CONTROLLING THE SHAPE OF PARAMETRIC B-SPLINE AND BETA-SPLINE CURVES

Brian A. Barsky

Berkeley Computer Graphics Laboratory
Computer Science Division
University of California
Berkeley, California 94720 USA
(415) 642-9838

John C. Beatty

Computer Graphics Laboratory
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
(519) 886-1351

ABSTRACT

Piecewise linear approximations to curves and surfaces have many disadvantages which representational techniques based on B-splines or Beta-splines alleviate. Unfortunately the very flexibility that makes these tools so useful can also lead to confusion about the effects that can be achieved by manipulating the parameters intrinsic to them. We first review pertinent definitions and properties of these techniques, and then discuss the effect of manipulating these parameters on the shape of parametric curves defined using B-splines and Beta-splines.

KEYWORDS: Beta-splines, B-splines, computer aided design, curves, geometric continuity.

1. Introduction

While irregularly curved lines and surfaces can, of course, be arbitrarily well approximated by straight line segments and polygons, there are many advantages to the use of modeling primitives that are themselves curved. Early work of this sort was done by Steven Coons [Coons64a, Coons67a] and Pierre Bézier [Bézier70a, Bézier77a], and the mathematical techniques that they developed are still used in production software. More recently Richard Riesenfeld has advocated the representation of piecewise polynomial curves and surfaces as linear combinations of B-splines [Riesenfeld73a, Gordon74a].

Among the advantages of parametric B-splines are: efficient computation; the ability to control the degree of continuity at the *joints* $\mathbf{Q}_i(\bar{u}_i)$ between the adjacent *curve segments* from which a curve is assembled (and at the borders between the adjacent *patches* forming a surface) independent of the order of the segments or the number of *control vertices* being approximated; the ability to independently vary the order of the polynomials forming these segments or patches and the number of control vertices; and *local control*. Indeed, the extensive flexibility inherent to the B-spline representation makes it desirable to distinguish the effects on a curve of manipulating each independent parameter separately from the others. Delineating such effects is the goal of the work reported here, much of which is based on [Gordon74a].

This work was supported by funds from the United States National Science Foundation (Grant ECS-82-04381), the Defense Advanced Research Projects Agency of the United States (Contract N00039-82-C-0235), the Natural Sciences and Engineering Research Council of Canada (Grants A3022 & G0651), and the National Research Council of Canada (Contract O5SU.31155-1-3103).

In the next section we will briefly recall pertinent definitions and theorems. Some of this material may also be found in [Newman73, pp 320-325], [Foley82a, pp 521-523], [Rogers76a, pp 144-155], [Faux79a, pp 162-169 & 178-183] and [Barsky82a]. A detailed treatment of splines, though not in the context of computer graphics, may be found in [deBoor78a]. A more intuitive treatment of the mathematics involved and the use of parametric B-splines for graphical applications is provided by [Barsky83a].

The remainder of this paper will be concerned with exploring the geometric effects of manipulating the various parameters available in B-spline and Beta-spline curves, the latter being a recent generalization of the uniform cubic B-splines which provides two additional local *shape parameters* $\beta 1$ and $\beta 2$ with which a piecewise cubic curve may be manipulated [Barsky81a, Barsky82b, Barsky83b].

2. Notation and Definitions

For our purposes a two-dimensional curve such as that shown in Figure 1 is represented parametrically as $\mathbf{Q}(\bar{u}) = (X(\bar{u}), Y(\bar{u}))$, where $X(\bar{u})$ and $Y(\bar{u})$ are single-valued *piecewise polynomial* functions of the parameter $\bar{u}$ which yield the x- and y-coordinates, respectively, of a point on the curve for any value of $\bar{u}$. That is, $X(\bar{u})$ and $Y(\bar{u})$ consist of some $m > 0$ pieces called *segments*, each of which is defined by separate polynomials $X_i(\bar{u})$ and $Y_i(\bar{u})$. If

$$\bar{u}_0 \leq \bar{u}_1 \leq \cdots \leq \bar{u}_{m-1} \leq \bar{u}_m \quad (1)$$

then we define the i^{th} segment $\mathbf{Q}_i(\bar{u})$ to be the set of points $(X_i(\bar{u}), Y_i(\bar{u}))$ for $\bar{u}_{i-1} \leq \bar{u} < \bar{u}_i$. The values $\bar{u}_0, \cdots, \bar{u}_m$ are called *parametric knots*; the distinct values in (1) are called *break-*

points. Usually $X(\bar{u})$ and $Y(\bar{u})$ are required to satisfy continuity constraints at the joints between the successive polynomial segments forming the curve; if the 0^{th} through d^{th} derivatives are everywhere continuous (in particular, at the joints), then $X(\bar{u})$ and $Y(\bar{u})$ are said to be C^d continuous. Sometimes the knots are equally spaced; this is called a *uniform knot sequence*, in which case we often have $\bar{u}_i = i$.

It will usually be simpler to write down $X(\bar{u}-\bar{u}_{i-1})$ rather than $X(\bar{u})$, and we shall generally do so; the reparametrization is easily accomplished by substitution. We shall write $Y_i(u)$ when we are parametrizing Y_i from the left end of the i^{th} interval (at which $u=0$ and $\bar{u}=u_{i-1}$), and $Y_i(\bar{u})$ when referring to a single parametrization of the entire curve. Thus both of the following

$$Y_i(u) = u^2$$

$$Y_i(\bar{u}) = (\bar{u}-\bar{u}_{i-1})^2 = \bar{u}^2 - 2\bar{u}_{i-1}\bar{u} + \bar{u}_{i-1}{}^2$$

define exactly the same segment, although the equations are written (and interpreted) differently.

For the techniques of interest here, one specifies a sequence of *control vertices* $V_i = (x_i, y_i)$, represented in our illustrations by a "+" sign or a dot "o", near which the curve is to pass. (See Figure 2.) Moving one or more of the control vertices then alters the curve, as in Figure 3. We have connected the control vertices together with lightly dotted lines in Figures 2 and 3 to form what is called the *control polygon*, indicating the order in which the control vertices are approximated.

2.1. B-splines

An important property of the B-splines (and of the Beta-splines discussed later) is their *locality*; movement of a control vertex alters the shape of only a small and well-defined portion of the curve. In this way one part of a curve may be manipulated without altering other parts which have already been satisfactorily designed. Locality is obtained because an order k (degree $k-1$) curve $Q(\bar{u})$ is represented as a "weighted sum"

$$\sum_i V_i B_{i,k}(\bar{u}) = \left(\sum_i x_i B_{i,k}(\bar{u}),\ \sum_i y_i B_{i,k}(\bar{u}) \right) \qquad (2)$$

of *basis functions* $B_{i,k}(\bar{u})$ such as the one shown in Figure 4, and each $B_{i,k}(\bar{u})$ has *local support*; *i.e.*, it is nonzero only for a small range of $\bar{u}$. Figure 5 shows, for example, how $Y(\bar{u})$ in Figure 2 is thus represented. For the B-splines of order k, with which we can represent any piecewise polynomial whose segments are order k polynomials, $B_{i,k}(\bar{u})$ is nonzero only for $\bar{u}$ in the open interval $(\bar{u}_i, \bar{u}_{i+k})$. For the i^{th} curve segment of a B-spline curve of order k (which is defined on $[\bar{u}_{i-1}, \bar{u}_i)$), only the k basis functions $B_{i-k,k}(\bar{u}), \cdots, B_{i-1,k}(\bar{u})$ are nonzero. (See Figure 6.) Since the weights in (2) are simply the coordinates of the control vertices, changing any particular control vertex alters only the k adjacent curve segments over which the support of the basis function weighted by that control vertex extends. Thus for $\bar{u}_{i-1} \le \bar{u} < \bar{u}_i$, equation (2) becomes

