UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT

Beta-Splines With a Difference

R.H. Bartels
J.C. Beatty

CS-83-40

May 1984

Beta-Splines With a Difference

Richard H. Bartels
John C. Beatty

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

Technical Report CS-83-40

ABSTRACT

Local control of the shape parameters β_1 and β_2 in a Beta-spline has previously relied on quintic Hermite interpolation of the distinct β-values associated with the joints in a geometrically continuous piecewise polynomial curve. Changing the value of β_1 or β_2 at a given joint affected only the two immediately adjacent curve segments. Such extreme locality was obtained at the cost of dealing with polynomials of unusually high degree, as these "continuously-shaped Beta-splines" are the quotients of 18^{th} and 15^{th} degree polynomials.

We here introduce an alternative means of obtaining local control of geometrically continuous piecewise polynomial curves. The essential idea is to generalize the truncated power functions suitably, from which the B-splines are obtained by differencing. The "discretely-shaped Beta-splines" which result from differencing these generalized functions are piecewise cubic, respond locally to changes in the shape parameters (as well as the control vertices), partition unity and possess a convex hull property. Moreover they are naturally defined over an irregularly-spaced knot sequence.

The research reported here was supported in part by the Natural Sciences and Engineering Research Council of Canada, the Province of Ontario's BILD program, the University of Waterloo and the Computer Graphics Laboratory.

Contents

1. Introduction

The Beta-splines introduced by Brian Barsky [Barsky1981a, Barsky1983a] are a generalization of the uniform cubic B-splines. They provide a means of constructing parametric spline curves in which the shape of a curve is controlled by two tension parameters β_1 and β_2 independently of the movement of control vertices. A *uniformly-shaped* Beta-spline curve results when a single value of each β parameter is used in evaluating an entire curve. In [Barsky1982a] and [Barsky1983b] a technique was introduced by which distinct values of β_1 and β_2, denoted $\beta_{1,i}$ and $\beta_{2,i}$, are associated with each control vertex so that altering a single such β value causes only a local change in the shape of the curve being defined.

This local control was accomplished by interpolating between any two successive values $\beta_{1,i-1}$ & $\beta_{1,i}$ of β_1 and $\beta_{2,i-1}$ & $\beta_{2,i}$ of β_2, followed by substitution of these interpolated values into the equations for a uniformly-shaped curve. The result is called a *continuously-shaped* Beta-spline curve. Very local control of shape is achieved in that changing the value of any particular $\beta_{1,i}$ or $\beta_{2,i}$ alters only two curve segments, but there is a substantial cost – continuously-shaped Beta-spline curve segments are actually the quotients of 18^{th} and 15^{th} degree polynomials.

In this paper, an expansion of the approach introduced in [Bartels1983a], we shall see that it is possible to obtain local control of the shape parameters in Beta-splines without resorting to polynomials of higher than cubic degree. The technique introduced here has the added advantage of generalizing the cubic Beta-spline curves to non-uniform knot sequences. The approach we take is to generalize the definition of B-splines as divided differences of truncated power functions.

In the next section we will sketch the basic ideas and terminology we need from the theory of B-splines; for a more leisurely development the reader is referred to [Bartels1983b].

2. B-Splines

The curves in which we are interested are *piecewise cubic parametric polynomials* $\mathbf{Q}(\bar{u}) = (X(\bar{u}), Y(\bar{u}))$ of the parameter $\bar{u} \in [\bar{u}_0, \bar{u}_m)$, where $\bar{u}_0$ and $\bar{u}_m$ are the first and last values in the *knot sequence*

$$\bar{u}_0 < \bar{u}_1 < \bar{u}_2 \cdots < \bar{u}_m \ .$$

For reasons which will become apparent later, we will need *auxiliary knots* $\bar{u}_{-3}$, $\bar{u}_{-2}$, $\bar{u}_{-1}$, and $\bar{u}_{m+1}$, $\bar{u}_{m+2}$, $\bar{u}_{m+3}$ satisfying

$$\bar{u}_{-3} < \bar{u}_{-2} < \bar{u}_{-1} < \bar{u}_0 \cdots \bar{u}_m < \bar{u}_{m+1} < \bar{u}_{m+2} < \bar{u}_{m+3} \ .$$

If the length of each *knot interval* $[\bar{u}_{i-1}, \bar{u}_i)$ is the same then the knot sequence, and all curves defined over that knot sequence, are said to be *uniform*.

The i^{th} *curve segment* $\mathbf{Q}_i(\bar{u}) = (X_i(\bar{u}), Y_i(\bar{u}))$ is generated as $\bar{u}$ runs from $\bar{u}_{i-1}$ to $\bar{u}_i$ ($u \in [\bar{u}_{i-1}, \bar{u}_i)$) and each parametric component $X_i(\bar{u})$ and $Y_i(\bar{u})$ is a single cubic polynomial. The knots $\bar{u}_1, \ldots, \bar{u}_{m-1}$ correspond to the *joints* between successive polynomial segments. (See Figure 1.) The curve $\mathbf{Q}(\bar{u})$ is a *cubic C^2* spline if the segments which meet at a joint do so with first and second derivative continuity; that is, if the first and second parametric derivatives of $\mathbf{Q}_{i-1}(\bar{u})$ and $\mathbf{Q}_i(\bar{u})$ at $\bar{u}_i$ are identical for $1 < i \leq m$.[1] It is a fact that any such curve can be represented as

$$\mathbf{Q}(\bar{u}) = \sum_i \mathbf{V}_i B_i(\bar{u}) = \sum_i (\, x_i B_i(\bar{u}), y_i B_i(\bar{u}) \,) \tag{1}$$

for some scaling values $\mathbf{V}_i = (x_i, y_i)$ and *cubic B-splines* $B_i(\bar{u})$.[2] The $B_i(\bar{u})$ are local functions, such as the one shown in Figure 2, which are non-zero only on four successive knot intervals. The $\mathbf{V}_i$ are called *control vertices* because the curve "passes near" them; generally one defines such a curve by placing and moving the control vertices. The term "B-spline" is short for *basis spline*. The $B_i(\bar{u})$ were given this name because any cubic spline can be represented as a scaled sum of the $B_i(\bar{u})$ as in (1), while no single $B_i(\bar{u})$ can itself be represented as a scaled sum of the remaining $B_j(\bar{u})$.

Because each $B_i(\bar{u})$ is nonzero on only four successive intervals, if $\bar{u}_{i-1} \leq \bar{u} < \bar{u}_i$ we may write (1) as

$$\begin{aligned} \mathbf{Q}_i(\bar{u}) &= \sum_{r=-4}^{-1} \mathbf{V}_{i+r} B_{i+r}(\bar{u}) \qquad (2) \\ &= \mathbf{V}_{i-4} B_{i-4}(\bar{u}) + \mathbf{V}_{i-3} B_{i-3}(\bar{u}) \\ &\qquad + \mathbf{V}_{i-2} B_{i-2}(\bar{u}) + \mathbf{V}_{i-1} B_{i-1}(\bar{u}) \ . \end{aligned}$$

The functions $B_i(\bar{u})$ are themselves C^2 continuous piecewise cubic polynomials, like the curves they define. Using the notation of Figure 2, on the interval $\bar{u}_{i-1} \leq \bar{u} < \bar{u}_i$ we may replace each B-spline with the *segment polynomial* which yields its value on that interval:

[1] More commonly one does not require that the knot sequence be strictly increasing, so that $\bar{u}_{i-1} = \bar{u}_i$ is allowed. Each time a knot is repeated the parametric continuity of the curve at that value of $\bar{u}$ is reduced by one. In this more general case any piecewise cubic curve is acceptable, even if there is a jump in position between curve segments. We will avoid this generality throughout.

[2] We will adhere to standard mathematical terminology, using the term "B-spline" to refer to a basis function. We will use the term "B-spline curve" to refer to a parametric curve such as (1) that is defined as a scaled sum of B-splines.

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

It is possible to show that

$$s_{i-1,-1}(\bar{u}) + s_{i-2,-2}(\bar{u}) + s_{i-3,-3}(\bar{u}) + s_{i-4,-4}(\bar{u}) = 1 \tag{3}$$

and

$$s_{i-1,-1}(\bar{u}) \ , \ s_{i-2,-2}(\bar{u}) \ , \ s_{i-3,-3}(\bar{u}) \ , \ s_{i-4,-4}(\bar{u}) \geq 0 \tag{4}$$

so that the i^{th} curve segment $\mathbf{Q}_i(\bar{u})$ lies within the convex hull of the vertices $\mathbf{V}_{i-4}$, $\mathbf{V}_{i-3}$, $\mathbf{V}_{i-2}$ and $\mathbf{V}_{i-1}$.

2.1. Divided Differences and the Truncated Power Function

There are a variety of ways in which the cubic B-splines $B_i(\bar{u})$ can be defined. An approach which leads to efficient and robust evaluation is to define them as "divided differences" of "truncated power functions." We shall first define these terms, and then explain why they are interesting.

