\nu\theta} \qquad (n = |\nu|). \tag{2}$$

We could equally well use the *real* eigenfunctions in which $e^{i\nu\theta}$ is replaced by $\sin n\theta$ or $\cos n\theta$; those functions are easier to visualize. In the drawing (cf. D. L. Powers, *Boundary Value Problems*, Fig. 5.10) the lines and curves indicate places where such a ψ equals 0, and the signs indicate how the solution $\operatorname{Re}\psi$ or $\operatorname{Im}\psi$ bulges above or below the plane $\psi = 0$. Such patterns may be seen in the surface of a cupful of coffee or other liquid when the container is tapped lightly.



For comparison, recall that we solved the heat equation in a rectangle some time ago and got products of sine functions. The corresponding patterns are rectangular grids:



$\sin\frac{x\pi}{a}\sin\frac{y\pi}{b}$ $\qquad$ $\sin\frac{2x\pi}{a}\sin\frac{y\pi}{b}$ $\qquad$ $\sin\frac{2x\pi}{a}\sin\frac{2y\pi}{b}$

The general solution of the heat problem should be a linear combination of the separated solutions:

$$u(t,r,\theta) = \sum_{\nu=-\infty}^{\infty} \sum_{i=1}^{\infty} c_{\nu i}\, \psi_{\nu i}(r,\theta)\, e^{-\omega_{ni}{}^2 t}. \tag{3}$$

The coefficients need to be calculated from the initial data:

$$\begin{aligned} g(r,\theta) &= u(0,r,\theta) \\ &= \sum_{\nu=-\infty}^{\infty} \sum_{i=1}^{\infty} c_{\nu i}\, \psi_{\nu i}(r,\theta) \\ &= \sum_{\nu=-\infty}^{\infty} \sum_{i=1}^{\infty} c_{\nu i}\, J_n(\omega_{ni} r)\, e^{i\nu\theta}. \end{aligned}$$

By the standard Fourier series formula,

$$\frac{1}{2\pi} \int_0^{2\pi} e^{-i\nu\theta}\, g(r,\theta)\, d\theta = \sum_{i=1}^{\infty} c_{\nu i}\, J_n(\omega_{ni} r).$$

We are left with a one-dimensional series in the eigenfunctions $R_n(r) \equiv J_n(\omega_{ni} r)$ (n fixed). We recall that these functions came out of the equation

$$R'' + \frac{1}{r}\, R' + \left(\omega^2 - \frac{n^2}{r^2}\right) R = 0$$

with the boundary condition $R(1) = 0$, which looks like a Sturm–Liouville problem. Unfortunately, it does not quite satisfy the technical conditions of the Sturm–Liouville theorem, because of the singular point in the ODE at $r = 0$. Nevertheless, it turns out that the conclusions of the theorem are still valid in this case: The eigenfunctions are complete (for each fixed n), and they are orthogonal with respect to the weight function r:

$$\int_0^1 J_n(z_{ni} r)\, J_n(z_{nj} r)\, r\, dr = 0 \quad \text{if } i \neq j.$$

Thus if $h(r)$ is an arbitrary function on $[0,1]$, it can be expanded as

$$h(r) = \sum_{i=1}^{\infty} c_i\, J_n(z_{ni} r),$$

and the coefficients are

$$c_i = \frac{\int_0^1 J_n(z_{ni} r)\, h(r)\, r\, dr}{\int_0^1 J_n(z_{ni} r)^2\, r\, dr}\,.$$

Furthermore, the experts on Bessel functions assure us that the integral in the denominator can be evaluated:

$$\int_0^1 J_n(z_{ni} r)^2\, r\, dr = \frac{1}{2}\, J_{n+1}(z_{ni})^2.$$

Applying this theorem to our problem, we get

$$c_{\nu i} = \left[\frac{1}{2}\, J_{n+1}(\omega_{ni})^2\right]^{-1} \int_0^1 r\,dr\, J_n(\omega_{ni} r)\, \frac{1}{2\pi} \int_0^{2\pi} d\theta\, e^{-i\nu\theta}\, g(r,\theta). \tag{4}$$

That is,

$$\begin{aligned} c_{\nu i} &= \left[\pi J_{n+1}(\omega_{ni})^2\right]^{-1} \int_{r=0}^1 \int_{\theta=0}^{2\pi} r\,dr\,d\theta\, \psi_{\nu i}(r,\theta)^*\, g(r,\theta) \\ &= \frac{1}{\|\psi_{\nu i}\|^2} \int_{r=0}^1 \int_{\theta=0}^{2\pi} r\,dr\,d\theta\, \psi_{\nu i}(r,\theta)^*\, g(r,\theta). \end{aligned}$$

(In the last version I have identified the constant factor as the normalization constant for the *two-dimensional* eigenfunction.) We now see that the mysterious weight factor r has a natural geometrical interpretation: It makes the r and θ integrations go together to make up the standard integration over the disc in polar coordinates!

