OK. Here’s the deal. I’ve been playing with the question in the title on and off for more than a year now. I’m interested in both the answer to the question and in finding the clearest possible derivation of that answer. My first efforts were very algebra-intensive, but had an aura of concreteness about them. As I played with the question and went through innumerable drafts of this column, I came up with more and more elegant ways to ward off the brute-force algebra and make the answer more intuitive without reams of calculation. For these elegant techniques to work, however, I needed to build up a collection of tools (or Lemmas for the mathematically erudite) that themselves took a little time to explain.

The tool that I described in part 1 of this column (the use of a particular matrix identity, which I will review here) popped up just before I started to write. I had to scrap my original idea for part 1 and rewrite it to describe this new tool. Using this tool to answer our big question has required major surgery to what you are now reading as part 2. I’ve already given you the answer to the big question: three. But deriving that answer in a clear and obvious manner is still a little ways away. But that’s OK. The reason we’re doing this is really not so much to answer the question as to build our intuition about the algebraic formulas and the geometry that they represent. Let’s see, isn’t there a phrase that encapsulates this idea? Oh yes: The journey is the reward.

**Definition**

So, let’s reiterate what we’re dealing with: parametric cubic curves defined by the formula

\[
\begin{bmatrix} x(T) \\ y(T) \\ w(T) \end{bmatrix} = T^3 \begin{bmatrix} x_3 & y_3 & w_3 \\ x_2 & y_2 & w_2 \\ x_1 & y_1 & w_1 \\ x_0 & y_0 & w_0 \end{bmatrix} \]

(1)

We’re interested in all points generated by this equation, for all values of \( T \) from \(-\infty\) to \( +\infty \). Infinite parameter values are easier to manipulate if we use a homogeneous parameterization, something like expressing the parameter in 1D homogeneous coordinates. I’ll denote the homogeneous parameter coordinates as \([t, s]\) where \( T = t/s \). In terms of the homogeneous parameterization, the curve equation is

\[
\begin{bmatrix} x(t,s) \\ y(t,s) \\ w(t,s) \end{bmatrix} = t^3 \begin{bmatrix} x_3 & y_3 & w_3 \\ x_2 & y_2 & w_2 \\ x_1 & y_1 & w_1 \\ x_0 & y_0 & w_0 \end{bmatrix}
\]

I’ll bounce back and forth between the more familiar nonhomogeneous parameter \( T \) and homogeneous parameter representation \([t, s]\) as clarity dictates.

According to the homogeneous convention, any scalar multiple of a \([t, s]\) coordinate represents the same parametric point and will generate the same geometric point. To generate the whole curve we must trace out a curve in \([t, s]\) space that looks roughly semicircular around the origin. Why go only halfway around the origin? Well, if the curve starts at \([t_0, s_0]\) and ends at \([-t_0, -s_0]\), it will have returned to the geometric starting point. If you went around the parametric origin through a whole circle, you would trace out the curve twice.

The matrix of coefficients contains all the information about the curve, so I will give it a name:

\[
\begin{bmatrix} x_3 & y_3 & w_3 \\ x_2 & y_2 & w_2 \\ x_1 & y_1 & w_1 \\ x_0 & y_0 & w_0 \end{bmatrix} = C
\]

**Different means not the same**

Next, let’s concretize what I mean by different curves by turning the question around and defining what makes two curves the same. We’ll then sort all the parametric cubic curves into groups of same curves and see how many groups we have.

**The same via transformation**

In projective geometry, we’re interested in properties of shapes that remain constant even if the shape is subjected to a perspective transformation. The sorts of geometric properties that remain unchanged by such a transformation include intersections, tangency, and inflection points.

Algebraically we typically express a perspective trans-
formation as a homogeneous matrix multiplication:

$$\begin{bmatrix} x & y & w \end{bmatrix} M = \begin{bmatrix} \tilde{x} & \tilde{y} & \tilde{w} \end{bmatrix}$$

The coefficient matrix transforms in the same way:

$$CM = \tilde{C}$$

Accordingly, if we can find a nonsingular transformation matrix $M$ that changes one coefficient matrix $C$ into another one $\tilde{C}$, we’ll say that the curves generated by those matrices are the same.

**The same via reparametrization**

There’s another algebraic transformation we can perform on the curve equation that doesn’t change the curve shape at all—we can change the parameterization. This change is possible because we’re interested in the whole curve generated by the infinite parameter range $\rightarrow -\infty \leq T \leq +\infty$. For example, if we define

$$T = \alpha \hat{T} + c,$$

and run $\hat{T}$ from $\rightarrow -\infty$ to $+\infty$, we’ll also generate all $T$s from $\rightarrow -\infty$ to $+\infty$. To see how this affects the algebra of the coefficient matrix, we calculate

$$T^2 = a^2 \hat{T}^2 + 2ac \hat{T} + c^2$$

$$T^3 = a^2 \hat{T}^3 + 3a^2c \hat{T}^2 + 3ac^2 \hat{T} + c^3$$

and write these equations in matrix form:

$$\begin{bmatrix} T^3 & T^2 & T & 1 \end{bmatrix} =$$

$$\begin{bmatrix} a^3 & 0 & 0 & 0 \\
3a^2c & a^2 & 0 & 0 \\
3ac^2 & 2ac & a & 0 \\
c^3 & c^2 & c & 1 \end{bmatrix}$$

Plugging this into Equation 1 we can see that

$$\begin{bmatrix} T^3 & T^2 & T & 1 \end{bmatrix} C =$$

$$\begin{bmatrix} a^3 & 0 & 0 & 0 \\
3a^2c & a^2 & 0 & 0 \\
3ac^2 & 2ac & a & 0 \\
c^3 & c^2 & c & 1 \end{bmatrix}$$

Evaluating each of these for all values of $T$ and all values of $\hat{T}$ will generate exactly the same set of points. In other words, the coefficient matrices

$$C$$

and

$$\begin{bmatrix} a^3 & 0 & 0 & 0 \\
3a^2c & a^2 & 0 & 0 \\
3ac^2 & 2ac & a & 0 \\
c^3 & c^2 & c & 1 \end{bmatrix}$$

generate exactly the same curve without any geometric distortions. When we go to the homogeneous parameterization, we can use an even more general transformation, a 1D perspective (rational linear) transform of the form:

$$T = \frac{a \hat{T} + c}{b \hat{T} + d}.$$  \hfill (3)

A plot of this type of function looks like either a line or a hyperbola. Note that the function is still a one-to-one mapping of $\hat{T}$ to $T$ (including infinite values) and that it’s monotonic, since the derivative $dT/d\hat{T}$ has a constant sign.

Recasting this function in homogeneous parameter space we get

$$\begin{bmatrix} t & s \end{bmatrix} =$$

$$\begin{bmatrix} \alpha \hat{t} + c & \alpha \hat{c} \\
\beta \hat{t} + d & \beta \hat{d} \end{bmatrix}$$

Writing this as a matrix gives

$$\begin{bmatrix} t & s \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\
\gamma & \delta \end{bmatrix} \begin{bmatrix} \hat{t} & \hat{c} \end{bmatrix}$$

In other words, the reparametrization is a simple linear transformation of $[t, s]$ space. This is OK as long as the transformation is nonsingular, which is as long as $ad - bc \neq 0$.

To see how this reparametrization affects the form of the coefficient matrix, we calculate the following

$$t^3 = (a \hat{t} + c)^2$$

$$t^2 s = (a \hat{t} + c)(\beta \hat{t} + \delta)$$

$$ts^2 = (a \hat{t} + c)(\beta \hat{c} + \gamma)$$

$$s^3 = (\beta \hat{t} + \gamma)^2$$

Expand this mess all out and write it as a matrix product and we get

$$\begin{bmatrix} t^3 & t^2 s & ts^2 & s^3 \end{bmatrix} = \begin{bmatrix} t^3 & t^2 \hat{s} & \hat{t} \hat{s}^2 & \hat{s}^3 \end{bmatrix} R$$  \hfill (4)

where

$$R = \begin{bmatrix} a^3 & a b & a b^2 & b^3 \\
3a^2c & a(2bc + ad) & b(2bc + ad) & 3b^2d \\
3ac^2 & c(2b + ad) & d(2bc + ad) & 3bd^2 \\
c^3 & c^2d & c d^2 & d^3 \end{bmatrix}$$

Now we combine Equations 2 and 4 to get

$$\begin{bmatrix} x & y & w \end{bmatrix} = \begin{bmatrix} t^3 & t^2 \hat{s} & \hat{t} \hat{s}^2 & \hat{s}^3 \end{bmatrix} R C$$

We will say that any coefficient matrix $C$ generates the same curve as the matrix $RC$ for any $R$ constructed with any values $a, b, c, d$ such that $ad - bc \neq 0$. 
The game
So here’s what we’re going to do. We’re going to develop an algorithm that takes an arbitrary coefficient matrix \( \mathbf{C} \) and transforms it via a (nonsingular) reparameterization matrix \( \mathbf{R} \) and a (nonsingular) geometric transformation matrix \( \mathbf{M} \) to turn it into a canonical form:

\[
\mathbf{RCM} = \tilde{\mathbf{C}}
\]

We’ll devise this canonical form to be as simple as possible, with lots of zeroes and ones as elements. It will turn out that six possible resulting types of \( \tilde{\mathbf{C}} \) matrices exist, three that represent true cubics, and three that come from lower order curves disguised as cubics.

The canonical geometric transform
As a warm-up exercise, let’s start with an arbitrary second-order curve and simplify it via geometric transformations. We start with

\[
\begin{bmatrix}
x \\
y \\
w
\end{bmatrix} = \begin{bmatrix} \alpha & \beta & \gamma \\
\delta & \epsilon & \zeta \\
\omega & \eta & \xi \\
\end{bmatrix}
\]

Transforming by the matrix \( \mathbf{M} \) gives

\[
\begin{bmatrix}
\tilde{x} \\
\tilde{y} \\
\tilde{w}
\end{bmatrix} = \begin{bmatrix} \alpha & \beta & \gamma \\
\delta & \epsilon & \zeta \\
\omega & \eta & \xi \\
\end{bmatrix} \begin{bmatrix} x \\
y \\
w
\end{bmatrix}
\]

We want to find a matrix \( \mathbf{M} \) that gives us as simple a product as possible with the coefficient matrix. Gee, what could that be? How about simply picking \( \mathbf{M} \) as the inverse of the coefficient matrix. You would get

\[
\begin{bmatrix}
\tilde{x} \\
\tilde{y} \\
\tilde{w}
\end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

Any second-order curve with a nonsingular coefficient matrix is the same as this canonical curve, which happens to be a parabola. All conic sections are therefore transformations of this canonical parabola. This means that, according to the rules of our game, there is exactly one second-order curve.

There’s a bit of a gotcha though. We need to worry about the case where the coefficient matrix is singular. As it happens, this turns out to be a linear curve (OK, a line) in disguise, not a true second-order curve. I’ll deal with such cases in a separate column.

Now let’s see what happens when we try to simplify a cubic in a similar manner. We want to find a transformation \( \mathbf{M} \) that will simplify the coefficient matrix

\[
\begin{bmatrix}
\tilde{x} \\
\tilde{y} \\
\tilde{w}
\end{bmatrix} = \begin{bmatrix} t^2 & t s & s^2 \\
\end{bmatrix} \mathbf{M}
\]

We could shave off the top three rows of the coefficient matrix and transform by the inverse of this 3 by 3 matrix. As long as this matrix is nonsingular, we’ll get

\[
\begin{bmatrix}
\tilde{x} \\
\tilde{y} \\
\tilde{w}
\end{bmatrix} = \begin{bmatrix} t^2 & t s & s^2 \\
\end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

So we’ve boiled our cubic curve down to what, at first, looks like a three-parameter class of curves, depending on what we get for \( \tilde{x}_0, \tilde{y}_0, \tilde{w}_0 \). To understand what to do next, we must remember our tool from last time.

Inflection points revisited
Last time we showed that the inflection points of a cubic curve are at parameter values that are the solutions to a homogeneous cubic polynomial

\[
D_3 t^3 - D_2 t^2 s + D_1 t s^2 - D_0 s^3 = 0
\]

where the coefficients \( D_i \) are the determinants of the various 3 by 3 submatrices of the coefficient matrix:

\[
D_3 = \det \begin{bmatrix} x_2 & y_2 & w_2 \\
x_1 & y_1 & w_1 \\
x_0 & y_0 & w_0 \\
\end{bmatrix}
\]

\[
D_2 = \det \begin{bmatrix} x_3 & y_3 & w_3 \\
x_1 & y_1 & w_1 \\
x_0 & y_0 & w_0 \\
\end{bmatrix}
\]

\[
D_1 = \det \begin{bmatrix} x_3 & y_3 & w_3 \\
x_2 & y_2 & w_2 \\
x_0 & y_0 & w_0 \\
\end{bmatrix}
\]

\[
D_0 = \det \begin{bmatrix} x_3 & y_3 & w_3 \\
x_2 & y_2 & w_2 \\
x_1 & y_1 & w_1 \\
\end{bmatrix}
\]

Notice that the matrix defining \( D_0 \) is the one we inverted for the canonical transform above.

Back to the game
Now we can relate the values of \( \tilde{x}_0, \tilde{y}_0, \tilde{w}_0 \) to the values of \( D_i \). We start by noting that each row of the coefficient matrix is a homogeneous point in the plane. I’ll give names to these row vectors as follows (I include the ellipses to emphasize that the points are row vectors):

\[
\begin{bmatrix} x_3 & y_3 & w_3 \\
x_2 & y_2 & w_2 \\
x_1 & y_1 & w_1 \\
x_0 & y_0 & w_0 \\
\end{bmatrix} = \begin{bmatrix} \cdots \mathbf{p}_3 \cdots \\
\cdots \mathbf{p}_2 \cdots \\
\cdots \mathbf{p}_1 \cdots \\
\cdots \mathbf{p}_0 \cdots \\
\end{bmatrix}
\]

It’s interesting to note that the top and bottom rows are points on the curve itself. \( \mathbf{p}_0 \) is the point at the homogeneous parameter value \( 0, 1 \) (or at \( T = 0 \) in the nonhomogeneous formulation) and \( \mathbf{p}_1 \) is the point at the homogeneous parameter value \( 1, 0 \) (or at \( T = \infty \) in the nonhomogeneous formulation). Furthermore, the line through \( \mathbf{p}_0 \) and \( \mathbf{p}_1 \) is tangent to the curve at \( \mathbf{p}_0 \). And the line through \( \mathbf{p}_2 \) and \( \mathbf{p}_3 \) is tangent to the curve at \( \mathbf{p}_3 \).
Anyway, we can now rewrite the definition of the elements of the $D$ vector from Equation 6 as

\[
D_0 = \mathbf{p}_2 \cdot \mathbf{p}_1 \times \mathbf{p}_0 \\
D_2 = -\mathbf{p}_3 \cdot \mathbf{p}_1 \times \mathbf{p}_0 \\
D_3 = \mathbf{p}_3 \cdot \mathbf{p}_2 \times \mathbf{p}_0 \\
D_0 = -\mathbf{p}_3 \cdot \mathbf{p}_2 \times \mathbf{p}_1
\]

And the geometric transformation matrix we were using as our canonical transformation is simply

\[
\mathbf{M} = \begin{bmatrix}
\mathbf{p}_1 \\
\mathbf{p}_2 \\
\mathbf{p}_3 \\
\end{bmatrix}^{-1}
\]

We can evaluate the matrix inverse from the cross products of the various row vectors. Each cross product gives a column vector, and the inverse is

\[
\begin{bmatrix}
\mathbf{p}_2 - \mathbf{p}_1 \\
\mathbf{p}_3 - \mathbf{p}_1 \\
\mathbf{p}_3 - \mathbf{p}_2 \\
\end{bmatrix}^{-1} = \frac{-1}{D_0}
\begin{bmatrix}
\mathbf{p}_2 \times \mathbf{p}_1 \\
\mathbf{p}_1 \times \mathbf{p}_3 \\
\mathbf{p}_2 \times \mathbf{p}_3 \\
\end{bmatrix}
\]

As long as this matrix is nonsingular (that is, the determinant $D_0$ is nonzero), we can see that our canonical transformation will generate

\[
\begin{bmatrix}
\mathbf{p}_2 \\
\mathbf{p}_3 \\
\mathbf{p}_1 \\
\mathbf{p}_0 \\
\end{bmatrix} = \frac{-1}{D_0}
\begin{bmatrix}
\mathbf{p}_2 \times \mathbf{p}_1 \\
\mathbf{p}_1 \times \mathbf{p}_3 \\
\mathbf{p}_2 \times \mathbf{p}_3 \\
\mathbf{p}_0 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

This, then, gives the meaning of the bottom row of the candidate canonical matrix. If, however, the original curve gave us a value of $D_0 = 0$, this is not a good choice for a canonical transformation. All is not lost, however. We can pick a transformation matrix as the inverse of the other subsets of three of the four rows and get the following possibilities for canonical coefficient matrices:

\[
\begin{bmatrix}
\mathbf{p}_3 \\
\mathbf{p}_2 \\
\mathbf{p}_1 \\
\mathbf{p}_0 \\
\end{bmatrix}^{-1} = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
D_3 & -D_2 & D_0 \\
D_1 & -D_2 & D_0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

Which of these is best to choose will depend on what values we have for the $D_i$'s.

We have therefore been able to recast the question “How many different parametric cubic curves are there?” as the question “How many different homogeneous cubic polynomials are there?” We’ve gone from needing to understand the 12 numbers in the coefficient matrix to needing to understand the four coefficients in a homogeneous cubic polynomial.

**How does the game affect $D$?**

The $D$ vector determines where the inflection points are. The reason that inflection points are interesting is that the number of them will not change due to perspective transformations or due to reparametrization. How does this fact show up in the algebra of transformation that we just defined?

**Geometric**

When we transform $\mathbf{C}$ by a nonsingular geometric transformation $\mathbf{M}$ to the new matrix $\tilde{\mathbf{C}} = \mathbf{C} \mathbf{M}$, what will be the elements of the $D$ vector for this new coefficient matrix? Remember that the determinant of a product is the product of the determinants. So the determinant of any three rows of $\tilde{\mathbf{C}}$ will equal the determinant of the same three rows of $\mathbf{C}$ times the determinant of $\mathbf{M}$. Appyling this to each 3 by 3 submatrix of $\mathbf{C}$ and $\tilde{\mathbf{C}}$ gives

\[
\begin{bmatrix}
\tilde{D}_3 & \tilde{D}_2 & \tilde{D}_1 & \tilde{D}_0 \\
\end{bmatrix} = (\det \mathbf{M})
\begin{bmatrix}
D_3 & D_2 & D_1 & D_0 \\
\end{bmatrix}
\]

In other words, a geometric transformation of the curve will just apply a uniform homogeneous scale of $\det \mathbf{M}$ to the $D$ vector. It will not change the location of the roots of the $D$ polynomial, the parameter values at the inflection points. This pretty much makes sense.

**Reparametrization**

Reparametrization doesn’t change the geometry of the curve at all, so inflection points stay put geometrically, while they may move around parametrically. To see where they move to, let’s rewrite Equation 5 as

\[
D_0 t^3 - 3D_1 t^2 s + 3D_2 t s^2 - D_3 s^3 = 0
\]

\[
\begin{bmatrix}
\mathbf{p}_3 \\
\mathbf{p}_2 \\
\mathbf{p}_1 \\
\mathbf{p}_0 \\
\end{bmatrix}
\begin{bmatrix}
D_0 \\
D_1 \\
D_2 \\
D_3 \\
\end{bmatrix}
\begin{bmatrix}
D_0 \\
D_1 \\
D_2 \\
D_3 \\
\end{bmatrix} = 0
\]

IEEE Computer Graphics and Applications 91
If we reparametrize this using Equation 4, we get

\[
\begin{bmatrix}
  t^3 & t^2 s & t s^2 & s^3
\end{bmatrix} R \begin{bmatrix}
  D_0 \\
  -3D_1 \\
  3D_2 \\
  -D_3
\end{bmatrix} = 0
\]

In other words, the elements of the reparametrized matrix \( D \) are

\[
\begin{bmatrix}
  D_0 \\
  -3D_1 \\
  3D_2 \\
  -D_3
\end{bmatrix} = \hat{R} \begin{bmatrix}
  \hat{D}_0 \\
  -3\hat{D}_1 \\
  3\hat{D}_2 \\
  -\hat{D}_3
\end{bmatrix}
\]

If you plot the function \( f(t, s) = D_0 t^3 - 3D_1 t^2 s + 3D_2 t s^2 - D_3 s^3 \) and consider what reparametrization does to it, you can see that it’s just a rotation and shear of the function in parameter space. It mangles the shape, but one thing it doesn’t do is to change the number and multiplicity of roots of the polynomial.

**Winning the game**

The next order of business is to immerse ourselves in what types of homogeneous cubic polynomials exist under reparametrization. Our goal will be to pick a reparametrization matrix that makes the new values \( \hat{D}_i \) as simple as possible. We’ll tackle that in the next millennium.

Readers may contact Blinn by e-mail at blinn@microsoft.com.