Let $f(t)$ be a suitable function. Then the *zeroth divided difference of $f(t)$ at $\bar{u}_i$*, denoted $[\bar{u}_i\!:\!t]f(t)$, is simply $f(\bar{u}_i)$. The *first divided difference of $f(t)$ at $\bar{u}_i$* is

$$[\bar{u}_i,\bar{u}_{i+1}\!:\!t]f(t) = \frac{[\bar{u}_{i+1}\!:\!t]f(t) - [\bar{u}_i\!:\!t]f(t)}{\bar{u}_{i+1} - \bar{u}_i}$$

and in general the l^{th} *divided difference of $f(t)$ at $\bar{u}_i$* is

$$[\bar{u}_i, \cdots ,\bar{u}_{i+l}\!:\!t]f(t) = \frac{[\bar{u}_{i+1}, \cdots ,\bar{u}_{i+l}\!:\!t]f(t) - [\bar{u}_i, \cdots ,\bar{u}_{i+l-1}\!:\!t]f(t)}{\bar{u}_{i+l} - \bar{u}_i} \ .$$

The *truncated power function of degree r at t* is defined to be

$$(\bar{u} - t)_+^r = \begin{cases} 0 & \bar{u} < t \\ (\bar{u} - t)^r & \bar{u} \geq t \ . \end{cases}$$

(Think of t as being a constant, although this is really a bivariate function of $\bar{u}$ and t.) This function behaves like $(\bar{u} - t)^r$ to the right of t, and has the constant value 0 to the left of t. The two functions $(\bar{u} - t)^r$ and 0 have the same value and first through $(r-1)^{\text{st}}$ derivative at t, namely 0, but their r^{th} derivatives at t do not agree, being the constants $r!$ and 0, respectively.

Finally, the cubic B-splines are defined by

$$B_i(\bar{u}) = (\bar{u}_{i+4} - \bar{u}_i)[\bar{u}_i, \bar{u}_{i+1}, \bar{u}_{i+2}, \bar{u}_{i+3}, \bar{u}_{i+4}; t](\bar{u} - t)_+^3 \ . \qquad (5)$$

It is essential for what follows that we motivate this definition, which we shall now do.

We have mentioned two important properties of the B-splines $B_{-3}(\bar{u})$, ..., $B_{m-1}(\bar{u})$ (for any underlying knot sequence $\bar{u}_{-3}$, ..., $\bar{u}_{m+3}$): they form a basis for the cubic splines (C^2 piecewise cubic polynomials on $[\bar{u}_0, \bar{u}_m)$ with joints at the knots), and they have local support (in fact $B_i(\bar{u}) > 0$ only on $(\bar{u}_i, \bar{u}_{i+4})$). However, it is not entirely obvious how definition (5) manages this.

On the other hand, the truncated power functions

$$(\bar{u} - \bar{u}_{-3})_+^3, \ (\bar{u} - \bar{u}_{-2})_+^3, \ \cdots, \ (\bar{u} - \bar{u}_{m-1})_+^3 \qquad (6)$$

are easily seen to provide a basis for the cubic splines on the knot sequence $\bar{u}_0$, $\bar{u}_1$, ... , $\bar{u}_{m-1}$, $\bar{u}_m$ over the interval $[\bar{u}_0, \bar{u}_m)$. This is established with rigour in [Schumaker1981a] using somewhat the following reasoning.

- Assume that the first segment is given by

  $$\mathbf{Q}_1(\bar{u}) = q_{10}\bar{u}^0 + q_{11}\bar{u}^1 + q_{12}\bar{u}^2 + q_{13}\bar{u}^3 \ . \qquad (7)$$

  That is, suppose that $\mathbf{Q}_1(\bar{u})$ is known in terms of powers of $\bar{u}$. If we can find constants γ_{ji} such that

  $$\gamma_{j0}(\bar{u} - \bar{u}_0)_+^3 + \gamma_{j1}(\bar{u} - \bar{u}_{-1})_+^3 + \gamma_{j2}(\bar{u} - \bar{u}_{-2})_+^3 + \gamma_{j3}(\bar{u} - \bar{u}_{-3})_+^3 = \bar{u}^j$$

  for $j = 0,1,2,3$ then these power functions can be expressed, in turn, as linear combinations of

  $$(\bar{u} - \bar{u}_{-3})_+^3, \ (\bar{u} - \bar{u}_{-2})_+^3, \ (\bar{u} - \bar{u}_{-1})_+^3, \ (\bar{u} - \bar{u}_0)_+^3 \ .$$

  For each j the problem of finding the four corresponding γ's can be expressed as a system of linear equations by expanding the cubics in (7) in powers of $\bar{u}$. For instance, for $j = 0$, the coefficients $\gamma_{00}, \ldots, \gamma_{03}$ are determined by solving the equations

  $$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -3\bar{u}_0 & -3\bar{u}_{-1} & -3\bar{u}_{-2} & -3\bar{u}_{-3} \\ +3\bar{u}_0^2 & +3\bar{u}_{-1}^2 & +3\bar{u}_{-2}^2 & +3\bar{u}_{-3}^2 \\ -\bar{u}_0^3 & -\bar{u}_{-1}^3 & -\bar{u}_{-2}^3 & -\bar{u}_{-3}^3 \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{02} \\ \gamma_{03} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} .$$

  For the other values of j, only the right–hand side of these equations will change; the system matrix remains the same. Clearly the γ's can be found if the system matrix is nonsingular. This matrix, however, is simply a scaled version of a 4×4 *Vandermonde* matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ \bar{u}_0 & \bar{u}_{-1} & \bar{u}_{-2} & \bar{u}_{-3} \\ \bar{u}_0^2 & \bar{u}_{-1}^2 & \bar{u}_{-2}^2 & \bar{u}_{-3}^2 \\ \bar{u}_0^3 & \bar{u}_{-1}^3 & \bar{u}_{-2}^3 & \bar{u}_{-3}^3 \end{bmatrix}$$

which is well-known to have the determinant

$$(\bar{u}_0-\bar{u}_{-1})(\bar{u}_0-\bar{u}_{-2})(\bar{u}_0-\bar{u}_{-3})(\bar{u}_{-1}-\bar{u}_{-2})(\bar{u}_{-1}-\bar{u}_{-3})(\bar{u}_{-2}-\bar{u}_{-3})$$

Hence, the system matrix is nonsingular if the knots are distinct, as we are assuming them to be.

- Because $\mathbf{Q}(\bar{u})$ is C^2 continuous, $\mathbf{Q}_2(\bar{u})-\mathbf{Q}_1(\bar{u})$ must have the value zero and have zero first and second derivatives at $\bar{u}_1$. By means of a Taylor expansion we therefore have

  $$\mathbf{Q}_2(\bar{u}) \;=\; \mathbf{Q}_1(\bar{u}) + \frac{1}{6}\Big[\,\mathbf{Q}^{(3)}(\bar{u}_1)-\mathbf{Q}_1^{(3)}(\bar{u}_1)\,\Big](\bar{u}-\bar{u}_1)^3$$

  for $\bar{u}_1 \le \bar{u} < \bar{u}_2$. Thus

  $$\mathbf{Q}(\bar{u}) \;\equiv\; \begin{cases} \mathbf{Q}_1(\bar{u}) & \text{for } \bar{u}_0 \le \bar{u} < \bar{u}_1 \\[2ex] \mathbf{Q}_2(\bar{u}) & \text{for } \bar{u}_1 \le \bar{u} < \bar{u}_2 \text{ (as above)} \end{cases}$$

  and this can clearly be written equivalently as

  $$\mathbf{Q}(\bar{u}) \;=\; \mathbf{Q}_1(\bar{u}) + \frac{1}{6}\Big[\,\mathbf{Q}^{(3)}(\bar{u}_1)-\mathbf{Q}_1^{(3)}(\bar{u}_1)\,\Big](\bar{u}-\bar{u}_1)_+^3$$

  for $\bar{u}_0 \le \bar{u} < \bar{u}_2$.

- Similarly

  $$\mathbf{Q}_3(\bar{u}) \;=\; \mathbf{Q}_2(\bar{u}) + \frac{1}{6}\Big[\,\mathbf{Q}^{(3)}(\bar{u}_2)-\mathbf{Q}_2^{(3)}(\bar{u}_2)\,\Big](\bar{u}-\bar{u}_2)^3 \;,$$

  for $\bar{u}_2 \le \bar{u} < \bar{u}_3$, whence

  $$\begin{aligned}\mathbf{Q}(\bar{u}) \;=\; \mathbf{Q}_1(\bar{u}) &+ \frac{1}{6}\Big[\,\mathbf{Q}^{(3)}(\bar{u}_1)-\mathbf{Q}_1^{(3)}(\bar{u}_1)\,\Big](\bar{u}-\bar{u}_1)_+^3 \\ &+ \frac{1}{6}\Big[\,\mathbf{Q}^{(3)}(\bar{u}_2)-\mathbf{Q}_2^{(3)}(\bar{u}_2)\,\Big](\bar{u}-\bar{u}_2)_+^3\end{aligned}$$

  for $\bar{u}_0 \le \bar{u} < \bar{u}_3$, and so on.

- None of the functions in (6) can be represented as a linear combination of the others on $[\bar{u}_0,\bar{u}_m)$. For the first four truncated powers, an argument involving a Vandermonde system like that given above establishes that they are linearly independent on the $[\bar{u}_0,\bar{u}_m)$ interval. For the remaining truncated functions independence follows inductively from the fact that

$(\bar{u}-\bar{u}_i)^3_+$ cannot be written as a linear combination of the $(\bar{u}-\bar{u}_{i+1})^3_+$, ..., $(\bar{u}-\bar{u}_{m-1})^3_+$ since the latter are all zero on $[\bar{u}_i,\bar{u}_{i+1})$ and $(\bar{u}-\bar{u}_i)^3_+$ is not.

The utility of the extra knots $\bar{u}_{-3}$, $\bar{u}_{-2}$, $\bar{u}_{-1}$, and $\bar{u}_0$ should now be clear. They enable us to define the initial segment $\mathbf{Q}_1(\bar{u})$ entirely in terms of truncated power functions, and this is desirable (for purposes of consistency) because the remaining segments are defined in terms of $\mathbf{Q}_1(\bar{u})$ by adding in more truncated power functions.