The formulas (1)–(4) give a complete solution to the problem with which we started.

The Wave Equation

The d'Alembert solution

The linear *wave equation* in one space and one time variable is

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}\,.$$

It can be solved by separation of variables just like the heat and Laplace equations. Since the equation is second-order in the time, it comes as no surprise that to determine a solution uniquely, one needs to prescribe initial values for both u and its time derivative:

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

The coefficients in the Fourier series (or transform) solution can be calculated routinely from the Fourier series (or transforms) of the data functions, f and g.

There is, however, a significant practical disadvantage to this procedure for the wave equation, which did not arise in the other cases. For the heat equation, for example, the separated solutions always contain factors of the type $e^{-{\omega_n}^2 t}$; therefore, as $n \to \infty$, the nth term of the Fourier series approaches 0 very rapidly, and the series converges very well. (Thus adding the first few terms on a computer may give a good approximation to the complete solution.) But for the wave equation, the time-dependent factors are $\sin \omega_n t$ and $\cos \omega_n t$; unless the initial data were very smooth, the series is likely to converge very slowly, and it may be cumbersome to work with in practice. It will be difficult to find a Green function from the Fourier solution (since a delta function — the initial data for a Green function — is certainly not "very smooth").

Fortunately, there is another way of obtaining the complete solution of the two-dimensional wave equation. It demonstrates that the solution depends on the initial data in a suprisingly simple and geometrically appealing way. So it tells us a lot more about the wave equation than the Fourier method does! Unfortunately, things are not so simple for higher-dimensional wave equations

$$\frac{\partial^2 u}{\partial t^2} = \nabla^2 u$$

or such a simple modification of the wave equation as

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - m^2 u$$

(called the *Klein–Gordon equation*). (See your homework, however, for a partial extension of the d'Alembert method to three space dimensions.)

Remark: Usually the wave equation appears with a constant coefficient:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \,.$$

Then c has the physical signficance of the *velocity* of travel of the waves (light, sound, or whatever) described by the equation. To speed up the blackboard presentation, I shall assume that $c = 1$. Clearly the units of length or time can always be chosen so that this is so.

To discover the d'Alembert solution, consider the change of variables

$$w \equiv x + t, \qquad z \equiv x - t.$$

(In the general case these would be $w = x + ct$, etc.) Then

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial w} - \frac{\partial}{\partial z} \,, \qquad \frac{\partial}{\partial x} = \frac{\partial}{\partial w} + \frac{\partial}{\partial z} \,.$$

$$\frac{\partial^2}{\partial t^2} = \frac{\partial^2}{\partial w^2} - 2\,\frac{\partial^2}{\partial w\,\partial z} + \frac{\partial^2}{\partial z^2} \,.$$

The result for $\dfrac{\partial^2}{\partial x^2}$ is the same with a $+2$. Thus the wave equation is

$$0 = \frac{1}{4}\left(\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2}\right) = \frac{\partial^2 u}{\partial w\,\partial z} \,.$$

Let's write this in the form

$$\frac{\partial}{\partial w}\left(\frac{\partial u}{\partial z}\right) = 0.$$

Then it just says that $\dfrac{\partial u}{\partial z}$ is a constant, as far as w is concerned. That is,

$$\frac{\partial u}{\partial z} = \gamma(z) \quad \text{(a function of } z \text{ only).}$$

Consequently,

$$u(w,z) = \int_{z_0}^{z} \gamma(\tilde{z})\, d\tilde{z} + C(w),$$

where z_0 is some arbitrary starting point for the indefinite integral. Note that the constant of integration will in general depend on w. Now since γ was arbitrary, its indefinite integral is an essentially arbitrary function too, and we can forget γ and just call the first term $B(z)$:

$$u(w,z) = B(z) + C(w).$$

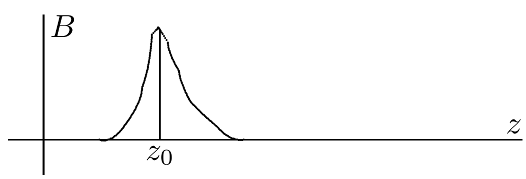
(The form of the result is symmetrical in z and w, as it must be, since we could equally well have worked with the equation in the form $\frac{\partial}{\partial z}\left(\frac{\partial u}{\partial w}\right) = 0$.)

