The Cubic Spline

James Emery

Version: 11/10/2009

Contents

1 Introduction

We shall compute the interpolating cubic spline curve that passes through a supplied set of interpolation points. A cubic spline function is a piecewise polynomial curve that has continuous derivatives to order two. A cubic spline curve is a parametric curve that has coordinate functions that are cubic spline functions. Hence a cubic spline curve has continuous curvature and tangent. The cubic spline may be determined by solving a system of linear tridiagonal equations. We shall derive these equations. The name spline refers to a thin length of wood used for drawing smooth curves. The thin beam takes the shape of a cubic spline. We shall derive the thin beam equation. The cubic spline has a minimum energy property. We shall give the classic proof.

A cubic spline, being a piecewise polynomial, may be represented using various piecewise polynomial basis functions. Some of these are (1) The power bases, (2) The Bezier bases (Bernstein basis), and (3) The B-spline basis. We shall not discuss these representations here.

2 The Thin Beam Equation

Let ρ be the radius of curvature of the neutral axis of the thin beam. Let y be the distance from the neutral axis to a point in the beam cross section. When the beam is bent, a length ℓ changes to a length $\ell + \delta\ell$. The length at the neutral axis does not change. The length ℓ on the neutral axis, and the radius of curvature ρ, are lengths of sides of a triangle. The distance from the neutral axis y and the length change $\delta\ell$, when the deflection is small, are sides of a similar triangle.

So using similar triangles, we find

$$\frac{\delta\ell}{y} = \frac{\ell}{\rho}.$$

The strain ε is the change in length divided by the length and so

$$\varepsilon = \frac{\delta\ell}{\ell} = \frac{y}{\rho}.$$

The stress is the force per unit cross sectional area and for linear materials is given by Hooke's Law

$$\sigma = E\frac{y}{\rho}.$$

The net normal force is zero over the cross section, so integrating the stress over the cross sectional area we get

$$0 = \int \sigma dA = \frac{E}{\rho}\int y dA.$$

Now this integral is the area centroid, so we see that the neutral axis passes through the area centroid.