We observe in passing that the $m+1$ knots $\bar{u}_0$, ..., $\bar{u}_m$ encompass m segments $\mathbf{Q}_1(\bar{u})$, ..., $\mathbf{Q}_m(\bar{u})$ and require $m+3$ coefficients for the basis functions (6) to represent an arbitrary cubic spline.

Unfortunately these truncated power functions are neither safe nor convenient to use. Near the end $\mathbf{Q}(\bar{u}_m)$ of a curve the power functions which become nonzero near its beginning $\mathbf{Q}(\bar{u}_0)$ have grown very large, and must be cancelled by large negative values of later power functions, with a consequent likelihood of overflow or loss of precision. (See Figure 3. A simple example illustrating this point appears in [Boor1978a, pp 104-105]). They are also non-local, in that altering the amplitude of a single truncated power function will cause a change in the entire remainder of the curve. The B-splines suffer from neither of these deficiencies, and in fact we can regard them as being obtained from the truncated power functions by "symbolic pre-cancellation" to avoid such loss of precision. The essential idea is to replace each truncated power function by a linear combination of several such, computed so as to avoid unbounded growth, and thus replace the truncated power functions with a more tractable basis. Let us see how this is done.

First we shall replace $(\bar{u}-\bar{u}_1)^3_+$. Consider $(\bar{u}-\bar{u}_1)^3_+$ and $(\bar{u}-\bar{u}_2)^3_+$. For $u \geq \bar{u}_2$ we have

$$(\bar{u}-\bar{u}_1)^3_+ \;=\; \bar{u}^3 - 3\bar{u}_1\bar{u}^2 + 3\bar{u}_1^2\bar{u} - \bar{u}_1^3$$

and

$$(\bar{u}-\bar{u}_2)^3_+ \;=\; \bar{u}^3 - 3\bar{u}_2\bar{u}^2 + 3\bar{u}_2^2\bar{u} - \bar{u}_2^3 \;.$$

We can cancel the cubic term by subtraction – for $\bar{u} \geq \bar{u}_2$ we have:

$$(\bar{u}-\bar{u}_2)^3_+ - (\bar{u}-\bar{u}_1)^3_+ \;= \qquad (8)$$

$$-\,(\bar{u}_2-\bar{u}_1)\left[\,3\bar{u}^2 - 3\bar{u}(\bar{u}_2+\bar{u}_1) + (\bar{u}_2^2+\bar{u}_2\bar{u}_1+\bar{u}_1^2)\right] \;.$$

Of course, for $\bar{u}_1 \leq \bar{u} < \bar{u}_2$ we still have a cubic term:

$$(\bar{u}-\bar{u}_2)^3_+ - (\bar{u}-\bar{u}_1)^3_+ \;=\; -\,(\bar{u}-\bar{u}_1)^3_+$$

$$= -\bar{u}^3 + 3\bar{u}_1\bar{u}^2 - 3\bar{u}_1^2\bar{u} + \bar{u}_1^3 \ .$$

Since $(\bar{u}-\bar{u}_1)_+^3$ and $(\bar{u}-\bar{u}_2)_+^3$ are linearly independent (neither is a scalar multiple of the other), this quadratic may replace $(\bar{u}-\bar{u}_1)_+^3$ in our basis.

We may deal with the remaining truncated power functions $(\bar{u}-\bar{u}_2)_+^3$, ..., $(\bar{u}-\bar{u}_{m-1})_+^3$ similarly: in each case $(\bar{u}-\bar{u}_{i+1})_+^3 - (\bar{u}-\bar{u}_i)_+^3$ may replace $(\bar{u}-\bar{u}_i)_+^3$. The resulting functions all behave quadratically for sufficiently large $\bar{u}$ (*i.e.*, for $\bar{u} \ge \bar{u}_{i+1}$).

We can repeat the process, replacing the quadratics with functions which are eventually linear, but to do this conveniently we must first modify our quadratics slightly. Notice that in equation (8) the coefficient of $\bar{u}^2$ depends on the knot spacing. In order to cancel quadratic terms in the same way that we cancelled cubics, we want each of the quadratic terms to have the same coefficient. Evidently we should replace $(\bar{u}-\bar{u}_1)_+^3$ not by $(\bar{u}-\bar{u}_2)_+^3 - (\bar{u}-\bar{u}_1)_+^3$, but by

$$\begin{aligned}\frac{(\bar{u}-\bar{u}_2)_+^3 - (\bar{u}-\bar{u}_1)_+^3}{\bar{u}_2-\bar{u}_1} &= [\bar{u}_1,\bar{u}_2{:}t](\bar{u}-t)_+^3 \\ &= -3\bar{u}^2 + 3\bar{u}(\bar{u}_2+\bar{u}_1) - (\bar{u}_2^2+\bar{u}_2\bar{u}_1+\bar{u}_1^2) \\ &\quad \text{for } \bar{u} \ge \bar{u}_2.\end{aligned}$$

Similarly we should replace $(\bar{u}-\bar{u}_2)_+^3$ by

$$\begin{aligned}\frac{(\bar{u}-\bar{u}_3)_+^3 - (\bar{u}-\bar{u}_2)_+^3}{\bar{u}_3-\bar{u}_2} &= [\bar{u}_2,\bar{u}_3{:}t](\bar{u}-t)_+^3 \\ &= -3\bar{u}^2 + 3\bar{u}(\bar{u}_3+\bar{u}_2) - (\bar{u}_3^2+\bar{u}_3\bar{u}_2+\bar{u}_2^2) \\ &\quad \text{for } \bar{u} \ge \bar{u}_3\end{aligned}$$

to obtain a constant coefficient of -3 for the quadratic term, which can then be cancelled by further subtraction. Thus we see that

$$\frac{(\bar{u}-\bar{u}_3)_+^3 - (\bar{u}-\bar{u}_2)_+^3}{\bar{u}_3-\bar{u}_2} - \frac{(\bar{u}-\bar{u}_2)_+^3 - (\bar{u}-\bar{u}_1)_+^3}{\bar{u}_2-\bar{u}_1}$$
$$= (\bar{u}_3-\bar{u}_1)(3\bar{u}-\bar{u}_3-\bar{u}_2-\bar{u}_1)$$

cancels the quadratic coefficient, but that to obtain a constant coefficient for the linear term we should in fact replace $[\bar{u}_1,\bar{u}_2{:}t](\bar{u}-t)_+^3$ by

$$\frac{[\bar{u}_2,\bar{u}_3{:}t](\bar{u}-t)_+^3 - [\bar{u}_1,\bar{u}_2{:}t](\bar{u}-t)_+^3}{\bar{u}_3-\bar{u}_1} = [\bar{u}_1,\bar{u}_2,\bar{u}_3{:}t](\bar{u}-t)_+^3$$

and so on.

The general scheme, then, is to:

- replace each cubic truncated power function $(\bar{u}-\bar{u}_i)_+^3$ by $[\bar{u}_i,\bar{u}_{i+1}:t](\bar{u}-t)_+^3$ to obtain a function that eventually grows quadratically, and in which the coefficient of $\bar{u}^2$ is identical for all i;
- replace each such "eventually quadratic" function $[\bar{u}_i,\bar{u}_{i+1}:t](\bar{u}-t)_+^3$ by $[\bar{u}_i,\bar{u}_{i+1},\bar{u}_{i+2}:t](\bar{u}-t)_+^3$ to obtain a function that eventually grows linearly, and in which the coefficient of $\bar{u}$ is identical for all i;
- replace each such "eventually linear" function $[\bar{u}_i,\bar{u}_{i+1},\bar{u}_{i+2}:t](\bar{u}-t)_+^3$ by a difference of itself with $[\bar{u}_{i+1},\bar{u}_{i+2},\bar{u}_{i+3}:t](\bar{u}-t)_+^3$, scaled so that each such difference eventually settles to the same constant value – in fact $[\bar{u}_i,\bar{u}_{i+1},\bar{u}_{i+2},\bar{u}_{i+3}:t](\bar{u}-t)_+^3$ works, resulting in a constant of -1;
- replace each such "eventually constant" function $[\bar{u}_i,\bar{u}_{i+1},\bar{u}_{i+2},\bar{u}_{i+3}:t](\bar{u}-t)_+^3$ by a difference of itself with $[\bar{u}_{i+1},\bar{u}_{i+2},\bar{u}_{i+3},\bar{u}_{i+4}:t](\bar{u}-t)_+^3$, to obtain a function that is eventually zero – in fact $[\bar{u}_i,\bar{u}_{i+1},\bar{u}_{i+2},\bar{u}_{i+3},\bar{u}_{i+4}:t](\bar{u}-t)_+^3$ works, although the division by $\bar{u}_{i+4}-\bar{u}_i$ is performed simply so that this can be expressed as a divided difference like the preceding steps.

These "eventually zero" functions, multiplied by an appropriate scale factor, are exactly the B-splines we wanted; the scale factor of $(\bar{u}_{i+4}-\bar{u}_i)$ in (5) is selected so that the cubic B-splines will be non-negative and sum to one.

