Chapter 1

Cubic Curves

1.1 Introduction

After quadratics, the next simplest curves are cubics. These are even more common than quadratics. They are used in many drawing programs, and serve as the foundation of the Postscript language used to describe fonts and other artwork.

We will see later, in chapter ???, that cubic curves can be strung together to form cubic splines, which are very common in many areas of CAD/CAM/CAE and graphics. But, before we discuss splines, let's get comfortable with the cubic curves that are used to build them.

1.2 Cubic Curves

Our discussion of cubic curves will closely parallel the handling of quadratic curves in the previous chapter. Suppose we have three vectors, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, and a point $\mathbf{D}$. Consider the equation

$$\mathbf{X}(u) = u^3\mathbf{A} + u^2\mathbf{B} + u\mathbf{C} + \mathbf{D}$$

Clearly this equation describes some sort of curve. It will be non-planar unless the vectors $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are coplanar. We can write it in the form

$$\mathbf{X}(u) = \begin{bmatrix} u^3 & u^2 & u \end{bmatrix} \begin{bmatrix} \leftarrow & \mathbf{A} & \rightarrow \\ \leftarrow & \mathbf{B} & \rightarrow \\ \leftarrow & \mathbf{C} & \rightarrow \end{bmatrix} + \mathbf{D}$$

This shows that the curve can be regarded as an affine transformation of the basic cubic curve $u \mapsto (u^3, u^2, u)$. If we restrict the parameter to the interval $0 \le u \le 1$, we get a bounded curve segment that might (we hope) be useful for design purposes.

As with the quadratic curve, this power basis representation of a cubic curve allows us to calculate points very efficiently by applying Horner's method. We have

$$\mathbf{X}(u) = \mathbf{D} + u\left\{\mathbf{C} + u(\mathbf{B} + u\mathbf{A})\right\}$$

so just three additions and three multiplications are needed to compute each coordinate. To understand the shape of this curve, let's compute a few of its properties. We have

$$\mathbf{X}(0) = \mathbf{D} \quad ; \quad \mathbf{X}'(0) = \mathbf{C} \quad ; \quad \mathbf{X}''(0) = 2\mathbf{B}$$

So $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ give us direct control over the properties of the curve at the point $u = 0$. But, on the other hand

$$\mathbf{X}(1) = \mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D} \quad ; \quad \mathbf{X}'(1) = 3\mathbf{A} + 2\mathbf{B} + \mathbf{C}$$

So the quantities **A**, **B**, **C**, **D** give us no easy way to control what happens at the other end of the curve, where $u = 1$. This is the typical situation with Taylor series expansions in general, of course — we get lots of information about *a single point* but only vague indications of what happens elsewhere. People doing interactive design will want to directly control *both* ends of a curve, of course, so we need some different “handles” that give us better intewractive control than the quantities **A**, **B**, **C**, **D**.

1.3 Bézier Form

Like the quadratic, the cubic curve can be generated by corner cutting. If we begin with four points $\mathbf{P}_0$, $\mathbf{P}_a$, $\mathbf{P}_b$, $\mathbf{P}_1$, then the corner-cutting process generates the curve with equation

$$\mathbf{X}(u) = (1-u)^3\mathbf{P}_0 + 3u(1-u)^2\mathbf{P}_a + 3u^2(1-u)\mathbf{P}_b + u^3\mathbf{P}_1$$

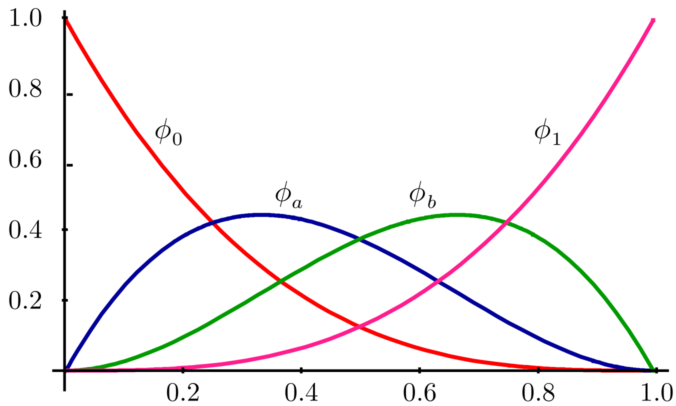
If we define four functions

$$\begin{aligned}
\phi_0(u) &= (1-u)^3 = 1 - 3u + 3u^2 - u^3 \\
\phi_a(u) &= 3u(1-u)^2 = 3u - 6u^2 + 3u^3 \\
\phi_b(u) &= 3u^2(1-u) = 3u^2 - 3u^3 \\
\phi_1(u) &= u^3 = u^3
\end{aligned}$$

then our curve equation can be written as

$$\mathbf{X}(u) = \phi_0(u)\mathbf{P}_0 + \phi_a(u)\mathbf{P}_a + \phi_b(u)\mathbf{P}_b + \phi_1(u)\mathbf{P}_1 = \sum \phi_i(u)\mathbf{P}_i$$

which is known as the **Bézier** form of the curve. The functions ϕ_0, ϕ_a, ϕ_b, ϕ_1 are the cubic Bernstein polynomials, which are well-known in many areas of mathematics. Their graphs look like this:



In matrix form, the curve is:

$$\mathbf{X}(u) = \begin{bmatrix} 1 & u & u^2 & u^3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ 1 & 3 & -3 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_a \\ \mathbf{P}_b \\ \mathbf{P}_1 \end{bmatrix}$$

It is easy to verify that $\phi_0(u) + \phi_a(u) + \phi_b(u) + \phi_1(u) = 1$ for all u, so this definition makes sense for all u. Furthermore, we have $0 \le \phi_i(u) \le 1$ for $i = 0, a, b, 1$ and $0 \le u \le 1$. This means that when $0 \le u \le 1$, the point $\mathbf{X}(u)$ lies inside convex hull of the four points $\mathbf{P}_0$, $\mathbf{P}_a$, $\mathbf{P}_b$, $\mathbf{P}_1$. Clearly, we have

$$\mathbf{X}(0) = \mathbf{P}_0 \quad ; \quad \mathbf{X}(1) = \mathbf{P}_1$$

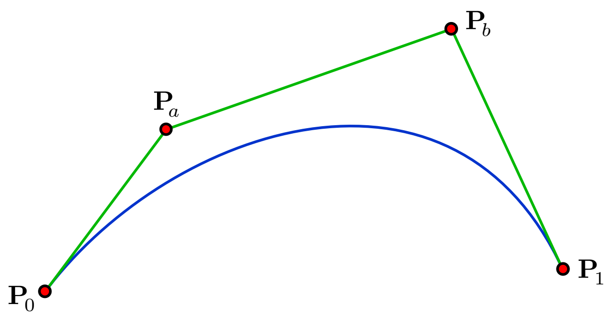
so the curve starts at the point $\mathbf{P}_0$ and ends at $\mathbf{P}_1$. Differentiating, we get

$$\mathbf{X}'(u) = 3\left\{(1-u)^2(\mathbf{P}_a - \mathbf{P}_0) + 2u(1-u)(\mathbf{P}_b - \mathbf{P}_a) + u^2(\mathbf{P}_1 - \mathbf{P}_b)\right\}$$

