Parametric Cubic Spline Tutorial

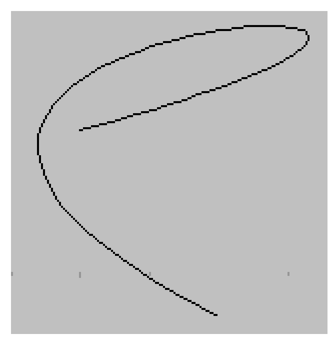
Parametric equations are powerful and flexible. Perhaps the most familiar example is the equation of a circle in the form $x = r*\cos(\theta)$, $y = r*\sin(\theta)$. In this case, the parameter θ is the independent variable and increases monotonically (i.e. each successive value is larger than the previous one, which is a requirement of parametric equations in general), and x and y are dependent variables. The resulting plot of y vs x can wrap on itself, which a circle indeed does. There are many forms of polynomial parametric equations (e.g. Bezier, B-Spline, NURBS, Catmull-Rom) that have entire books devoted to their understanding and use. In this tutorial, we offer a brief introduction by way of examples to provide a basic understanding of the relative simplicity of parametric equations. Specifically, we will build upon the tools learned in the cubic spline tutorial to extend our knowledge to parametric cubic splines.

Recall that a cubic spline is nothing more than a sequence of 3^{rd} order polynomials joined at the endpoints with enforced 1^{st} and 2^{nd} derivative compatibility at interior points and specified boundary conditions at the end points. Extending this concept to parametric splines just means formulating two sets of equations instead of one using the exact same methodology as a standard (non-parametric) cubic spline. In the case of parametric cubic splines, each spline segment is represented by two equations in the independent variable s:

$$\mathbf{x = f_1(s) = a_x(s\text{-}s_0)^3 + b_x(s\text{-}s_0)^2 + c_x(s\text{-}s_0) + d_x \qquad x = a_xt^3 + b_xt^2 + c_xt + d_x \qquad t = s\text{-}s_0}$$
$$\mathbf{y = f_2(s) = a_y(s\text{-}s_0)^3 + b_y(s\text{-}s_0)^2 + c_y(s\text{-}s_0) + d_y \qquad y = a_yt^3 + b_yt^2 + c_yt + d_y \qquad t = s\text{-}s_0}$$

where s_0 represents the value of the independent variable s at the beginning of the segment. For convenience, we've made a variable substitution $t = s\text{-}s_0$. Though not required, it is customary to have t vary from 0 to 1. To get an idea of how powerful these equations are, consider the following simple equations for x, y:

$$\boldsymbol{x = 26t^3 - 40t^2 + 15t - 1}$$
$$\boldsymbol{y = -4t^2 + 3t}$$



The resulting plot of y vs x (for t = 0 to 1) is truly amazing. It almost seems magical that a single cubic polynomial (well, two actually) can generate such as complicated shape. That's the power of parametric equations.

To gain insight into the reason for additional flexibility of parametric curves, it is instructive to think in terms of constraints. Since there are two cubic equations, there are a total of eight polynomial coefficients, or alternatively, eight degrees of freedom or constraints that can be specified. Considering the case where only two points are specified (i.e. the end points), four of the constraints (i.e. x,y for each point) are used, leaving four constraints to choose. By comparison, standard (non-parametric) cubic polynomials have only two constraints to be chosen. Thus, the added flexibility of parametric cubic curves can be thought of as two extra degrees of freedom versus the standard cubic polynomial. Does this mean that the 1^{st} derivative (i.e. dy/dx) and 2^{nd} derivative (d^2y/dx^2) at each end can be specified? The answer is a bit complicated and will be addressed later.