So, we have found the *general solution* of the wave equation to be

$$\boxed{u(t,x) = B(x-t) + C(x+t),}$$

where B and C are arbitrary functions. (Technically speaking, we should require that the second derivatives B'' and C'' exist and are continuous, to make all our calculus to this point legal. However, from the modern point of view of the theory of *distributions* (recall the discussion of delta functions), it turns out that the d'Alembert formula remains meaningful and correct for choices of B and C that are much rougher than that.)

Interpretation



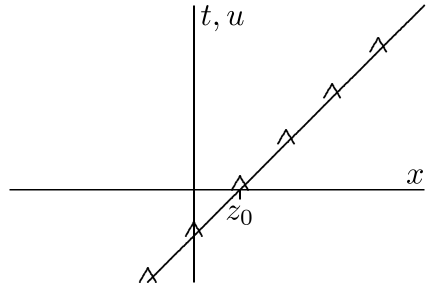
What sort of function is $B(x-t)$? It is easiest to visualize if $B(z)$ has a peak around some point $z = z_0$. Contemplate $B(x-t)$ as a function of x for a fixed t: It will have a peak in the neighborhood of a point x_0 satisfying $x_0 - t = z_0$, or

$$x_0 = z_0 + t.$$

That is, the "bump" moves to the right with velocity 1, keeping its shape exactly. (If we had kept the factor c^2 in the equation, the velocity would be c.)

(Note that in the drawing we have to plot u on the same axis as t. Such pictures should be thought of as something like a strip of movie film which we are forced to look at without the help of a projector.)*

* In advanced physics, especially relativistic physics, it is standard to plot t on the *vertical* axis and x on the horizontal, even though for particle motion t is the independent variable and x the dependent one.



Similarly, the term $C(x+t)$ represents a wave pattern which moves rigidly to the *left* at the wave velocity -1 (or $-c$). If both terms are present, and the functions are sharply peaked, we will see the two bumps collide and pass through each other. If the functions are not sharply peaked, the decomposition into left-moving and right-moving parts will not be so obvious to the eye.

Remark: $B(x-t)$ is the general solution of the equation

$$\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial t}\right)u=0.$$

Similarly, $C(x+t)$ solves

$$\left(\frac{\partial}{\partial x}-\frac{\partial}{\partial t}\right)u=0.$$

Thus the solutions of one of these first-order equations gives "half" the solutions of the wave equation, those solutions that move in one direction only. For this reason, the first-order nonlinear equation

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

can be used to model more difficult second-order nonlinear wave equations.

Initial conditions

Let's consider the wave equation in the domain

$$-\infty<x<\infty, \qquad -\infty<t<\infty,$$

with initial data at $t=0$:

$$u(0,x)=f(x), \quad \frac{\partial u}{\partial t}(0,x)=g(x).$$

The d'Alembert solution implies

$$f(x)=B(x)+C(x), \qquad g(x)=-B'(x)+C'(x).$$

The second condition implies

$$-B(x) + C(x) = \int \frac{g(x)}{c}\,dx = G(x) + A,$$

where G *is any antiderivative of* g, and A is an unknown constant of integration. Solve the two equations for B and C:

$$B(x) = \tfrac{1}{2}[f(x) - G(x) - A], \qquad C(x) = \tfrac{1}{2}[f(x) + G(x) + A].$$

We note that A cancels out of the total solution, $B(x-t) + C(x+t)$. (Being constant, it qualifies as both left-moving and right-moving; so to this extent, the decomposition of the solution into left and right parts is ambiguous.) So we can set $A = 0$ without losing any solutions. Now our expression for the solution in terms of the initial data is

$$u(t,x) = \tfrac{1}{2}[f(x+t) + f(x-t)] + \tfrac{1}{2}[G(x+t) - G(x-t)].$$

Note that this is much simpler than what we would get by separation of variables, which would involve two Fourier integrals in succession. (The Fourier transform of the solution is expressed in terms of the Fourier transform of the data; then you need an inverse Fourier transform to get the solution itself.)

