

Monotonic Bézier Curves for Animation

Edmund Kapusniak
March 2013

Cubic Bézier curves are often used in computer graphics as animation curves, which relate time to the value of some animated parameter. For an animation curve to be useful, it must produce a unique parameter value at each time value. We can refer to such curves as monotonic curves, since between the curve endpoints they generate a monotonically increasing time value. This paper derives the constraints that monotonic animation curves must satisfy. A solution for evaluating such a curve to find the parameter value at a particular time is also presented.

Bézier Curves

A cubic Bézier curve with endpoints P_0 and P_3 , and control points P_1 and P_2 has the familiar equation:

$$P(t) = (1-t)^3 P_0 + 3(1-t)^2 t P_1 + 3(1-t)t^2 P_2 + t^3 P_3$$

With its first derivative given by a quadratic Bézier curve called the hodograph:

$$\begin{aligned} H_0 &= 3(P_1 - P_0) \\ H_1 &= 3(P_2 - P_1) \\ H_2 &= 3(P_3 - P_2) \\ H(t) &= (1-t)^2 H_0 + 2(1-t)t H_1 + t^2 H_2 \end{aligned}$$

And second derivative by a line:

$$\begin{aligned} L_0 &= 2(H_1 - H_0) \\ L_1 &= 2(H_2 - H_1) \\ L(t) &= (1-t)L_0 + tL_1 \end{aligned}$$

Typically animation curves are displayed with time along the horizontal axis and the animated parameter along the vertical axis. We will refer to the time parameter as x and the animated parameter as y .

If we have a monotonic animation curve, there must be just one y value for each value of x . This means that the equation for x must be monotonic between P_0 and P_3 . For such curves, there is a unique value of the parametric variable t between 0 and 1 for each value of x between P_0 and P_3 . Given a particular time x , if we can find the corresponding value of t then we can simply evaluate the curve to find the value of the animated parameter y at time x .

We are therefore concerned with two problems:

- Is the curve monotonic in x between P_0 and P_3 ?
- Find the parametric variable t corresponding to a particular value of x .

A Simplifying Transformation

To simplify our analysis, we can reversibly transform any non-degenerate curve to place P_0 at the origin and P_3 on the line $x = 1$. The only variables remaining are the positions of the two control points P_1 and P_2 , which can be expressed using the horizontal distance from the curve endpoints. Letting f be the horizontal distance between the transformed P_0 and P_1 , and g the horizontal distance between the transformed P_2 and P_3 , our equations become:

$$P(t) = 3(1-t)^2tf + 3(1-t)t^2(1-g) + t^3$$

$$H_0 = 3f$$

$$H_1 = 3(1-f-g)$$

$$H_2 = 3g$$

$$L_0 = 6(1-2f-g)$$

$$L_1 = 6(f+2g-1)$$

Monotonicity

For our curve to be monotonic, the first derivative must be greater than zero at all points along the curve. As the first derivative is a quadratic Bézier curve, we know that:

- At $t = 0$, the derivative equals H_0 , or $3f$.
- At $t = 1$, the derivative equals H_2 , or $3g$.
- The derivative is a quadratic curve. Its graph takes the shape of a parabola. The derivative's maximum or minimum value occurs at its sole inflection point.

From the first two properties we can trivially construct our first two conditions for a monotonic animation curve: both f and g must be greater than zero. Looking at the control points of our original curve, we can see that this constraint forces the control points P_1 and P_2 to lie between the curve endpoints.

$$f > 0 \tag{1}$$

$$g > 0 \tag{2}$$

Next, we check whether the value of the derivative at its inflection point is a maximum or a minimum. At this inflection point, the second derivative of the original curve is zero. If the inflection point is a maximum, the second derivative is positive before the inflection point and negative afterwards. Since the second derivative is a line we check for a negative gradient:

$$\begin{aligned} L_1 - L_0 &< 0 \\ 3f + 3g - 2 &< 0 \\ g &< \frac{2}{3} - f \end{aligned} \tag{3}$$

If this condition holds the inflection point is a maximum and either H_0 or H_1 is the minimum value of the derivative between $0 < t < 1$. Since these values are greater than zero due to conditions 1 and 2, the derivative is greater than zero and the original curve is monotonic.

Otherwise, the derivative has a minimum value at the inflection point. We know that at the inflection point the second derivative is zero. We set $L(t) = 0$ and solve for t :

$$\begin{aligned} (1-t)L_0 + tL_1 &= 0 \\ 6(1-2f-g)(1-t) + 6(f+2g-1)t &= 0 \\ (1-2f-g) + (3f+3g-2)t &= 0 \\ t &= \frac{2f+g-1}{3f+3g-2} \end{aligned}$$

Using the location of the inflection point, we can construct two new conditions for a monotonic animation curve. If the minimum point lies outside the range $0 < t < 1$ then the value of the derivative inside that range is again bounded by the endpoints H_0 and H_1 and the curve is monotonic.

First, for $t < 0$. Notice that the denominator is the same expression as condition 3. Since we have already excluded the possibility of this expression being negative, it must be positive. Therefore the sign of the result depends solely on the sign of the numerator:

$$\begin{aligned}\frac{2f + g - 1}{3f + 3g - 2} &< 0 \\ 2f + g - 1 &< 0 \\ g &< 1 - 2f\end{aligned}\tag{4}$$

Similarly for $t > 1$:

$$\begin{aligned}\frac{2f + g - 1}{3f + 3g - 2} &> 1 \\ \frac{2f + g - 1}{3f + 3g - 2} - 1 &> 0 \\ \frac{1 - f - 2g}{3f + 3g - 2} &> 0 \\ 1 - f - 2g &> 0 \\ -2g &> f - 1 \\ g &< \frac{1}{2} - \frac{1}{2}f\end{aligned}\tag{5}$$

If either of these conditions is met, the curve is monotonic.

Finally, we consider the case where the derivative has a minimum point between $0 < t < 1$. If the value of the derivative at its minimum point is greater than zero, then the derivative is everywhere greater than zero and the curve is monotonic. Substituting our value for t back into $H(t)$ and simplifying:

$$\begin{aligned}H(t) &> 0 \\ (1 - t)^2 H_0 + 2(1 - t)tH_1 + t^2 H_2 &> 0 \\ 3f(1 - t)^2 + 6(1 - f - g)(1 - t)t + 3gt^2 &> 0 \\ (3f + 3g - 2)t^2 - 2(2f + g - 1)t + f &> 0 \\ (3f + 3g - 2)\left(\frac{2f + g - 1}{3f + 3g - 2}\right)^2 - 2(2f + g - 1)\left(\frac{2f + g - 1}{3f + 3g - 2}\right) + f &> 0 \\ \frac{f(3f + 3g - 2) - (2f + g - 1)^2}{3f + 3g - 2} &> 0 \\ \frac{-f^2 - g^2 + 2f + 2g - fg - 1}{3f + 3g - 2} &> 0\end{aligned}$$

And remembering that we have already checked $3f + 3g - 2$ for condition 3 and found it to be positive:

$$\begin{aligned}-f^2 - g^2 + 2f + 2g - fg - 1 &> 0 \\ f^2 + g^2 - 2f - 2g + fg + 1 &< 0 \\ g^2 + (f - 2)g + f^2 - 2f + 1 &< 0\end{aligned}$$

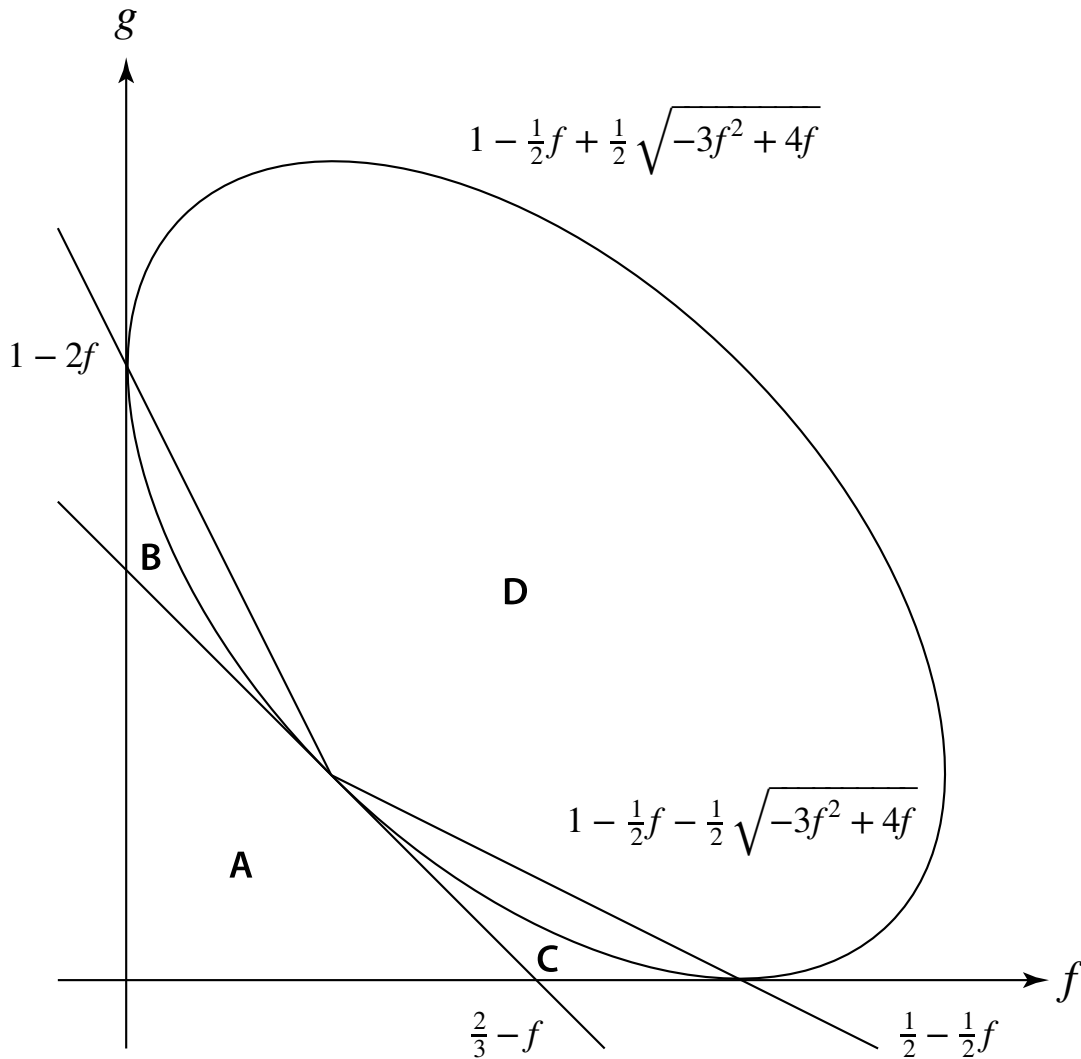
This equation is a conic section. We can use the quadratic formula to find conditions in terms of g . Notice that since the coefficient of g^2 is positive this equation has a minimum and is therefore less than zero between the two roots:

$$\begin{aligned}\frac{2 - f - \sqrt{(f - 2)^2 - 4(f^2 - 2f + 1)}}{2} &< g < \frac{2 - f + \sqrt{(f - 2)^2 - 4(f^2 - 2f + 1)}}{2} \\ 1 - \frac{1}{2}f - \frac{1}{2}\sqrt{-3f^2 + 4f} &< g < 1 - \frac{1}{2}f + \frac{1}{2}\sqrt{-3f^2 + 4f}\end{aligned}\tag{6a}$$

It is possible to express this as a single inequality:

$$(f + 2g - 2)^2 < 4f - 3f^2\tag{6b}$$

Drawing these constraints as lines on a graph, we can identify the limits of f and g for a monotonic curve:



Summarizing our conditions for a monotonic curve:

$$f > 0 \quad (1)$$

$$g > 0 \quad (2)$$

$$g < \frac{2}{3} - f \quad (3)$$

$$g < 1 - 2f \quad (4)$$

$$g < \frac{1}{2} - \frac{1}{2}f \quad (5)$$

$$1 - \frac{1}{2}f - \frac{1}{2}\sqrt{-3f^2 + 4f} < g < 1 - \frac{1}{2}f + \frac{1}{2}\sqrt{-3f^2 + 4f} \quad (6)$$

- Outside the first quadrant, conditions 1 or 2 are not met and the curve is not monotonic.
- In the triangle A, condition 3 is true. The derivative has a maximum and therefore the curve is monotonic.
- The area marked B represents condition 4. The derivative's inflection point lies at a point $t < 0$ and the curve is monotonic.
- The area marked C represents condition 5., The derivative's inflection point lies at a point $t > 1$ and the curve is monotonic.
- The final constraint forms the ellipse D. Here, the derivative's minimum is greater than zero, and the curve is monotonic.

Evaluating the Curve

We have our monotonic animation curve. We have a time, x . We want to know the value of the animated parameter, y , at this time. Essentially we need to find t when the horizontal component of our curve equals x . Once we have t it is simple to evaluate the vertical component of the the curve to find y .

Plugging the appropriate values into the equation for our Bézier curve:

$$\begin{aligned}
 P(t) &= x \\
 3(1-t)^2tf + 3(1-t)t^2(1-g) + t^3 &= x \\
 3(1-2t+t^2)tf + 3(1-t)t^2 - 3(1-t)t^2g + t^3 &= x \\
 3tf - 6t^2f + 3t^3f + 3t^2 - 3t^3 - 3t^2g + 3t^3g + t^3 &= x \\
 -2t^3 + 3t^3f + 3t^3g + 3t^2 - 6t^2f - 3t^2g + 3tf &= x \\
 (3f + 3g - 2)t^3 - (6f + 3g - 3)t^2 + 3ft &= x
 \end{aligned}$$

From here we can make things easier for ourselves by defining two derived values d and n . Note that a point $\begin{bmatrix} f & g \end{bmatrix}$ can be transformed to the point $\begin{bmatrix} d & n \end{bmatrix}$ by an affine transformation, and transformed back by the inverse transformation. dn space is a transformed version of fg space, which is useful to know. I will not present it here, but we could for example apply the transformation to our graph of the limits of f and g to find the limits in terms of d and n .

$$\begin{aligned}
 d &= 3f + 3g - 2 \\
 n &= 2f + g - 1
 \end{aligned}$$

$$\begin{bmatrix} d \\ n \\ 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 & -2 \\ 2 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f \\ g \\ 1 \end{bmatrix} \quad \begin{bmatrix} f \\ g \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} & 1 & \frac{1}{3} \\ \frac{2}{3} & -1 & \frac{1}{3} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} d \\ n \\ 1 \end{bmatrix}$$

Using d and n our equation can be expressed as:

$$dt^3 - 3nt^2 + 3ft = x$$

And to find t we must find the appropriate root of the cubic equation:

$$\begin{aligned}
 dt^3 - 3nt^2 + 3ft - x &= 0 \\
 t^3 - \frac{3n}{d}t^2 + \frac{3f}{d}t - \frac{x}{d} &= 0
 \end{aligned}$$

This equation can be reduced to a depressed cubic by substituting:

$$\begin{aligned}
 t &= u + \frac{n}{d} \\
 (u + \frac{n}{d})^3 - \frac{3n}{d}(u + \frac{n}{d})^2 + \frac{3f}{d}(u + \frac{n}{d}) - \frac{x}{d} &= 0 \\
 (u^2 + \frac{2nu}{d} + \frac{n^2}{d^2})(u + \frac{n}{d}) - \frac{3n}{d}(u^2 + \frac{2nu}{d} + \frac{n^2}{d^2}) + \frac{3fu}{d} + \frac{3fn}{d^2} - \frac{x}{d} &= 0 \\
 u^3 + \frac{3nu^2}{d} + \frac{3n^2u}{d^2} + \frac{n^3}{d^3} - \frac{3nu^2}{d} - \frac{6n^2u}{d^2} - \frac{3n^3}{d^3} + \frac{3fu}{d} + \frac{3fn}{d^2} - \frac{x}{d} &= 0 \\
 u^3 - \frac{3n^2u}{d^2} - \frac{2n^3}{d^3} + \frac{3fu}{d} + \frac{3fn}{d^2} - \frac{x}{d} &= 0 \\
 u^3 - (\frac{3n^2}{d^2} - \frac{3f}{d})u + (\frac{3fn}{d^2} - \frac{2n^3}{d^3} - \frac{x}{d}) &= 0 \\
 u^3 - (\frac{3n^2 - 3fd}{d^2})u + (\frac{3fdn - 2n^3}{d^3} - \frac{x}{d}) &= 0
 \end{aligned}$$

And once again we simplify our notation by introducing derived values:

$$r = \frac{n^2 - fd}{d^2}, \quad q = \frac{3fdn - 2n^3}{d^3} - \frac{x}{d}$$

$$u^3 - 3ru + q = 0$$

Our next step is to perform Vieta's substitution:

$$u = w + \frac{r}{w}$$

$$(w + \frac{r}{w})^3 - 3r(w + \frac{r}{w}) + q = 0$$

$$(w^2 + 2r + \frac{r^2}{w^2})(w + \frac{r}{w}) - 3r(w + \frac{r}{w}) + q = 0$$

$$w^3 + 2rw + \frac{r^2}{w} + rw + \frac{2r^2}{w} + \frac{r^3}{w^3} - 3rw - \frac{3r^2}{w} + q = 0$$

$$w^3 + q + \frac{r^3}{w^3} = 0$$

$$w^6 + qw^3 + r^3 = 0$$

This is a quadratic equation in w^3 and we can solve it using the quadratic formula:

$$w^3 = \frac{-q \pm \sqrt{q^2 - 4r^3}}{2}$$

In the case where the discriminant $q^2 - 4r^3$ is positive, then our equation has only one real root, which we can calculate directly by reversing our transformations. Using a non-zero solution for w^3 :

$$w = \sqrt[3]{w^3}$$

$$u = w + \frac{r}{w}$$

$$t = u + \frac{n}{d}$$

Alternatively, the discriminant may be negative, meaning that the equation has three real roots and making w^3 a complex number. This is the *casus irreducibilis*, as solving it involves taking the cube root of a complex number. Finding the cube root of a complex number is equivalent to angular trisection and there is no algebraic solution.

We can however find the cube root of a complex number trigonometrically. Given a complex number:

$$z = x + iy = r(\cos \theta + i \sin \theta)$$

There are three cube roots, each with the same absolute value of $\sqrt[3]{r}$. The principal cube root has argument $\frac{1}{3}\theta$, with the other two roots having arguments which split the circle centred on zero into three equal sectors. Letting $\tau = 2\pi$:

$$\sqrt[3]{z} = \sqrt[3]{r}(\cos(\frac{1}{3}\theta) + i \sin(\frac{1}{3}\theta))$$

$$\sqrt[3]{z} = \sqrt[3]{r}(\cos(\frac{1}{3}\theta + \frac{1}{3}\tau) + i \sin(\frac{1}{3}\theta + \frac{1}{3}\tau))$$

$$\sqrt[3]{z} = \sqrt[3]{r}(\cos(\frac{1}{3}\theta + \frac{2}{3}\tau) + i \sin(\frac{1}{3}\theta + \frac{2}{3}\tau))$$

It is also useful to review the following result for the reciprocal of a complex number:

$$\frac{1}{z} = \frac{x - iy}{x^2 + y^2} = \frac{r(\cos \theta - i \sin \theta)}{r^2}$$

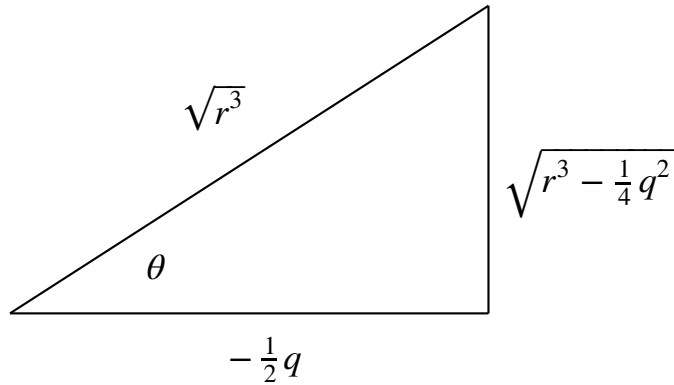
Armed with these tools, we first put w^3 into standard form:

$$w^3 = -\frac{1}{2}q \pm \frac{1}{2}\sqrt{q^2 - 4r^3}$$

$$w^3 = -\frac{1}{2}q \pm \sqrt{\frac{1}{4}\sqrt{-1}\sqrt{-q^2 + 4r^3}}$$

$$w^3 = -\frac{1}{2}q \pm i\sqrt{r^3 - \frac{1}{4}q^2}$$

And then into polar form:



$$w^3 = \sqrt{r^3}(\cos \theta + i \sin \theta)$$

$$\cos \theta = \frac{-\frac{1}{2}q}{\sqrt{r^3}}$$

$$\cos \theta = -\frac{q}{2\sqrt{r^3}}$$

$$\theta = \arccos\left(-\frac{q}{2\sqrt{r^3}}\right)$$

Now we can find the cube roots of w^3 :

$$w^3 = \sqrt{r^3}(\cos \theta + i \sin \theta)$$

$$\phi = \frac{1}{3}\theta, \quad \phi = \frac{1}{3}\theta + \frac{1}{3}\tau, \quad \phi = \frac{1}{3}\theta + \frac{2}{3}\tau$$

$$w = \sqrt[3]{\sqrt{r^3}}(\cos \phi + i \sin \phi)$$

$$w = \sqrt[3]{r}(\cos \phi + i \sin \phi)$$

Picking one value for ϕ we can find the corresponding root in terms of u . Even though w is complex, our root is real:

$$u = w + \frac{r}{w}$$

$$u = \sqrt[3]{r}(\cos \phi + i \sin \phi) + r\left(\frac{1}{\sqrt[3]{r}(\cos \phi + i \sin \phi)}\right)$$

$$u = \sqrt[3]{r}(\cos \phi + i \sin \phi) + r\left(\frac{\sqrt[3]{r}(\cos \phi - i \sin \phi)}{\sqrt{r^2}}\right)$$

$$u = \sqrt[3]{r}(\cos \phi + i \sin \phi) + \sqrt[3]{r}(\cos \phi - i \sin \phi)$$

$$u = 2\sqrt[3]{r}\cos \phi$$

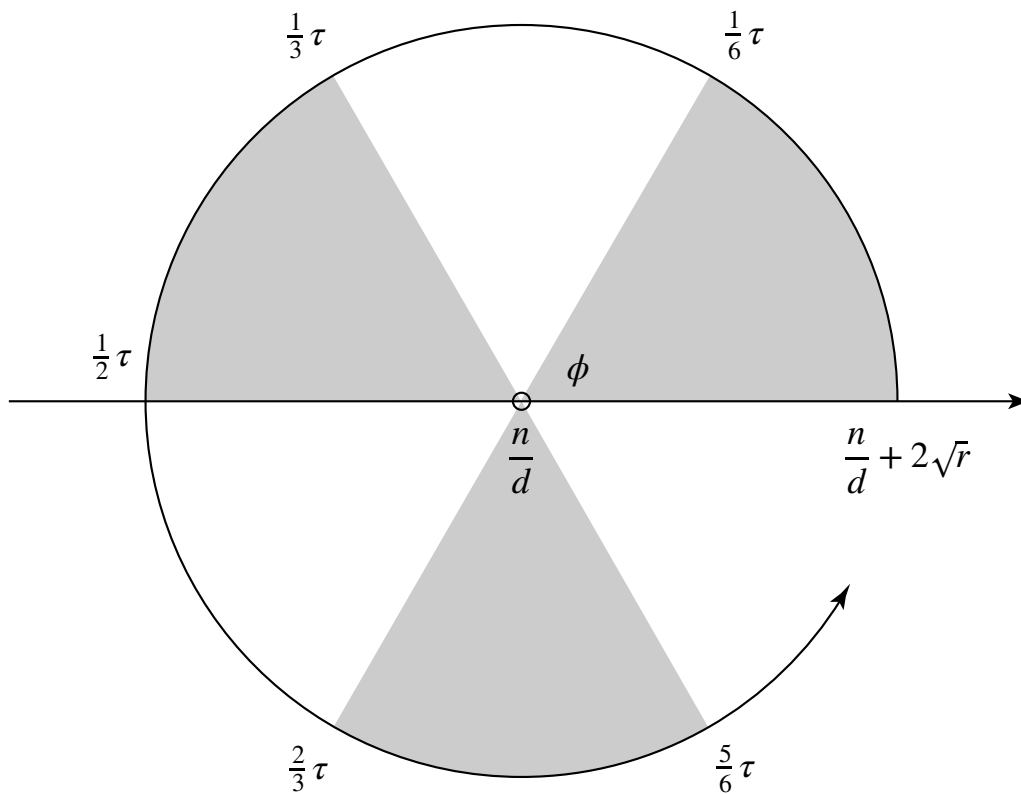
And again we can recover t:

$$t = u + \frac{n}{d}$$

There are some properties of the presented solution which require further analysis:

- If $d = 0$ then our solutions involve a division by zero. However in this case the original curve is quadratic or linear rather than cubic, and can be solved using the appropriate technique.
- Above, we picked the value of w^3 with positive i . The other value is the complex conjugate of this value and has the effect of flipping the sign of θ and ϕ . It doesn't matter which is picked as either gives the same value for the root - remember $\cos(-\theta) = \cos \theta$.
- In the case with three real roots, for u to be real $r > 0$. This follows directly from the discriminant $q^2 - 4r^3$. This discriminant is negative in the case with three real roots, and q^2 is obviously positive.

There is a final question: which of the three possible values for ϕ will lead to the desired root $0 < t < 1$? By experiment, it appears that the third value $\frac{1}{3}\theta + \frac{2}{3}\tau$ is the correct value to pick. We can build some intuitive support for this by representing the root visually. Notice that as θ is the result of an arccosine operation it lies in the range $0 < \theta < \frac{1}{2}\tau$, and therefore $0 < \frac{1}{3}\theta < \frac{1}{6}\tau$:



You can see that picking the third value produces roots centered around $\frac{n}{d}$, with the other two values producing more extreme solutions. It requires further work to prove that this is the desired root in all cases.

To summarize our procedure:

$$\begin{aligned}
 d &= 3f + 3g - 2 \\
 n &= 2f + g - 1 \\
 r &= \frac{n^2 - fd}{d^2} \\
 q &= \frac{3fdn - 2n^3}{d^3} - \frac{x}{d} \\
 \text{discriminant} &= q^2 - 4r^3
 \end{aligned}$$

If the discriminant is positive proceed, picking a non-zero solution for w^3 :

$$\begin{aligned}w^3 &= -\frac{1}{2}q \pm \frac{1}{2}\sqrt{q^2 - 4r^3} \\w &= \sqrt[3]{w^3} \\u &= w + \frac{r}{w} \\t &= u + \frac{n}{d}\end{aligned}$$

Otherwise, for a negative discriminant:

$$\begin{aligned}\theta &= \arccos\left(-\frac{q}{2\sqrt{r^3}}\right) \\ \phi &= \frac{1}{3}\theta + \frac{2}{3}\tau \\ u &= 2\sqrt{r}\cos\phi \\ t &= u + \frac{n}{d}\end{aligned}$$

Conclusion

Cubic Bézier curves used as animation curves must be monotonic. The analysis of such curves can be simplified by transforming them such that the first control point lies at zero and the last control point lies on the line $x = 1$. The two remaining control points must be constrained to lie within certain limits in order for the resulting curve to be horizontally monotonic.

Given a monotonic animation curve, standard approaches to the solution of cubic equations can be applied in order to find the parametric parameter at a particular time. In the case where the resulting equation has a single root, it can be solved by taking the cube root of a real number. In the case where the resulting equation has three real roots - the *casus irreducibilis* - it can be solved by finding the cube root of a complex number trigonometrically.

The problem of which of the three roots corresponds to the desired point on the original curve has been addressed experimentally.

Cubic Bézier curves are not often solved analytically, and it is hoped that presenting these solutions will give greater understanding and allow more options when approaching problems involving such curves.