FINITE DIFFERENCING

The basic method for solving differential equations in a computer is ***finite differencing***. The nicest feature of the method is that it allows analysis of objects of almost any shape, such as earth topography or geological structure. Ordinarily, finite differencing is a fairly straightforward task. The main pitfall is instability. It rather often happens that a seemingly reasonable approach to a reasonable physical problem leads to wildly oscillatory, divergent calculations. Luckily, there is a fairly small body of important tricks, which can be quickly learned, that should solve most stability problems.

Of secondary concern are the matters of cost and accuracy. These must be considered together since improved accuracy can always be achieved simply by paying the higher price of a more refined computational mesh. Although the methods to be considered on the next several pages have not been chosen to be best in the sense of accuracy or efficiency it turns out that they are excellent methods. To my knowledge and experience, some cannot be improved upon at all while others can be improved upon only in small ways. By "small" I mean an improvement in efficiency of a factor of five or less. Such an improvement is rarely of consequence in research or experimental work; however, its importance in matters of production and finance will justify pursuit of the literature far beyond the succeeding pages.

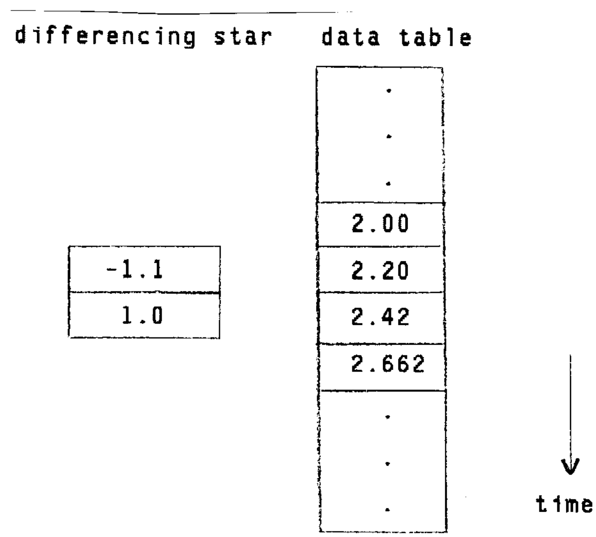
First Derivatives, Explicit Methods

The inflation of money q at a 10% rate can be described by the difference equation

$$q_{t+1} - q_t = .10\ q_t \tag{1a}$$

$$(1.0)\ q_{t+1} + (-1.1)q_t = 0 \tag{1b}$$

This one-dimensional calculation can be re-expressed in the form of a differencing star and a data table. As such it provides a prototype for the organization of calculations on two-dimensional partial-differential equations. Consider



(2)

Since the data in the data table actually satisfy the difference equation (1), the differencing star may be laid anywhere on top of the data table, the numbers in the star multiplied by those in the underlying table, and the resulting cross products will sum to zero. On the other hand, if all but one number (the initial condition) in the data table were missing then the rest of the numbers could be filled in, one at a time, by sliding the star along, taking the difference equation to be true, and solving for the unknown data value at each stage.

Less trivial examples utilizing the same differencing star arise when the numerical constant .10 is replaced by a complex number. Such examples exhibit oscillation as well as growth and decay. The heat-flow equation is a partial-differential equation which has the same structure as the 15-degree seismogram extrapolation equation. A two-sentence derivation of it is to say: (1) The heat flow H in the x-direction equals the negative of the gradient $\partial q/\partial x$ of temperature q times the heat conductivity σ. (2) The decrease $-dq/dt$ of temperature q is proportional to the divergence of heat flow $\partial H/\partial x$ divided by the heat storage capacity C of the material. Combining these, and for convenience neglecting the case of C or σ being a function of x or t, we get the usual one-dimensional heat-flow equation

$$\frac{\partial q}{\partial t} = \frac{\sigma}{C}\frac{\partial^2 q}{\partial x^2} \tag{3}$$