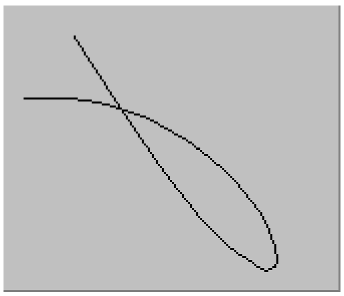
It is instructive to consider a single parametric cubic polynomial before addressing the task of connecting them together. We'll start by constructing a curve by specifying two points plus 2^{nd} derivative boundary conditions as was done in the (non-parametric) cubic spline tutorial. Following are the equations that will be used:

$$\begin{bmatrix} 6t_1 & 2 & 0 & 0 \\ t_1^3 & t_1^2 & t_1 & 1 \\ t_2^3 & t_2^2 & t_2 & 1 \\ 6t_2 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_x \\ b_x \\ c_x \\ d_x \end{bmatrix} = \begin{bmatrix} x_1'' \\ x_1 \\ x_2 \\ x_2'' \end{bmatrix} \qquad \begin{bmatrix} 6t_1 & 2 & 0 & 0 \\ t_1^3 & t_1^2 & t_1 & 1 \\ t_2^3 & t_2^2 & t_2 & 1 \\ 6t_2 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_y \\ b_y \\ c_y \\ d_y \end{bmatrix} = \begin{bmatrix} y_1'' \\ y_1 \\ y_2 \\ y_2'' \end{bmatrix}$$

Note that there are two equations instead of one (one for x and another for y) using t as the independent variable. The important point is that the structure is identical to that of the standard (non-parametric) cubic polynomial.

Let's pick $p_1 = (x_1,y_1) = (0,0)$ and $p_2 = (x_2,y_2) = (1,1)$. Before continuing, the values of the independent variable need to be determined. Actually, it is rather arbitrary how the values are chosen. For this example we will pick $t_1 = 0$ to correspond to p_1 and $t_2 = 1$ to correspond to p_2. Next, the boundary conditions must be chosen. As was the case with standard (non-parametric) cubic polynomials, the choice can be arbitrary. We will choose $[x_1'', x_2'', y_1'', y_2''] = [-30, -45, -45, 90]$. Given that we've chosen to specify the same boundary conditions (i.e. 2nd derivatives) for both x and y, the base matrix is identical for both. Had we chosen to specify, for example, 1st derivatives for x and 2nd derivatives for y, the base matrix would be different for each. In this case, however, we can represent the equations in more compact matrix form on the right. Solving the equation for the polynomial coefficients: $[a_x, b_x, c_x, d_x] = [-2.5, -15, 18.5, 0]$ and $[a_y, b_y, c_y, d_y] = [22.5, -22.5, 1, 0]$. The resulting plot is on the left.

$$\begin{bmatrix} . & 2 & . & . \\ . & . & . & 1 \\ 1 & 1 & 1 & 1 \\ 6 & 2 & . & . \end{bmatrix} \begin{bmatrix} a_x & a_y \\ b_x & b_y \\ c_x & c_y \\ d_x & d_y \end{bmatrix} = \begin{bmatrix} -30 & -45 \\ 0 & 0 \\ 1 & 1 \\ -45 & 90 \end{bmatrix}$$



It is interesting to note that generating points to plot the curve (i.e. evaluating the parametric curve) can be represented rather compactly in matrix form. Suppose we wish to generate n points (p_1 to p_n) for plotting purposes. The matrix equation is:

$$\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_n \end{bmatrix} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_{n-1} & y_{n-1} \\ x_n & y_n \end{bmatrix} = \begin{bmatrix} t_1^3 & t_1^2 & t_1 & 1 \\ t_2^3 & t_2^2 & t_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ t_{n-1}^3 & t_{n-1}^2 & t_{n-1} & 1 \\ t_n^3 & t_n^2 & t_n & 1 \end{bmatrix} \begin{bmatrix} a_x & a_y \\ b_x & b_y \\ c_x & c_y \\ d_x & d_y \end{bmatrix}$$

This simple demonstration shows that the creation of a parametric cubic polynomial can be no more complicated than creating two standard cubic polynomials.