$$Q_i(\bar{u}) = \sum_{r=-k}^{r=-1} V_{i+r} B_{i+r,k}(\bar{u}) \qquad (3)$$

and for a cubic B-spline ($k=4$),

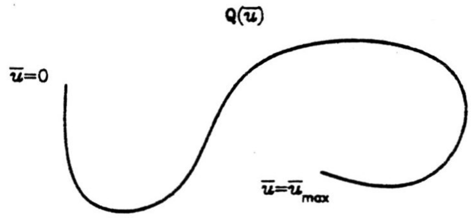

Figure 1. A piecewise polynomial cubic curve.

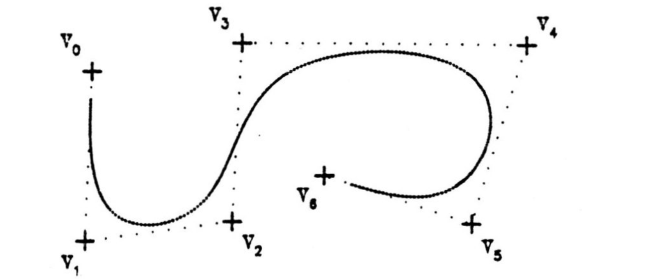

Figure 2. Using control vertices to define a curve. The control vertices are shown with a "+" and the control polygon is shown as lightly dotted lines connecting the control vertices. Every other segment is shown dotted.

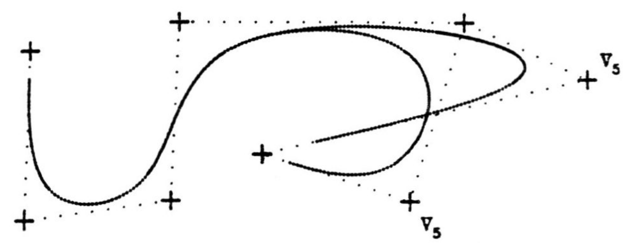

Figure 3. The effect of moving a control vertex.

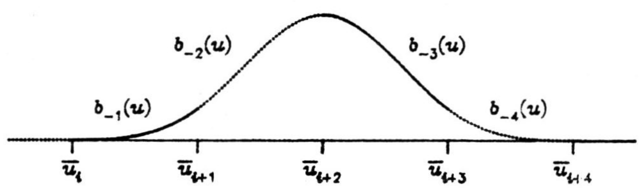

Figure 4. A uniform cubic (order 4) B-spline basis function. It is zero for $\bar{u} \le \bar{u}_i$ and for $\bar{u} \ge \bar{u}_{i+4}$.

$$Q_i(\bar{u}) = V_{i-4}B_{i-4,4}(\bar{u}) + V_{i-3}B_{i-3,4}(\bar{u}) + V_{i-2}B_{i-2,4}(\bar{u}) + V_{i-1}B_{i-1,4}(\bar{u})$$

If we replace each basis function $B_{i+r,k}(\bar{u})$ by the particular polynomial piece $b_{i+r,k,r}(u)$ (which we call a *basis segment*) of $B_{i+r,k}(\bar{u})$ pertaining to the interval $[\bar{u}_{i-1}, \bar{u}_i)$, then (3) can be written as

$$\mathbf{Q}_i(u) = \mathbf{V}_{i-4}b_{i-4,4,-4}(u) + \mathbf{V}_{i-3}b_{i-3,4,-3}(u) \quad (4)$$
$$+ \mathbf{V}_{i-2}b_{i-2,4,-2}(u) + \mathbf{V}_{i-1}b_{i-1,4,-1}(u)$$

For simplicity we will drop the k from $B_{i,k}$, and k or i,k from $b_{i,k,r}$, when no confusion can result because all of the basis functions have the same shape, and in the last case when the order of the curve is clear.

For an arbitrary knot sequence

$$\bar{u}_0 \leq \bar{u}_1 \leq \cdots \leq \bar{u}_m \leq \cdots \leq \bar{u}_{m+k} \quad (5)$$

(the knots $u_{m+1}, \cdots, u_{m+k}$ are needed to properly define $B_{m,k}(\bar{u})$), the i^{th} B-spline basis function of order k, written $B_{i,k}(\bar{u})$, is given recursively by

$$\frac{\bar{u} - \bar{u}_i}{\bar{u}_{i+k-1} - \bar{u}_i} B_{i,k-1}(\bar{u}) + \frac{\bar{u}_{i+k} - \bar{u}}{\bar{u}_{i+k} - \bar{u}_{i+1}} B_{i+1,k-1}(\bar{u}) \quad (6)$$

where

$$B_{i,1}(\bar{u}) = \begin{cases} 1 & \text{if } \bar{u}_i \leq \bar{u} < \bar{u}_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

Here $(\bar{u} - \bar{u}_i)/(\bar{u}_{i+k-1} - \bar{u}_i)$ is agreed to be 0 if $\bar{u}_i = \bar{u}_{i+k-1}$ (and similarly for the second term); $B_{i,k-1}(\bar{u})$ is in any case 0 for $\bar{u} \leq \bar{u}_i$ and $\bar{u} \geq \bar{u}_{i+k-1}$, while for $\bar{u}_i \leq \bar{u} \leq \bar{u}_{i+k-1}$, $(\bar{u} - \bar{u}_i)/(\bar{u}_{i+k-1} - \bar{u}_i) = 0/0$, so that this is reasonable. Definition (6) lends itself readily to efficient and accurate evaluation [deBoor78a, Barsky83a]. Its expansion for a particular $B_{i,k}(\bar{u})$ results in a sequence of k order k polynomials $b_{i,k,-1}(u), \cdots, b_{i,k,-k}(u)$, one for each of the half-open intervals $[\bar{u}_i, \bar{u}_{i+1}), \cdots, [\bar{u}_{i+k-1}, \bar{u}_{i+k})$, As de Boor points out [deBoor78a, p 131], it is a remarkable fact that if the knots in (5) are all distinct, so that the sequence is strictly increasing, then equation (6) defines a sequence of k non-trivial order k polynomials that meet in such a way as to be C^{k-2} continuous. Moreover, if a particular knot value is *repeated* some number $\rho < k$ times in (5) (has *muliplicity* $\rho + 1 \leq k$) so that for some i,

$$\bar{u}_{i-1} < \bar{u}_i = \bar{u}_{i+1} = \cdots = \bar{u}_{i+\rho} < \bar{u}_{i+\rho+1}$$

then any of the curves which can be defined *via* (2), with the particular knot sequence in question, are only $C^{k-2-\rho}$ continuous at $\bar{u} = \bar{u}_i$. This follows from the fact that if $\bar{u}_i$ is repeated $\sigma \leq \rho$ times within the $k+1$ knots $\bar{u}_j, \cdots, \bar{u}_{j+k}$ embracing the support of some $B_{j,k}(\bar{u})$, then $B_{j,k}(\bar{u})$ is only $C^{k-2-\sigma}$ continuous at $\bar{u} = \bar{u}_i$. For $\rho < k$ it follows that at least $B_{i-1,k}(\bar{u})$ and $B_{i,k}(\bar{u})$ are only $C^{k-2-\rho}$ continuous at $\bar{u} = \bar{u}_i$.