To implement (3) in a computer we need some difference approximations for the partial differentials. The most obvious (but not the only) approach is the basic definition of elementary calculus. For the time derivative

$$\frac{\partial q}{\partial t} \approx \frac{q(t+\Delta t) - q(t)}{\Delta t} \tag{4a}$$

It is convenient to use a subscript notation so that this is compacted into

$$\frac{\partial q}{\partial t} \approx \frac{q_{t+1} - q_t}{\Delta t} \tag{4b}$$

In this subscript notation $t+\Delta t$ is abbreviated by $t+1$, a convenience when we get to complicated equations. The second-derivative formula may be obtained by doing the first derivative twice. This leads to $q_{t+2} - 2q_{t+1} + q_t$. The formula is usually treated more symmetrically by shifting it to $q_{t+1} - 2q_t + q_{t-1}$. They are equivalent as Δt tends to zero but the more symmetrical arrangement will be more accurate when Δt is not zero. Using superscripts to describe x-dependence we have a finite-difference approximation to the second space derivative:

$$\frac{\partial^2 q}{\partial x^2} \approx \frac{q^{x+1} - 2q^x + q^{x-1}}{\Delta x^2} \tag{5}$$

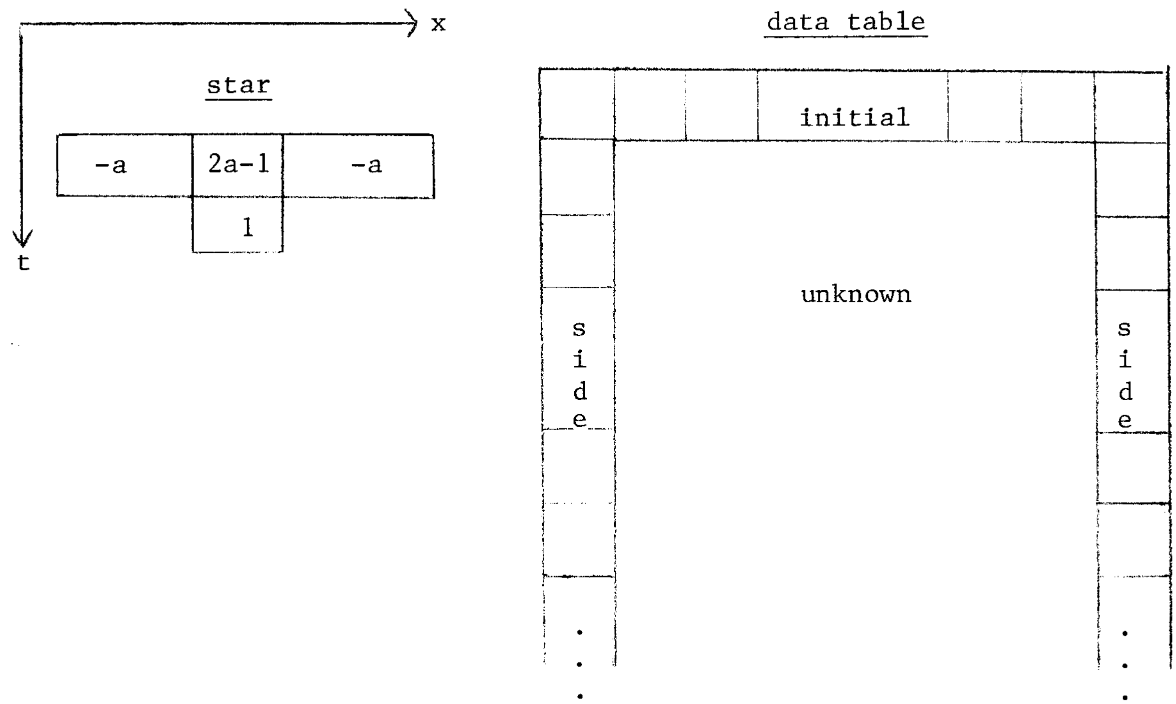
Inserting the last two equations into the heat-flow equation (and using = to denote $\approx$) we get

$$\frac{q_{t+1}^x - q_t^x}{\Delta t} = \frac{\sigma}{C}\frac{q_t^{x+1} - 2q_t^x + q_t^{x-1}}{(\Delta x)^2} \tag{6a}$$