However, it is useful to look more closely at boundary conditions. In the case of standard (non-parametric) cubic polynomials, the boundary conditions (1st and 2nd derivatives) had a simple geometric interpretation in relation to the curve. The 1st derivative dy/dx is the slope and the 2nd derivative d^2y/dx^2 approximates the curvature. In the case of parametric equations, the interpretation of the derivatives is more complicated. Consider the 1st and 2nd derivatives of x and y in the independent variable t:

$x' = dx/dt = 3a_xt^2 + 2b_xt + c_x$ $\qquad$ $x'' = d^2x/dt^2 = 6a_xt + 2b_x$

$y' = dy/dt = 3a_yt^2 + 2b_yt + c_y$ $\qquad$ $y'' = d^2y/dt^2 = 6a_yt + 2b_y$

Analogous to the standard (non-parametric) cubic polynomial, we can formulate the equations in terms of specifying two points plus, in this case, four derivatives, which can be some combination of 1st or 2nd derivatives at either end point (like was done in the simple demonstration above). The question is, how do x', y', x'', y'' relate to the derivatives of the actual curve, in other words, dy/dx and d^2y/dx^2? The answer to this question requires application of the derivative chain and product rules:

dy/dx = (dy/dt)(dt/dx). But, dy/dt = y' and dx/dt = x', so dy/dx = y'/x'

Defining h(t) = dy/dt = y' and g(t) = dt/dx = 1/(dx/dt) = 1/x' = $(x')^{-1}$, we get dy/dx = h(t)g(t) = hg

Therefore, d^2y/dx^2 = d(dy/dx)/dx = d(hg)/dx = g(dh/dx) + h(dg/dx) = g(dh/dt)(dt/dx) + h(dg/dt)(dt/dx)

$dh/dt = d(dy/dt)/dt = d^2y/dt^2 = y''$

$dg/dt = d[(x')^{-1}]/dt = (-1)(x')^{-2}dx'/dt = (-1)(x')^{-2}x'' = -x''/(x')^2$

$d^2y/dx^2 = gy''/x' - h[x''/(x')^2](1/x') = y''/(x')^2 - y'x''/(x')^3 = (x'y'' - y'x'')/(x')^3$

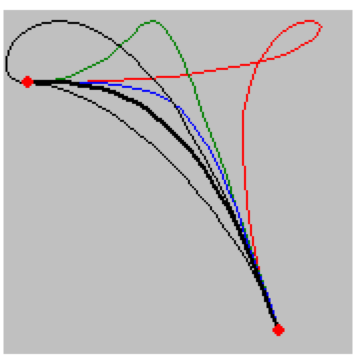
Thus, the 1st derivative of the parametric curve (i.e. dy/dx) is a simple ratio of the 1st derivatives of each dependent variable with respect to the independent variable t. This illustrates some extremely important points:

- A horizontal slope (i.e. zero slope) can be represented by setting $y' = 0$
- A vertical slope (i.e. infinite slope) can be represented by setting $x' = 0$. Standard (non-parametric) cubic polynomials can't have vertical slopes.
- A fixed slope requires only that the ratio of y'/x' be constant, meaning that either can still be freely chosen.

The 2nd derivative of the parametric curve (i.e. d^2y/dx^2) is a complicated non-linear equation involving x', y', x'', y''. What does this really mean?

- Natural boundary conditions can be specified by setting $y'' = x'' = 0$ or $y'' = y' = 0$.
- Specifying x' and y' at both points explicitly sets dy/dx at both points. The extra freedom to additionally choose the specific value of x' or y' indirectly determines d^2y/dx^2 at both points.
- Specifying y'' and x'' at both points indirectly determines dy/dx and d^2y/dx^2 at both points. The only exception is $y'' = x'' = 0$ as already mentioned.
- d^2y/dx^2 can only be explicitly set (other than zero and degenerating to non-parametric form) by iterative means since only four of eight possible boundary conditions (x', y', x'', y'' at the 1st point and x', y', x'', y'' at the 2nd point) can be specified.