So, in particular, the derivatives at the start and end points of the curve are

$$\mathbf{X}'(0) = 3(\mathbf{P}_a - \mathbf{P}_0) \quad ; \quad \mathbf{X}'(1) = 3(\mathbf{P}_1 - \mathbf{P}_b)$$

which means that the end tangents of the curve are parallel to the polygon "legs" $\mathbf{P}_a\mathbf{P}_0$ and $\mathbf{P}_b\mathbf{P}_1$. So, again, the four points $\mathbf{P}_0$, $\mathbf{P}_a$, $\mathbf{P}_b$, $\mathbf{P}_1$ give us an obvious and direct way to control the end points and end tangents of the curve.



It is easy to show that the second derivative at the start of the curve is given by

$$\mathbf{X}''(0) = 6(\mathbf{P}_b - \mathbf{P}_a) - (\mathbf{P}_a - \mathbf{P}_0)$$

and its curvature is

$$\kappa(0) = \frac{2\,\|(\mathbf{P}_a - \mathbf{P}_0) \times (\mathbf{P}_b - \mathbf{P}_a)\|}{3\,\|\mathbf{P}_a - \mathbf{P}_0\|^3}$$

Similarly, at the end of the curve, we have the second derivative

$$\mathbf{X}''(1) = 6(\mathbf{P}_1 - \mathbf{P}_b) - (\mathbf{P}_b - \mathbf{P}_a)$$

and curvature

$$\kappa(1) = \frac{2\,\|(\mathbf{P}_b - \mathbf{P}_a) \times (\mathbf{P}_1 - \mathbf{P}_b)\|}{3\,\|\mathbf{P}_1 - \mathbf{P}_b\|^3}$$

1.4 Cubic Curves in Hermite Form

Suppose we are given two points $\mathbf{P}_0$ and $\mathbf{P}_1$, and two vectors $\mathbf{Q}_0$ and $\mathbf{Q}_1$, and we wish to construct a curve $\mathbf{X}(u)$ with the properties

$$\mathbf{X}(0) = \mathbf{P}_0 \quad ; \quad \mathbf{X}'(0) = \mathbf{Q}_0 \tag{1.1}$$

$$\mathbf{X}(1) = \mathbf{P}_1 \quad ; \quad \mathbf{X}'(1) = \mathbf{Q}_1 \tag{1.2}$$

Let's assume that the curve is to be a blended curve of the form

$$\mathbf{X}(u) = \phi_0(u)\mathbf{P}_0 + \phi_1(u)\mathbf{P}_1 + \psi_0(u)\mathbf{Q}_0 + \psi_1(u)\mathbf{Q}_1$$

Then we can satisfy the constraints (1.1) by requiring that

$$\begin{array}{ccccccc}
\phi_0(0) = 1 & ; & \phi_0(1) = 0 & ; & \phi_0'(0) = 0 & ; & \phi_0'(1) = 0 \\
\phi_1(0) = 0 & ; & \phi_1(1) = 1 & ; & \phi_1'(0) = 0 & ; & \phi_1'(1) = 0 \\
\psi_0(0) = 0 & ; & \psi_0(1) = 0 & ; & \psi_0'(0) = 1 & ; & \psi_0'(1) = 0 \\
\psi_1(0) = 0 & ; & \psi_1(1) = 0 & ; & \psi_1'(0) = 0 & ; & \psi_1'(1) = 1
\end{array}$$

This does not require that the blending functions be polynomials, but, in practice, they almost always are. If we wish to use polynomial blending functions, we may assume that

$$\mathbf{X}(u) = \mathbf{A}_0 + u\mathbf{A}_1 + u^2\mathbf{A}_2 + u^3\mathbf{A}_3 = [1 \quad u \quad u^2 \quad u^3][\mathbf{A}_0 \quad \mathbf{A}_1 \quad \mathbf{A}_2 \quad \mathbf{A}_3]^{\mathsf{T}}$$

where the vectors $\mathbf{A}_0$, $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ are as yet unknown. Then

$$\mathbf{X}'(u) = \mathbf{A}_1 + 2u\mathbf{A}_2 + 3u^2\mathbf{A}_3$$

and the given constraints (1.1) imply that

$$\begin{aligned}
\mathbf{P}_0 &= \mathbf{A}_0 \\
\mathbf{P}_1 &= \mathbf{A}_0 + \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3 \\
\mathbf{Q}_0 &= \mathbf{A}_1 \\
\mathbf{Q}_1 &= \mathbf{A}_1 + 2\mathbf{A}_2 + 3\mathbf{A}_3
\end{aligned}$$

Inverting the matrix to solve these equations for $\mathbf{A}_0$, $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, we get

$$\begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & 3 & -2 & -1 \\ 2 & -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{Q}_0 \\ \mathbf{Q}_1 \end{bmatrix}$$

Using $\mathbf{M}$ to denote the 4×4 matrix on the right-hand side, our blended curve becomes

$$\mathbf{X}(u) = [1 \quad u \quad u^2 \quad u^3] \cdot \mathbf{M} \cdot [\mathbf{P}_0 \quad \mathbf{P}_1 \quad \mathbf{Q}_0 \quad \mathbf{Q}_1]^{\mathsf{T}}$$

In other words

$$\mathbf{X}(u) = \phi_0(u)\mathbf{P}_0 + \phi_1(u)\mathbf{P}_1 + \psi_0(u)\mathbf{Q}_0 + \psi_1(u)\mathbf{Q}_1$$