For example, if (5) is the uniform knot sequence in which $u_i = i$ and $k = 4$ (the *uniform cubic B-splines*), then all of the basis functions have the shape shown in Figure 4, and the four cubic polynomials $b_{-1}(u)$, $b_{-2}(u)$, $b_{-3}(u)$ and $b_{-4}(u)$ which (6) yields for $B_{i,4}(\bar{u})$ are given by

$$b_{-1}(u) = \frac{1}{6}\left(u^3\right) \quad (7)$$
$$b_{-2}(u) = \frac{1}{6}\left(1 + 3u + 3u^2 - 3u^3\right)$$
$$b_{-3}(u) = \frac{1}{6}\left(4 - 6u^2 + 3u^3\right)$$
$$b_{-4}(u) = \frac{1}{6}\left(1 - 3u + 3u^2 - u^3\right)$$

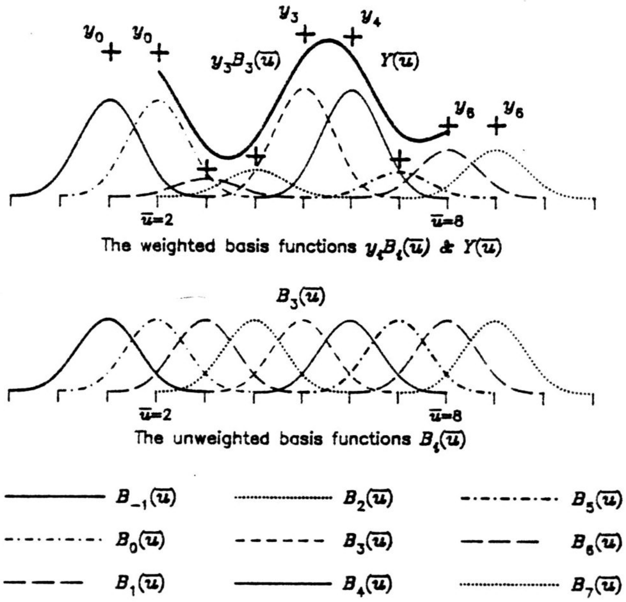

Figure 5. $Y(\bar{u})$ as a weighted sum of basis functions, the weights being the y coordinates of the control vertices. $X(\bar{u})$ is represented analogously, as is $Z(\bar{u})$ for three dimensional curves.

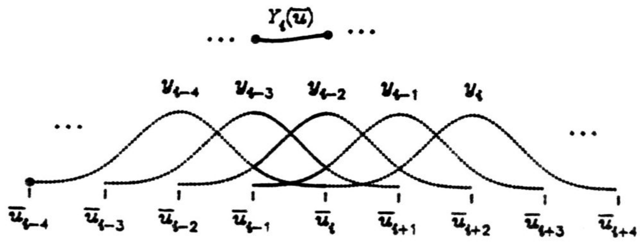

Figure 6. Illustrating our indexing conventions for $B_{i,4}(\bar{u})$.

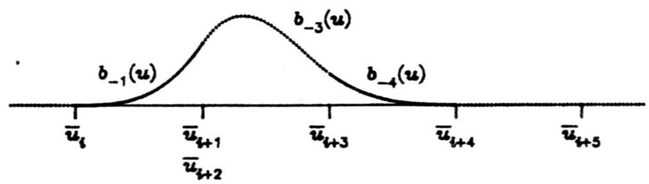

Figure 7. A knot of multiplicity 2.

and are C^2 continuous. The reader may easily verify, for example, that the second derivative of $b_{-1}(u)$ at $\bar{u}_{i+1}$ (written $b_{-1}^{(2)}(1)$) and of $b_{-2}(u)$ at $\bar{u}_{i+1}$ (written $b_{-2}^{(2)}(0)$) both have the value 1.0, while the third derivatives at $\bar{u}_{i+1}$ are not equal. If we alter the uniform knot sequence so that $u_{i+1} = u_{i+2}$ then we obtain the basis function shown in Figure 7. The basis segments which (6) now yields for $B_{i,4}(\bar{u})$ are

$$b_{i,-1}(u) = \frac{1}{2}\left(u^3\right)$$

$b_{i,-2}(u)$ *is defined on a vacuous interval*

$$b_{i,-3}(u) = \frac{1}{4}\left(2 + 6u - 12u^2 + 5u^3\right)$$
$$b_{i,-4}(u) = \frac{1}{4}\left(1 - 3u + 3u^2 - u^3\right)$$

These segments define a basis function which is $C^{4-2-1} = C^1$ continuous at $\bar{u} = \bar{u}_{i+1}$, but not C^2 continuous. The first derivatives are

$$b^{(1)}_{i,-1}(u) = \frac{3}{2}\left(u^2\right)$$

$b^{(1)}_{i,-2}(u)$ *is defined on a vacuous interval*

$$b^{(1)}_{i,-3}(u) = \frac{1}{4}\left(6 - 24u + 15u^2\right)$$

$$b^{(1)}_{i,-4}(u) = \frac{1}{4}\left(-3 + 6u - 3u^2\right)$$

and, as we promised, $b^{(1)}_{i,-1}(1) = b^{(1)}_{i,-3}(0) = 1.5$ and $b^{(1)}_{i,-3}(1) = b^{(1)}_{i,-4}(0) = -0.75$. The second derivatives are

$$b^{(2)}_{i,-1}(u) = 3u$$

$b^{(2)}_{i,-2}(u)$ *is defined on a vacuous interval*

$$b^{(2)}_{i,-3}(u) = \frac{1}{4}\left(-24 + 30u\right)$$

$$b^{(2)}_{i,-4}(u) = \frac{1}{4}\left(6 - 6u\right)$$

We see that $b^{(2)}_{i,-1}(1) = 3$ while $b^{(2)}_{i,-3}(0) = -6$, so that $B_{i,k}(\bar{u})$ has a discontinuous second derivative at $\bar{u} = \bar{u}_{i+1}$. Because $\bar{u}_{i+3}$ is not repeated, $B_{i,4}(\bar{u})$ does have a continuous second derivative at $\bar{u} = \bar{u}_{i+3}$ (the next breakpoint), and indeed $b^{(2)}_{i,-3}(1) = b^{(2)}_{i,-4}(0) = 1.5$.

Figure 8 shows a fourth order basis function containing a knot of multiplicity three (C^0, or positional continuity) and Figure 9 shows a fourth order basis function containing a knot of multiplicity four (no continuity). Notice that in each case the basis function, which is piecewise cubic, is nonzero over four intervals, namely for $u \, \varepsilon \, [\bar{u}_i, \bar{u}_{i+4})$, although if a knot has been repeated ρ times then ρ of these intervals are vacuous. In the remaining intervals the basis segments are still cubic, although the two cubics which meet at the repeated knot meet only with $C^{4-2-\rho}$ continuity. Each additional time a knot is repeated the parametric continuity of the underlying basis functions, and hence the parametric continuity of any curve they define, is reduced by one degree.