The extra degrees of freedom that parametric cubic polynomials enjoy versus their non-parametric cousins is embodied in the fact that (1) four constraints need to be chosen, and (2) dy/dx is specified by the ratio y'/x', leaving total freedom to choose the value of either y' or x'. An example will clearly illustrate. Create a standard cubic polynomial with $p_1 = (1,0)$, $p_2 = (0,1)$, $y_1' = 0$ (i.e. zero slope = horizontal), and $y_2' = -3$. This plot is shown in the bold black line in the plot. There is only one possible curve that meets the criteria since only two constraints can be specified. However, in the case of a parametric cubic polynomial, there are infinitely many curves that meet the criteria, because all that's needed is $y_1' = 0$ to satisfy $dy/dx = 0$ at p_1 and $y_2' = -3x_2'$ to satisfy $dy/dx = -3$ at p_2. This means x_1' and x_2' can be set to any desired values. The plot shows just a few of the possibilities. The choices of x_1' and x_2' indirectly determine dy^2/dx^2 and therefore the overall shape of the curve.



An interesting exercise is to find the right x_1' and x_2' to force the parametric cubic polynomial to be identical to the non-parametric version. Actually, it is quite easy. To force a parametric polynomial to degenerate to its non-parametric equivalent, all that is needed is to force the polynomial coefficients $a = b = 0$ for $x = f(t)$. A bit of algebra reveals that this condition is met if $x_1' = x_2' = (x_2 - x_1) / (t_2 - t_1)$.

To further illustrate some of the above points, an example of fitting a quarter circle is given.

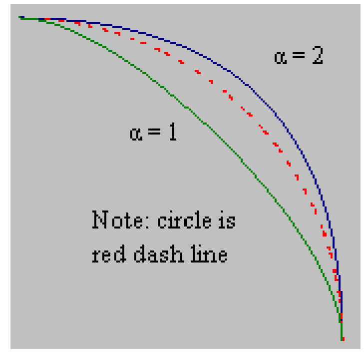
Example Problem #1 (Quarter Circle)

$$\begin{bmatrix} 3x_1^2 & 2x_1 & 1 & 0 \\ x_1^3 & x_1^2 & x_1 & 1 \\ x_2^3 & x_2^2 & x_2 & 1 \\ 3x_2^2 & 2x_2 & 1 & 0 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} p_1' \\ p_1 \\ p_2 \\ p_2' \end{bmatrix}$$

Suppose it is desired to fit a single cubic polynomial to a quarter circle. The end points will be chosen as $p_1 = (1,0)$ and $p_2 = (0,1)$. Since the slope is vertical (i.e. infinite) at p_1 and horizontal (i.e. zero) at p_2, a good choice of formulations is one specifying two points and two slopes. From the above discussion, it is clear that $x_1' = 0$ will result in vertical slope at p_1 and $y_2' = 0$ will result in horizontal slope at p_2.

$$\begin{bmatrix} . & . & 1 & . \\ . & . & . & 1 \\ 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & . \end{bmatrix}\begin{bmatrix} a_x & a_y \\ b_x & b_y \\ c_x & c_y \\ d_x & d_y \end{bmatrix} = \begin{bmatrix} 0 & y_1' \\ 1 & 0 \\ 0 & 1 \\ x_2' & 0 \end{bmatrix}$$

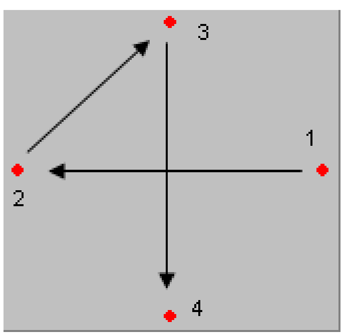
But, that means y_1' and x_2' need to be chosen (equation to the left). By simple trial and error, it becomes evident that $y_1' = - x_2' = \alpha$ and that $1 < \alpha < 2$. This illustrates the indirect influence of the boundary conditions. We can set the slope explicitly, but only indirectly the 2nd derivative (or curvature).