It is now easy to see why the extra knots $\bar{u}_{m+1}$, $\bar{u}_{m+2}$ and $\bar{u}_{m+3}$ are convenient – they allow us to transform the truncated cubics at $\bar{u}_{m-3}$, $\bar{u}_{m-2}$ and $\bar{u}_{m-1}$ in the same way, avoiding the need to discuss special cases. We are free to adopt such a notational convenience, as our curve is defined only on $[\bar{u}_0,\bar{u}_m)$. Thus

- in computing the eventually zero function beginning at $\bar{u}_{m-1}$ we need the constant function beginning at $\bar{u}_m$;
- which in turn requires the eventually linear function beginning at $\bar{u}_{m+1}$,
- which in turn requires the eventually quadratic function beginning at $\bar{u}_{m+2}$,
- which in turn requires the truncated cubic beginning at $\bar{u}_{m+3}$.

We are left, as one would expect, with $m+3$ basis functions $B_{-3}(\bar{u})$, ..., $B_{m-1}(\bar{u})$, linear combinations (1) of which yield an arbitrary C^2 piecewise polynomial curve on $[\bar{u}_0,\bar{u}_m)$ of m pieces.

3. Uniformly-Shaped Beta-Splines

The details of what follows may be found in [Barsky1981a].

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

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

and the *curvature vector* is

$$\mathbf{K}(\bar{u}) = \kappa(\bar{u})\,\hat{\mathbf{N}}(\bar{u}) = \kappa(\bar{u})\,\frac{\hat{\mathbf{T}}^{(1)}(\bar{u})}{|\hat{\mathbf{T}}^{(1)}(\bar{u})|} \tag{10}$$

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 $\mathbf{Q}(\bar{u})$.[3] $\hat{\mathbf{T}}(\bar{u})$ and $\mathbf{K}(\bar{u})$ capture the physically meaningful notions of the direction of motion and curvature 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 $\mathbf{K}(\bar{u})$ to be continuous also at the joints between consecutive curve segments of $\mathbf{Q}(\bar{u})$ (which is called G^2 or *second degree geometric continuity*), it is sufficient that

$$\mathbf{Q}_{i-1}(\bar{u}_i) = \mathbf{Q}_i(\bar{u}_i) \tag{11}$$

$$\beta_1 \mathbf{Q}_{i-1}^{(1)}(\bar{u}_i) = \mathbf{Q}_i^{(1)}(\bar{u}_i) \tag{12}$$

$$\beta_1^2 \mathbf{Q}_{i-1}^{(2)}(\bar{u}_i) + \beta_2 \mathbf{Q}_{i-1}^{(1)}(\bar{u}_i) = \mathbf{Q}_i^{(2)}(\bar{u}_i) \tag{13}$$

at every knot $\bar{u}_i$ and for any $\beta_1 > 0$ and β_2 [Barsky1981a]. These equations are by definition less restrictive than simple continuity of position and parametric derivatives ($\beta_1 = 1$ and $\beta_2 = 0$), the special case corresponding to the C^2 splines.

Equation (11) enforces positional continuity. Equation (12) requires that the first parametric derivative vectors from the left and right at a joint be colinear, but allows their magnitudes to differ. There is an instantaneous change in "velocity" at the joint, but not a change in direction.

A sufficient condition for curvature continuity is that $\mathbf{Q}_i^{(2)}(\bar{u}_i) = \beta_1^2 \mathbf{Q}_{i-1}^{(2)}(\bar{u}_i)$, the factor of β_1^2 arising from the assumption that equation (12) holds (see [Barsky1981a, Barsky1983a, Bartels1983b] for details.) However, $\mathbf{Q}_i^{(2)}(\bar{u}_i)$ may have an additional component directed along the tangent $\mathbf{Q}_{i-1}^{(1)}(\bar{u}_i)$ since acceleration along the tangent does not "deflect" a point traveling along the curve, and so does not affect the curvature there.

Curves possessing G^2 continuity result if they are defined using equation (1) in

[3] The *osculating circle* at $\mathbf{Q}(\bar{u})$ is the circle whose first and second derivative vectors agree with those of $\mathbf{Q}$ at $\bar{u}$. The curvature $\kappa(\bar{u})$ is then the reciprocal of the radius of this osculating circle.

terms of G^2 continuous basis functions whose segment polynomials $s(\bar{u})$ themselves satisfy scalar versions of the vector constraint equations (11), (12) and (13):

$$s_{left}(\bar{u}_i) = s_{right}(\bar{u}_i) \tag{14}$$

$$\beta_1 s_{left}^{(1)}(\bar{u}_i) = s_{right}^{(1)}(\bar{u}_i) \tag{15}$$

$$\beta_1{}^2 s_{left}^{(2)}(\bar{u}_i) + \beta_2 s_{left}^{(1)}(\bar{u}_i) = s_{right}^{(2)}(\bar{u}_i) \ . \tag{16}$$

For a uniform knot sequence ($\bar{u}_i - \bar{u}_{i-1} = 1$) and fixed values of β_1 and β_2, such a basis function is defined by

$$\begin{aligned}
s_{i,-1}(u) &= \frac{1}{\delta}\left(2u^3 \right) \\
s_{i,-2}(u) &= \frac{1}{\delta}\Big(2 + (6\beta_1)u + (3\beta_2 + 6\beta_1{}^2)u^2 \\
&\qquad\qquad - (2\beta_2 + 2\beta_1{}^2 + 2\beta_1 + 2)u^3 \Big) \\
s_{i,-3}(u) &= \frac{1}{\delta}\Big((\beta_2 + 4\beta_1{}^2 + 4\beta_1) + (6\beta_1{}^3 - 6\beta_1)u \\
&\qquad - (3\beta_2 + 6\beta_1{}^3 + 6\beta_1{}^2)u^2 \\
&\qquad + (2\beta_2 + 2\beta_1{}^3 + 2\beta_1{}^2 + 2\beta_1)u^3 \Big) \\
s_{i,-4}(u) &= \frac{1}{\delta}\left((2\beta_1{}^3) - (6\beta_1{}^3)u + (6\beta_1{}^3)u^2 - (2\beta_1{}^3)u^3 \right)
\end{aligned} \tag{17}$$

where

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

For simplicity these expressions have been individually parametrized via the substitutions $u = \bar{u} - \bar{u}_j$ for $j = i,\ i+1,\ i+2,\ i+3$; $u = 0$ yields the left end of each segment and $u = 1$ yields the right end of each segment. The composite

$$G_i(\bar{u}) \equiv \begin{cases} 0 & \text{for } \bar{u} < \bar{u}_i \\ s_{i,-1}(u) & \text{for } u = \bar{u} - \bar{u}_i \text{ and } \bar{u}_i \le \bar{u} < \bar{u}_{i+1} \\ s_{i,-2}(u) & \text{for } u = \bar{u} - \bar{u}_{i+1} \text{ and } \bar{u}_{i+1} \le \bar{u} < \bar{u}_{i+2} \\ s_{i,-3}(u) & \text{for } u = \bar{u} - \bar{u}_{i+2} \text{ and } \bar{u}_{i+2} \le \bar{u} < \bar{u}_{i+3} \\ s_{i,-4}(u) & \text{for } u = \bar{u} - \bar{u}_{i+3} \text{ and } \bar{u}_{i+3} \le \bar{u} < \bar{u}_{i+4} \\ 0 & \text{for } \bar{u}_{i+4} \le \bar{u} \end{cases}$$

is the ***uniformly-shaped Beta-spline*** introduced in [Barsky1981a], which was constructed over a uniform knot sequence. Because the knots are equally spaced the uniformly-shaped Beta-splines are translates of one another, so that $s_{i,r}(u) = s_{j,r}(u)$, and it is perfectly safe to write $s_r(u)$. Use of the basis function defined by (17) results in a *uniformly-shaped Beta-spline curve.*

4. Discretely-Shaped Beta-Splines

We seek a simple and computationally efficient means: (a) to attach distinct values of β_1 and β_2 to each joint in a piecewise cubic polynomial curve, in such a way that changing a single β parameter will alter only a local portion of the curve being defined; and (b) to generalize the uniformly-shaped Beta-splines to non-uniform knot sequences. Our approach is analogous to the development of cubic B-splines sketched earlier.

4.1. A Truncated Power Basis for the Beta-Splines

Our first task is to define an analog of the truncated power function. $(\bar{u}-t)_+^3$ itself will not do, since its first and second derivatives are continuous across all knots. What we want is a function that undergoes a jump in its first and second derivatives as it crosses each knot, sufficient to satisfy (14), (15) and (16). Consider a function of the form

$$p(\bar{u}) + a_{i,i+1}(\bar{u}-\bar{u}_{i+1})_+^1 + b_{i,i+1}(\bar{u}-\bar{u}_{i+1})_+^2 \ .$$

Its first and second derivatives from the left at $\bar{u}_{i+1}$ are simply $p^{(1)}(\bar{u}_i)$ and $p^{(2)}(\bar{u}_i)$. (We assume that these exist.) Its first and second derivatives from the right at $\bar{u}_{i+1}$ are

$$p^{(1)}(\bar{u}_{i+1}) + a_{i,i+1} \tag{18}$$

$$p^{(2)}(\bar{u}_{i+1}) + 2b_{i,i+1} \ . \tag{19}$$

Thus there is a jump of $a_{i,i+1}$ in the first derivative and of $2b_{i,i+1}$ in the second derivative. If we want to satisfy (15) then we must have

$$\beta_{1,i+1}p^{(1)}(\bar{u}_{i+1}) = p^{(1)}(\bar{u}_{i+1}) + a_{i,i+1}$$

or

$$a_{i,i+1} = (\beta_{1,i+1}-1)\,p^{(1)}(\bar{u}_{i+1}) \ . \tag{20}$$

To satisfy (16) we must have

$$\beta_{1,i+1}^2p^{(2)}(\bar{u}_{i+1}) + \beta_{2,i+1}p^{(1)}(\bar{u}_{i+1}) = p^{(2)}(\bar{u}_{i+1}) + 2b_{i,i+1}$$

or

$$b_{i,i+1} = \frac{1}{2}\left[(\beta_{1,i+1}^2-1)p^{(2)}(\bar{u}_{i+1}) + \beta_{2,i+1}p^{(1)}(\bar{u}_{i+1})\right] \ . \tag{21}$$

