Fundamental Concepts: Linearity and Homogeneity

This is probably the most abstract section of the course, and also the most important, since the procedures followed in solving PDEs will be simply a bewildering welter of magic tricks to you unless you learn the general principles behind them. We have already seen the tricks in use in a few examples; it is time to extract and formulate the principles. (These ideas will already be familiar if you have had a good linear algebra course.)

Linear equations and linear operators

I think that you already know how to recognize linear and nonlinear equations, so let's look at some examples before I give the official definition of "linear" and discuss its usefulness.

Algebraic equations:

Linear

$$\left\{ \begin{array}{l} x + 2y = 0, \\ x - 3y = 1 \end{array} \right\}$$

Nonlinear

$$x^5 = 2x$$

Ordinary differential equations:

Linear

$$\frac{dy}{dt} + t^3\, y = \cos 3t$$

Nonlinear

$$\frac{dy}{dt} = t^2 + e^y$$

Partial differential equations:

Linear *Nonlinear*

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \qquad\qquad \frac{\partial u}{\partial t} = \left(\frac{\partial u}{\partial x}\right)^2$$

What distinguishes the linear equations from the nonlinear ones? The most visible feature of the linear equations is that they involve the unknown quantity (the dependent variable, in the differential cases) *only to the first power*. The unknown does not appear inside transcendental functions (such as sin and ln), or in a denominator, or squared, cubed, etc. This is how a linear equation is usually recognized by eye. Notice that there may be terms (like $\cos 3t$ in one example) which don't involve the unknown at all. Also, as the same example term shows, there's no rule against nonlinear functions of the *independent* variable.

The formal definition of "linear" stresses not what a linear equation looks like, but the properties that make it easy to describe all its solutions. For concreteness let's assume that the unknown in our problem is a (real-valued) function of one or more (real) variables, $u(x)$ or $u(x, y)$. The fundamental concept is not "linear equation" but "linear operator":

Definition: An operation, L, on functions is *linear* if it satisfies

$$L(u + v) = L(u) + L(v) \qquad \text{and} \qquad L(\lambda u) = \lambda L(u) \tag{$*$}$$

for all functions u and v and all numbers λ.

Examples of linear operations are

- differentiation of u: $\quad L(u) \equiv \dfrac{du}{dx}$,
- multiplication of u by a given function of x: $\quad L(u) \equiv x^2 u(x)$,

- evaluation of u at a particular value of x: $\quad L(u) \equiv u(2)$,
- integration of u $\quad L(u) \equiv \int_0^1 u(x)\,dx$.

In each example it's easy to check that $(*)$ is satisfied, and we also see the characteristic *first-power* structure of the formulas (*without* u-independent terms this time). In each case L is a *function on functions*, a mapping which takes a function as input and gives as output either another function (as in the first two examples) or a number (as in the last two). Such a superfunction, considered as a mathematical object in its own right, is called an *operator*.

Now we can return to equations:

Definition: A *linear equation* is an equation of the form

$$L(u) = g,$$

where L is a linear operator, g is a "given" or "known" function (or number, as the case may be), and u is the unknown to be solved for.

So the possible u-independent terms enter the picture in the role of g. This leads to an absolutely crucial distinction:

Homogeneous vs. nonhomogeneous equations

Definition: A linear equation, $L(u) = g$, is *homogeneous* if $g = 0$ (i.e., *all* terms in the equation are *exactly* of the first degree in u); it is *nonhomogeneous* if $g \neq 0$ (i.e., "constant" terms also appear).

In the second parenthetical clause, "constant" means independent of u. The "constant" term g may be a nontrivial function of the *independent* variable(s) of the problem.

Among our original examples, the linear ODE example was nonhomogeneous (because of the $\cos 3t$) and the PDE example was homogeneous. The algebraic example is nonhomogeneous because of the 1. Here we are thinking of the system of simultaneous equations as a *single* linear equation in which the unknown quantity is a two-component vector,

$$\vec{u} \equiv \begin{pmatrix} x \\ y \end{pmatrix}.$$

The linear operator L maps $\vec{u}$ onto another vector,

$$\vec{g} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

As you probably know, the system of equations can be rewritten in matrix notation as

$$\begin{pmatrix} 1 & 2 \\ 1 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The linear operator is described by the square matrix

$$M = \begin{pmatrix} 1 & 2 \\ 1 & -3 \end{pmatrix}.$$

In solving a differential equation we usually need to deal with initial or boundary conditions in addition to the equation itself. The main reason is that initial or boundary data need to be specified to give the problem a unique answer. Usually these conditions are themselves linear equations — for example, a standard initial condition for the heat equation:

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

Often the differential equation will be homogeneous but at least one of the boundary conditions will be nonhomogeneous. (The reverse situation also occurs.) Therefore, I think it's helpful to introduce one more bit of jargon:

Definitions: A *linear problem* consists of one or more linear conditions (equations) to be satisfied by the unknown, u. A linear problem is *homogeneous* if all of its conditions are homogeneous, *nonhomogeneous* if one or more of the conditions are nonhomogeneous.