For an order k B-spline we may usefully repeat a knot at most $k-1$ times, so that it has multiplicity at most k. Since the basis functions and curves initially have $k-2$ continuity, a knot of multiplicity k, which is repeated $k-1$ times, will result in $C^{k-2-(k-1)} = C^{-1}$ continuity — that is, no continuity whatsoever.

It is a fact that for an arbitrary $\bar{u}$, $\bar{u}_{i-1} \le \bar{u} < \bar{u}_i$,

$$\sum_j B_{j,k}(\bar{u}) = \sum_{r=-k}^{r=-1} B_{i+r,k}(\bar{u}) = 1 \tag{8}$$

and each of the $B_{j,k}(\bar{u})$ is always non-negative, so that equation (3) defines $Q_i(\bar{u})$ to be a *convex combination* of k control vertices. It is then easy to show that the B-splines have the *convex hull property*: any particular curve segment lies within the convex hull of the k control vertices that define it. Intuitively we obtain the convex hull for a set of points by "shrink wrapping" a boundary around them. The straight line segment joining any two points inside the convex hull itself lies entirely inside the convex hull.

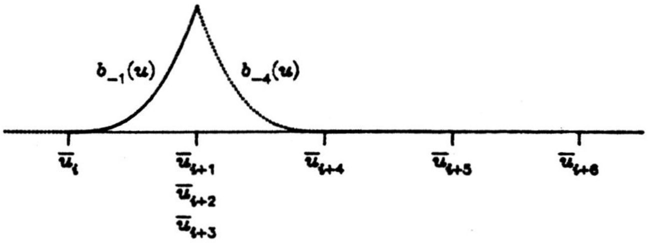

Figure 8. A knot of multiplicity 3.

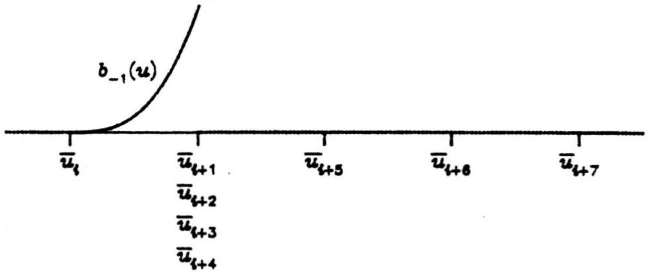

Figure 9. A knot of multiplicity 4.

The way in which repeated knots affect continuity is best explained by defining the B-splines in terms of divided differences of the truncated power functions $(\bar{u} - \bar{u}_i)^d_+$, which are 0 when $\bar{u} < \bar{u}_i$ and simply $(\bar{u} - \bar{u}_i)^d$ when $\bar{u} \ge \bar{u}_i$. (See [deBoor78a].) The divided difference can be thought of as a means for symbolically "pre-computing" the differences of several such truncated power functions so as to avoid cancellation errors and overflow. For details see [Barsky83a].

2.2. Beta-splines

Beta-spline curves and surfaces were introduced in [Barsky81a]. They are a generalization of the uniform cubic B-splines in which continuity of the parametric first and second derivatives is replaced by the requirement that the "unit tangent" and "curvature" vectors be continuous, which is called G^2 continuity.

The *unit tangent vector* of a curve $\mathbf{Q}(\bar{u})$ is

$$\hat{\mathbf{T}}(\bar{u}) = \frac{\mathbf{Q}^{(1)}(u)}{|\mathbf{Q}^{(1)}(u)|} \tag{9}$$

and the *curvature vector* is

$$\mathbf{K}(\bar{u}) = \kappa(\bar{u})\hat{\mathbf{N}}(\bar{u}) = \kappa(\bar{u})\frac{\hat{\mathbf{T}}^{(1)}(\bar{u})}{|\hat{\mathbf{T}}^{(1)}(\bar{u})|}$$

where $\kappa(\bar{u})$ is the curvature of $\mathbf{Q}$ at $\bar{u}$ and $\hat{\mathbf{N}}(\bar{u})$ is a unit vector pointing from $\mathbf{Q}(\bar{u})$ towards the center of the osculating circle at $\mathbf{Q}(\bar{u})$. ($\kappa(\bar{u})$ is the reciprocal of the radius of the osculating circle at $\mathbf{Q}(\bar{u})$, which is the circle whose first and second derivative vectors agree with those of $\mathbf{Q}$ at $\bar{u}$.) $\hat{\mathbf{T}}(\bar{u})$ and $\mathbf{K}(\bar{u})$ capture the physically meaningful notions of the direction and curvature, respectively, at a point on the curve. $\mathbf{Q}(\bar{u})$, $\hat{\mathbf{T}}(\bar{u})$ and $\mathbf{K}(\bar{u})$ are easily seen to be continuous away from the joints of a piecewise polynomial; it is possible to show that in order for $\mathbf{Q}(\bar{u})$, $\hat{\mathbf{T}}(\bar{u})$ and $\mathbf{K}(\bar{u})$ also to be continuous at the joint between two consecutive curve segments $\mathbf{Q}_{i-1}(\bar{u})$ and $\mathbf{Q}_i(\bar{u})$ of $\mathbf{Q}(\bar{u})$ it is sufficient that

$$Q_{i-1}(\bar{u}_i) = Q_i(\bar{u}_i) \tag{10}$$

$$\beta_1 Q_{i-1}^{(1)}(\bar{u}_i) = Q_i^{(1)}(\bar{u}_i) \tag{11}$$

$$\beta_1^2 Q_{i-1}^{(2)}(\bar{u}_i) + \beta_2 Q_{i-1}^{(1)}(\bar{u}_i) = Q_i^{(2)}(\bar{u}_i) \tag{12}$$

at every knot $\bar{u}_i$ and for any scalars β_1 and β_2. These equations are, by definition, less restrictive than simple continuity of position and parametric derivatives, which is the special case in which $\beta_1 = 1$ and $\beta_2 = 0$ and corresponds to the uniform cubic B-splines.

Equation (10) enforces positional continuity. Equation (11) requires that the first parametric derivative vectors from the left and right at a joint be colinear, but allows their magnitudes to differ by the factor β_1. There is an instantaneous change in "velocity" at the joint, but not a change in direction. Equation (12) reflects the fact that $Q_i^{(2)}(\bar{u}_i)$ may have a component of arbitrary magnitude directed along the tangent since acceleration along the tangent does not "deflect" a point traveling along the curve, and so does not affect the curvature there. The following basis segments can be shown, for fixed β_1 and β_2 and a uniform knot sequence, to define a basis function $B(\bar{u})$ which satisfies equations (10), (11) and (12).

$$b_{-1}(u) = \frac{1}{\delta}\left(2u^3 \right) \tag{13}$$

$$b_{-2}(u) = \frac{1}{\delta}\Big(2 + (6\beta_1)u + (3\beta_2 + 6\beta_1^2)u^2 - (2\beta_2 + 2\beta_1^2 + 2\beta_1 + 2)u^3 \Big)$$

$$b_{-3}(u) = \frac{1}{\delta}\Big((\beta_2 + 4\beta_1^2 + 4\beta_1) + (6\beta_1^3 - 6\beta_1)u - (3\beta_2 + 6\beta_1^3 + 6\beta_1^2)u^2 + (2\beta_2 + 2\beta_1^3 + 2\beta_1^2 + 2\beta_1)u^3 \Big)$$

$$b_{-4}(u) = \frac{1}{\delta}\Big((2\beta_1^3) - (6\beta_1^3)u + (6\beta_1^3)u^2 - (2\beta_1^3)u^3 \Big)$$

where

$$\delta = \beta_2 + 2\beta_1^3 + 4\beta_1^2 + 4\beta_1 + 2 \neq 0$$

$B(\bar{u})$ is simply translated to form each of the $B_i(\bar{u})$, all of which have the same shape. (The same is true of the B-spline basis functions for a uniform knot sequence, though not for an arbitrary knot sequence.)