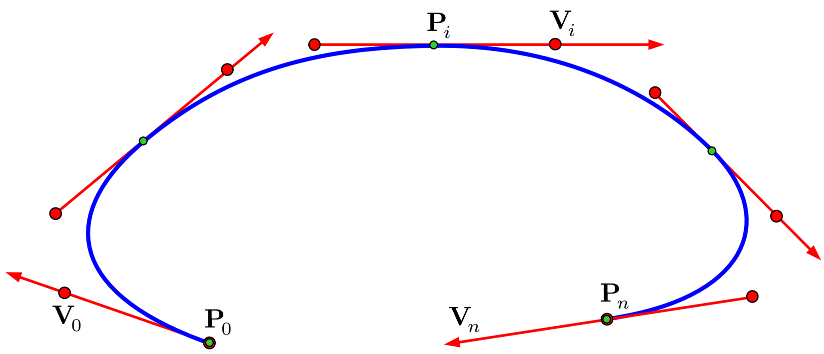
where

$$[\phi_0(u) \quad \phi_1(u) \quad \psi_0(u) \quad \psi_1(u)] = [1 \quad u \quad u^2 \quad u^3] \cdot \mathbf{M}$$

The functions ϕ_0, ϕ_1, ψ_0, ψ_1 are called the **Hermite cubic blending functions**. Since the matrix $\mathbf{M}$ is invertible, they form a basis for $\mathcal{P}^3$. Writing them out explicitly, they are

$$\begin{array}{ccc}
\phi_0(u) = 1 - 3u^2 + 2u^3 & ; & \phi_1(u) = 3u^2 - 2u^3 \\
\psi_0(u) = u - 2u^2 + u^3 & ; & \psi_1(u) = u^3 - u^2
\end{array}$$

We note that $\phi_0(u) + \phi_1(u) = 1$, so the definition of $\mathbf{X}(u)$ makes sense. By stringing together Hermite cubic curves that share end-points and first derivative vectors, we can construct a smooth composite curve that interpolates a given sequence of points



1.5 The Timmer Form

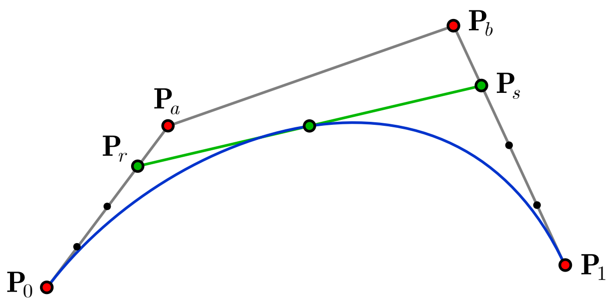
One problem with the Bézier form of the cubic curve is that the curve does not follow the control polygon very closely. In an attempt to improve things, we define two points $\mathbf{P}_r$ and $\mathbf{P}_s$ by

$$\mathbf{P}_r = \tfrac{1}{4}\mathbf{P}_0 + \tfrac{3}{4}\mathbf{P}_a \quad ; \quad \mathbf{P}_s = \tfrac{3}{4}\mathbf{P}_b + \tfrac{1}{4}\mathbf{P}_1$$

Then, routine calculations show that

$$\mathbf{X}(\tfrac{1}{2}) = \tfrac{1}{2}\left(\mathbf{P}_r + \mathbf{P}_s\right)$$

so the parametric mid-point of the curve is the mid-point of the line $\mathbf{P}_r\mathbf{P}_s$, as shown in the diagram below



From the picture, we might think that the curve is tangent to the line $\mathbf{P}_r\mathbf{P}_s$, but in fact this is not the case, in general. In fact, we have

$$\mathbf{X}'(\tfrac{1}{2}) = \mathbf{P}_s - \mathbf{P}_r + \tfrac{1}{2}\left(\mathbf{P}_1 - \mathbf{P}_0\right)$$

Next, consider the special case where $\mathbf{P}_r\mathbf{P}_s$ is parallel to $\mathbf{P}_0\mathbf{P}_1$. Then we can find a number λ such that

$$\mathbf{P}_1 - \mathbf{P}_0 = \lambda\left(\mathbf{P}_s - \mathbf{P}_r\right)$$

so we have

$$\mathbf{X}'(\tfrac{1}{2}) = \left(1 + \tfrac{1}{2}\lambda\right)\left(\mathbf{P}_s - \mathbf{P}_r\right)$$

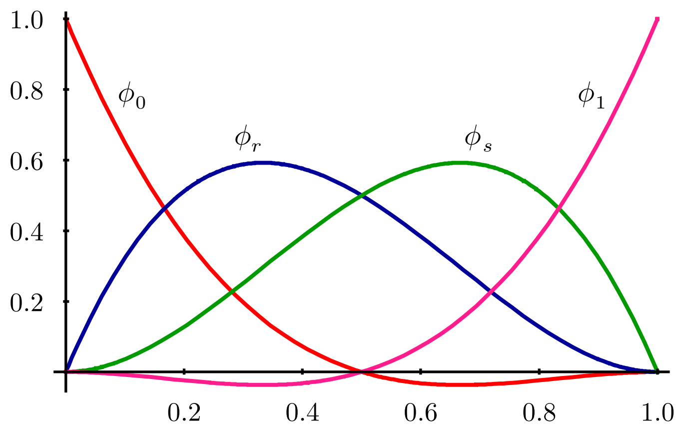
So, in this case, the curve actually *is* tangent to the line $\mathbf{P}_r\mathbf{P}_s$ at its mid-point. If we define four functions

$$\phi_0(u) = 1 - 4u + 5u^2 - 2u^3$$
$$\phi_r(u) = 4u(1-u)^2 = 4u - 8u^2 + 4u^3$$
$$\phi_s(u) = 4u^2(1-u) = 4u^2 - 4u^3$$
$$\phi_1(u) = 2u^3 - u^2$$

then our curve equation can be written as

$$\mathbf{X}(u) = \phi_0(u)\mathbf{P}_0 + \phi_r(u)\mathbf{P}_r + \phi_s(u)\mathbf{P}_s + \phi_1(u)\mathbf{P}_1 = \sum \phi_i(u)\mathbf{P}_i$$

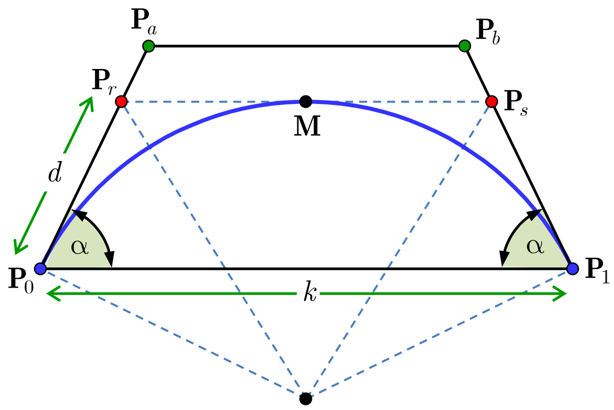
It is easy to verify that $\phi_0(u) + \phi_r(u) + \phi_s(u) + \phi_1(u) = 1$ for all u, so this definition makes sense for all u. The graphs of the functions ϕ_0, ϕ_r, ϕ_s, ϕ_1 look like this:



We note that $\phi_0(\frac{2}{3}) = \phi_1(\frac{1}{3}) = -\frac{1}{27}$. Because there are regions where $\phi_0(u) < 0$ and regions where $\phi_1(u) < 0$, the curve does not lie inside the convex hull of the the four points $\mathbf{P}_0$, $\mathbf{P}_r$, $\mathbf{P}_s$, $\mathbf{P}_1$.