The bending moment is given by

$$M = \int y\sigma dA = \frac{E}{\rho}\int y^2 dA = \frac{E}{\rho}I.$$

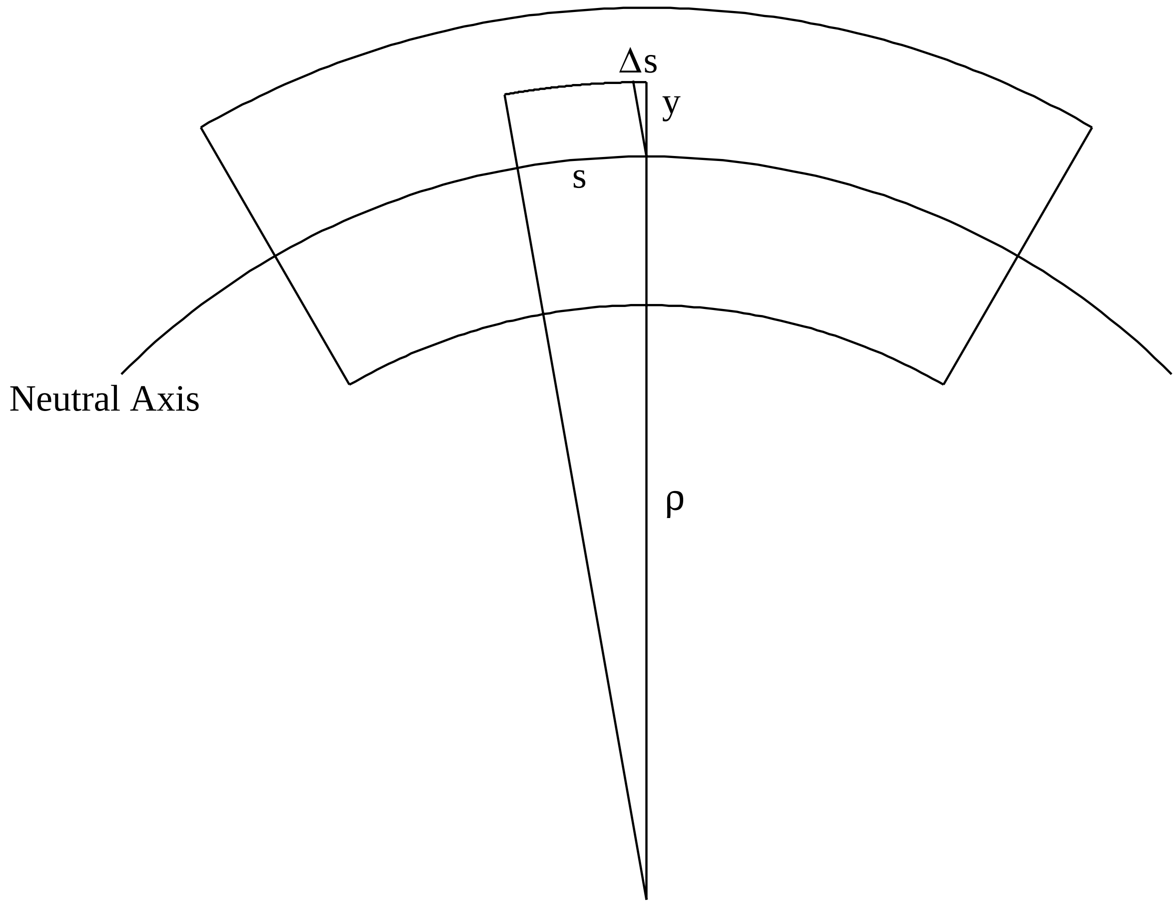

Figure 1: **Bending Beam.** *The neutral axis of the beam has radius of curvature* ρ*. A length* s *on the neutral axis is stretched to length* $s + \Delta s$ *at a height* y *above the neutral axis. The triangle with sides* ρ *and* s *is similar to the triangle with sides* y *and* Δs*. So the strain at distance* y *from the neutral axis is* $\frac{\Delta s}{s} = \frac{y}{\rho}$.

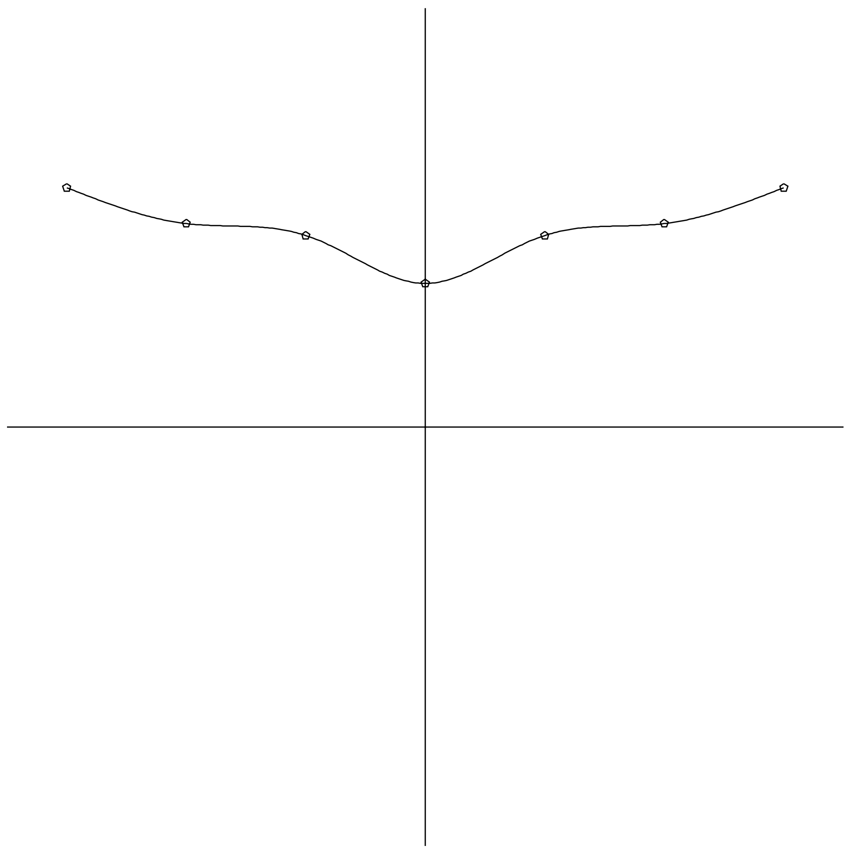

Figure 2: **A Parametric Two Dimensional Cubic Spline.** Here the interpolation points through which the spline is to pass are $(-3,2),(-2,1.7),(-1,1.6),(0,1.2),(1,1.6),(2,1.7),(3,2)$. The natural end conditions are used, which models a zero bending moment at the spline ends. This approximates a physical spline, which is a thin piece of wood or plastic, that is forced to pass through interpolation points, using weights called "ducks" that produce forces at the interpolation points, and thus produce bending moments at these points. This figure was generated using the program **cspln.ftn .**