These equations tell us how to modify an arbitrary function so that it will satisfy our G^2 continuity conditions as it crosses a knot. To construct a truncated power basis for the Beta-splines, we begin with a truncated power function $(\bar{u}-\bar{u}_i)_+^3$ and modify it as above each time it crosses a knot. Consider the function

$$\begin{aligned} g_i(\bar{u}) &= (\bar{u}-\bar{u}_i)_+^3 + a_{i,i+1}(\bar{u}-\bar{u}_{i+1})_+^1 + \cdots + a_{i,m+3}(\bar{u}-\bar{u}_{m+3})_+^1 \\ &\qquad + b_{i,i+1}(\bar{u}-\bar{u}_{i+1})_+^2 + \cdots + b_{i,m+3}(\bar{u}-\bar{u}_{m+3})_+^2 \ . \end{aligned}$$

Since (14), (15) and (16) will necessarily be satisfied by any linear combination of functions individually satisfying (14), (15) and (16), it is sufficient to ensure that the functions $g_i(\bar{u})$ each do so.

$(\bar{u}-\bar{u}_i)_+^3$ itself has zero value, as well as zero first and second derivatives, at $\bar{u}_i$ and at all knots left of $\bar{u}_i$, and so trivially satisfies our G^2 constraints. It is therefore sufficient to define the $a_{i,j}$ and $b_{i,j}$ from left to right, for $i < j \leq m+3$, using equations (20) and (21). Thus when computing $a_{i,i+1}$ and $b_{i,i+1}$, $p(\bar{u})$ is simply $(\bar{u}-\bar{u}_i)_+^3$; more generally, when computing $a_{i,j}$ and $b_{i,j}$, $p(\bar{u})$ has the value

$$(\bar{u}-\bar{u}_i)^3 + \sum_{k=i+1}^{j-1} a_{i,k}(\bar{u}-\bar{u}_k)^1 + \sum_{k=i+1}^{j-1} b_{i,k}(\bar{u}-\bar{u}_k)^2 \ ,$$

the preceding a_i's and b_i's having already been computed. Consequently the first derivative $P_{left}^{(1)}$ of $p(\bar{u})$ at $\bar{u}_j$ is

$$3(\bar{u}_j-\bar{u}_i)^2 + \sum_{k=i+1}^{j-1} a_{i,k} + 2\sum_{k=i+1}^{j-1} b_{i,k}(\bar{u}_j-\bar{u}_k)^1$$

and the second derivative $P_{left}^{(2)}$ is

$$6(\bar{u}_j-\bar{u}_i)^1 + 2\sum_{k=i+1}^{j-1} b_{i,k} \ .$$

The following algorithm computes the $a_{i,j}$ and $b_{i,j}$.

Algorithm I

for $i \leftarrow 0$ **step** 1 **until** $m+2$ **do**

$Sa \leftarrow 0$

$Sb \leftarrow 0$

for $j \leftarrow i+1$ **step** 1 **until** $m+3$ **do**

$P_{left}^{(1)} \leftarrow 3(\bar{u}_j - \bar{u}_i)^2 + Sa + 2 \sum_{k=i+1}^{j-1} b_{i,k}(\bar{u}_j - u_k)$

$P_{left}^{(2)} \leftarrow 6(\bar{u}_j - \bar{u}_i) + Sb$

$a_{i,j} \leftarrow (\beta_{1,j} - 1) P_{left}^{(1)}$

$b_{i,j} \leftarrow \frac{1}{2}\left[(\beta_{1,j}^2 - 1) P_{left}^{(2)} + \beta_{2,j} P_{left}^{(1)} \right]$

$Sa \leftarrow Sa + a_{i,j}$

$Sb \leftarrow Sb + 2b_{i,j}$

endfor

endfor

The outer loop steps through the $g_i(\bar{u})$ in turn. For each $g_i(\bar{u})$ the inner loop computes the $a_{i,j}$'s and $b_{i,j}$'s; Sa and Sb keep a running total of the $a_{i,j}$'s and $b_{i,j}$'s which have been computed thus far.

It is not hard to see that the functions $g_i(\bar{u})$ form a basis for the G^2 splines over some particular knot sequence and associated shape parameters $\beta_{1,i}$ and $\beta_{2,i}$ – the argument is very much analogous to that given in the case of C^2 splines for the truncated cubics, and is therefore omitted.

4.2. A Local Basis for the Beta-Splines

The $g_i(\bar{u})$ have the same deficiencies – namely rapid growth and non-locality – that the truncated power basis for the C^2 splines suffers from. The obvious next step, then, is to see whether some form of differencing can be applied to the $g_i(\bar{u})$ so as to abtain a local basis.

Just as when constructing the B-splines, the cubic term in each $g_i(\bar{u})$ is easily cancelled for $\bar{u}$ sufficiently far to the right. We need only compute

$$g_{i+1}(\bar{u}) - g_i(\bar{u}) \ , \tag{22}$$

$$g_{i+2}(\bar{u}) - g_{i+1}(\bar{u}) \ , \tag{23}$$

and so on. In order to cancel the quadratic terms in (22) and (23) by computing a further difference we need to arrange for the coefficient of $\bar{u}^2$ in (22) and (23) to have the same value. Unfortunately, these coefficients depend not only on the knot spacing (as was true for the B-splines), but also on the particular knot interval containing $\bar{u}$ since we pick up an additional $a_{i,j}$ and $b_{i,j}$ each time we move rightwards across a knot. In particular, if $\bar{u}_j \leq \bar{u} < \bar{u}_{j+1}$ and $i < j$ then

$$\begin{aligned} g_i(\bar{u}) \ = \ & \bar{u}^3 + \bar{u}^2\Big[(b_{i,i+1} + \cdots + b_{i,j}) - 3\bar{u}_i \Big] \\ & + \bar{u} \Big[(a_{i,i+1} + \cdots + a_{i,j}) \\ & \qquad - 2(b_{i,i+1}\bar{u}_{i+1} + \cdots + b_{i,j}\bar{u}_j) + 3\bar{u}_i^2 \Big] \\ & + \Big[(b_{i,i+1}\bar{u}_{i+1}^2 + \cdots + b_{i,j}\bar{u}_j^2) \\ & \qquad - (a_{i,i+1}\bar{u}_{i+1} + \cdots + a_{i,j}\bar{u}_j) - \bar{u}_i^3 \Big] \ , \end{aligned}$$

while the truncated power basis used for the B-splines is simply

$$\begin{aligned} (\bar{u} - \bar{u}_i)^3 \ = \ & \bar{u}^3 + \bar{u}^2\Big[-3\bar{u}_i \Big] \\ & + \bar{u} \Big[+ 3\bar{u}_i^2 \Big] \\ & + \Big[-\bar{u}_i^3 \Big] \end{aligned}$$

for all $\bar{u} > \bar{u}_i$. Thus for $(\bar{u} - \bar{u}_i)_+^3$ the coefficient of $\bar{u}^2$ is a constant, while for $g_i(\bar{u})$ the coefficient of $\bar{u}^2$ alters each time we move rightwards across a knot.

This difficulty can be overcome, however. For the B-splines we needed to take a fourth difference in order to obtain a local function, and the $B_i(\bar{u})$ became zero for $\bar{u} \geq \bar{u}_{i+4}$. At each step we arranged for the leading coefficients to be identical for $\bar{u} \geq \bar{u}_{i+4}$ so that they would cancel when performing the next difference.

For the Beta-splines, then, we will normalize the leading coefficients after each difference so that for $\bar{u}_{i+4} \leq \bar{u} < \bar{u}_{i+5}$ these coefficients will be identical, except for the fourth difference, which will be identically zero "on this interval". Let

$$\begin{aligned} A_{i,i} \ = \ & (b_{i,i+1} + b_{i,i+2} + b_{i,i+3} + b_{i,i+4}) \\ & - 3\bar{u}_i \\ B_{i,i} \ = \ & (a_{i,i+1} + a_{i,i+2} + a_{i,i+3} + a_{i,i+4}) \end{aligned}$$

$$
\begin{aligned}
& - 2(b_{i,i+1}\bar{u}_{i+1} + b_{i,i+2}\bar{u}_{i+2} + b_{i,i+3}\bar{u}_{i+3} + b_{i,i+4}\bar{u}_{i+4}) \\
& + 3\bar{u}_i^2 \\
C_{i,i} \;=\; & (b_{i,i+1}\bar{u}_{i+1}^2 + b_{i,i+2}\bar{u}_{i+2}^2 + b_{i,i+3}\bar{u}_{i+3}^2 + b_{i,i+4}\bar{u}_{i+4}^2) \\
& - (a_{i,i+1}\bar{u}_{i+1} + a_{i,i+2}\bar{u}_{i+2} + a_{i,i+3}\bar{u}_{i+3} + a_{i,i+4}\bar{u}_{i+4}) \\
& - \bar{u}_i^3
\end{aligned}
$$

$$
\begin{aligned}
A_{i,i+1} \;=\; & (b_{i+1,i+2} + b_{i+1,i+3} + b_{i+1,i+4}) \\
& - 3\bar{u}_{i+1} \\
B_{i,i+1} \;=\; & (a_{i+1,i+2} + a_{i+1,i+3} + a_{i+1,i+4}) \\
& - 2(b_{i+1,i+2}\bar{u}_{i+2} + b_{i+1,i+3}\bar{u}_{i+3} + b_{i+1,i+4}\bar{u}_{i+4}) \\
& + 3\bar{u}_{i+1}^2 \\
C_{i,i+1} \;=\; & (b_{i+1,i+2}\bar{u}_{i+2}^2 + b_{i+1,i+3}\bar{u}_{i+3}^2 + b_{i+1,i+4}\bar{u}_{i+4}^2) \\
& - (a_{i+1,i+2}\bar{u}_{i+2} + a_{i+1,i+3}\bar{u}_{i+3} + a_{i+1,i+4}\bar{u}_{i+4}) \\
& - \bar{u}_{i+1}^3
\end{aligned}
$$

