Spline Interpolation

We have seen that an increase in the number of interpolation points (i.e. increasing the degree of the interpolating points) may not lead to better approximation for a function $f(x)$. Let $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ are the interpolating points. A better approach may be to partition the interval $[a, b]$ into small subintervals and approximate $f(x)$ in each subinterval by piece of small degree polynomials. A function $S(x)$ is called a spline of degree k if

1. The domain of S is $[a, b]$
2. $S, S', S'', \cdots, S^{(k-1)}$ are all continuous in $[a, b]$
3. The points x_i's are called knots and S is a polynomial of degree less than or equal to k in each subinterval $[x_i, x_{i+1}]$ for $i = 0, 1, 2, \cdots, n-1$.

We discuss the more widely used spline of degree 3 which is also called cubic spline. Let

$$S(x) = \Big\{ S_i(x), x \in [x_i, x_{i+1}] \Big\}, i = 0, 1, 2, \cdots, n-1,$$

where S_i is a polynomial of degree less equal to 3. We can write S_i as

$$S_i(x) = a_i + b_i x + c_i x^2 + d_i x^3.$$

Thus there are in total $4n$ unknown constants. The interpolation conditions are

$$\begin{aligned}
S(x_i) &= f_i, \quad 0 \le i \le n \\
\lim_{x \to x_i - 0} S(x) &= \lim_{x \to x_i + 0} S(x), \quad 1 \le i \le n-1 \\
\lim_{x \to x_i - 0} S'(x) &= \lim_{x \to x_i + 0} S'(x), \quad 1 \le i \le n-1 \\
\lim_{x \to x_i - 0} S''(x) &= \lim_{x \to x_i + 0} S''(x), \quad 1 \le i \le n-1.
\end{aligned}$$

Thus we have $4n - 2$ conditions. Hence, we need to specify another 2 conditions and that can be done in various ways. Since S'' is continuous, $z_i = S''(x_i)$ for $i = 0, 1, 2, \cdots, n$ are defined and S'' is a linear polynomial in each $[x_i, x_{i+1}]$. Thus

$$S_i''(x) = \frac{x_{i+1} - x}{h_i} z_i + \frac{x - x_i}{h_i} z_{i+1}$$

Integrating this twice, we get

$$S_i(x) = \frac{(x_{i+1} - x)^3}{6h_i} z_i + \frac{(x - x_i)^3}{6h_i} z_{i+1} + rx + s.$$

We adjust constant r and s such that

$$S_i(x) = \frac{(x_{i+1} - x)^3}{6h_i} z_i + \frac{(x - x_i)^3}{6h_i} z_{i+1} + C(x - x_i) + D(x_{i+1} - x),$$

where C and D are again constants. The first two sets of interpolation conditions can be taken care of by $S_i(x_i) = f_i$ and $S_i(x_{i+1}) = f_{i+1}$. Hence, we find

$$S_i(x) = \frac{(x_{i+1} - x)^3}{6h_i} z_i + \frac{(x - x_i)^3}{6h_i} z_{i+1} + \left(\frac{f_{i+1}}{h_i} - \frac{z_{i+1} h_i}{6} \right)(x - x_i) + \left(\frac{f_i}{h_i} - \frac{z_i h_i}{6} \right)(x_{i+1} - x)$$

Hence, the only unknown values are z_i's. To find these, we use continuity condition of $S'(x)$ at $x_1, x_2, \cdots, x_{n-1}$. These conditions are $S_i'(x_i + 0) = S_{i-1}'(x_i - 0)$. This gives

$$h_{i-1}z_{i-1} + 2(h_i + h_{i-1})z_i + h_i z_{i+1} = 6(b_i - b_{i-1}), \quad i = 1, 2, \cdots, n-1,$$

where $b_i = (f_{i+1} - f_i)/h_i$. These are $n-1$ equations in $n+1$ unknowns. We can choose z_0 and z_n (by some choice) and then solve the resulting tridiagonal system to find $z_1, z_2, \cdots, z_{n-1}$. One choice is $z_0 = z_n = 0$ and the resulting spline is called natural cubic spline. The resulting system is symmetric, tridiagonal and diagonally dominant:

$$\begin{bmatrix} d_1 & h_1 & & & & \\ h_1 & d_2 & h_2 & & & \\ & h_2 & d_3 & h_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & h_{n-3} & d_{n-2} & h_{n-2} \\ & & & & h_{n-2} & d_{n-1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{n-2} \\ z_{n-1} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_{n-2} \\ r_{n-1} \end{bmatrix},$$

where $d_i = 2(h_i + h_{i-1})$, $r_i = 6(b_i - b_{i-1})$. This can be solved by Gaussian elimination without scaled pivoting.