where I is the area moment of inertia. The curvature is the reciprocal of the radius of curvature and we have

$$\frac{1}{\rho} = \kappa = \frac{d^2y/dx^2}{(1+(dy/dx)^2)^{3/2}}.$$

When the slope is small the curvature is approximately equal to the second derivative, hence the beam equation is

$$\frac{d^2y}{dx^2} = \kappa = \frac{M}{EI}.$$

When the beam is subjected to point loads , the bending moment is linear between the point loads, hence on each segment of the beam we have an equation of the form

$$\frac{d^2y}{dx^2} = \frac{ax+b}{EI},$$

where a and b are constants depending on the point loads. This integrates to a cubic polynomial. Hence the solution deflection curve is a piecewise cubic polynomial with continuous first and second derivatives, and zero second derivatives at the ends. The deflection curve is a natural cubic spline. A natural cubic spline is one which has zero second derivatives at the ends and so becomes straight at the ends.

3 Derivation of The Equations

First we shall derive the equation of a cubic polynomial in terms of the function values y_a and y_b and the second derivative values m_a and m_b taken at a and b. Then we will find the equations that define a cubic spline. The spline calculations, which we shall define, are implemented in the Fortran program **cspln.ftn**.

Since the function is a cubic, its second derivative is linear and is defined by the second derivative values at the interval endpoints a and b. Thus

$$f'' = (x-a)m_b/h - (x-b)m_a/h,$$

where $h = b - a$. Integrating twice we find

$$f = (x-a)^3 m_b/6h - (x-b)^3 m_a/6h + dx + e$$

where d and e are constants of integration. Differentiating we get

$$f' = d - \frac{ma(x-b)^2}{2h} + \frac{m_b(x-a)^2}{2h}.$$

Then

$$f'(a) = d - \frac{m_a h}{2},$$

and

$$f'(b) = d - \frac{m_b h}{2}.$$

Using the supplied values at a and b we get two equations

$$e + ad + \frac{m_a h^2}{6} = y_a,$$

and

$$e + bd + \frac{m_b h^2}{6} = y_b.$$

Solving these two equations we find

$$d = \frac{y_b - y_a + m_a h^2/6 - m_b h^2/6}{h},$$

and

$$e = \frac{b(y_a - m_a h^2/6) - a(y_b - m_b h^2/6)}{h}.$$

Simplifying

$$d = \frac{y_b}{h} - \frac{y_a}{h} + \frac{m_a h}{6} - \frac{m_b h}{6}.$$

Substituting this expression for d we find

$$f'(a) = \frac{y_b}{h} - \frac{y_a}{h} - \frac{m_a h}{3} - \frac{m_b h}{6}.$$

Similarly

$$f'(b) = \frac{y_b}{h} - \frac{y_a}{h} + \frac{m_a h}{6} - \frac{m_b h}{3}.$$

The cubic may be conveniently stored and evaluated in Hermite form. The Hermite form is

$$f(x) = f(a)(1 - 3u^2 + 2u^3) + f(b)(3u^2 - 2u^3) +$$

$$f'(a)(u - 2u^2 + u^3)h + f'(b)(-u^2 + u^3)h,$$

where $u = (x - a)/h$. This representation, for the case of a c_1 piecewise cubic, has the advantage of reduced storage. The storage is $3n$ for an n point piecewise cubic (n x values, n y values, and n derivative values). On the other hand, if the four coefficients of each cubic are retained, then the storage needed is $5n - 4$ (four coefficients for the $n - 1$ segments and n x values). The disadvantage of this representation is that more computation is needed for evaluation of the spline.

Let us compute the four power coefficients. We write

$$f(x) = A + Bu + Cu^2 + Du^3,$$

for $a \leq x \leq b$. Then with $h = b - a$,

$$A = f(a),$$

$$B = f'(a)h,$$

$$C = -3f(a) + 3f(b) - 2f'(a)h - f'(b)h,$$

$$D = 2f(a) - 2f(b) + f'(a)h + f'(b)h.$$

If we let

$$s = x - a = uh,$$

then

$$f(x) = A + Bs/h + Cs^2/h^2 + Ds^3/h^3 = A' + B's + C's^2 + D's^3,$$

where

$$A' = f(a),$$

$$B' = f'(a),$$

$$C' = 3(f(b) - f(a))/h^2 - (2f'(a) + f'(b))/h,$$

$$D' = 2(f(a) - f(b)/h^3 + (f'(a) + f'(b))/h^2.$$

This last representation is the IGES (Initial Graphics Exchange Specification) representation for a parametric cubic.