1.6 Approximating Circles

There are many ways to approximate circular arcs with cubic curves. The simplest approach is to use the Timmer form of the cubic described in the previous section. Suppose we have a circular arc of radius r having an angle α between each end-tangent and its chord, as shown in the picture below:



A little trigonometry shows that

$$d = r \tan \tfrac{1}{2}\alpha = r\sqrt{\frac{1-\cos\alpha}{1+\cos\alpha}} = \frac{r(1-\cos\alpha)}{\sin\alpha} \tag{1.3}$$

This tells us where the Timmer points $\mathbf{P}_r$ and $\mathbf{P}_s$ should be placed, and we can then calculate the Bézier control points $\mathbf{P}_a$ and $\mathbf{P}_b$ from

$$\mathbf{P}_a = \tfrac{4}{3}\mathbf{P}_r - \tfrac{1}{3}\mathbf{P}_0 \quad ; \quad \mathbf{P}_b = \tfrac{4}{3}\mathbf{P}_s - \tfrac{1}{3}\mathbf{P}_1$$

Stating this another way, we should position the Bézier control points $\mathbf{P}_a$ and $\mathbf{P}_b$ so that the lengths of the "legs" $\mathbf{P}_0\mathbf{P}_a$ and $\mathbf{P}_b\mathbf{P}_1$ are given by

$$\|\mathbf{P}_a - \mathbf{P}_0\| = \|\mathbf{P}_1 - \mathbf{P}_b\| = \tfrac{4}{3} r \tan \tfrac{1}{2}\alpha$$

The cubic curve constructed this way will match the circular arc in position and tangent direction at the end points $\mathbf{P}_0$ and $\mathbf{P}_1$ and also at the mid-point $\mathbf{M}$. It always lies outside the circular arc.

The maximum error of approximation [Dokken, 1990] is given by

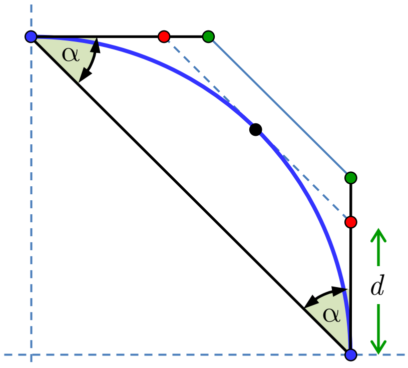
$$\epsilon(\alpha) = \frac{4\sin^6 \frac{1}{2}\alpha}{27\cos^2 \frac{1}{2}\alpha}$$

The formulae can be written in terms of the chord-length, rather than the arc radius. If we let k be the length of the chord $\mathbf{P}_0\mathbf{P}_1$, then $r = \frac{1}{2}k \operatorname{cosec}\alpha$. Substituting into (1.3), we get

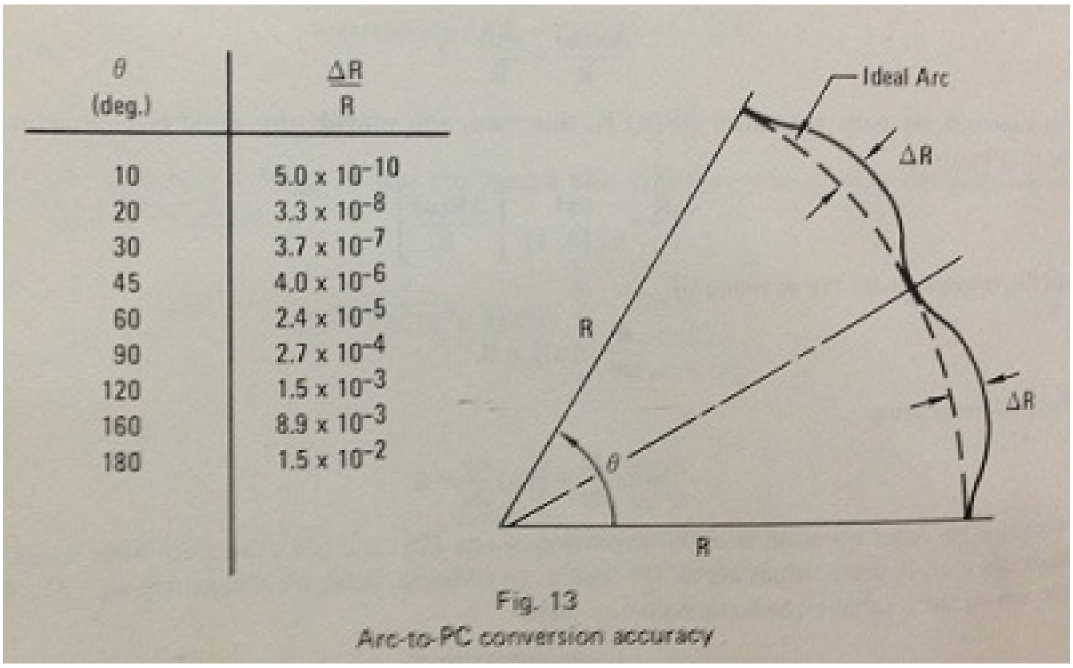
$$d = \tfrac{1}{2}k \sec^2 \tfrac{1}{2}\alpha = \frac{k}{2(1+\cos\alpha)} \tag{1.4}$$

1.7 Example: Circular Quadrant

As an example, let's consider the approximation of a quadrant of a unit circle, as shown below



We have $r = 1$ and $\alpha = \frac{1}{4}\pi$, so $\cos\alpha = \frac{1}{2}\sqrt{2}$ and a little arithmetic shows that $d = \sqrt{2} - 1$. The radial error is roughly 0.00027, which means that this approximation is good enough for many purposes. In fact, drawing programs often use four quadrants of this form to represent circles.