If we substitute $\beta_1 = 1$ and $\beta_2 = 0$ into the Beta-spline basis segments (13) we obtain the B-spline basis segments (7). For other values of β_1 and β_2 the Beta-spline basis segments fail to be C^2 continuous at joints, although they are G^2 continuous.

Like the B-spline basis segments, the Beta-spline basis segments are non-negative and sum to one, and so possess the convex hull property.

Because β_1 and β_2 affect the shape of a curve, it is desirable to be able to specify distinct values of these two shape parameters at each joint. This can be accomplished without destroying the convex hull or other properties of the Beta-splines by properly interpolating between them at intermediate points within a segment and then evaluting equations (13) with the interpolated values. The details are given in [Barsky83a, Barsky83b].

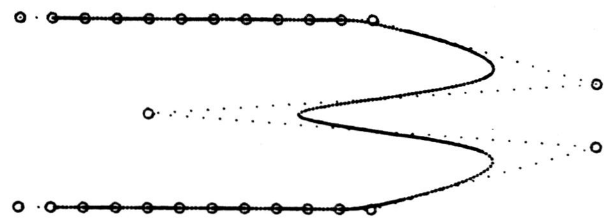

Figure 10. A uniform cubic B-spline (order 4). The control vertices are circled.

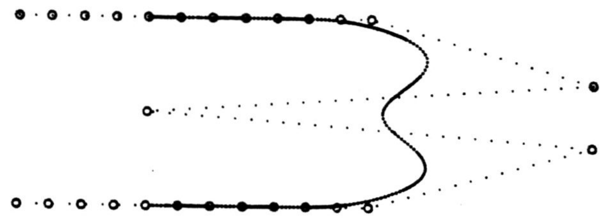

Figure 11. A uniform B-spline of order 10.

3. Shaping B-splines

In the diagrams which appear subsequently we will often distinguish between successive curve segments by alternately drawing them as solid and as dotted lines.

3.1. Changing the Order of the Curve

Let's begin with a uniform cubic B-spline curve like the one shown in Figure 10. Because it requires four vertices — and basis functions — to define a segment, there are three fewer segments than there are control vertices in a cubic B-spline curve; for a curve of order k there are $k-1$ fewer segments than control vertices. The first segment $Q_k(\bar{u})$ is defined on $[\bar{u}_{k-1}, \bar{u}_k)$ and the last segment Q_m is defined on $[\bar{u}_{m-1}, \bar{u}_m)$, the knot sequence being $\bar{u}_0, \cdots, \bar{u}_{m+k}$ and the vertex sequence being $V_0, \cdots, V_m$. If we increase the order of the B-spline for a fixed set of control vertices we therefore reduce the number of segments, and we have consequently placed a large number of initial and final vertices in the control polygon for Figure 10.

Figure 11 illustrates what happens when we use a tenth order B-spline. Firstly, there are six fewer segments than in the fourth order curve of Figure 10: three fewer at the beginning and three fewer at the end. Also, the curve "oscillates less;" the influence of a given control vertex on any particular segment has been reduced. This follows from the fact there are more vertices influencing the segment, since each basis function has larger support. Each segment also lies within the convex hull of a larger number of vertices — within the convex hull of ten control vertices now instead of four.

Conversely, if we reduce the order of the curve then each vertex has more influence. For a second order B-spline, each segment is determined by two control vertices. Since it must lie within the convex hull of the two vertices, and the basis functions go to zero at either end, the segment is simply a straight line from the first vertex to the second. Figure 12 illustrates this. To facilitate comparison we show several curves of differing order in Figure 13.

3.2. Colinear Control Vertices

Figures 10-12 illustrate a second useful fact about the B-spline curves of order k. If k successive control vertices are colinear then they define a straight line segment. This follows easily from the convex hull property. If the knot sequence is such as to require the adjoining segments to meet this straight line with C^2 continuity (*i.e.* the breakpoint corresponding to that joint has multiplicity at most $k-3$) then the ends of those segments must have zero curvature there. Hence if a control polygon ends in $k-1$ colinear control vertices and the terminating breakpoint has multiplicity at most $k-3$ (one for a cubic) then the curve will end with zero curvature.

3.3. Multiple Knots

We have already seen that repeating a knot reduces the parametric continuity of the basis functions whose support contains that knot, and Figures 7-9 illustrate the effect this has on an isolated basis function. Figures 14-21 illustrate the effect of knot multiplicity on a parametric curve, as well as showing the simultaneous effect on adjacent basis functions. (In Figures 15, 17, 19, and 21 we alternate between solid and dotted lines in drawing adjacent basis functions .)

Figures 18 and 19 are particularly interesting. Consider the second curve segment ($Q_5(\bar{u})$), which is defined by $B_1(\bar{u})$, $B_2(\bar{u})$, $B_3(\bar{u})$ and $B_4(\bar{u})$, which in turn involve the knots

$$\begin{matrix} 1, & 2, & 3, & 4, & 5, & 5, & 5, & 6 \\ \bar{u}_1, & \bar{u}_2, & \bar{u}_3, & \bar{u}_4, & \bar{u}_5, & \bar{u}_6, & \bar{u}_7, & \bar{u}_8 \end{matrix} .$$

$B_1(\bar{u})$, which is defined over the knots 1, 2, 3, 4 and 5, falls to zero at $\bar{u}=5$ with C^2 continuity since that is the right hand side of its support and its support contains no repeated knots. Similarly, $B_2(\bar{u})$ must also fall to zero at $\bar{u}=5$, though with only C^1 continuity since it is defined over the knots 2, 3, 4, 5 and 5. And $B_3(\bar{u})$ must fall to zero there, with only C^0 continuity since it is defined over the knots 3, 4, 5, 5, and 5. It therefore follows from equation (8) that $B_4(\bar{u})$, which contains a knot of multiplicity three and is weighted by $\mathbf{V}_4$ in drawing Figure 18 (B_4 is also shown isolation in Figure 8), must rise to the value one at $u=5$. Hence the end of the second segment must interpolate $\mathbf{V}_4$. By a similar argument we can show that the beginning of the third segment must also interpolate $\mathbf{V}_4$, as it does in Figure 18; $B_4(\bar{u})$ also converges to one as $\bar{u}$ approaches 5 from the right. A generalization of this argument establishes that a knot of multiplicity $k-1$ results in interpolation of the appropriate control vertex by an order k B-spline, at which point the curve has only $C^{k-1-(k-1)}=C^0$ continuity.