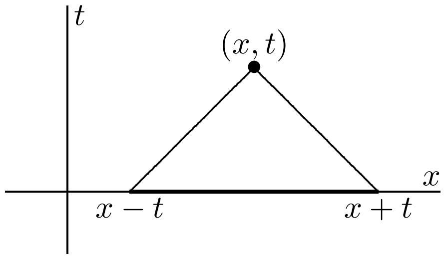
By the fundamental theorem of calculus, the G term can be rewritten as an integral over g:

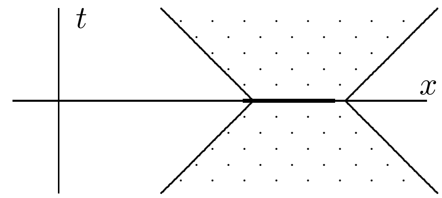
$$u(t,x) = \tfrac{1}{2}[f(x+t) + f(x-t)] + \tfrac{1}{2}\int_{x-t}^{x+t} g(w)\,dw. \qquad (\dagger)$$

This formula demonstrates that the value of u at a point (x,t) depends only on the initial data that has had time to reach x while traveling at speed 1 (or c, in general) — that is, the data $f(w,0)$ and $g(w,0)$ on the *interval of dependence*

$$x - ct < w < x + ct \qquad (\text{for } t > 0).$$

Conversely, any interval on the initial data "surface" (the line $t = 0$, in the two-dimensional case) has an expanding *region of influence* in space-time, beyond which its initial data are irrelevant. In other words, "signals" or "information" are carried by the waves with a finite maximum speed. These properties continue to hold for higher-dimensional wave equations and Klein–Gordon equations, even though in those cases the simple d'Alembert formula for the solution is lost and the waves no longer keep exactly the same shape as they travel.





From (†) we can extract a Green function for the wave equation. Since we already have a G, I will name it $W(x-y;t)$. Let

$$W(z;t) \equiv \begin{cases} \frac{1}{2} & \text{if } t > |z|, \\ -\frac{1}{2} & \text{if } t < -|z|, \\ 0 & \text{if } |t| < |z|. \end{cases}$$

Thus $W(x-y;t)$, as a function of (t,x), is nonzero only in the region of influence of $(0,y)$. Since the derivative of a step function is a delta function, we can write

$$\frac{\partial W}{\partial t}(z;t) = \tfrac{1}{2}[\delta(t-z) + \delta(t+z)].$$

Then (†) is equivalent to

$$u(t,x) = \int_{-\infty}^{\infty} dy \left[W(x-y;t) g(y) + \frac{\partial W}{\partial t}(x-y;t) f(y) \right].$$

Incidentally, if you try to derive this formula from the separation-of-variables (Fourier-transform) solution, you will get an integral for W whose convergence is rather dubious, and some distribution theory will be necessary to interpret it correctly.

Boundary conditions

When a wave hits a boundary, it *reflects*, or "bounces off". Let's see this mathematically. Consider the interval $0 < x < \infty$ and the Dirichlet condition

$$u(t,0) = 0.$$

Of course, we will have initial data, f and g, defined for $x \in (0,\infty)$.

We know that

$$u(t,x) = B(x-t) + C(x+t) \tag{1}$$

and

$$B(w) = \tfrac{1}{2}[f(w) - G(w)], \qquad C(w) = \tfrac{1}{2}[f(w) + G(w)], \tag{2}$$

where f and G' are the initial data. However, if we try to calculate u from (1) for $t > x$, we find that (1) directs us to evaluate $B(w)$ for *negative* w; this is not defined in our present problem! To see what is happening, start at (t, x) and trace a right-moving ray backwards in time: It will run into the wall (the positive t-axis), not the initial-data surface (the positive x-axis).

Salvation is at hand through the boundary condition, which gives us the additional information

$$B(-t) = -C(t). \tag{3}$$

For $t > 0$ this condition determines B(negative argument) in terms of C(positive argument). For $t < 0$ it determines C(negative argument) in terms of B(positive argument). Thus B and C are uniquely determined for all arguments by (2) and (3) together.