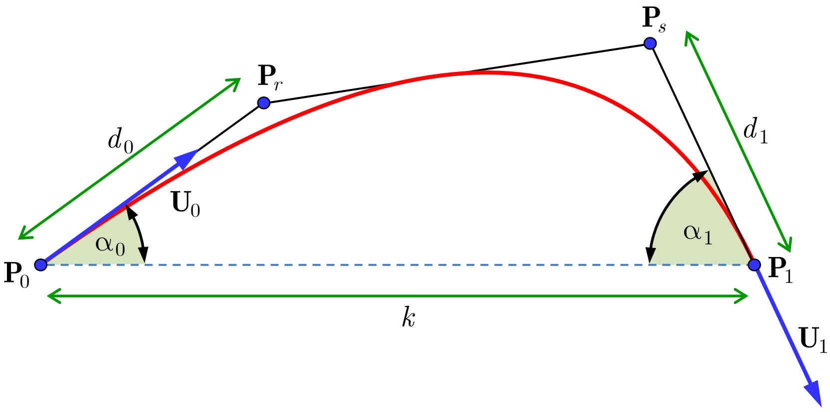
θ (deg.)	$\frac{\Delta R}{R}$
10	5.0×10^{-10}
20	3.3×10^{-8}
30	3.7×10^{-7}
45	4.0×10^{-6}
60	2.4×10^{-5}
90	2.7×10^{-4}
120	1.5×10^{-3}
160	8.9×10^{-3}
180	1.5×10^{-2}

Fig. 13
Arc-to-PC conversion accuracy

Better circle approximations are possible if we are willing to sacrifice interpolation conditions.

1.8 Cubics from Points and Directions

There are many situations where we want to construct a cubic curve from two points and two directions, where the directions are described by unit vectors. In addition to the two end-points $\mathbf{P}_0$ and $\mathbf{P}_1$, suppose we are given two unit vectors $\mathbf{U}_0$ and $\mathbf{U}_1$, or, equivalently, the angles α_0 and α_1 that these vectors form with the curve's chord. Let k be the length of the chord $\mathbf{P}_0\mathbf{P}_1$.



To construct a cubic curve, we will decide the locations of the Timmer control points $\mathbf{P}_r$ and $\mathbf{P}_s$. We can choose any two numbers d_0 and d_1, and define the Timmer control points by

$$\mathbf{P}_r = \mathbf{P}_0 + d_0 \mathbf{U}_0 \quad ; \quad \mathbf{P}_s = \mathbf{P}_1 - d_1 \mathbf{U}_1$$

The curve $\mathbf{P}(t)$ defined this way has $\mathbf{P}'(0) = 4d_0\mathbf{U}_0$ and $\mathbf{P}'(1) = 4d_1\mathbf{U}_1$, so its end tangents are parallel to the given unit vectors $\mathbf{U}_0$ and $\mathbf{U}_1$ no matter what values of d_0 and d_1 we choose. But different choices will give dramatically different curves; our challenge is to find "good" values of d_0 and d_1 that will produce "nice" curves.

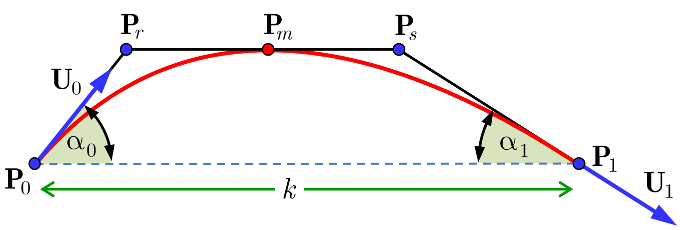
One approach that has been used for many years is to define

$$d_0 = \frac{k}{2(1+\cos\alpha_1)} \quad ; \quad d_1 = \frac{k}{2(1+\cos\alpha_0)}$$

Note the "swapping" in these equations: α_0 is used to define d_1 and α_1 is used to define d_0. Comparing this with (1.3), we see that this scheme reproduces circles, when possible.

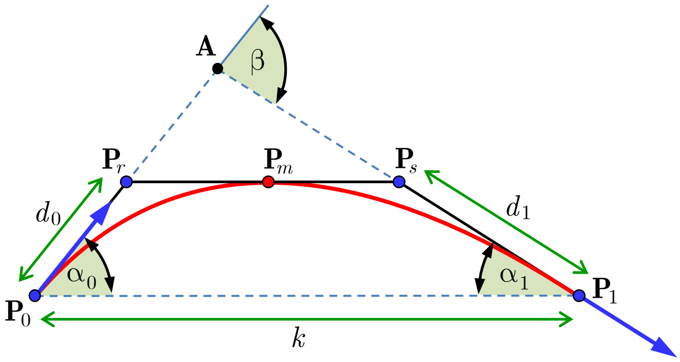
1.9 Rho Cubics

We can control the shape of a cubic curve using the same sort of ρ parameter that we used with conics in section (**??**). Assume we are given two points $\mathbf{P}_0$ and $\mathbf{P}_1$, and two unit vectors $\mathbf{U}_0$ and $\mathbf{U}_1$, as shown below, plus a value of ρ.



We will construct Timmer control points $\mathbf{P}_r$ and $\mathbf{P}_s$ so that $\mathbf{P}_r\mathbf{P}_s$ is parallel to $\mathbf{P}_0\mathbf{P}_1$. If we do this, then we know from section (1.5) that the cubic curve will be tangent to $\mathbf{P}_r\mathbf{P}_s$ at its mid-point.

We know the tangent lines at $\mathbf{P}_0$ and $\mathbf{P}_1$, and, if the input data do not imply an inflexion, we can intersect these tangent lines to get the apex point $\mathbf{A}$.



Then the Timmer control points are

$$\mathbf{P}_r = (1-\rho)\mathbf{P}_0 + \rho\mathbf{A}$$

$$\mathbf{P}_s = (1-\rho)\mathbf{P}_1 + \rho\mathbf{A}$$

However, it is often more convenient to calculate $\mathbf{P}_r$ and $\mathbf{P}_s$ without computing the apex point $\mathbf{A}$. Using the sine rule in triangle $\mathbf{P}_0\mathbf{A}\mathbf{P}_1$, we have

$$\frac{\|\mathbf{A}-\mathbf{P}_0\|}{\sin\alpha_1} = \frac{\|\mathbf{A}-\mathbf{P}_1\|}{\sin\alpha_0} = \frac{k}{\sin(\pi-\beta)}$$

But $\sin(\pi-\beta) = \sin\beta = \sin(\alpha_0+\alpha_1)$, so we get

$$\|\mathbf{A}-\mathbf{P}_0\| = \frac{k\sin\alpha_1}{\sin(\alpha_0+\alpha_1)}$$

$$\|\mathbf{A}-\mathbf{P}_1\| = \frac{k\sin\alpha_0}{\sin(\alpha_0+\alpha_1)}$$

Finally, $d_0 = \rho\,\|\mathbf{A}-\mathbf{P}_0\|$ and $d_1 = \rho\,\|\mathbf{A}-\mathbf{P}_1\|$, so

$$d_0 = \frac{\rho k\sin\alpha_1}{\sin(\alpha_0+\alpha_1)} \quad ; \quad d_1 = \frac{\rho k\sin\alpha_0}{\sin(\alpha_0+\alpha_1)} \tag{1.5}$$

Note that these same formulae do not depend on the apex point $\mathbf{A}$, so they can be used even if the input data $\mathbf{P}_0$, $\mathbf{P}_1$, $\mathbf{U}_0$ and $\mathbf{U}_1$ imply an inflexion in the curve.