In this example, by choosing some criteria and employing an iterative solving process, it can determined that the ideal value of α is approximately 1.657. This was done by using the criteria that the midpoint of the curve (i.e. t = 0.5) have a length of 1. The result is a curve that is visually indistinguishable from a quarter circle. Actually, there is an analytic way to determine the value of α using Bezier formulation, but that's beyond the scope of this tutorial.

Now that we have explored the characteristics of single parametric cubic polynomials, let's join several together to form a shape. The methodology for determining the coefficients of the polynomials is exactly the same as the non-parametric case. The trick is to determine the values of the parameter, s. (We regress to the expressions that use $s\text{-}s_0$ instead of t). One method is to use distance along the curve (arc length), or an approximation thereof. An example will illustrate.

Example Problem #2 (fit parametric cubic splines to a set of points)



Consider the following four points: [(1,0), (0,1), (-1,0), (0,-1)]. There are several ways the points can be connected to form a shape. We wish to connect them in a crossing fashion as illustrated in the picture.

The values of the parameter s will be formed by summing the distances between the points, $(\Delta x^2 + \Delta y^2)^{1/2}$. The key is to sum the distances in the sequence that the curve is to follow. In this case:

$p_1 = [1,0]$ $\quad$ $p_2 = [-1, 0]$ $\quad$ $p_3 = [0, 1]$ $\quad$ $p_4 = [0, -1]$

Calculating the distances between points:

$d_{12} = \text{sqrt}[(-1 - 1)^2 + (0 – 0)^2] = \text{sqrt}[4 + 0] = 2$ $\qquad$ $s_2 = 2$
$d_{23} = \text{sqrt}[(-1 – 0)^2 + (0 – 1)^2] = \text{sqrt}[1 + 1] = 1.414$ $\qquad$ $s_3 = s_2 + 1.414 = 3.414$
$d_{34} = \text{sqrt}[(0 – 0)^2 + (-1 – 1)^2] = \text{sqrt}[0 + 4] = 2$ $\qquad$ $s_4 = s_3 + 2 = 5.414$

And the starting point is $s_1 = 0$. As stated before, though not required, it is convenient (mainly for plotting purposes) to scale the values of s to range from 0 to 1. In this case, we just divide each value by s_4 (i.e. 5.414):

$s_1 = 0$ $\quad$ $s_2 = 0.369$ $\quad$ $s_3 = 0.631$ $\quad$ $s_4 = 1$

This gives us two sets of points, each of which will be fit with a cubic spline:

[s, x] = [(0, 1), (0.369, -1), (0.631, 0), (1, 0)] $\qquad$ [s, y] = [(0, 0), (0.369, 0), (0.631, 1), (1, -1)]

Recalling the matrix equation of a cubic spline, we have the following for n = 4:

$$\begin{bmatrix} ? & & & \\ h_1 & 2(h_1+h_2) & h_2 & . \\ . & h_2 & 2(h_2+h_3) & h_3 \\ & & & ? \end{bmatrix} \begin{bmatrix} x_1'' & y_1'' \\ x_2'' & y_2'' \\ x_3'' & y_3'' \\ x_4'' & y_4'' \end{bmatrix} = 6 \begin{bmatrix} ? & ? \\ (x_3-x_2)/h_2-(x_2-x_1)/h_1 & (y_3-y_2)/h_2-(y_2-y_1)/h_1 \\ (x_4-x_3)/h_3-(x_3-x_2)/h_2 & (y_4-y_3)/h_3-(y_3-y_2)/h_2 \\ ? & ? \end{bmatrix}$$