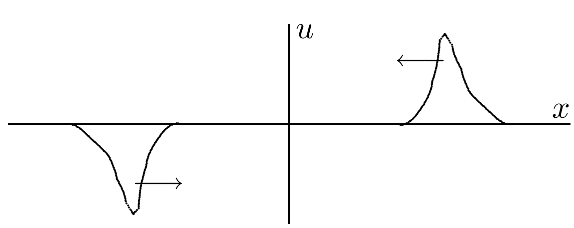
Now that we have B and C, it is convenient to require that (2) hold for negative values of w as well as positive; this is a *definition* of $f(w)$ and $G(w)$ for negative w. A bit of algebra shows that the result is the *odd extension* of f and the *even extension* of G:

$$f(-x) = -f(x), \qquad G(-x) = +G(x).$$

Since the derivative of an even function is odd, the second equation here is equivalent to dealing with the odd extension of g. We can now solve the wave equation in all of $\mathbf{R}^2$ ($-\infty < x < \infty$, $-\infty < t < \infty$) with these odd functions f and g as initial data. The solution is given by d'Alembert's formula,

$$u(t, x) = \tfrac{1}{2}[f(x + t) + f(x - t)] + \tfrac{1}{2}\int_{x-t}^{x+t} g(w)\, dw,$$

and it is easy to see that the boundary condition, $u(t, 0) = 0$, is satisfied, because the data functions are odd. Only the part of the solution in the region $x > 0$ is physical; the other region is fictitious. In the latter region we have a "ghost" wave which is an *inverted mirror image* of the physical solution.



This odd extension of the solution and the data should come as no surprise to us, since we already know that the solution of such a Dirichlet problem by *Fourier sine transforms* leads to odd extensions.

It should also be no surprise that the Neumann boundary condition,

$$\frac{\partial u}{\partial x}(0,t) = 0,$$

is solved by considering the *even* extensions of the data functions. The result is that the pulse reflects *without* turning upside down. Approximations to the "ideal" Dirichlet and Neumann boundary conditions are provided by a standard high-school physics experiment with Slinky$^{\text{TM}}$ springs. A small, light spring and a large, heavy one are attached end to end. When a wave traveling along the light spring hits the junction, the heavy spring remains almost motionless and the pulse reflects inverted. When the wave is in the heavy spring, the light spring serves merely to stabilize the apparatus; it carries off very little energy and barely constrains the motion of the end of the heavy spring. The pulse, therefore, reflects without inverting.

If we have two boundary conditions, say

$$u(t,0) = 0, \qquad u(t,L) = 0,$$

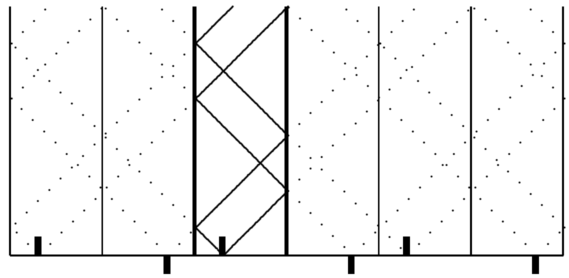
then we know that the Fourier series solution will correspond to *odd periodic* extensions of the data (and the solution), with period $2L$. Given

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

for $0 < x < L$, define extended data functions f and g for $-L < x < 0$ by the requirement of oddness, and then for all other x by the requirement

$$f(x+2L) = f(x), \qquad g(x+2L) = g(x) \quad \text{(for all } x\text{)}.$$

Then $G(x) \equiv \int g(x)\,dx$ will be an even, periodic function. The d'Alembert formula then gives a solution to the wave equation that satisfies the desired boundary and initial conditions. If the original initial data describe a single "bump", then the extended initial data describe an infinite sequence of image bumps, of alternating sign, as if space were filled with infinitely many parallel mirrors reflecting each other's images. Part of each bump travels off in each direction at speed 1. What this really means is that the two wave pulses from the original, physical bump will suffer many reflections from the two boundaries. When a "ghost" bump penetrates into the physical region, it represents the result of one of these reflection events.



Nonlinear waves; shock waves

It would be nice to include some study of nonlinear wave equations and the associated new phenomena, such as shock waves. Unfortunately, that does not fit into the semester.

Classification of Second-Order Linear Equations

We have looked at three fundamental partial differential equations:

$$\text{Laplace:}\quad \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

$$\text{wave:}\quad \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = 0$$

$$\text{heat:}\quad \frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} = 0.$$

Each of these turned out to have its own characteristic properties, which we want to review here and put in a more general context. Of particular interest for each equation are

(1) what sort of data (initial or boundary conditions) is needed to constitute a *well-posed problem* — one with exactly one solution;

(2) smoothness of the solutions;

(3) how the influence of the data spreads (*causality* or *finite propagation speed*).