If no ρ value is given as input, we can use either the “least tension” value from section (**??**) or the minimum eccentricity value from section (**??**).

1.10 Hobby’s Cubic

John Hobby proposed yet another scheme for interpolating two points and two directions with a cubic curve. His method is used in the Metafont system for designing parameterized fonts. The calculations are as follows. First, we define three constants that Hobby determined experimentally:

$$a = \sqrt{2} \quad ; \quad b = \tfrac{1}{16} \quad ; \quad c = \tfrac{1}{2}(3-\sqrt{5})$$

Next, we define

$$\mu = a(\sin\alpha_0 - b\sin\alpha_1)(\sin\alpha_1 - b\sin\alpha_0)(\cos\alpha_0 - \cos\alpha_1)$$

Finally, the distances that serve to locate the Timmer control points are:

$$d_0 = \frac{(2+\mu)k}{4(1+(1-c)\cos\alpha_0 + c\cos\alpha_1)}$$

$$d_1 = \frac{(2-\mu)k}{4(1+(1-c)\cos\alpha_1 + c\cos\alpha_0)}$$

In the symmetric case where $\alpha_0 = \alpha_1 = \alpha$, we have $\mu = 0$, and the formulae for d_0 and d_1 above reduce to the ones for circular arcs. So, like our other interpolation methods, the Hobby method again has circular precision. Hobby's method is considerably more complex than the other ones described earlier, and, in my experience, it doesn't work any better, so it's hard to justify using it.

1.11 Cubic Through Four Points

Let $\mathbf{A}$ be a point, and let $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ be three vectors, all of which are unknown, as yet, and consider the parametric cubic curve

$$\mathbf{X}(t) = \mathbf{A} + t\mathbf{B} + t^2\mathbf{C} + t^3\mathbf{D} = [1 \quad t \quad t^2 \quad t^3] \cdot [\mathbf{A} \quad \mathbf{B} \quad \mathbf{C} \quad \mathbf{D}]^\mathsf{T}$$

Let $\mathbf{P}_0$, $\mathbf{P}_m$, $\mathbf{P}_n$, $\mathbf{P}_1$ be four points in $\mathbb{R}^3$, and suppose we want our curve to pass through these points, with

$$\mathbf{X}(0) = \mathbf{P}_0 \quad ; \quad \mathbf{X}(\tfrac{1}{3}) = \mathbf{P}_m \quad ; \quad \mathbf{X}(\tfrac{2}{3}) = \mathbf{P}_n \quad ; \quad \mathbf{X}(1) = \mathbf{P}_1$$

From these conditions, we can construct four equations

$$\begin{aligned}
\mathbf{X}(0) = \mathbf{P}_0 &\Longrightarrow & \mathbf{A} &= \mathbf{P}_0 \\
\mathbf{X}(\tfrac{1}{3}) = \mathbf{P}_m &\Longrightarrow & \mathbf{A} + \tfrac{1}{3}\mathbf{B} + \tfrac{1}{9}\mathbf{C} + \tfrac{1}{27}\mathbf{D} &= \mathbf{P}_m \\
\mathbf{X}(\tfrac{2}{3}) = \mathbf{P}_n &\Longrightarrow & \mathbf{A} + \tfrac{2}{3}\mathbf{B} + \tfrac{4}{9}\mathbf{C} + \tfrac{8}{27}\mathbf{D} &= \mathbf{P}_n \\
\mathbf{X}(1) = \mathbf{P}_1 &\Longrightarrow & \mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D} &= \mathbf{P}_1
\end{aligned}$$

Hence

$$\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \\ \mathbf{D} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & \frac{1}{3} & \frac{1}{9} & \frac{1}{27} \\ 1 & \frac{2}{3} & \frac{4}{9} & \frac{8}{27} \\ 1 & 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_m \\ \mathbf{P}_n \\ \mathbf{P}_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{11}{2} & 9 & -\frac{9}{2} & 2 \\ 9 & -\frac{45}{2} & 18 & -\frac{9}{2} \\ -\frac{9}{2} & \frac{27}{2} & -\frac{27}{2} & \frac{9}{2} \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_m \\ \mathbf{P}_n \\ \mathbf{P}_1 \end{bmatrix}$$

If we denote the matrix on the right by $\mathbf{M}$, then our equation can again be written as

$$\mathbf{X}(t) = [1 \quad t \quad t^2 \quad t^3] \cdot \mathbf{M} \cdot [\mathbf{P}_0 \quad \mathbf{P}_m \quad \mathbf{P}_n \quad \mathbf{P}_1]^\mathsf{T}$$

1.12 Lagrange Polynomials

Lagrange interpolation is a standard technique in numerical analysis that we can use to construct a curve passing through given points, as an alternative to the technique described in section (1.11) above. Here we will deal only with cubic polynomials, although the extension to other degrees is straightforward. Let $t_0 < t_m < t_n < t_1$. Then the cubic Lagrange polynomials determined by t_0, t_m, t_n, t_1, are

$$\begin{aligned}
\phi_0(t) &= \frac{(t-t_m)(t-t_n)(t-t_1)}{(t_0-t_m)(t_0-t_n)(t_0-t_1)} \\
\phi_m(t) &= \frac{(t-t_0)(t-t_n)(t-t_1)}{(t_m-t_0)(t_m-t_n)(t_m-t_1)} \\
\phi_n(t) &= \frac{(t-t_0)(t-t_m)(t-t_1)}{(t_n-t_0)(t_n-t_m)(t_n-t_1)} \\
\phi_1(t) &= \frac{(t-t_0)(t-t_m)(t-t_n)}{(t_1-t_0)(t_1-t_m)(t_1-t_n)}
\end{aligned}$$