This is identical to the non-parametric case, except now we have two sets of 2nd derivatives to determine and two right hand sides. The base matrix is the same for both. Note in this case, the h_i are formed by subtracting successive values of s_i (e.g. $h_1 = s_2 - s_1$), the independent variable. We can calculate the needed quantities:

$h_1 = s_2 - s_1 = 0.369 - 0 = 0.369$
$h_2 = s_3 - s_2 = 0.631 - 0.369 = 0.261$
$h_3 = s_4 - s_3 = 1 - 0.631 = 0.369$
$x_2 - x_1 = -1 - 1 = -2$
$x_3 - x_2 = 0 - (-1) = 1$
$x_4 - x_3 = 0 - 0 = 0$
$(x_2 - x_1) / h_1 = (-2) / 0.369 = -5.414$
$(x_3 - x_2) / h_2 = 1 / 0.261 = 3.828$
$(x_4 - x_3) / h_3 = 0 / 0.369 = 0$
$(y_2 - y_1) / h_1 = 0 / 0.369 = 0$
$(y_3 - y_2) / h_2 = 1 / 0.261 = 3.828$
$(y_4 - y_3) / h_3 = (-2) / 0.369 = -5.414$

$2(h_1 + h_2) = 2(0.369 + 0.261) = 1.261$
$2(h_2 + h_3) = 2(0.261 + 0.369) = 1.261$
$y_2 - y_1 = 0 - 0 = 0$
$y_3 - y_2 = 1 - 0 = 1$
$y_4 - y_3 = (-1) - 1 = -2$

$(x_3-x_2)/h_2 - (x_2-x_1)/h_1 = 1/0.261 - (-2)/0.369 = 9.243$
$(x_4-x_3)/h_3 - (x_3-x_2)/h_2 = 0/0.369 - 1/0.261 = -3.828$

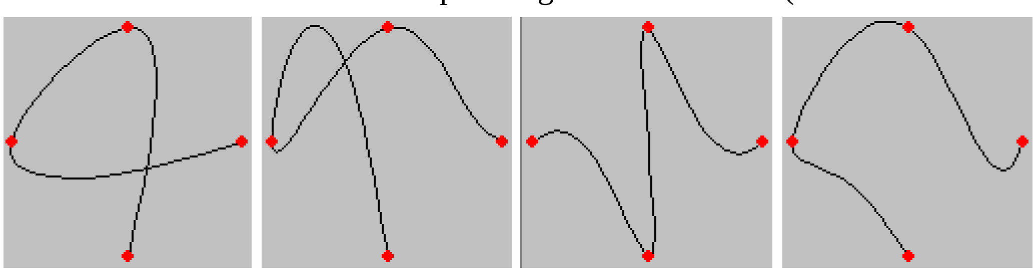
$(y_3-y_2)/h_2 - (y_2-y_1)/h_1 = 1/0.261 - 0/0.369 = 3.828$
$(y_4-y_3)/h_3 - (y_3-y_2)/h_2 = (-2)/0.369 - 1/0.261 = -9.243$

Choosing natural boundary conditions for x'' and specific values for y'' (to obtain a more pleasing plot), we get:

$$\begin{bmatrix} 1 & . & . & . \\ 0.369 & 1.261 & 0.261 & . \\ . & 0.261 & 1.261 & 0.369 \\ . & . & . & 1 \end{bmatrix} \begin{bmatrix} x_1'' & y_1'' \\ x_2'' & y_2'' \\ x_3'' & y_3'' \\ x_4'' & y_4'' \end{bmatrix} = 6 \begin{bmatrix} 0 & 1 \\ 9.243 & 3.828 \\ -3.828 & -9.243 \\ 0 & 1 \end{bmatrix}$$

$x'' = [0, 49.882, -28.544, 0,]^T$
$y'' = [6, 27.088, -51.338, 6]^T$

from which we can calculate the spline segment coefficients (one set for x and another set for y) and plot the result (left hand picture). By varying the sequence in which the points are connected and playing with the end conditions, interesting results can be obtained.