Letting $a = \sigma\Delta t/(C\Delta x^2)$ this can be arranged thus

$$q_{t+1}^{x} - q_{t}^{x} - a\left(q_{t}^{x+1} - 2q_{t}^{x} + q_{t}^{x-1}\right) = 0 \qquad (6b)$$

Now we can interpret (6b) geometrically in terms of a computational star in the (x,t)-plane as



(7)

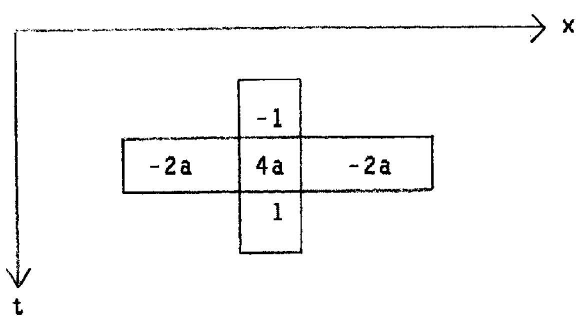
By moving the star around in the data table you will note that it can be positioned so that only one number at a time (the 1) lies over an unknown element in the data table. This enables the computation of subsequent rows beginning from the top. By doing this you are solving the partial-differential equation by the finite-difference method. There are other possible arrangements of initial and side conditions such as zero-slope side conditions.

The difficulty with the above method is that it doesn't work for all possible numerical values of a. It turns out that when a is too large (Δx too small) that the solution in the interior region in the data table will contain oscillations growing with time like (1,-2,4,-8,16,...). What is happening is that the low-frequency part of the solution is o.k. (for a while) but that the high-frequency part is diverging. The precise reason why the divergence occurs is the subject of some mathematical analysis which will be done later. At wavelengths long compared to Δx or Δt we expect our difference approximation to agree with the true heat-flow equation, which smooths out irregularities in temperature. At short wavelengths the wild oscillation shows that the difference equation can behave in a way somewhat opposite to the way the differential equation behaves. The short wavelength discrepancy arises because difference operators become equal to differential operators only at long wavelengths. The divergence of the solution is a fatal problem because the subsequent round-off error will eventually destroy the low frequencies too.

If we believe that the instability arises because the time derivative is centered at a slightly different time $t+\frac{1}{2}$ than the second x-derivative at time t, then we are led to consider the so called *leapfrog* method in which the time derivative is taken as a difference between t-1 and t+1

$$\frac{\partial q}{\partial t} \approx \frac{q_{t+1} - q_{t-1}}{2\Delta t} \tag{8}$$

The resulting leapfrog differencing star is



(9)

Here the result is even worse. A later analysis shows that the solution is now divergent for *all* real numerical values of a. Although it was a good idea to center both derivatives in the same place it turns out that it was a bad idea to express a first derivative over a span of more mesh points. The enlarged operator has two solutions in time instead of just the familiar one.

The numerical solution is the sum of the two theoretical solutions, one of which, unfortunately (in this case), grows and oscillates for all real values of a.

To avoid all these problems (and get more accurate answers as well) we now turn to slightly more complicated solution methods known as ***implicit*** methods.

