Mathematics 307—November 5, 1995

Cubic splines

In now ancient days—say ten years ago—when a draftsman wanted to draw a smooth curve through a number of points he would use a flexible metal ruler called a **spline**. Very roughly, if the spline were forced to pass through a fixed set of points, it would adopt the shape in which it had the least energy, and in which it avoided severe curvature. The energy involved here is of course the potential energy stored in the flexing of the ruler. Modern draftsmen use splines, too, but—naturally—in software. The idea is to approximate a curve by smooth pieces which join together smoothly. It is the basis of nearly all computer-generated curve drafting.

I will be following roughly the discussion in §§5.8–5.9 of *Essentials of Numerical Analysis* by Peter Henrici.

Cubic polynomials on an interval

Given $a < b$ and points (a, y_a), (b, y_b) in the plane, there is exactly one line segment in the plane from one point to the other. In other words, we can find a unique linear function $f(x)$ such that

$$f(a) = y_a, \quad f(b) = y_b \ .$$

This is essentially because the equation $y = c_1 x + c_0$ has two parameters to choose in order to match the endpoints.

A cubic polynomial $c_3x^3 + c_2x^2 + c_1x + c_0$ has four parameters, and we would therefore expect to be able to match four independent conditions at the end points of the interval $[a, b]$. In fact, we can specify both values and slopes at the endpoints.

Proposition. *Given $a < b$ and numbers*

$$y_a, \quad s_a, \quad y_b, \quad s_b$$

there exists a unique cubic polynomial $f(x)$ such that

$$\begin{aligned} f(a) &= y_a \\ f'(a) &= s_a \\ f(b) &= y_b \\ f'(b) &= s_b \end{aligned}$$

It is simple to find an explicit formula for $f(x)$. Instead of writing it as a combination of monomials x^i, however, it will be simpler to write it as a linear combination

$$f(x) = c_{3,0}(x-a)^3 + c_{2,1}(x-a)^2(x-b) + c_{1,2}(x-a)(x-b)^2 + c_{0,3}(x-b)^3 \ .$$

Then

$$f'(x) = (x-a)^2(3c_{3,0} + c_{2,1}) + 2(x-a)(x-b)(c_{2,1} + c_{1,2}) + (x-b)^2(c_{1,2} + 3c_{0,3})$$

so that the conditions to be satisfied can be written

$$\begin{aligned} c_{0,3}(a-b)^3 &= y_a \\ c_{3,0}(b-a)^3 &= y_b \\ (c_{1,2} + 3c_{0,3})(a-b)^2 &= s_a \\ (c_{2,1} + 3c_{3,0})(b-a)^2 &= s_b \end{aligned}$$

or

$$\begin{aligned}
c_{0,3} &= -y_a/\Delta^3 \\
c_{3,0} &= \quad y_b/\Delta^3 \\
c_{1,2} + 3c_{0,3} &= \quad s_a/\Delta^2 \\
c_{1,2} &= \quad s_a/\Delta^2 + y_a/\Delta^3 \\
c_{2,1} + 3c_{3,0} &= \quad s_b/\Delta^2 \\
c_{2,1} &= \quad s_b/\Delta^2 - y_b/\Delta^3
\end{aligned}$$

where $\Delta = (b - a)$.

Since a cubic polynomial is determined completely by its values and its slopes at the endpoints of an interval $[a, b]$, its second derivatives at the endpoints are in principle determined by those values and slopes. Explicitly, we have the formula

$$\frac{f''(x)}{2} = (x - a)(3c_{3,0} + 2c_{2,1} + c_{1,2}) + (x - b)(c_{2,1} + 2c_{1,2} + 3c_{0,3})$$