Figures 20 and 21 are also interesting. Here we have a knot of multiplicity $k=4$, resulting in a discontinuity; the curve jumps from $\mathbf{V}_4$ to $\mathbf{V}_5$. $B_4(\bar{u})$ — which in the previous example had support (4,6), rose from zero at $\bar{u}=4$ to one at $\bar{u}=5$ and then fell to zero at $\bar{u}=6$, with C^0 continuity — now has support (4,5), converges to one from the left at $\bar{u}=5$, and then jumps to zero at $\bar{u}=5$. $B_5(\bar{u})$ in Figure 21 behaves similarly, except that it jumps from zero to one as it crosses $\bar{u}=5$ and then falls to zero at $\bar{u}=6$. (As was the case for Figure 19, the multiple knot causes all other basis functions of interest here to take the value zero at $\bar{u}=5$.) Since $B_4(\bar{u})$ is weighted by $\mathbf{V}_4$ and $B_5(\bar{u})$ is weighted by $\mathbf{V}_5$, a jump from $\mathbf{V}_4$ to $\mathbf{V}_5$ results as we cross $\bar{u}=5$.

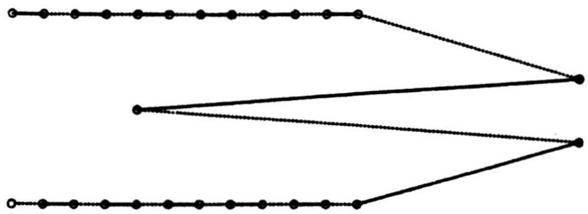

Figure 12. A uniform B-spline of order 2 (the control polygon is not shown). A B-spline of order 1 would consist simply of the control vertices.

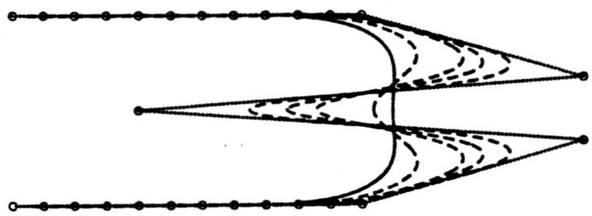

Figure 13. Uniform B-spline curves of orders 2, 3, 4, 5, 10 and 20 for the same control vertices. The second order curve is shown with a dotted line, and connects the control vertices with straight line segments. The curve of order 20 is shown with a solid line. Intermediate curves are drawn dashed.

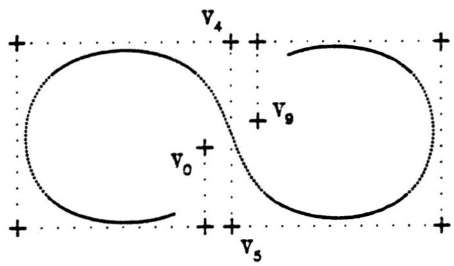

Figure 14. A uniform cubic B-spline. The knot sequence here is 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13.

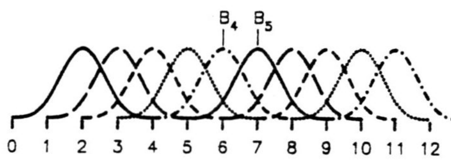

Figure 15. The basis functions for Figure 14, scaled vertically by 3 for legibility.

3.4. Multiple Vertices

Just as we can repeat knots in the knot sequence underlying a B-spline curve, we can repeat points in the sequence of control vertices. First let us recall that the i^{th} segment in a B-spline curve of order k, according to equation (3), is defined by control vertices $\mathbf{V}_{i-k}, \cdots, \mathbf{V}_{i-1}$. Furthermore, since $B_{i-k}(\bar{u})$ is zero at $\bar{u}_i$, $\mathbf{V}_{i-k}$ does not effect the last point $\mathbf{Q}_i(\bar{u}_i)$ of the i^{th} segment. $\mathbf{Q}_i(\bar{u}_i)$ is therefore entirely determined by $\mathbf{V}_{i-k+1}, \cdots, \mathbf{V}_{i-1}$. If these $k-1$ vertices are identical, then substituting into equation (3) and using (8) we obtain

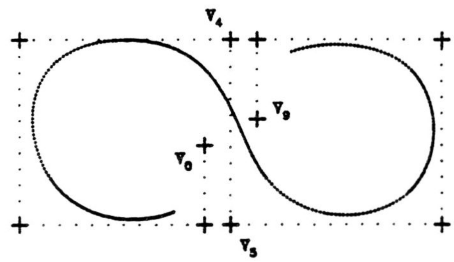

Figure 16. A knot of multiplicity 2 at $\bar{u}=5$. The knot sequence is 0, 1, 2, 3, 4, 5, 5, 6, 7, 8, 9, 10, 11, 12.

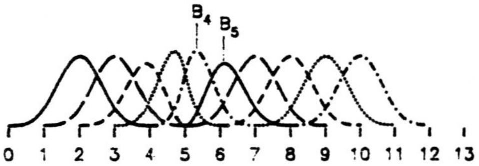

Figure 17. The basis functions for Figure 16, scaled vertically by 3.

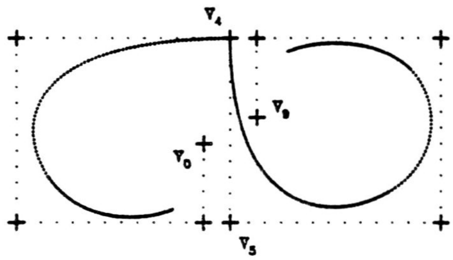

Figure 18. A knot of multiplicity 3 at $\bar{u}=5$. The knot sequence is 0, 1, 2, 3, 4, 5, 5, 5, 6, 7, 8, 9, 10, 11.

Figure 19. The basis functions for Figure 18, scaled vertically by 3.

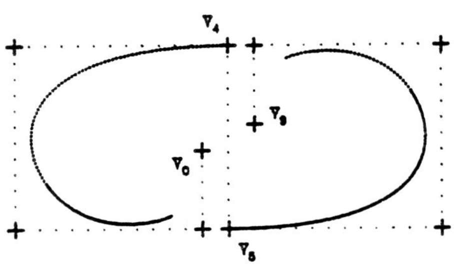

Figure 20. A knot of multiplicity 4 at $\bar{u}=5$. The knot sequence here is 0, 1, 2, 3, 4, 5, 5, 5, 5, 6, 7, 8, 9, 10.

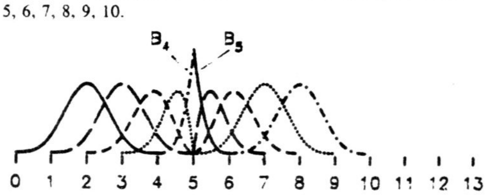

Figure 21. The basis functions for Figure 20, scaled vertically by 3.