Another choice of endpoint conditions are $S'(x_0) = f'(a)$ and $S'(x_n) = f'(b)$, which is also known as clamped conditions. In this case $S'(x_0) = f'(a)$ gives

$$2h_0 z_0 + h_0 z_1 = 6(b_0 - f'(a))$$

and $S'(x_n) = f'(b)$ gives

$$h_{n-1}z_{n-1} + 2h_{n-1}z_n = 6(f'(b) - b_{n-1})$$

Thus the system can be writeen as $n+1$ equations in $(n+1)$ unknowns as

$$\begin{bmatrix} d_0 & h_0 & & & & \\ h_0 & d_1 & h_1 & & & \\ & h_2 & d_3 & h_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & h_{n-2} & d_{n-1} & h_{n-1} \\ & & & & h_{n-1} & d_n \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ z_n \end{bmatrix} = \begin{bmatrix} r_0 \\ r_1 \\ r_2 \\ \vdots \\ r_{n-1} \\ r_n \end{bmatrix},$$

where $d_i = 2(h_i + h_{i-1})$, $r_i = 6(b_i - b_{i-1})$ for $1 \le i \le n-1$ and $d_0 = 2h_0$, $d_n = 2h_{n-1}$, $r_0 = 6(b_0 - f'(a))$ and $r_n = 6(f'(b) - b_{n-1})$.

Alternatively, we may also eliminate z_0 and z_n and obtain $n-1$ equations in $n-1$ unknowns. This gives the system

$$\begin{bmatrix} e_1 & h_1 & & & & \\ h_1 & e_2 & h_2 & & & \\ & h_2 & e_3 & h_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & h_{n-3} & e_{n-2} & h_{n-2} \\ & & & & h_{n-2} & e_{n-1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{n-2} \\ z_{n-1} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_{n-2} \\ s_{n-1} \end{bmatrix},$$

where $e_i = d_i$, $s_i = r_i$ (defined earlier) for $2 \le i \le n-2$ and $e_1 = 3h_0/2 + 2h_1$, $e_{n-1} = 2h_{n-2} + 3h_{n-1}/2$, $s_1 = r_1 - 3(b_0 - f'(a))$ and $s_{n-1} = r_{n-1} - 3(f'(b) - b_{n-1})$

Next we prove that a theorem on the optimality of natural cubic spline. Let f'' be continuous in $[a, b]$ and $a = x_0 < x_1 < x_2 < \cdots < x_n$ be the knots. If S is the natural cubic spline, then

$$\int_a^b \Big(S''(x)\Big)^2 dx \le \int_a^b \Big(f''(x)\Big)^2 dx$$

This result is also true for spline with clamped conditions. To prove this, let $g = f - S$. Clearly, $g(x_i) = 0$ for $i = 0, 1, 2, \cdots, n$. Then

$$\int_a^b \Big(f''(x)\Big)^2 dx = \int_a^b \Big(S''(x)\Big)^2 dx + \int_a^b \Big(g''(x)\Big)^2 dx + 2\int_a^b \Big(S''(x)g''(x)\Big)\, dx$$

We need to show that the last integral in the RHS is greater or equal to zero. Now, using integration by parts, we find (Note that S''' is a constant, c_i say, in $[x_{i-1}, x_i]$)

$$\begin{aligned}
\int_a^b \Big(S''(x)g''(x)\Big)\, dx &= \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} S''g''\, dx \\
&= \sum_{i=1}^{n} \left\{ (S''g')|_{x_i} - (S''g')|_{x_{i-1}} - \int_{x_{i-1}}^{x_i} S'''g'\, dx \right\} \\
&= (S''g')|_b - (S''g')|_a - \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} S'''g'\, dx \\
&= (S''g')|_b - (S''g')|_a - \sum_{i=1}^{n} c_i \int_{x_{i-1}}^{x_i} g'\, dx \\
&= S''(b)[f'(b) - S'(b)] - S''(a)[f'(a) - S'(a)] - \sum_{i=1}^{n} c_i[g(x_i) - g(x_{i-1})] \\
&= S''(b)[f'(b) - S'(b)] - S''(a)[f'(a) - S'(a)]
\end{aligned}$$

In case of natural spline, $S''(b) = S''(a) = 0$ and for the clamped boundaries, $f'(b) - S'(b)$ and $S'(a) = f'(a)$. Hence, for both the natural and clamped boundaries, the required integral is zero.

Note that the curvature of a curve described by $y = f(x)$ is the quantity

$$\kappa = \frac{|f''(x)|}{[1 + \{f'(x)\}^2]^{3/2}}$$

If the nonlinear term is dropped, then $|f''(x)|$ is a measure of approximate curvature. So the natural cubic spline or the cubic spline with clamped boundary conditions has minimal curvature among all functions having continuous second derivative and passing through the knots.

Theorem (no proof): If $f(x)$ is four times continuously differentiable and S is a cubic spline, then for $x \in [a, b]$

$$|f(x) - S(x)| \le \frac{5}{384} h^4 \max_{x \in [a,b]} |f^{(4)}(x)|,$$

where $h = \max_{0 \le i \le n-1} h_i$. It can also be shown that for $x \in [a, b]$

$$|f'(x) - S'(x)| \le \frac{1}{24} h^3 \max_{x \in [a,b]} |f^{(4)}(x)|,$$

This process can be continued upto 3rd derivative and the power of h decreases by one in each step.