so that

$$\begin{aligned}
\frac{f''(a)}{2} &= (a - b)(c_{2,1} + 2c_{1,2} + 3c_{0,3}) \\
&= -\Delta(c_{2,1} + c_{1,2} + c_{1,2} + 3c_{0,3}) \\
&= -\Delta\left(\frac{2s_a + s_b}{\Delta^2} + \frac{3(y_a - y_b)}{\Delta^3}\right) \\
&= -\frac{3(y_a - y_b)}{\Delta^2} - \frac{2s_a + s_b}{\Delta} \\
&= -\frac{3y_a}{\Delta^2} + \frac{3y_b}{\Delta^2} - \frac{2s_a}{\Delta} - \frac{s_b}{\Delta} \\
\frac{f''(b)}{2} &= (b - a)(3c_{3,0} + 2c_{2,1} + c_{1,2}) \\
&= \quad \Delta(c_{1,2} + c_{2,1} + c_{2,1} + 3c_{3,0}) \\
&= \quad \Delta\left(\frac{s_a + 2s_b}{\Delta^2} + \frac{3(y_a - y_b)}{\Delta^3}\right) \\
&= \quad \frac{3(y_a - y_b)}{\Delta^2} + \frac{s_a + 2s_b}{\Delta} \\
&= \quad \frac{3y_a}{\Delta^2} - \frac{3y_b}{\Delta^2} + \frac{s_a}{\Delta} + \frac{2s_b}{\Delta}
\end{aligned}$$

Splining

The mathematical problem is to be solved is this:

We are given a series of intervals

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

and y-values

$$y_0, \quad y_1, \quad \ldots, \quad y_{n-1}, \quad y_n$$

and we want to find a function $f(x)$ with these properties:

(1) *We have $f(x_i) = y_i$ for $i = 0, 1, \ldots, n$.*

That is to say the curve $y = f(x)$ passes through all the points (x_i, y_i).

(2) *On each interval* $[x_i, x_{i+1}]$ *the function* $f(x)$ *is a cubic polynomial.*

In other words, it is *piecewise cubic*.

(3) The cubic polynomials involved may very well be different on different intervals, but *at any intermediate point* x_i *with* $(0 < i < n)$ *the cubic polynomials on the left and right intervals must agree in both value and slope.*

In other words, the function $f(x)$ is to be reasonably smooth, with no obvious breaks.

(4) *Among all functions satisfying the first two conditions,* $f(x)$ *has the least* **bending energy**, which is to say that

$$\int_a^b f''(x)^2\, dx$$

has the least possible value.

Proposition. *There is a unique function* $f(x)$ *satisfying these three conditions, and in addition with specified values of the second derivative at the points* a, b.

The key point in constructing it is that any function $f(x)$ satisfying conditions (1)–(3) will satisfy condition (4) precisely when it satisfies the new condition

(4 *bis*) *At intermediate points* x_i *the polynomials on the left and right intervals must have the same second derivatives.*

In other words, the condition of minimum energy is equivalent to an extra smoothness condition at the intermediate points.

According to the formula at the end of the previous section, condition (4 *bis*) together with an assignment of values for the second derivatives at a and b is equivalent to the system of equations

$$\begin{aligned}
\frac{d_a}{2} &= -\frac{3y_0}{\Delta_{0,1}^2} + \frac{3y_1}{\Delta_{0,1}^2} - \frac{2s_0}{\Delta_{0,1}} - \frac{s_1}{\Delta_{0,1}} \\
\frac{3y_{i-1}}{\Delta_{i-1,i}^2} - \frac{3y_i}{\Delta_{i-1,i}^2} + \frac{s_{i-1}}{\Delta_{i-1,i}} + \frac{2s_i}{\Delta_{i-1,i}} &= -\frac{3y_i}{\Delta_{i,i+1}^2} + \frac{3y_{i+1}}{\Delta_{i,i+1}^2} - \frac{2s_i}{\Delta_{i,i+1}} - \frac{s_{i+1}}{\Delta_{i,i+1}} \\
\frac{d_b}{2} &= \frac{3y_{n-1}}{\Delta_{n-1,n}^2} - \frac{3y_n}{\Delta_{n-1,n}^2} + \frac{s_{n-1}}{\Delta_{n-1,n}} + \frac{2s_n}{\Delta_{n-1,n}}
\end{aligned}$$