Example A: The ODE problem

$$u'' + 4u = 0, \qquad u(0) = 1, \quad u'(0) = 0$$

is a nonhomogeneous linear problem. The ODE by itself is homogeneous, however.

Example B: The PDE problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + j(x), \qquad u(x,0) = 0, \quad u(0,t) = 0, \quad u(t,1) = 0$$

is a nonhomogeneous linear problem. The boundary conditions *and* the initial condition are homogeneous, but the heat equation itself is nonhomogeneous in this case; the function j represents generation of heat inside the bar (perhaps by combustion or radioactivity), a possibility not considered in the discussion of the heat-conduction problem in Appendix A.

Remark: It is easy to see that every homogeneous linear equation has $u = 0$ as a solution. (One proof: $L(0) = L(u - u)$ (for any u) $= L(u) - L(u) = 0$, QED.) Therefore, any homogeneous linear problem has 0 as a solution. Therefore, *if a linear problem has a unique solution and that solution is nontrivial (not just the 0 function), then that linear problem must be nonhomogeneous.* That is, an interesting, well-posed problem always has at least one nonhomogeneous condition.

Solving linear problems

The importance of linear problems is that solving them is made easy by the *superposition principles* (which don't apply to nonlinear problems):

Principles of Superposition:

1. A linear combination of solutions of a homogeneous problem is a new solution of that problem. That is, if $L(u_1) = 0$ and $L(u_2) = 0$, then $L(c_1u_1 + c_2u_2) = 0$ for any numbers c_1 and c_2 (and similarly for more than two solutions, and for more than one homogeneous linear equation defining the problem).

    Example: Let **Problem 1** be the homogeneous ODE $u'' + 4u = 0$. Two solutions of this problem are

    $$u_1 \equiv \cos 2x, \qquad u_2 \equiv \sin 2x.$$

    Then $u = u_1 + 3u_2$, for example, is also a solution. (In fact, we know that the *most general* solution is $c_1u_1 + c_2u_2$ where the c's are arbitrary constants. But for this we need a deeper existence-and-uniqueness theorem for second-order ODEs; it doesn't just follow from linearity.)

2. The sum of a solution of a nonhomogeneous problem and a solution of the *corresponding homogeneous problem* is a new solution of the original nonhomogeneous problem. ("Corresponding homogeneous problem" means the one with the same L's, but with all g's replaced by 0.)

   Example: Let **Problem 2** be the nonhomogeneous equation $u'' + 4u = e^x$. One solution is $u_{\mathrm{p}} \equiv \frac{1}{5}e^x$. (This has to be found by the method of undetermined coefficients, or by luck. Again, general principles of linearity by themselves can't solve the whole problem.) Now if we add a solution of Problem 1 we get a new solution of Problem 2: $u_3 \equiv \frac{1}{5}e^x + \cos 2x$.

3. The difference of two solutions of a nonhomogeneous problem is a solution of the corresponding homogeneous problem. Therefore, *every* solution of a nonhomogeneous problem can be obtained from *one particular* solution of that problem by adding some solution of the homogeneous problem.

Example: The general solution of Problem 2 is

$$u = \tfrac{1}{5}e^x + c_1 \cos 2x + c_2 \sin 2x.$$

4. The sum of solutions to two nonhomogeneous problems with the same L's is a solution of a *new* nonhomogeneous problem, for which the g's are the *sums* of the corresponding g's of the two original problems. (Similarly for more than two nonhomogeneous problems.)

   Example 1: The sum of two solutions of Problem 2, u_{p} and u_3, is $z \equiv \tfrac{2}{5}e^x + \cos 2x$, which is a solution of $z'' + 4z = 2e^x$. The important lesson to be learned from this example is that the right-hand side of this new equation is *not* e^x, the nonhomogeneous term of the two old equations. *Do not superpose solutions of a* NONHOMOGENEOUS *problem in the hope of getting a solution of that* SAME *problem.*