Example Problem #3 (Closed Loop)

Let us connect the points in the same sequence as before, but close the loop. In this case, a 5th point is needed in the sequence that is identical to the 1st:

$p_1 = [1,0]$ $p_2 = [-1, 0]$ $p_3 = [0, 1]$ $p_4 = [0, -1]$ $p_5 = p_1 = [1,0]$

For s, arc length approximation could be used as before, but an easier approach will be used so the simplified equations can be employed. Equal spacing will be chosen: s = [0, 0.25, 0.5, 0.75, 1].

Boundary conditions require special consideration. In the case of a closed loop, the 1st and last points are joined, which makes them interior points subject to the same 1st and 2nd derivative compatibility constraints as the other interior points. Thus, with closed loops, there are no choices to be made for boundary conditions! To derive the 1st derivative compatibility equation, simply equate 1st derivatives of p_1 and p_5:

$p_1' = (p_2 - p_1)/h_1 - p_2''h_1/6 - p_1''h_1/3 = p_5' = (p_5 - p_4)/h_4 + p_5''h_4/3 + p_4''h_4/6$

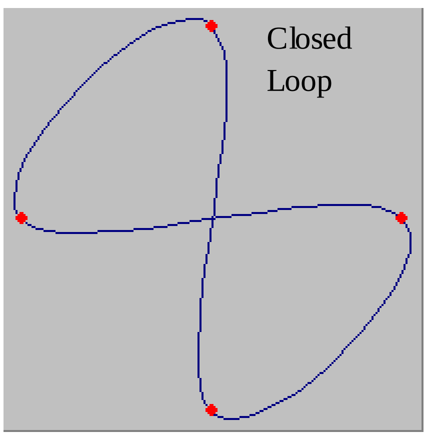
$6(p_2 - p_1)/h_1 - p_2''h_1 - 2p_1''h_1 = 6(p_5 - p_4)/h_4 + 2p_5''h_4 + p_4''h_4$

$6[(p_2 - p_1)/h_1 - (p_5 - p_4)/h_4] = (2h_1)p_1'' + (h_1)p_2'' + (h_4)p_4'' + (2h_4)p_5''$ $\qquad p_1' = p_5'$

The 2nd derivative boundary conditions are easy: $p_1'' - p_5'' = 0$. (*Note:* using p just means there are two sets of equations, one for x and another for y). Since $h_1 = h_2 = h_3 = h_4 = h = s_2 = 0.25$, the resulting matrix equation to be solved is:

$$\begin{bmatrix} 2 & 1 & . & 1 & 2 \\ 1 & 4 & 1 & . & . \\ . & 1 & 4 & 1 & . \\ . & . & 1 & 4 & 1 \\ 1 & . & . & . & -1 \end{bmatrix} \begin{bmatrix} x_1'' & y_1'' \\ x_2'' & y_2'' \\ x_3'' & y_3'' \\ x_4'' & y_4'' \\ x_5'' & y_5'' \end{bmatrix} = 6/(h*h) \begin{bmatrix} x_2 - x_1 - x_5 + x_4 & y_2 - y_1 - y_5 + y_4 \\ x_3 - 2x_2 + x_1 & y_3 - 2y_2 + y_1 \\ x_4 - 2x_3 + x_2 & y_4 - 2y_3 + y_2 \\ x_5 - 2x_4 + x_3 & y_5 - 2y_4 + y_3 \\ 0 & 0 \end{bmatrix} = 96 \begin{bmatrix} -3 & -1 \\ 3 & 1 \\ -1 & -3 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}$$

The solution is $x'' = [-120, 120, -72, 72, -120]^T$ and $y'' = [-72, 72, -120, 120, -72]^T$, from which we can calculate the coefficients of the spline segments and generate a plot.



:

Conclusion:

We have extended our knowledge of cubic splines to include the powerful and flexible parametric version and have fit various shapes (including approximating a circle) to a set of given points.