The most general *second-order linear* differential operator in two variables, say x and y, looks like

$$L[u] = A(x,y)\,\frac{\partial^2 u}{\partial x^2} + B(x,y)\,\frac{\partial^2 u}{\partial x\,\partial y} + C(x,y)\frac{\partial^2 u}{\partial y^2} + D(x,y)\,\frac{\partial u}{\partial x} + E(x,y)\,\frac{\partial u}{\partial y} + F(x,y)u,$$

where $A, \ldots, F$ are functions of x and y. Suppose just for a moment that these coefficients are *constants*. Then the long expression is reminiscent of the formula for the most general *conic section*. Indeed, if we replace each $\partial/\partial x$ by a new variable, X, and replace each $\partial/\partial y$ by Y, and replace L by 0, then we get exactly the conic section equation:

$$0 = AX^2 + BXY + CY^2 + DX + EY + F.$$

Now recall from analytic geometry that it is always possible to make a *rotation of axes* in the X–Y space after which the cross-term coefficient B is zero. Suppose that this has been done:

$$0 = AX^2 + CY^2 + DX + EY + F.$$

Then recall that (if certain "degenerate cases" are ignored) the curve described by this equation is an

ellipse	if A and C have the same sign,
hyperbola	if A and C have opposite signs,
parabola	if one of them (A or C) is 0.

We assign the same terminology to the partial differential equations that result when X is replaced by $\partial/\partial x$, etc. Thus Laplace's equation is elliptic, the wave equation is hyperbolic, and the heat equation is parabolic. (In the latter two cases y is called t for physical reasons.)

Now suppose that A, etc., do depend on x and y. Then at each point (x, y) it is possible to find a rotation

$$\frac{\partial}{\partial x'} = \cos\theta \frac{\partial}{\partial x} - \sin\theta \frac{\partial}{\partial y}\,,$$
$$\frac{\partial}{\partial y'} = \sin\theta \frac{\partial}{\partial x} + \cos\theta \frac{\partial}{\partial y}\,,$$

which eliminates the $B(x, y)$ term. (The angle θ may depend on x and y, so B is not necessarily zero at other points.) The character of the PDE at that point is defined to be elliptic, hyperbolic, or parabolic depending on the signs of the coefficients of the *new* coefficients A and C there. The *discriminant*

$$\Delta \equiv B(x, y)^2 - 4A(x, y)C(x, y)$$

is not changed by a rotation of coordinates. Therefore, it is easy to see that the equation is

elliptic	if $\Delta < 0$,
hyperbolic	if $\Delta > 0$,
parabolic	if $\Delta = 0$.

For most equations of practical interest, the operator will be of the same type throughout the domain.

The classification can be extended to *nonlinear* equations, provided they are *linear in their dependence on the second derivatives of u*. Example:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + u^3 = 0$$

is elliptic.

Remark: Those of you who know some linear algebra will realize that what we are doing here is diagonalizing the matrix

$$\begin{pmatrix} A & \frac{1}{2}B \\ \frac{1}{2}B & C \end{pmatrix},$$

that the *new* A and C are the eigenvalues of that matrix, and that Δ is -4 times its determinant. This is the secret to extending the classification to equations in more than two variables, such as

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = 0.$$

This example counts as hyperbolic, since it has one coefficient with sign opposite to the others. More generally, there is a coefficient matrix (quadratic form) which has to be diagonalized, and the signs of its eigenvalues are what counts: The operator is elliptic if all the signs are the same, hyperbolic if one is different, and parabolic if one eigenvalue is zero and the others have the same sign. (There are other possibilities, such as two positive and two negative eigenvalues, but they seldom arise in applications.)

Now let's discuss the three matters listed at the beginning. The facts I'm about to state are generalizations of things we already know about the heat, wave, and Laplace equation.

(1) In a hyperbolic or parabolic equation, we identify the "special" coordinate as the *time*. (That is the coordinate with the strange sign in the hyperbolic case or the zero in the parabolic case. In the latter case we assume that first-order derivatives with respect to t do occur, although by hypothesis second-order ones don't.) Then the fact is that these equations behave just like ordinary differential equations as to initial data: The parabolic equation is first-order, so a solution is uniquely determined by its initial value, $u(x, \ldots, t = 0)$. The hyperbolic equation is second-order, so you need also the initial value of the time derivative. Boundary conditions at the edges of the spatial domain may also be necessary to specify the solution, as we well know from examples. (These boundary conditions are of the same type as needed to produce a well-posed problem for an *elliptic* equation on that spatial domain — see below. This is not surprising, since the spatial equation we get when a hyperbolic or parabolic equation (in 3 or more variables) is solved by separation of variables is an elliptic equation, such as $\nabla^2\phi = -\omega^2\phi$.)