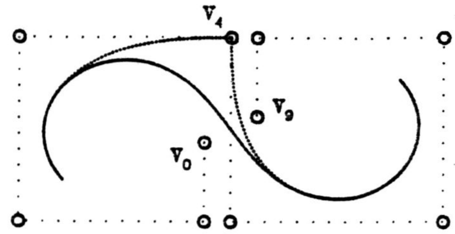

Figure 22. The solid curve is a uniform B-spline of order six defined by the ten control vertices shown, each with multiplicity one. The dotted line results when the V_4 appears five times times in succession (resulting in a total of fourteen control vertices), for the uniform knot sequence 0, 1, $\cdots$, 18, 19. Multiplicities between one and five result in curves intermediate between these two.

$$Q_i(\bar{u}_i) = \sum_{r=-k}^{r=-1} V_{i+r} B_{i+r,k}(\bar{u}) = \sum_{r=-k+1}^{r=-1} V_{i-1} B_{i+r,k}(\bar{u})$$

$$= V_{i-1} \sum_{r=-k}^{r=-1} B_{i+r,k}(\bar{u}) = V_{i-1}$$

Thus a control vertex of multiplicity $k-1$ is interpolated, regardless of the knot sequence at hand. (See Figure 22 for an example.) Moreover, $Q_i(\bar{u})$ is guaranteed to be a straight line segment, since in this case we can factor equation (3) for $Q_i(\bar{u}_i)$ as

$$V_{i-k} B_{i-k,k}(\bar{u}) + V_{i-1} \sum_{r=-k+1}^{r=-1} B_{i+r,k}(\bar{u}) \qquad (14)$$

which is a convex combination of V_{i-k} and V_{i-1} and so defines a straight line segment. V_{i-k} and V_{i-1} are not, in general, interpolated.

Unless the knot $\bar{u}_{i-1}$ has multiplicity $k-2$ or greater, the previous curve segment, namely $Q_{i-1}(\bar{u})$, will be at least C^2 continuous with $Q_i(\bar{u})$ at $\bar{u}_{i-1}$, and must therefore have a zero second derivative and zero curvature there since $Q_i(\bar{u})$ is a straight line segment. Moreover, $Q_{i-1}(\bar{u})$ must terminate somewhere on the line segment connecting V_{i-k} and $V_{i-k+1} = \cdots = V_{i-1}$, again because (14) is a convex combination. A similar argument establishes that $Q_{i+1}(\bar{u})$ begins with zero curvature at a point lying between $V_{i-k+1} = \cdots = V_{i-1}$ and V_i.

3.5. Knot Spacing

We can also affect the shape of a curve by moving the knots. It is probably most productive to think in terms of the space between knots, so that when we "increase the size" of an interval $[\bar{u}_{i-1}, \bar{u}_i)$ by some amount du, each of the knots to the right of $\bar{u}_i$ is moved right by du also. Figures 23 and 24 demonstrate that altering the knot spacing can have a significant and useful affect on the curve. By setting the length of a knot interval to a small value we can produce curves which to a greater or lesser extent approximate the effect of having repeated knots, which is the case in which one or more intervals have actually gone to zero.

3.6. End Conditions

Our description of multiple knots and multiple vertices is applicable to any part of the curve, but is particularly useful in controlling the behaviour at the beginning and end of a curve [Barsky82c]. In general the most we can say about these end-points is that they lie within the convex hull of the first and last $k-1$ control vertices, respectively. From earlier remarks it follows that either an initial knot of multiplicity $k-1$, or an initial control vertex of multiplicity $k-1$, will cause the curve to interpolate the first control vertex. In the latter case the first curve segment is a short straight line. Moreover, an initial vertex of multiplicity $k-2$ will cause the curve to begin somewhere on the line segment joining the first and second control vertices with zero curvature. The end of a curve may be similarly controlled.

4. Shaping Beta-splines

In generalizing the uniform cubic B-splines to form the Beta-splines the effect of repeating a vertex, as described in section 3.3, is preserved (more detail may be found in [Barsky82b, Barsky83a]). As the knot spacing is fixed uniformly at one, we need only discuss the effect of changing β_1 and β_2.

When we interpolate the shape parameters β_1 and β_2 from one joint to the next in evaluating (13), the effect of changing β_1 or β_2 at some joint is limited to the two segments which meet at that joint. Thus the control they provide is "more local" than that provided by the B-splines of order three or higher.

4.1. Changing β_2

β_2 serves to control *tension* in the curve. If β_2 goes to $+\infty$ at every joint the curve flattens, converging to the control polygon. (See Figure 25.) Small negative values result in "repulsion". (See Figure 26.)

Tension is controlled locally by interpolating the β_2's between joints. The i^{th} basis function, which reaches its maximum value at u_{i+2}, is weighted by $\mathbf{V}_i$. Altering β_2 at $\bar{u}_{i+2}$ moves that joint along a straight line passing through $\mathbf{V}_i$. For example, increasingly positive values move this joint towards $\mathbf{V}_i$, pulling the curve segments on either side of the joint towards the control polygon. (See Figure 27.) Other segments are completely unaffected. (In this paper we have indexed the Beta-spline basis functions from the left end of their support, rather than from the middle as we have done elsewhere, so as to be compatible with the usual convenient notation for B-splines.)

4.2. Changing β_1

Increasing β_1 at a joint increases the "velocity" with which we traverse a curve (say from left to right) immediately to the right of the joint, with respect to the "velocity" just left of the joint. Values in excess of one cause the unit tangent vector at the joint to have greater influence to the right than to the left, in that the curve will lie close to the tangent longer in the right-hand segment. Values of β_1 ranging between one and zero have the reverse effect, causing the curve to lie close to the tangent longer to the left of a joint than to the right. Figure 28 illustrates how β_1 thus *biases* the curve.

β_1 can also be thought of as an "asymmetric tension parameter:" as β_1 goes to $+\infty$ (for any β_2) the curve flattens in much the same way it does as β_2 goes to $+\infty$, except that the joint at $\bar{u}_{i+2}$ converges to $\mathbf{V}_{i-1}$ rather than $\mathbf{V}_i$. This effect is illustrated in Figure 29. If β_2 is zero then as β_1 goes to zero the joint at $\bar{u}_{i+2}$ converges to $\mathbf{V}_{i+1}$ instead. (See figure 30). If β_2 is not zero, then the limiting behaviour as β_1 goes to zero is more complicated. The details are discussed in [Barsky83a].

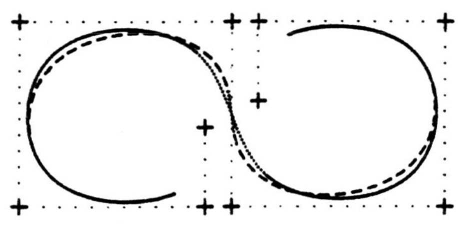

Figure 23. The solid curve is a uniform cubic B-spline defined by the ten (multiplicity one) control vertices shown, and by a knot sequence in which the spacing from one knot to the next is 1.0. The dashed curve is defined by the same control vertices and knot spacings except that the spacing between the sixth and seventh knots is 0.2 instead of 1.0. The portion of each curve corresponding to values of $\bar{u}$ between the sixth and seventh knots is shown dotted.

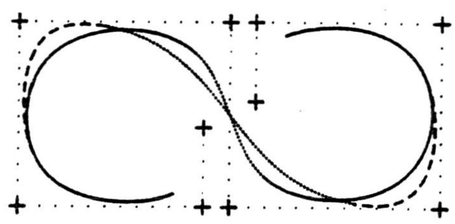

Figure 24. The solid curve is as in Figure 23. For the dashed curve we have changed the length of the interval between the sixth and seventh knots from 1 to 5. Otherwise the knot spacings and vertex multiplicities are all one.