A cubic spline is a piecewise cubic that matches derivative values at the knots (up to and including the second derivative). In order to get the equations for the spline, let us introduce a second cubic defined on the interval from b to c. We obtain the equations for the derivatives by making the following substitution

$$y_b \to y_c, b \to c, m_b \to m_c, h \to k,$$

followed by

$$y_a \to y_b, a \to b, m_a \to m_b.$$

We equate the left derivative to the right derivative at b, and obtain the following equation

$$\frac{m_a h}{6} + \frac{m_b h}{3} + \frac{m_b k}{3} + \frac{m_c k}{6} = \frac{y_a - y_b}{h} + \frac{y_c - y_b}{k}.$$

When we have several cubic pieces joined at $n+1$ knots, $x_0, x_1, ..., x_n$, the corresponding equation at the ith knot is obtained by making the following substitutions:

$$a = x_{i-1}, b = x_i, c = x_{i+1},$$

$$y_a = y_{i-1}, y_b = y_i, y_c = y_{i+1}.$$

$$m_a = m_{i-1}, m_b = m_i, m_c = m_{i+1},$$

$$h = h_i, k = h_{i+1}.$$

At the jth knot we have the equation

$$h_j m_{j-1} + 2(h_j + h_{j+1}) m_j + h_{j+1} m_{j+1} =$$

$$6[(y_{j+1} - y_j)/h_{j+1} - (y_j - y_{j-1})/h_j].$$

We shall introduce coefficient vectors

$$\{a_j\}, \{b_j\}, \{c_j\}, \{d_j\}.$$

We divide the last equation by $h_j + h_{j+1}$ and get

$$a_j m_{j-1} + b_j m_j + c_j m_{j+1} = d_j$$

where

$$a_j = h_j/(h_j + h_{j+1}),$$

$$b_j = 2,$$

$$c_j = (h_{j+1} + h_j - h_j)/(h_{j+1} + h_j) = 1 - a(j),$$

and

$$d_j = 6[(y_{j+1} - y_j)/h_{j+1} - (y_j - y_{j-1})/h_j].$$

We have an equation at each internal knot and so $n-1$ equations for the $n+1$ values

$$m_i, i = 0, ..., n$$

of the second derivative. We need two more conditions. We get what is called a natural cubic spline by taking the two additional equations to be $m_0 = 0$ and $m_n = 0$. The natural physical spline has no bending moments at the ends and so zero end curvature. This is the origin of "natural." We may specify the end tangents and get the equations

$$f'(x_0) = (y_1 - y_0)/h_1 - h_1(m_0/3 + m_1/6)$$

and

$$f(x_n) = (y_n - y_{n-1})/h_{n-1} + h_{n-1}(m_{n-1}/6 + m_n/3).$$

A third end condition is defined by the two equations $m_0 = m_1$ and $m_{n-1} = m_n$. This condition causes the first and last cubic to be quadratic. Lastly if $y_0 = y_n$, we can make the spline periodic. The first and second derivatives are to agree at x_0 and x_n. The two additional equations are $m_0 = m_n$ and

$$(h_{n-1}/6)m_{n-1} + (h_{n-1}/3 + h_1/3)m_0 + h_1 m_1/6 = (y_1 - y_2)/h_1 + (y_{n-1} - y_n)/h_{n-1}.$$

The first three end conditions are special cases of the following two equations

$$2m_0 + c_0 m_1 = d_0$$

and

$$a_n m_{n-1} + 2m_n = d_n.$$

If we use this general boundary formulation, the system is tridiagonal of the form

$$a_j m_{j-1} + b_j m_j + c_j m_{j+1} = d_j, (j = 0, ..., n),$$

where each $b_j = 2, a_0 = 0$, and $c_n = 0$. The periodic system does not fit into this scheme because it is not tridiagonal. However it can be treated in a like manner. We can solve the tridiagonal system in the following way. We convert each equation into an equivalent equation, which has the form

$$m_i = q_i m_{i+1} + r_i,$$

by doing a forward substitution. The first equation can be written as

$$m_0 = -c_0 m_1/2 + d_0/2.$$

Hence $q_0 = -c_0/2$ and $r_0 = d_0/2$. Suppose that the kth equation has been put in the form