Example 2: Note that u_{p} is the unique solution of **Problem 3**:

$$u'' + 4u = e^x, \qquad u(0) = \tfrac{1}{5}, \quad u'(0) = \tfrac{1}{5}\,.$$

Suppose that we really want to solve **Problem 4**:

$$u'' + 4u = e^x, \qquad u(0) = 0, \quad u'(0) = 0.$$

Recalling Principles 2 and 3 as applied to the differential equation alone (not the initial conditions), we see that $u = u_{\mathrm{p}} + y$, where y is some solution of $y'' + 4y = 0$. A moment's further thought shows that the correct y is the solution of **Problem 5**:

$$y'' + 4y = 0, \qquad y(0) = -\tfrac{1}{5}, \quad y'(0) = -\tfrac{1}{5}.$$

A standard calculation shows that $y = -\frac{1}{5}\cos 2x - \frac{1}{10}\sin 2x$, and from this and u_{p} we can get the solution of Problem 4. (Of course, in solving such problems we usually don't write out Problem 5 as an intermediate step; the standard procedure is to impose the initial data of Problem 4 on the general solution found earlier. That is just a different way of organizing the same algebra. However, consciously splitting a nonhomogeneous problem into two nonhomogeneous problems, as I've demonstrated here for an ODE, is a common technique for solving PDEs.)

In summary, these principles provide the basic strategies for solving linear problems. If the problem is nonhomogeneous and complicated, you split it into simpler nonhomogeneous problems and add the solutions. If the solution is not unique, the nonuniqueness resides precisely in the possibility of adding a solution of the corresponding homogeneous problem. (In particular, if the original problem is homogeneous, then you seek the general solution as a linear combination of some list of basic solutions.) If the problem statement contains enough initial and boundary conditions, the solution will be unique; in that case, the only solution of the homogeneous problem is the zero function.

An important example application of this strategy is the solution of the heat-conduction problem in a bar with fixed end temperatures:*

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

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

BC: $$u(0,t) = T_1\,, \qquad u(1,t) = T_2\,.$$

Here we have a homogeneous PDE, a nonhomogeneous initial condition, and two nonhomogeneous boundary conditions. The trick is to treat the two types of nonhomogeneity separately. One writes $u = v + w$, where

* See Appendix A, or Haberman's book.

(1) v is to be a solution of the problem consisting of the PDE and the nonhomogeneous BC, with *no* particular IC assumed. It is possible to find a solution of this problem which is *independent of* t: $v(x,t) = V(x)$.

(2) w is to be a solution of the problem consisting of the PDE, the *homogeneous* Dirichlet boundary conditions

$$w(0,t) = 0, \qquad w(1,t) = 0,$$

and the initial condition needed to make u satisfy the original IC. Namely,

$$w(x,0) = f(x) - V(x).$$

It is very important that the only nonhomogeneity in this second problem is the IC. This makes it possible to solve for w by the method of separation of variables and then add the solutions without falling into the trap I warned you against earlier (Example 1). The solution is completed by finding the Fourier series of the function $f - V$.

The details of steps (1) and (2) are carried out in Appendix A.

For the processes of separating variables and calculating Fourier coefficients to work here, it was absolutely crucial to make the boundary conditions homogeneous first. In the calculation of normal modes, *no* nonhomogeneous conditions at all are imposed. The appropriate nonhomogeneous IC is imposed on a *superposition* (w) of normal modes. Then still another term, v, is added to satisfy the nonhomogeneous BC.

One more time:

Impose only HOMOGENEOUS conditions on normal modes (separated solutions).

Impose nonhomogeneous conditions only on a SUPERPOSITION (sum or integral) of normal modes.

A related principle is

Handle only one nonhomogeneity at a time!

This principle is handled in practice by different strategies in different problems. Let's consider a doubly nonhomogeneous problem with the structure

$$L_1(u) = f_1\,, \qquad L_2(u) = f_2\,.$$

The two principal strategies are these:

1. **"Zero out" the other condition.** Solve

$$\begin{aligned} L_1(u_1) &= f_1\,, \qquad & L_2(u_1) &= 0, \\ L_1(u_2) &= 0, \qquad & L_2(u_2) &= f_2\,. \end{aligned}$$

Then $u = u_1 + u_2$.

Examples where this strategy is used include

(a) treatment of the initial data u and $\frac{\partial u}{\partial t}$ in the wave equation;

(b) Laplace's equation in a rectangle with boundary values given on two perpendicular sides.

2. **Temporarily ignore the other condition.** Solve $L_1(u_1) = f_1$ and let $L_2(u_1)$ be whatever it turns out to be, say $L_2(u_1) \equiv h$. Next solve

$$L_1(u_2) = 0, \qquad L_2(u_2) = f_2 - h.$$

Then $u = u_1 + u_2$.

Examples where this strategy is used include

(a) the method of undetermined coefficients for an ordinary differential equation with initial conditions;

(b) finding a *steady-state solution* for the wave or heat equation with nonzero, but time-independent, boundary conditions.