$\ldots$

$$
\begin{aligned}
A_{i,i+4} &= -3\bar{u}_{i+4} \\
B_{i,i+4} &= +3\bar{u}_{i+4}^2 \\
C_{i,i+4} &= -\bar{u}_{i+4}^3 \; .
\end{aligned}
$$

Then we may write

$$g_i(\bar{u}) = \bar{u}^3 + A_{i,i}\bar{u}^2 + B_{i,i}\bar{u} + C_{i,i} \qquad \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \; .$$

Similarly we have

$$
\begin{aligned}
g_{i+1}(\bar{u}) &= \bar{u}^3 + A_{i,i+1}\bar{u}^2 + B_{i,i+1}\bar{u} + C_{i,i+1} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \\
g_{i+2}(\bar{u}) &= \bar{u}^3 + A_{i,i+2}\bar{u}^2 + B_{i,i+2}\bar{u} + C_{i,i+2} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \\
g_{i+3}(\bar{u}) &= \bar{u}^3 + A_{i,i+3}\bar{u}^2 + B_{i,i+3}\bar{u} + C_{i,i+3} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \\
g_{i+4}(\bar{u}) &= \bar{u}^3 + A_{i,i+4}\bar{u}^2 + B_{i,i+4}\bar{u} + C_{i,i+4} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \; .
\end{aligned}
$$

From these we form the four functions $\Delta_i^1 g_i(\bar{u})$, $\Delta_i^1 g_{i+1}(\bar{u})$, $\Delta_i^1 g_{i+2}(\bar{u})$ and $\Delta_i^1 g_{i+3}(\bar{u})$ defined by

$$\begin{aligned}
\Delta_i^1 g_j(\bar{u}) &= \frac{g_{j+1}(\bar{u}) - g_j(\bar{u})}{A_{i,j+1} - A_{i,j}} \qquad \text{for all } \bar{u} \text{ and } j = i, i+1, i+2, i+3 \\
&= \bar{u}^2 + \frac{B_{i,j+1} - B_{i,j}}{A_{i,j+1} - A_{i,j}}\bar{u} + \frac{C_{i,j+1} - C_{i,j}}{A_{i,j+1} - A_{i,j}} \qquad \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \ , \\
&= \bar{u}^2 + D_{i,j}\bar{u} + E_{i,j}
\end{aligned}$$

thus implicitly defining the $D_{i,j}$ and $E_{i,j}$. The index i with which we subscript Δ reminds us that we are eventually going to replace $g_i(\bar{u})$ with an appropriate linear combination $G_i(\bar{u})$ of $g_i(\bar{u})$, $g_{i+1}(\bar{u})$, $g_{i+2}(\bar{u})$, $g_{i+3}(\bar{u})$ and $g_{i+4}(\bar{u})$, computed in such a way as to ensure that $G_i(\bar{u})$ will be zero on $\bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5}$.

We can now cancel the quadratic term by forming the three functions $\Delta_i^2 g_i(\bar{u})$, $\Delta_i^2 g_{i+1}(\bar{u})$ and $\Delta_i^2 g_{i+2}(\bar{u})$ as

$$\begin{aligned}
\Delta_i^2 g_j(\bar{u}) &= \frac{\Delta_i^1 g_{j+1}(\bar{u}) - \Delta_i^1 g_j(\bar{u})}{D_{i,j+1} - D_{i,j}} \qquad \text{for all } \bar{u} \text{ and } j = i, i+1, i+2 \\
&= \bar{u} + \frac{E_{i,j+1} - E_{i,j}}{D_{i,j+1} - D_{i,j}} \qquad \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \\
&= \bar{u} + F_{i,j} \qquad \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5}
\end{aligned}$$

and then cancel the linear term by forming the two functions $\Delta_i^3 g_i(\bar{u})$ and $\Delta_i^3 g_{i+1}(\bar{u})$ as

$$\begin{aligned}
\Delta_i^3 g_j(\bar{u}) &= \frac{\Delta_i^2 g_{j+1}(\bar{u}) - \Delta_i^2 g_i(\bar{u})}{F_{i,j+1} - F_{i,j}} \qquad \text{for all } \bar{u} \text{ and } j = i, i+1 \\
&= 1 \qquad \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \ .
\end{aligned}$$

Finally we compute the function

$$\begin{aligned}
\Delta_i^4 g_i(\bar{u}) &= -\left[\Delta_i^3 g_{i+1}(\bar{u}) - \Delta_i^3 g_i(\bar{u})\right] \\
&= 0 \qquad \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \ ,
\end{aligned}$$

with which we replace $g_i(\bar{u})$. The pattern of this computation is shown in the following diagram.

$$
\begin{array}{ccccccccc}
g_i(\bar{u}) & & g_{i+1}(\bar{u}) & & g_{i+2}(\bar{u}) & & g_{i+3}(\bar{u}) & & g_{i+4}(\bar{u}) \\
& \Delta_i^1 g_i(\bar{u}) & & \Delta_i^1 g_{i+1}(\bar{u}) & & \Delta_i^1 g_{i+2}(\bar{u}) & & \Delta_i^1 g_{i+3}(\bar{u}) & \\
& & \Delta_i^2 g_i(\bar{u}) & & \Delta_i^2 g_{i+1}(\bar{u}) & & \Delta_i^2 g_{i+2}(\bar{u}) & & \\
& & & \Delta_i^3 g_i(\bar{u}) & & \Delta_i^3 g_{i+1}(\bar{u}) & & & \\
& & & & \Delta_i^4 g_i(\bar{u}) & & & &
\end{array}
$$

Now $\Delta_i^4 g_i(\bar{u})$ is defined for any value of $\bar{u}$, but we have only ensured that it is zero when $\bar{u}$ lies between $\bar{u}_{i+4}$ and $\bar{u}_{i+5}$, or is less than $\bar{u}_i$. To arrange for locality we simply define our *discretely-shaped Beta-splines* $G_i(\bar{u})$ to be

$$
G_i(\bar{u}) \equiv \begin{cases} 0 & \bar{u} < \bar{u}_i \ \ or \ \ \bar{u} \geq \bar{u}_{i+4} \\ \Delta_i^4 g_i(\bar{u}) & \bar{u}_i \leq \bar{u} < \bar{u}_{i+4} \ . \end{cases}
$$

Since by construction $\Delta_i^4 g_i(\bar{u})$ is zero on $[\bar{u}_4, \bar{u}_5)$, the rightwards extension by zero leaves us with a function satisfying the G^2 continuity constraints.

We have still to argue that the $G_i(\bar{u})$ are a basis for the Beta-splines. Since each of the $G_i(\bar{u})$ is a linear combination of G^2 functions, each of the $G_i(\bar{u})$ is a G^2 function, and there are as many $G_i(\bar{u})$ in $[\bar{u}_0, \bar{u}_m)$ as there are $g_i(\bar{u})$. It remains to be shown that the $G_i(\bar{u})$ are linearly independent, which we will leave to a later section.

Discretely-shaped Beta-spline curves are now defined by the obvious analog of equation (1):

$$
\mathbf{Q}(\bar{u}) = \sum_i \mathbf{V}_i \, G_i(\bar{u}) = \sum_i \left(x_i \, G_i(\bar{u}), \, y_i \, G_i(\bar{u}) \right) \ .
$$

The i^{th} curve segment is

$$
\begin{aligned}
\mathbf{Q}_i(\bar{u}) &= \sum_{r=-4}^{-1} \mathbf{V}_{i+r} \, G_{i+r}(\bar{u}) \\
&= \mathbf{V}_{i-4} G_{i-4}(\bar{u}) + \mathbf{V}_{i-3} G_{i-3}(\bar{u}) \\
&\qquad + \mathbf{V}_{i-2} G_{i-2}(\bar{u}) + \mathbf{V}_{i-1} G_{i-1}(\bar{u}) \ .
\end{aligned}
$$

4.3. Evaluation

For the C^2 splines one defines basis functions $B_{i,k}(\bar{u})$ of arbitrary order k, and develops a recursive definition of $B_{i,k}(\bar{u})$ in terms of $B_{i,k-1}(\bar{u})$ and $B_{i+1,k-1}(\bar{u})$. This provides both an efficient means of computing the value of a basis function, and (it turns out) of computing its derivatives. Given the latter one can then develop an efficient algorithm for converting from a "control vertex" representation such as (1) to a power representation

$$c_0 + c_1(\bar{u} - \bar{u}_i)^1 + c_2(\bar{u} - \bar{u}_i)^2 + c_3(\bar{u} - \bar{u}_i)^3 ,$$

from which one can efficiently compute points along a curve segment by using forward differences. Unfortunately we have not been able to develop such a recursive definition of the Beta-splines, and indeed we rather doubt that a natural such definition exists. This is not, however, a fatal problem. One can simply pre-compute the coefficients $a_{i,j}$, $b_{i,j}$, $A_{i,j}$, $B_{i,j}$, $C_{i,j}$, $D_{i,j}$, $E_{i,j}$, $F_{i,j}$ and then compute the difference $\Delta_i^4 g_i(\bar{u})$ directly whenever a point on the curve is required. Doing so does not require an $a_{i,j}$ or $b_{i,j}$ for any value of j other than $i+1$, $i+2$, $i+3$ or $i+4$. Hence Algorithm I can be made somewhat more efficient by replacing the expression $m+3$ in line 4 by $min(i+4,m+3)$.