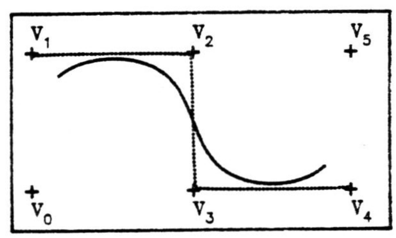

Figure 25. The solid curve is a uniform cubic B-spline; $\beta_1 = 1$ and $\beta_2 = 0$ at each joint. The dotted line is produced by setting β_2 to the value 100 throughout the curve. The control polygon is not shown.

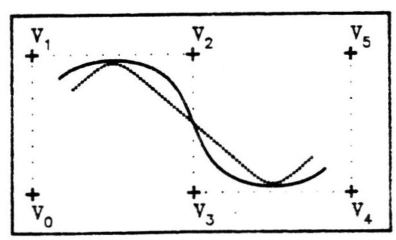

Figure 26. The solid curve is a uniform cubic B-spline; $\beta_1 = 1$ and $\beta_2 = 0$ at each joint. The dotted line is produced by setting β_2 to the value -4 throughout the curve.

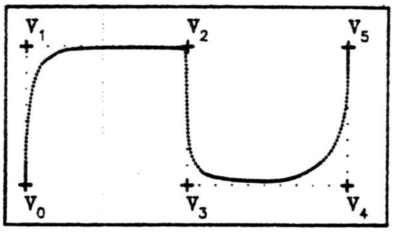

Figure 27. Here β_1 is uniformly one and (in left to right order) β_2 has the values 0, 0, 6, 100, 9, 0, 0 and 0 at successive joints. The first and last vertices of both curves have been tripled so as to cause the curves to interpolate them.

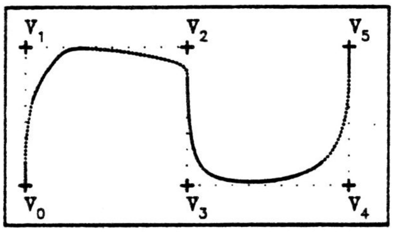

Figure 28. Here β_2 is uniformly zero and (in left to right order) β_1 has the values 1, 1, 2, 4, 8, 1, 1 and 1 at successive joints. The first and last vertices of both curves have been tripled.

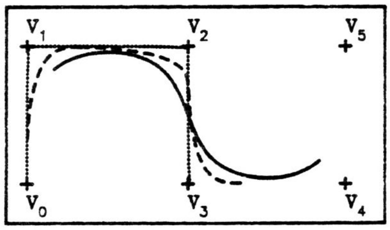

Figure 29. The solid curve is a uniform cubic B-spline; $\beta_1 = 1$ and $\beta_2 = 0$ at each joint. The dashed curve is produced by setting β_1 to the value 5 throughout the curve, β_2 being 0. For the dotted curve β_1 is 100.

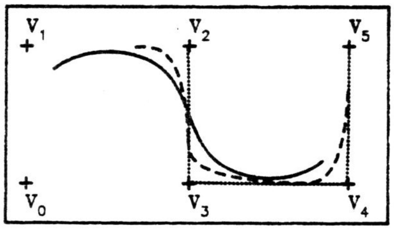

Figure 30. The solid curve is a uniform cubic B-spline; $\beta_1 = 1$ and $\beta_2 = 0$ at each joint. The dashed curve is produced by setting β_1 to the value 0.2 throughout the curve, β_2 being 0. For the dotted curve β_1 is 0.01.

5. Conclusions

We have sketched the way in which changing vertex position, the order k, knot multiplicity, vertex multiplicity and knot spacing may affect the shape of B-splines; for the Beta-splines we have demonstrated how β_1 and β_2 allow us to alter curve shape by means of controlled discontinuities in the first and second parametric derivatives of a piecewise polynomial while preserving geometric continuity. In the process we have also generalized some of the results appearing in [Barsky82c], although we have not by any means supplied an exhaustive analysis of the B-splines and Beta-splines. We particularly emphasize that there is a difference between changing the multiplicity of a knot and changing the multiplicity of a vertex, although in the literature the two are often changed together and the distinction is therefore sometimes blurred.

6. References

[Barsky81a] Brian A. Barsky (1981), The Beta-spline: A Local Representation Based On Shape Parameters and Fundamental Geometric Measures, PhD dissertation, Department of Computer Science, University of Utah, December.

[Barsky82b] Brian A. Barsky (1982), The Beta-spline: A Curve and Surface Representation for Computer Graphics and Computer Aided Geometric Design [submitted for publication].

[Barsky82c] Brian A. Barsky (1982), End Conditions and Boundary Conditions for Uniform B-Spline Curve and Surface Representations, *Computers in Industry* 3(1/2), March/June, 17-29.

[Barsky83a] Brian A. Barsky, Richard H. Bartels, John C. Beatty, and Kellogg S. Booth (1983), An Introduction to the Use of Splines in Computer Graphics [in preparation].

[Barsky83b] Brian A. Barsky and John C. Beatty (1983), Local Control of Bias and Tension in Beta-splines, *Computer Graphics* 17(3), July [(proceedings of the 1983 SIGGRAPH Conference)].

[Barsky82a] Brian A. Barsky and Alain Fournier (1982), Computational Techniques for Parametric Curves and Surfaces, *Proceedings of the Graphics Interface 82 Conference*, May, Toronto, Ontario, 57-71.

[Bézier70a] Pierre E. Bézier (1970), *Emploi des machines à commande numérique*, Masson et Cie., Paris [Translated by A. Robin Forrest and Anne F. Pankhurst as *Numerical Control – Mathematics and Applications*, John Wiley and Sons, Ltd., London, 1972].

[Bézier77a] Pierre E. Bézier (1977), Essai de définition numérique des courbes et des surfaces expérimentales, l'Université Pierre et Marie Curie, Paris, February.

[deBoor78a] Carl de Boor (1978), *A Practical Guide to Splines*, Applied Mathematical Sciences Volume 27, Springer-Verlag.

[Coons64a] Steven A. Coons (1964), Surfaces for Computer Aided Design, Design Division, Mech. Engin. Dept., M.I.T., Cambridge, Massachusetts.

[Coons67a] Steven A. Coons (1967), Surfaces for Computer-Aided Design of Space Forms, MAC-TR-41, Project MAC, M.I.T., Cambridge, Massachusetts, June [Available as AD-663 504 from NTIS, Springfield, Virginia.].

[Faux79a] I. D. Faux and M. J. Pratt (1979), *Computational Geometry for Design and Manufacture*, John Wiley & Sons.

[Foley82a] James D. Foley and Andries van Dam (1982), *Fundamentals of Interactive Computer Graphics*, Addison Wesley.

[Gordon74a] William J. Gordon and Richard F. Riesenfeld (1974), B-spline Curves and Surfaces, *Computer Aided Geometric Design*, Robert E. Barnhill & Richard F. Riesenfeld (eds.), Academic Press, 95-126.

[Newman73] William M. Newman and Robert F. Sproull (1973), *Principles of Interactive Computer Graphics*, McGraw-Hill.

[Riesenfeld73a] Richard F. Riesenfeld (1973), Applications of B-spline Approximation to Geometric Problems of Computer-Aided Design, PhD dissertation, Department of Systems and Information Science, Syracuse University, May.

[Rogers76a] David F. Rogers and J. Alan Adams (1976), *Mathematical Elements for Computer Graphics*, McGraw-Hill.