Note that

$$\phi_i(t_j) = \delta_{ij} \qquad (i, j = 0, m, n, 1)$$

and so the parametric cubic curve expressed in Lagrange form

$$\mathbf{X}(t) = \phi_0(t)\mathbf{P}_0 + \phi_m(t)\mathbf{P}_m + \phi_n(t)\mathbf{P}_n + \phi_1(t)\mathbf{P}_1$$

has the property that $\mathbf{X}(t_j) = \mathbf{P}_j$ for $j = 0, m, n, 1$.

picture

The values of t_0, t_m, t_n, t_1 can have a significant effect on the shape of the curve. A very common choice is $t_0 = 0$, $t_m = \frac{1}{3}$, $t_m = \frac{2}{3}$, $t_1 = 1$. Then the cubic Lagrange polynomials become

$$\begin{aligned}
\phi_0(t) &= -\tfrac{1}{2}(3t-1)(3t-2)(t-1) = -\tfrac{9}{2}t^3 + 9t^2 - \tfrac{11}{2}t + 1 \\
\phi_m(t) &= -\tfrac{9}{2}t(3t-2)(t-1) = \tfrac{27}{2}t^3 - \tfrac{45}{2}t^2 + 9t \\
\phi_n(t) &= -\tfrac{9}{2}t(3t-1)(t-1) = -\tfrac{27}{2}t^3 + 18t^2 - \tfrac{9}{2}t \\
\phi_1(t) &= \tfrac{1}{2}t(3t-1)(3t-2) = \tfrac{9}{2}t^3 - \tfrac{9}{2}t^2 + t
\end{aligned}$$

As always, these basis functions can be written using a transition matrix

$$\begin{bmatrix} \phi_0(t) & \phi_m(t) & \phi_n(t) & \phi_1(t) \end{bmatrix} = \begin{bmatrix} 1 & t & t^2 & t^3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{11}{2} & 9 & -\frac{9}{2} & 2 \\ 9 & -\frac{45}{2} & 18 & -\frac{9}{2} \\ -\frac{9}{2} & \frac{27}{2} & -\frac{27}{2} & \frac{9}{2} \end{bmatrix}$$

Thus, if we denote the matrix shown by $\mathbf{M}$, we have

$$\begin{bmatrix} \phi_0(t) & \phi_m(t) & \phi_n(t) & \phi_1(t) \end{bmatrix} = \begin{bmatrix} 1 & t & t^2 & t^3 \end{bmatrix} \cdot \mathbf{M}$$

and the curve equation can be written as

$$\mathbf{X}(t) = \begin{bmatrix} 1 & t & t^2 & t^3 \end{bmatrix} \cdot \mathbf{M} \cdot \begin{bmatrix} \mathbf{P}_0 & \mathbf{P}_m & \mathbf{P}_n & \mathbf{P}_1 \end{bmatrix}^t$$

The matrix $\mathbf{M}$ is invertible, as we would expect.

1.13 Summary of Forms of Parametric Cubics

The Bézier-Bernstein Basis

The basis matrix $\mathbf{M}_b$ for the cubic Bézier-Bernstein basis is

$$\mathbf{M}_b = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix} \quad ; \quad \mathbf{M}_b^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

The Hermite Basis

The basis matrix $\mathbf{M}_h$ for the cubic Hermite basis is

$$\mathbf{M}_h = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & 3 & -2 & -1 \\ 2 & -2 & 1 & 1 \end{bmatrix} \quad ; \quad \mathbf{M}_h^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & \frac{1}{3} & 1 & 0 \\ 1 & \frac{2}{3} & \frac{1}{3} & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

The B-Spline Basis

The basis matrix $\mathbf{M}_s$ for the uniform cubic b-spline basis is

$$\mathbf{M}_s = \begin{bmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{3} & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{6} \end{bmatrix} \quad ; \quad \mathbf{M}_s^{-1} = \begin{bmatrix} 1 & -1 & \frac{2}{3} & 0 \\ 1 & 0 & -\frac{1}{3} & 0 \\ 1 & 1 & \frac{2}{3} & 0 \\ 1 & 2 & \frac{11}{3} & 6 \end{bmatrix}$$

The Lagrange Basis

The basis matrix $\mathbf{M}_l$ for the Lagrange cubic basis is

$$\mathbf{M}_l = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{11}{2} & 9 & -\frac{9}{2} & 1 \\ 9 & -\frac{45}{2} & 18 & -\frac{9}{2} \\ -\frac{9}{2} & \frac{27}{2} & -\frac{27}{2} & \frac{9}{2} \end{bmatrix} \quad ; \quad \mathbf{M}_l^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & \frac{1}{3} & \frac{1}{9} & \frac{1}{27} \\ 1 & \frac{2}{3} & \frac{4}{9} & \frac{8}{27} \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

1.14 Table of Conversion Matrices

Various matrices can now be computed to convert geometric data from one basis to another. In general, the matrix $\mathbf{M}_Y^{-1}\mathbf{M}_X$ converts geometry from the X form to the Y form. We have

$$\mathbf{M}_h^{-1}\mathbf{M}_b = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -3 & 3 & 2 & 0 \\ 0 & 0 & -3 & 3 \end{bmatrix} \quad ; \quad \mathbf{M}_b^{-1}\mathbf{M}_h = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{M}_s^{-1}\mathbf{M}_b = \begin{bmatrix} 6 & -7 & 2 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 2 & -7 & 6 \end{bmatrix} \quad ; \quad \mathbf{M}_b^{-1}\mathbf{M}_s = \begin{bmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{bmatrix}$$

$$\mathbf{M}_l^{-1}\mathbf{M}_b = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{8}{27} & \frac{4}{9} & \frac{2}{9} & \frac{1}{27} \\ \frac{1}{27} & \frac{2}{9} & \frac{4}{9} & \frac{8}{27} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad ; \quad \mathbf{M}_b^{-1}\mathbf{M}_l = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{5}{6} & 3 & -\frac{3}{2} & \frac{1}{3} \\ \frac{1}{3} & -\frac{3}{2} & 3 & -\frac{5}{6} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{M}_s^{-1}\mathbf{M}_h = \begin{bmatrix} -1 & 2 & -\frac{7}{3} & -\frac{2}{3} \\ 2 & -1 & \frac{2}{3} & \frac{1}{3} \\ -1 & 2 & -\frac{1}{3} & -\frac{2}{3} \\ 2 & -1 & \frac{2}{3} & \frac{7}{3} \end{bmatrix} \quad ; \quad \mathbf{M}_h^{-1}\mathbf{M}_s = \begin{bmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ -\frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$\mathbf{M}_l^{-1}\mathbf{M}_h = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{20}{27} & \frac{7}{27} & \frac{4}{27} & -\frac{2}{27} \\ \frac{7}{27} & \frac{20}{27} & \frac{2}{27} & -\frac{4}{27} \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad ; \quad \mathbf{M}_h^{-1}\mathbf{M}_l = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{11}{2} & 9 & -\frac{9}{2} & 1 \\ -1 & \frac{9}{2} & -9 & \frac{11}{2} \end{bmatrix}$$