Implicit Methods

Let us consider solving the equation

$$\frac{dq}{dt} = 2\ r\ q \tag{10}$$

by numerical methods. First of all, note that the inflation of money equation (1) where $2r = .1$ provides an approximation. But then note that in the inflation-of-money equation the expression of dq/dt is centered at $t+\frac{1}{2}$, whereas the expression of q by itself is at time t. There is no reason why the q on the right side of equation (10) cannot be averaged at time t with time t+1, thus centering the whole equation at $t+\frac{1}{2}$. Specifically, a centered approximation of (10) is

$$q_{t+1} - q_t = 2r\Delta t\ \frac{q_{t+1} + q_t}{2} \tag{11a}$$

Letting $b = r\Delta t$ this becomes

$$(1-b)\ q_{t+1} - (1+b)\ q_t = 0 \tag{11b}$$

which is representable as the difference star

$$t \downarrow \quad \begin{array}{|c|} \hline -1 - b \\ \hline 1 - b \\ \hline \end{array} \tag{11c}$$

For a fixed Δt this star gives a more accurate solution to the differential equations (10) than equation (1) for the inflation of money. It turns out that we can achieve an analogous accuracy improvement as well as solve the stability problem by applying the same idea to the heat-flow equation. We had represented the heat-flow equation (6b) as

$$q_{t+1}^{x} - q_{t}^{x} = a\left(q_{t}^{x+1} - 2q_{t}^{x} + q_{t}^{x-1}\right)$$

Now, instead of expressing the right-hand side entirely at time t, we average it at t and t+1, obtaining

$$q_{t+1}^{x} - q_{t}^{x} = \frac{a}{2}\left[\left(q_{t}^{x+1} - 2q_{t}^{x} + q_{t}^{x-1}\right) + \left(q_{t+1}^{x+1} - 2q_{t+1}^{x} + q_{t+1}^{x-1}\right)\right] \qquad (12a)$$

This is called the Crank-Nicolson method. Letting $b = a/2$ the difference star is

-b	2b - 1	-b
-b	2b + 1	-b

(12b)

When placing this star over the data table you will note that typically there are three elements at a time covering unknowns. Saying the same thing with equations, we move all the t+1 terms in (12a) to the left, and the t terms to the right, obtaining

$$-bq_{t+1}^{x+1} + (1+2b)q_{t+1}^{x} - bq_{t+1}^{x-1} = bq_{t}^{x+1} + (1-2b)q_{t}^{x} + bq_{t}^{x-1} \qquad (13a)$$

Taking all the t+1 values to be unknown while all the t values are known, the right side of (13a) is known, say k_t^x, and the left side is a set of simultaneous equations for the unknown q_{t+1}. In other words (13a) does not give us each q_{t+1}^{x} ***explicitly*** but they are ***implicitly*** given by virtue of our ability to solve simultaneous equations. If the x-axis is limited to 5 points these equations are

$$\begin{bmatrix} end1 & -b & 0 & 0 & 0 \\ -b & 1+2b & -b & 0 & 0 \\ 0 & -b & 1+2b & -b & 0 \\ 0 & 0 & -b & 1+2b & -b \\ 0 & 0 & 0 & -b & end2 \end{bmatrix} \begin{bmatrix} q^1_{t+1} \\ q^2_{t+1} \\ q^3_{t+1} \\ q^4_{t+1} \\ q^5_{t+1} \end{bmatrix} = \begin{bmatrix} k^1_t \\ k^2_t \\ k^3_t \\ k^4_t \\ k^5_t \end{bmatrix} \qquad (13b)$$

The values "end1" and "end2" are adjustable and have to do with the side boundary conditions. The important thing to notice is that the matrix is tridiagonal - that is, except for three central diagonals, all the elements of the matrix in (13b) are zero. The solution to such a set of simultaneous equations may be very economically obtained. It turns out that the cost is very little (about a factor of two) more than the cost of the explicit method given by (6). In fact, it turns out to be cheaper since the increased accuracy of (13) over (6) allows the use of a much larger numerical choice of Δt. The solution technique for the tridiagonal simultaneous equations is in all the standard textbooks on the solution of partial-differential equations, including FGDP. It will be omitted here.