In the parabolic case, a solution is guaranteed to exist only in one direction of time from the initial data surface. (With the usual choice of signs, this is the positive time direction.) If you try to solve the heat equation in the negative direction, a solution may not exist for the given data; when solutions do exist, they are *unstable* in the sense that a small change in the data creates drastic changes in the solution.

Since real science and engineering deal with only approximately measured data, this makes the solution in the backward direction almost useless in practice.

For an elliptic equation, one might expect to have a well-posed problem given the value of u and its normal derivative on an "initial" surface, since the equation is second-order in every variable. However, it turns out that a solution may not exist for all data; it will exist in a neighborhood of the surface, but it will "blow up" somewhere else. When solutions exist, they may be unstable. Instead, the proper and natural boundary condition for an elliptic equation (as we know from physical applications of Laplace's equation) is to prescribe the function *or* its derivative (but not both) at every point on a *closed curve* surrounding a region. (Conversely, this sort of boundary condition will *not* give a well-posed problem for a hyperbolic or parabolic equation.)

I have been using the term *well-posed* without formally defining it. It means, above all, that the problem (consisting, typically, of a differential equation plus boundary conditions) has been stated so that it has exactly one solution. Stating too few conditions will make the solution nonunique; too many conditions, and it will not exist; try to use the wrong kind of conditions (e.g., initial data for an elliptic equation), and there will be no happy medium! In addition, it is customary to require *stability*; that is, that the solution depends continuously on the data.

(2) Elliptic and parabolic equations (with smooth coefficients) have solutions that are *smooth* (that is, differentiable arbitrarily many times), regardless of how rough their data (boundary values) are. But solutions of hyperbolic equations may be nondifferentiable, discontinuous, or even distributions — such as $\delta(x - ct)$ for the wave equation. In other words, singularities in the initial data are propagated by a hyperbolic equation into the solution region.

(3) Hyperbolic equations spread the initial data out into space at a finite "wave" speed. In contrast, the initial data of the heat equation can instantly affect the solution arbitrarily far away.

Validity of the WKB approximation

(some unfinished business)

Months ago, we searched for solutions of the differential equation

$$\frac{d^2y}{dt^2} + \omega^2 V(t) y = 0 \qquad (\omega \to \infty) \tag{†}$$

and arrived by several different arguments at the approximations

$$y_\pm = V(t)^{-\frac{1}{4}} e^{\pm i\omega \int_0^t \sqrt{V(\tilde{t})}\, d\tilde{t}}.$$

All that we actually showed at that time is that y_+ and y_- are *exact solutions to approximate problems.* That is, if you substitute $y_\pm$ into the left side of (†), you get

$$\frac{d^2y_\pm}{dt^2} + \omega^2 V(t) y_\pm = \left[\frac{5}{16} \left(\frac{V'}{V} \right)^2 - \frac{1}{4} \frac{V''}{V} \right] y_\pm\,, \tag{‡}$$

where the important point about the right side of (‡) is that it does not grow with ω. In fact, we *constructed* $y_\pm$ precisely so that this would be true: The phase function $\int \sqrt{V}\, dt$ is precisely what's needed to make the terms that grow as ω^2 cancel each other, and the amplitude function $V^{-\frac{1}{4}}$ similarly conspires to make all terms that grow as the first power of ω cancel out. Therefore, although the right side of (‡) is not zero, it is "small" in a certain sense; thus (‡) is an *approximation* to the equation (†) that we really want to solve. I showed how you could calculate higher-order WKB approximations, which satisfy equations similar to (‡) whose right sides actually *decrease* like some negative power of ω as $\omega \to \infty$.

What we did not show (in this or any other ODE perturbation problem) is that the approximations we constructed are *approximate solutions to the exact problem.* That is, given any linear combination, y_0, of the two functions $y_\pm$, is there some exact solution, y, of (†) for which the *error*, $y - y_0$, is small? This is the really crucial question about the approximation. Unfortunately, answering it requires more sophisticated theoretical techniques than the formal construction of the approximation did. For many perturbation problems, especially nonlinear ones, proofs of the asymptotic validity of the formal series solutions are not known. For the linear equation (†), however, there is a fairly easy proof, which I shall outline now.

Given a WKB function $y_0(t)$, let $y(t)$ be the exact solution of (†) with the same initial data:

$$y(0) = y_0(0), \qquad y'(0) = y_0'(0).$$

We know that

$$\frac{d^2y_0}{dt^2} + \omega^2 V(t) y_0 = -j(t),$$

where $j = O(\omega^0)$ as $\omega \to \infty$. (This is just a restatement of (‡). Note that all these functions depend on ω as well as t, but we won't write ω as an argument.) Now let

$$E(t) \equiv y(t) - y_0(t).$$

Then

$$\frac{d^2E}{dt^2} + \omega^2 V(t) E = +j(t,\omega),$$

and

$$E(0) = 0 = E'(0).$$

But this type of problem is one for which we constructed a Green function recently!

$$E(t) = \int_0^t G(t,z)\, j(z)\, dz,$$

where

$$G(t,z) \equiv \begin{cases} 0 & \text{for } t < z, \\ \dfrac{1}{W}\big(y_1(z) y_2(t) - y_2(z) y_1(t)\big) & \text{for } t > z. \end{cases}$$

Here y_1 and y_2 are any two independent solutions of (†), and W is their Wronskian.

If we know how big G is, then we know how big E is:

$$\begin{aligned} \max |E(t)| &\le \max |G(t,z)| \int_0^t |j(\tilde z)|\, d\tilde z \\ &\le \text{const. } \max |G(t,z)|. \end{aligned}$$

(I'm assuming that t is confined to some finite interval $[0,T]$, so that integrals of smooth functions over t are are always bounded by constants.) Suppose that

$$G(t,z) = O(\omega^{-1}). \qquad (*)$$

Then the same is true of E, and we will have achieved our goal of showing that $y \approx y_0$. (Recall that $E = y - y_0$.) More generally, *the error in a higher-order approximation to y will fall off as an appropriately high negative power of ω* (one power faster than the corresponding j does). If V were a constant, then a short calculation would show that

$$G(t,z) = \frac{1}{\omega} V^{-\frac{1}{2}} \sin\!\big[\omega\sqrt{V}\,(z-t)\big],$$

so that $(*)$ would be satisfied. This gives us some hope that $(*)$ is true in the real problem, where V is t-dependent.

Unfortunately, we don't know explicitly what G is, since it is built out of solutions of the exact equation, which we can't solve exactly. (That's why we're hunting an approximation, after all!) However, there is a cute way of lifting ourselves by our bootstraps to show that G obeys $(*)$. Recall that on the interval $t > z$ (which is all that matters in the integral for E), G itself satisfies (†):

$$\frac{\partial^2}{\partial t^2}G(t,z) + \omega^2 V(t)G(t,z) = 0. \tag{1}$$

The relevant initial conditions are the continuity and jump conditions at $t = z$:

$$G(z,z) = 0, \qquad \frac{\partial G}{\partial t}(z,z) = 1. \tag{2}$$

We can therefore construct a WKB approximation for G itself:

$$G(t,z) \approx \frac{1}{\omega}[V(z)V(t)]^{-\frac{1}{4}} \sin\left[\omega \int_z^t \sqrt{V(\tilde{z})}\, d\tilde{z}\right].$$

(The easiest way to arrive at this expression is to substitute WKB approximations for y_1 and y_2 into the previous formula for G.) A calculation too long to write out here verifies that this expression satisfies (2) exactly and satisfies (1) up to a term

$$j_G(t,z) = O(\omega^{-1}).$$

Denote the error term in the Green function by E_G:

$$G(t,z) = \frac{1}{\omega}[V(z)V(t)]^{-\frac{1}{4}} \sin\left[\omega \int_z^t \sqrt{V(\tilde{z})}\, d\tilde{z}\right] + E_G(t,z).$$

We estimate it by the same reasoning as before:

$$E_G(t,z) = \int_z^t G(t,\tilde{z})\, j_G(\tilde{z},z)\, d\tilde{z};$$

$$\max|E_G| \le \frac{\text{const.}}{\omega} \max|G|.$$

Consequently,

$$\max|G| \le \frac{C_1}{\omega} + \frac{C_2}{\omega} \max|G|$$

for some constants C_1 and C_2. So far we still seem to be arguing in a circle, since $|G|$ appears on both sides of the inequality. But now comes the bootstrap trick: Rewrite the inequality as

$$\left(1 - \frac{C_2}{\omega}\right) \max|G| \le \frac{C_1}{\omega}.$$

If $\omega > C_2$ — and recall that we are interested only in what happens as ω becomes large — the first factor here is positive, and we can divide by it:

$$\max|G| \le \frac{C_1}{\omega}\left(1 - \frac{C_2}{\omega}\right)^{-1} = O(\omega^{-1}).$$

THE END