$$\mathbf{M}_l^{-1}\mathbf{M}_s = \begin{bmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 \\ \frac{4}{81} & \frac{31}{54} & \frac{10}{27} & \frac{1}{162} \\ \frac{1}{162} & \frac{10}{27} & \frac{31}{54} & \frac{4}{81} \\ 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{bmatrix} \quad ; \quad \mathbf{M}_s^{-1}\mathbf{M}_l = \begin{bmatrix} \frac{25}{2} & -24 & \frac{33}{2} & -4 \\ -2 & \frac{15}{2} & -6 & \frac{3}{2} \\ \frac{3}{2} & -6 & \frac{15}{2} & -2 \\ -4 & \frac{33}{2} & -24 & \frac{25}{2} \end{bmatrix}$$

1.15 Example

Suppose we have four points $\mathbf{Q}_0$, $\mathbf{Q}_m$, $\mathbf{Q}_n$, $\mathbf{Q}_1$, and we want to calculate the Bézier control points of the cubic curve passing through them. Specifically, we want to find the Bézier control points $\mathbf{P}_0$, $\mathbf{P}_a$, $\mathbf{P}_b$, $\mathbf{P}_1$ of a curve $\mathbf{X}(u)$ that has $\mathbf{X}(0) = \mathbf{Q}_0$, $\mathbf{X}(\frac{1}{3}) = \mathbf{Q}_m$, $\mathbf{X}(\frac{2}{3}) = \mathbf{Q}_n$, and $\mathbf{X}(1) = \mathbf{Q}_1$. The equation of the curve can be written directly in Lagrange form:

$$\mathbf{X}(u) = [1 \quad u \quad u^2 \quad u^3] \cdot \mathbf{M}_l \cdot [\mathbf{Q}_0 \quad \mathbf{Q}_m \quad \mathbf{Q}_n \quad \mathbf{Q}_1]^\mathsf{T}$$

Alternatively, in Bézier form, it is

$$\mathbf{X}(u) = [1 \quad u \quad u^2 \quad u^3] \cdot \mathbf{M}_b \cdot [\mathbf{P}_0 \quad \mathbf{P}_a \quad \mathbf{P}_b \quad \mathbf{P}_1]^\mathsf{T}$$

Equating coefficients of powers of u, we get

$$\mathbf{M}_l \cdot [\mathbf{Q}_0 \quad \mathbf{Q}_m \quad \mathbf{Q}_n \quad \mathbf{Q}_1]^\mathsf{T} = \mathbf{M}_b \cdot [\mathbf{P}_0 \quad \mathbf{P}_a \quad \mathbf{P}_b \quad \mathbf{P}_1]^\mathsf{T}$$

so

$$[\mathbf{P}_0 \quad \mathbf{P}_a \quad \mathbf{P}_b \quad \mathbf{P}_1]^\mathsf{T} = \mathbf{M}_b^{-1}\mathbf{M}_l \cdot [\mathbf{Q}_0 \quad \mathbf{Q}_m \quad \mathbf{Q}_n \quad \mathbf{Q}_1]^\mathsf{T}$$

Using the numerical values of $\mathbf{M}_b^{-1}\mathbf{M}_l$ from section (1.14) above, we get $\mathbf{P}_0 = \mathbf{Q}_0$ and $\mathbf{P}_1 = \mathbf{Q}_1$, which was obvious from the beginning, and also

$$\begin{aligned} \mathbf{P}_a &= -\tfrac{5}{6}\mathbf{Q}_0 + 3\mathbf{Q}_m - \tfrac{3}{2}\mathbf{Q}_n + \tfrac{1}{3}\mathbf{Q}_1 \\ \mathbf{P}_b &= \tfrac{1}{3}\mathbf{Q}_0 - \tfrac{3}{2}\mathbf{Q}_m + 3\mathbf{Q}_n - \tfrac{5}{6}\mathbf{Q}_1 \end{aligned}$$

Note that the coefficients on the right-hand sides add up to 1, so these formulae make sense. From these formulae, we can see how changing $\mathbf{Q}_m$ and $\mathbf{Q}_n$ will affect $\mathbf{P}_a$ and $\mathbf{P}_b$. First of all, modifying either $\mathbf{Q}_m$ or $\mathbf{Q}_n$ will affect *both* $\mathbf{P}_a$ and $\mathbf{P}_b$. Also, since the coefficient of $\mathbf{Q}_n$ in the first formula is negative, this means that moving $\mathbf{Q}_n$ in one direction will cause $\mathbf{P}_a$ to move in the opposite direction. So, in short, trying to edit the curve by moving $\mathbf{Q}_m$ and $\mathbf{Q}_n$ might give somewhat surprising results.

1.16 Planar Rational Cubic Curves

It turns out that parametric cubic curves are not capable of exactly representing any conic section curves except parabolas. However, as suggested earlier, all conic section curves have rational quadratic parameterisations. Therefore, by using rational cubic parameterisations, we can represent both conics and parametric cubics in a single form. Planar rational cubic curves will take the form

$$x(t) = \frac{h(t)}{w(t)} \quad ; \quad y(t) = \frac{k(t)}{w(t)}$$

where h, k, w are cubic polynomials. If $w(t) = 1$ for all t, then the curve reduces to a polynomial cubic, of course. Any rational cubic curve (hence any polynomial cubic) can be implicitised, i.e. we can obtain an implicit equation that also represents it. For example, the analysis in [SED84] shows that the rational cubic curve

$$x(t) = \frac{-52t^3 + 63t^2 - 15t + 7}{-37t^2 + 3t + 1} \quad ; \quad y(t) = \frac{4}{-37t^2 + 3t + 1}$$

can also be represented by the implicit equation

$$448y^3 - 3936y^2 - 27x^2y + 270xy + 10713y - 10816 = 0$$

Parametric cubic and rational cubic curve segments are often joined end-to-end to form composite curves that are usually known as splines. Most of the spline curves used in contemporary CAD/CAM systems can be constructed this way, including Wilson-Fowler splines, β-splines, N-splines, and cubic NURB (non-uniform rational B-spline) curves. Thorough descriptions of spline curves and their uses are provided in [FAR88] and [BAR88].

1.17 Cubic Curves in Hermite Form

1.18 Problems With Lagrange and Bézier Blending