$$m_{k-1} = q_{k-1} m_k + r_{k-1}.$$

We substitute this value for m_{k-1} in the original $k+1$ equation and get

$$(b_k + a_k q_{k-1}) m_k = -c_k m_{k+1} + d_k - a_k r_{k-1}.$$

Thus $m_k = q_k m_{k+1} + r_k$, where $q_k = -c_k/(b_k + a_k q_{k-1})$ and

$$r_k = (d_k - a_k r_{k-1})/(b_k + a_k q_{k-1}).$$

Notice that $c_n = 0$ and thus $x_n = r_n$. Now we may do backwards substitution to solve for each m_k using

$$m_k = q_k m_{k+1} + r_k.$$

4 Bezier Representation

The value of the spline function at parameter s,on the ith segment, using the Hermite form of the cubic, is

$$f(s) = f(s_i)(1 - 3u^2 + 2u^3) + f(s_{i+1})(3u^2 - 2u^3) +$$

$$f'(s_i)(u - 2u^2 + u^3)h + f'(s_{i+1})(-u^2 + u^3)h,$$

where $u = (s - s_i)/h$, and $h = s_{i+1} - s_i$. Writing this in power form

$$f(s) = a_0 + a_1 u + a_2 u^2 + a_3 u^3,$$

where

$$a_0 = f(s_i),$$

$$a_1 = f'(s_i)h,$$

$$a_2 = 3f(s_{i+1}) - (2f'(s_i)h + f'(s_{i+1})h + 3f(s_i)),$$

$$a_3 = 2f(s_i) + f'(s_i)h - 2f(s_{i+1}) + f'(s_{i+1})h.$$

The Bernstein coefficients are

$$c_i = \sum_{j=0}^{3} A_{ij} a_j,$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1/3 & 0 & 0 \\ 1 & 2/3 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Hence

$$c_0 = a_0 = f(s_i),$$

$$c_1 = a_0 + a_1/3,$$

$$c_2 = a_0 + 2a_1/3 + a_2/3,$$

$$c_3 = a_0 + a_1 + a_2 + a_3 = f(s_{i+1}).$$

Then

$$f(s) = c_0 b_0^3(u) + c_1 b_1^3(u) + c_2 b_2^3(u) + c_3 b_3^3(u),$$

where b_i^3 is the ith cubic Bernstein polynomial.

5 The Minimum Property of Cubic Splines

The cubic spline interpolate minimizes the following energy integral.

$$\int \left(\frac{d^2 f}{dx^2} \right)^2 dx.$$

For suppose a second function g has the same end derivatives and interpolates the same points as the cubic spline function f. We write

$$g = f + (g - f).$$

We have

$$\int \left(\frac{d^2 g}{dx^2} \right)^2 dx = \int \left(\frac{d^2 f}{dx^2} \right)^2 dx + 2 \int \frac{d^2 f}{dx^2} \left(\frac{d^2 g}{dx^2} - \frac{d^2 f}{dx^2} \right) dx + \int \left(\frac{d^2 g}{dx^2} - \frac{d^2 f}{dx^2} \right)^2 dx.$$

Evaluating the second integral on the left by parts, we have

$$\int \frac{d^2f}{dx^2}\left(\frac{d^2g}{dx^2}-\frac{d^2f}{dx^2}\right)dx =$$

$$\left[\frac{d^2f}{dx^2}\left(\frac{dg}{dx}-\frac{df}{dx}\right)\right]_{x_1}^{x_n} - \int \frac{d^3f}{dx^3}\left(\frac{dg}{dx}-\frac{df}{dx}\right)dx.$$

The first term is zero because the derivatives of f and g agree at the end points t_1 and t_n. The third derivative of f is constant on each interpolation interval. Hence on each subinterval the integral is a constant times the integral of the difference of the f and g derivatives. Carrying out the integration we get $f - g$ evaluated at the two interpolation endpoints. These two functions by hypothesis agree at the interpolation points , so the integral is zero. Hence the energy of the cubic spline f equals the energy of the g plus a term that is greater than or equal to 0. Hence the energy of f is less than g. So f gives the minimum energy.

Note. The application of integration by parts above appears to be invalid because d^2f/dx^2 is not differentiable. However, the integral can be broken up into a sum of integrals over each interpolation interval and the first term on the right becomes

$$\sum_{i=1}^{n-1}\left[\frac{d^2f}{dx^2}\left(\frac{dg}{dx}-\frac{df}{dx}\right)\right]_{x_i}^{x_{i+1}}.$$

Because d^2f/dx^2 is continuous, this sum telescopes and equals

$$\left[\frac{d^2f}{dx^2}\left(\frac{dg}{dx}-\frac{df}{dx}\right)\right]_{x_1}^{x_n}.$$