or

$$\begin{aligned}
\frac{2s_0 + s_1}{\Delta_{0,1}} &= \frac{3(y_1 - y_0)}{\Delta_{0,1}^2} - \frac{d_a}{2} \\
\frac{s_{i-1} + 2s_i}{\Delta_{i-1,i}} + \frac{2s_i + s_{i+1}}{\Delta_{i,i+1}} &= \frac{3(y_i - y_{i-1})}{\Delta_{i-1,i}^2} + \frac{3(y_{i+1} - y_i)}{\Delta_{i,i+1}^2} \quad (0 < i < n) \\
\frac{s_{n-1} + 2s_n}{\Delta_{n-1,n}} &= \frac{3(y_n - y_{n-1})}{\Delta_{n-1,n}^2} + \frac{d_b}{2}
\end{aligned}$$

which is a system of equations in the unknown slopes s_i, in matrix form $As = t$ with

$$s = \begin{bmatrix} s_0 \\ s_1 \\ \dots \\ s_n \end{bmatrix}$$

$$A = \begin{bmatrix} \frac{2}{\Delta_{0,1}} & \frac{1}{\Delta_{0,1}} & 0 & \ldots & & & \\ \frac{1}{\Delta_{0,1}} & \frac{2}{\Delta_{0,1}} + \frac{2}{\Delta_{1,2}} & \frac{1}{\Delta_{1,2}} & 0 & & \ldots & \\ 0 & \frac{1}{\Delta_{1,2}} & \frac{2}{\Delta_{1,2}} + \frac{2}{\Delta_{2,3}} & \frac{1}{\Delta_{2,3}} & & 0 & \ldots \\ & \ldots & & & & & \\ & \ldots & 0 & \frac{1}{\Delta_{n-2,n-1}} & \frac{2}{\Delta_{n-2,n-1}} + \frac{2}{\Delta_{n-1,n}} & & \frac{1}{\Delta_{n-1,n}} \\ & & \ldots & 0 & \frac{1}{\Delta_{n-1,n}} & & \frac{2}{\Delta_{n-1,n}} \end{bmatrix}$$

$$t = \begin{bmatrix} -\frac{d_a}{2} + \frac{3(y_1 - y_0)}{\Delta_{0,1}^2} \\ \ldots \\ \frac{3(y_i - y_{i-1})}{\Delta_{i-1,i}^2} + \frac{3(y_{i+1} - y_i)}{\Delta_{i,i+1}^2} \\ \ldots \\ \frac{d_b}{2} + \frac{3(y_n - y_{n-1})}{\Delta_{n-1,n}^2} \end{bmatrix}$$

We know from the previous section that the values y_i and the slopes s_i determine the interpolating cubic polynomials completely.

The system has a unique solution. Let x be the column vector which is the transpose of $[\,x_0 \quad x_1 \quad \ldots \quad x_n\,]$, and consider the expression

$$^t x\, A\, x = \frac{2}{\Delta_{0,1}}(x_0^2 + x_0 x_1 + x_1^2) + \frac{2}{\Delta_{1,2}}(x_1^2 + x_2 x_1 + x_2^2) + \frac{2}{\Delta_{n-1,n}}(x_{n-1}^2 + x_{n-1} x_n + x_n^2)$$

Since

$$x^2 + xy + y^2 = (x + y/2)^2 + 3y^2/4$$

it is always positive unless $x = 0$, $y = 0$, and this ensures the same property for the full expression. Therefore if $Ax = 0$, x itself must be 0, and this guarantees that A is non-singular.

Furthermore, solving an equation with A as coefficient matrix involves no row swapping, and requires very little effort because the coefficient matrix is concentrated along its diagonal.

An example

Splining usually does a great job, as long as you space your points closely enough. But if the set of points you want to plot lie on an underlying discontinuous curve, you will get interesting effects. Here is what happens when you spline a step function: