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) are 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 equation in two variables, say x and y, looks like

$$L[u] \equiv A(x,y)\,\frac{\partial^2 u}{\partial x^2} + B(x,y)\,\frac{\partial^2 u}{\partial x\,\partial y} + C(x,y)\frac{\partial^2 u}{\partial y^2} + D(x,y)\,\frac{\partial u}{\partial x} + E(x,y)\,\frac{\partial u}{\partial y} + F(x,y)u = 0,$$

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

$$\begin{aligned} \text{elliptic} \quad & \text{if } \Delta < 0, \\ \text{hyperbolic} \quad & \text{if } \Delta > 0, \\ \text{parabolic} \quad & \text{if } \Delta = 0. \end{aligned}$$

For most equations of practical interest, the operator will be of the same type at all points.

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

called *quasilinear*. Example:

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

is quasilinear and elliptic.

Remark: From linear algebra you may recall that what we are doing here is diagonalizing the matrix (quadratic form)

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

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 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 the first-order derivative with respect to t does occur (multiplied by a real coefficient), although

by hypothesis the second-order one doesn'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 (let's call it D) may also be necessary to specify the solution, as we well know from examples. (These latter boundary conditions are of the same type as needed to produce a well-posed problem for an *elliptic* equation on D — 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 or surface* 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 applications, this is the speed of sound, the speed of light, etc.) In contrast, the initial data of the heat equation can instantly affect the solution arbitrarily far away.

There is one more type of second-order linear equation: the *(time-dependent) Schrödinger equations* of quantum mechanics, of which the simplest case is

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

This was overlooked in the classification above, because we were tacitly assuming that all the quantities were real. The Schrödinger equation does not fit into any of the three standard categories; instead, it shares some of the features of parabolic and hyperbolic equations.

- Like the heat equation, it is first-order in time. Therefore, $u(x,0)$ (by itself) is appropriate initial data.

- Unlike the heat equation, but like the wave equation, its solutions are not necessarily smooth. Unlike the wave equation, the singularities in the solutions can disappear and then reappear at later times; this happens most notoriously for the Green function of the *harmonic oscillator* equation

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

  which contains the periodic factor $\csc(2t)$.

- Unlike the wave equation, but like the heat equation, its solutions are not limited by a finite propagation speed.

- Its nicest property is *unitarity*: The L^2 norm of the solution at any fixed t is the same as the L^2 norm of the initial data. That is,

$$\int_D |u(x,t)|^2\,dx = \int_D |u(x,0)|^2\,dx.$$

(Here it is assumed that the differential operator in the spatial variables (the Hamiltonian) is self-adjoint.)

The maximum principle

Consider an *elliptic* or *parabolic* PDE (second-order, linear, homogeneous),

$$A(x,y)\,\frac{\partial^2 u}{\partial x^2} + B(x,y)\,\frac{\partial^2 u}{\partial x\,\partial y} + C(x,y)\,\frac{\partial^2 u}{\partial y^2} + D(x,y)\,\frac{\partial u}{\partial z} + E(x,y)\,\frac{\partial u}{\partial y} + F(x,y)u = \epsilon\,\frac{\partial u}{\partial t}$$

where $\epsilon = 0$ or 1. (In this discussion u is a function of two variables or of three, depending on the context, but we'll usually suppress t in the notation.) If a solution has a maximum or minimum, it occurs either on the *boundary* of the

region considered or at a *critical point* (where all first-order partial derivatives of u are zero). Consider the latter possibility. We may assume for simplicity that the critical point is at the origin and that $B(0,0)$ has been eliminated by a rigid rotation of the x and y axes. So, at the origin the equation reduces to

$$A(0,0)\,\frac{\partial^2 u}{\partial x^2} + C(0,0)\,\frac{\partial^2 u}{\partial y^2} + F(0,0)u = 0,$$

where A and C are *both positive*, by definition of elliptic or parabolic.

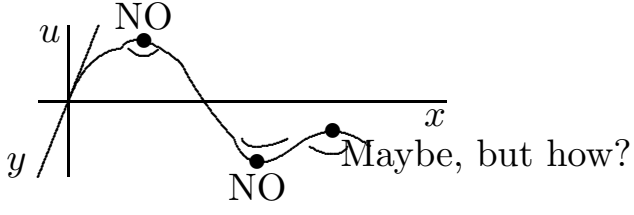
If $F(0,0)$ is negative and $u(0,0)$ is positive, then

$$A(0,0)\,\frac{\partial^2 u}{\partial x^2}(0,0) + C(0,0)\,\frac{\partial^2 u}{\partial y^2}(0,0) > 0,$$

so at least one of the second-order partials is positive. Therefore, $u(0,0)$ *cannot be a local maximum*. Similarly, if $F(0,0)$ and $u(0,0)$ are both negative, then

at least one of the second-order partials is negative, so $u(0,0)$ *cannot be a local minimum*. Putting these facts together, we can conclude:

Theorem (a simple maximum principle): If $F(x,y)$ is always negative, then u cannot have a positive maximum nor a negative minimum in the interior of its domain.



We have seen this principle at work in the eigenvalue problem for ∇^2. If $-\nabla^2 u = \lambda u$, then we found that λ must be positive for Dirichlet boundary conditions and nonnegative for Neumann boundary conditions; thus F is nonnegative

and the theorem does *not* apply. That allows us to have eigenfunctions like those pictured on pp. 43 and 113, which blatantly violate the conclusion of the theorem. On the other hand, when $\lambda < 0$, or $\lambda = 0$ and u is not constant, then the theorem applies; indeed, the solutions (of the PDE) that we found in such cases were always *concave away* from the coordinate plane in at least one dimension (like e^x), and that was why we could never find solutions for such λ that satisfied all the homogeneous boundary conditions needed to be eigenfunctions.

A somewhat more technical proof yields a somewhat simpler sounding theorem:

Theorem (a maximum principle): If $F(x, y) = 0$ everywhere, then u cannot have an interior local extremum of *either* kind, except in the trivial case where u is a constant function.

For the special case of Laplace's equation, $\nabla^2 u = 0$, this maximum principle

follows from the theorem that

$$u(\mathbf{x}) = (\text{average of } u \text{ over a circle centered at } \mathbf{x}),$$

which in turn is easy to see from the expansion of u in a Fourier series (in the polar angle) inside the circle (see p. 103). The same thing is true for Laplace's equation in 3 dimensions, with the circle replaced by a sphere and the Fourier series by the expansion in spherical harmonics.

As we've seen, the maximum principle holds only for a rather restricted class of differential equations: not only must the equation be elliptic or parabolic, but also there is a sign condition on the terms without derivatives. Nevertheless, various forms of the maximum principle are important tools in proving theorems about the properties of solutions. Here are two examples:

Corollary 1: if $F = 0$ everywhere and the domain of u is bounded, then the global maximum and minimum values of u *occur on the boundary*.

Corollary 2: If $F \leq 0$ everywhere, then $u = 0$ everywhere on the boundary implies that $u = 0$ *everywhere*.

COROLLARY TO THE COROLLARY: **Uniqueness Theorem:** If a homogeneous linear problem satisfies the maximum principle in the sense of Corollary 2, then the solution of an associated nonhomogeneous problem is unique.

EXAMPLE: Consider the Poisson equation $-\nabla^2 u = j$ in a bounded region with the nonhomogeneous boundary condition that $u(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x}$ on the boundary of the region. (The functions j and f are fixed, as part of the statement of the problem.) Then the solution (if we assume that it exists) is unique: If there were two of them, u_1 and u_2, then $v \equiv u_1 - u_2$ would satisfy $\nabla^2 v = 0$ in the region and $v = 0$ on the boundary, so v would be identically zero, a contradiction.