Moreover, since differencing and differentiation commute, we may compute derivatives of the $G_i(\bar{u})$ by differencing derivatives of the $g_i(\bar{u})$, and so obtain a power representation of the basis segments which can be evaluated by using forward differences.

4.4. Properties

Practically speaking the most important properties of spline basis functions are summation to one (because this ensures translation invariance) and positivity (because together with summation to one this provides a convex hull property). From a theoretical point of view we must also show that the basis functions we have constructed are, in fact, linearly independent.

4.5. Linear Independence

To consider the question of the linear independence of the functions we are constructing, let us take a look at the differencing process from another point of view. We have established that

$$\begin{aligned}
g_i(\bar{u}) &= \bar{u}^3 + A_{i,i}\bar{u}^2 + B_{i,i}\bar{u} + C_{i,i} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \\
g_{i+1}(\bar{u}) &= \bar{u}^3 + A_{i,i+1}\bar{u}^2 + B_{i,i+1}\bar{u} + C_{i,i+1} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \\
g_{i+2}(\bar{u}) &= \bar{u}^3 + A_{i,i+2}\bar{u}^2 + B_{i,i+2}\bar{u} + C_{i,i+2} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \\
g_{i+3}(\bar{u}) &= \bar{u}^3 + A_{i,i+3}\bar{u}^2 + B_{i,i+3}\bar{u} + C_{i,i+3} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} \\
g_{i+4}(\bar{u}) &= \bar{u}^3 + A_{i,i+4}\bar{u}^2 + B_{i,i+4}\bar{u} + C_{i,i+4} && \text{for } \bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5} .
\end{aligned}$$

The differencing process seeks to find nontrivial coefficients

$\gamma_i\,,\ \gamma_{i+1}\,,\ \gamma_{i+2}\,,\ \gamma_{i+3}\,,\ \gamma_{i+4}$

such that

$$\gamma_i g_i(\bar{u}) + \gamma_{i+1} g_{i+1}(\bar{u}) + \gamma_{i+2} g_{i+2}(\bar{u}) + \gamma_{i+3} g_{i+3}(\bar{u}) + \gamma_{i+4} g_{i+4}(\bar{u}) = 0$$

for all $\bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5}$. This implies a separate equation for each power of $\bar{u}$, which yields, in matrix format,

$$\mathbf{M}^{(0)}\gamma = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ A_{i,i} & A_{i,i+1} & A_{i,i+2} & A_{i,i+3} & A_{i,i+4} \\ B_{i,i} & B_{i,i+1} & B_{i,i+2} & B_{i,i+3} & B_{i,i+4} \\ C_{i,i} & C_{i,i+1} & C_{i,i+2} & C_{i,i+3} & C_{i,i+4} \end{bmatrix} \begin{bmatrix} \gamma_i \\ \gamma_{i+1} \\ \gamma_{i+2} \\ \gamma_{i+3} \\ \gamma_{i+4} \end{bmatrix} = 0 \ .$$

Since this system of equations is underdetermined, we are assured that a nontrivial solution will exist.

The differencing computations are, by inspection, equivalent to performing elementary column operations on the matrix of this system so as to bring it to echelon form. Computing the first differences is equivalent to subtracting the j^{th} column of $\mathbf{M}^{(0)}$ from the $j+1^{st}$ column, replacing the j^{th} column by the result, for $j=0,1,2,3$. The matrix representation of this is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ A_{i,i+1}-A_{i,i} & A_{i,i+2}-A_{i,i+1} & A_{i,i+3}-A_{i,i+2} & A_{i,i+4}-A_{i,i+3} & A_{i,i+4} \\ B_{i,i+1}-B_{i,i} & B_{i,i+2}-B_{i,i+1} & B_{i,i+3}-B_{i,i+2} & B_{i,i+4}-B_{i,i+3} & B_{i,i+4} \\ C_{i,i+1}-C_{i,i} & C_{i,i+2}-C_{i,i+1} & C_{i,i+3}-C_{i,i+2} & C_{i,i+4}-C_{i,i+3} & C_{i,i+4} \end{bmatrix}$$

The left four columns are then scaled by their respective topmost nonzero elements to yield

$$\mathbf{M}^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & A_{i,i+4} \\ D_{i,i} & D_{i,i+1} & D_{i,i+2} & D_{i,i+3} & B_{i,i+4} \\ E_{i,i} & E_{i,i+1} & E_{i,i+2} & E_{i,i+3} & C_{i,i+4} \end{bmatrix} .$$

In the second step the j^{th} column of $\mathbf{M}^{(1)}$ is subtracted from from the $j+1^{st}$ for $j=0,1,2$ to yield

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & A_{i,i+4} \\ D_{i,i+1}-D_{i,i} & D_{i,i+2}-D_{i,i+1} & D_{i,i+3}-D_{i,i+2} & D_{i,i+3} & B_{i,i+4} \\ E_{i,i+1}-E_{i,i} & E_{i,i+2}-E_{i,i+1} & E_{i,i+3}-E_{i,i+2} & E_{i,i+3} & C_{i,i+4} \end{bmatrix}$$

and then the left three columns are each scaled by their respective topmost nonzero elements to yield

$$\mathbf{M}^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & A_{i,i+4} \\ 1 & 1 & 1 & D_{i,i+3} & B_{i,i+4} \\ F_{i,i} & F_{i,i+1} & F_{i,i+2} & E_{i,i+3} & C_{i,i+4} \end{bmatrix} .$$

In the third step $\mathbf{M}^{(2)}$ is similarly transformed to

$$\mathbf{M}^{(3)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & A_{i,i+4} \\ 0 & 0 & 1 & D_{i,i+3} & B_{i,i+4} \\ 1 & 1 & F_{i,i+2} & E_{i,i+3} & C_{i,i+4} \end{bmatrix} .$$

In the fourth and final step a difference of the first and second columns replaces the first to yield

$$\mathbf{M}^{(4)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & A_{i,i+4} \\ 0 & 0 & 1 & D_{i,i+3} & B_{i,i+4} \\ 0 & 1 & F_{i,i+2} & E_{i,i+3} & C_{i,i+4} \end{bmatrix} . \tag{24}$$

The matrix of this system clearly has rank 4.

Every step of the process has taken place using a nonsingular elementary transformation on the right since we have assumed that the denominators $A_{i,i+1} - A_{i,i}$, etc. are non zero. Thus

$$\mathbf{M}^{(1)} = \mathbf{M}^{(0)}\mathbf{S}^{(0)}\mathbf{T}^{(0)}$$

$$\mathbf{M}^{(2)} = \mathbf{M}^{(1)}\mathbf{S}^{(1)}\mathbf{T}^{(1)}$$

where

$$\mathbf{S}^{(0)} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\mathbf{T}^{(0)} = \begin{bmatrix} \frac{1}{A_{i,i+1}-A_{i,i}} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{A_{i,i+2}-A_{i,i+1}} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{A_{i,i+3}-A_{i,i+2}} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{A_{i,i+4}-A_{i,i+3}} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{S}^{(1)} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{T}^{(1)} = \begin{bmatrix} \frac{1}{D_{i,i+1}-D_{i,i}} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{D_{i,i+2}-D_{i,i+1}} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{D_{i,i+3}-D_{i,i+2}} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} .$$

and so on. In total:

$$\mathbf{M}^{(4)} = \mathbf{M}^{(0)}\mathbf{S}^{(0)}\mathbf{T}^{(0)}\mathbf{S}^{(1)}\mathbf{T}^{(1)}\mathbf{S}^{(2)}\mathbf{T}^{(2)}\mathbf{S}^{(3)}\mathbf{T}^{(3)}\mathbf{S}^{(4)} \tag{25}$$

where the matrices **S** provide the column subtractions and the matrices **T** do the scaling. The original system becomes

$$\mathbf{M}^{(4)}\phi = 0$$

where

$$\phi = \begin{bmatrix} \phi_i \\ \phi_{i+1} \\ \phi_{i+2} \\ \phi_{i+3} \\ \phi_{i+4} \end{bmatrix}$$

$$= \mathbf{S}^{(4)-1}\mathbf{T}^{(3)-1}\mathbf{S}^{(3)-1}\mathbf{T}^{(2)-1}\mathbf{S}^{(2)-1}\mathbf{T}^{(1)-1}\mathbf{S}^{(1)-1}\mathbf{T}^{(0)-1}\mathbf{S}^{(0)-1} \begin{bmatrix} \gamma_i \\ \gamma_{i+1} \\ \gamma_{i+2} \\ \gamma_{i+3} \\ \gamma_{i+4} \end{bmatrix} .$$

It is easily determined that

$$\phi_i = \gamma_i (A_{i,i+1} - A_{i,i})(D_{i,i+1} - D_{i,i})(F_{i,i+1} - F_{i,i}) .$$

A number of observations are possible from the above:

- From (24) and (25) it is clear that $\phi_{i+4} = 0$, whence $\phi_{i+3} = 0$, whence $\phi_{i+2} = 0$, whence $\phi_{i+1} = 0$ (by successive substitution).
- If the solution is to be nontrivial it must therefore be the case that γ_i is nonzero, which means that $G_i(\bar{u})$ has a nontrivial contribution from $g_i(\bar{u})$.
- The right-hand four columns of $\mathbf{M}^{(0)}$ are transformed in a way which is completely independent of the first column. That is, if we were to carry out the same steps on the submatrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ A_{i,i+1} & A_{i,i+2} & A_{i,i+3} & A_{i,i+4} \\ B_{i,i+1} & B_{i,i+2} & B_{i,i+3} & B_{i,i+4} \\ C_{i,i+1} & C_{i,i+2} & C_{i,i+3} & C_{i,i+4} \end{bmatrix}$$

  then the same sequence of operations restricted to this submatrix would yield the nonsingular matrix

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & A_{i,i+4} \\ 0 & 1 & D_{i,i+3} & B_{i,i+4} \\ 1 & F_{i,i+2} & E_{i,i+3} & C_{i,i+4} \end{bmatrix} .$$

As a result of this last observation, we may conclude that

$$g_{i+1}(\bar{u}) \; , \; g_{i+2}(\bar{u}) \; , \; g_{i+3}(\bar{u}) \; , \; g_{i+4}(\bar{u})$$

are linearly independent on the interval $[\bar{u}_{i+4}, \bar{u}_{i+5})$, since if constants

$\delta_{i+1}, \ldots, \delta_{i+4}$ are to be found satisfying

$$\delta_{i+1} g_{i+1}(\bar{u}) + \cdots + \delta_{i+4} g_{i+4}(\bar{u}) = 0$$

for $\bar{u}_{i+4} \le \bar{u} < \bar{u}_{i+5}$, then these constants must solve the system

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & A_{i,i+4} \\ 0 & 1 & D_{i,i+3} & B_{i,i+4} \\ 1 & F_{i,i+2} & E_{i,i+3} & C_{i,i+4} \end{bmatrix} \begin{bmatrix} \delta_{i+1} \\ \delta_{i+2} \\ \delta_{i+3} \\ \delta_{i+4} \end{bmatrix} = 0$$

and the only possible solution is the trivial one.

In this fashion the success of the differencing process itself verifies the linear independence of

$$g_i(\bar{u}) \ , \ g_{i+1}(\bar{u}) \ , \ g_{i+2}(\bar{u}) \ , \ g_{i+3}(\bar{u}) \quad \text{on } [\bar{u}_{i+3}, \bar{u}_{i+4})$$

for $i = -3, \ldots, m-1$. Since $G_{-3}(\bar{u})$ is a linear combination of $g_{-3}(\bar{u}), \ldots, g_0(\bar{u})$ with a nonzero contribution from $g_{-3}(\bar{u})$, we may conclude that

$$G_{-3}(\bar{u}) \ , \ g_{-2}(\bar{u}) \ , \ g_{-1}(\bar{u}) \ , \ g_0(\bar{u})$$

are linearly independent on $[\bar{u}_0, \bar{u}_1)$. It may be shown similarly that $G_{-2}(\bar{u})$ is a linear combination of $g_{-2}(\bar{u}), g_{-1}(\bar{u})$, and $g_0(\bar{u})$ on $[\bar{u}_0, \bar{u}_1)$ with a nonzero contribution from $g_{-2}(\bar{u})$. Hence

$$G_{-3}(\bar{u}) \ , \ G_{-2}(\bar{u}) \ , \ g_{-1}(\bar{u}) \ , \ g_0(\bar{u})$$

are linearly independent on $[\bar{u}_0, \bar{u}_1)$. Proceeding in this fashion, we find that

$$G_{-3}(\bar{u}) \ , \ G_{-2}(\bar{u}) \ , \ G_{-1}(\bar{u}) \ , \ G_0(\bar{u})$$

are linearly independent on $[\bar{u}_0, \bar{u}_1)$, and by a generalization of this argument, that

$$G_i(\bar{u}) \ , \ G_{i+1}(\bar{u}) \ , \ G_{i+2}(\bar{u}) \ , \ G_{i+3}(\bar{u})$$

are linearly independent on $[\bar{u}_{i+3}, \bar{u}_{i+4})$ for $i = -3, \ldots, m-4$. The locality of the functions $G_i(\bar{u})$ allows us to argue that they are linearly independent on the entire interval $[\bar{u}_0, \bar{u}_m)$ as follows: if

$$\delta_{-3} G_{-3}(\bar{u}) + \cdots + \delta_{m-1} G_{m-1}(\bar{u}) = 0$$

for all $\bar{u}$ in $[\bar{u}_0, \bar{u}_m)$, then when $\bar{u}$ falls in the subinterval $[\bar{u}_i, \bar{u}_{i+1})$, this sum reduces to

$$\delta_{i-3} G_{i-3}(\bar{u}) + \delta_{i-2} G_{i-2}(\bar{u}) + \delta_{i-1} G_{i-1}(\bar{u}) + \delta_i G_i(\bar{u}) = 0$$

and linear independence of $G_{i-3}(\bar{u}), \ldots, G_i(\bar{u})$ on this subinterval implies that

$$\delta_{i-3} = \cdots = \delta_i = 0 \ .$$

4.6. Partition of Unity

Because the constant polynomial 1 is trivially a G^2 spline for every knot sequence, no matter what values of β_1 and β_2 are required at the joints, it is clear that scale factors c_i exist such that $\sum c_i G_i(\bar{u}) = 1$.[4] We are certain from computational experience, although we do not yet have a proof, that in fact $c_i = 1$ for all i.

4.7. Locality

Consider a β value at the joint corresponding to the knot $\bar{u}_i$ (See Figure 4). By construction it is clear that no basis function prior to $G_{i-4}(\bar{u})$ or subsequent to $G_i(\bar{u})$ could possibly be affected by a change in $\beta_{1,i}$ or $\beta_{2,i}$ because no use is made of them in the truncated power functions from which $G_{i-4}(\bar{u})$ and $G_i(\bar{u})$ are formed (on the intervals of interest). Hence we know immediately that the effect of changing β values at $\bar{u}_i$ must be restricted at least to the eight intervals comprising $[\bar{u}_{i-4},\bar{u}_{i+4})$.

We can show substantially greater locality if we assume that

$$G_{i-3}(\bar{u}) + G_{i-2}(\bar{u}) + G_{i-1}(\bar{u}) + G_i(\bar{u}) = 1$$

on $[\bar{u}_i,\bar{u}_{i+1})$ without the need of further scale factors, as is almost certainly the case. We make this assumption throughout the remainder of this subsection.

It is clear from our construction that no use is made of $\beta_{1,i}$ or $\beta_{2,i}$ in constructing $G_i(\bar{u})$ and $G_i(\bar{u})$ must therefore be independent of $\beta_{1,i}$ and $\beta_{2,i}$. With somewhat more effort we can also show that $G_{i-4}(\bar{u})$ is independent of $\beta_{1,i}$ and $\beta_{2,i}$.

- We are assuming that

$$G_{i-7}(\bar{u}_{i-3}) + G_{i-6}(\bar{u}_{i-3}) + G_{i-5}(\bar{u}_{i-3}) + G_{i-4}(\bar{u}_{i-3}) = 1$$

- We already know that $G_{i-7}(\bar{u})$, $G_{i-6}(\bar{u})$ and $G_{i-5}(\bar{u})$ are independent of $\beta_{1,i}$ and $\beta_{2,i}$.
- Hence $G_{i-4}(\bar{u}_{i-3})$ has some fixed value K, regardless of the value assigned to $\beta_{1,i}$ or $\beta_{2,i}$.
- But $G_{i-4}(\bar{u})$ is composed of the four segment polynomials $s_{i-4,-1}(\bar{u})$, $s_{i-4,-2}(\bar{u})$, $s_{i-4,-3}(\bar{u})$ and $s_{i-4,-4}(\bar{u})$ (having sixteen coefficients) which are the necessarily unique solution to the fifteen equations obtained by applying the constraints (14), (15) and (16) at $\bar{u}_{i-4}$, $\bar{u}_{i-3}$, $\bar{u}_{i-2}$, $\bar{u}_{i-1}$ and

[4] We are indebted to Tony DeRose for pointing this out.

$\bar{u}_i$ together with the (scaling) constraint that $G_{i-4}(\bar{u}_{i-3}) = K$.

- Hence $G_{i-4}(\bar{u})$ cannot change for any value of $\bar{u}$ if $\beta_{1,i}$ or $\beta_{2,i}$ is altered.

Since neither $G_{i-4}(\bar{u})$ not $G_i(\bar{u})$ is dependent on the shape parameters at $\bar{u}_i$, we may conclude that the effect of changing these values at $\bar{u}_i$ must be restricted at least to the six intervals comprising $[\bar{u}_{i-3},\bar{u}_{i+3})$.

By a similar argument we can easily show that $G_{i-3}(\bar{u})$ is independent of $\beta_{1,i}$ and $\beta_{2,i}$ on $[\bar{u}_{i-3},\bar{u}_{i-2})$, and that $G_{i-1}(\bar{u})$ is independent of $\beta_{1,i}$ and $\beta_{2,i}$ on $[\bar{u}_{i+2},\bar{u}_{i+3})$. Finally, then, we conclude that the affect of changing β values at $\bar{u}_i$ is restricted to the four intervals comprising $[\bar{u}_{i-2},\bar{u}_{i+2})$, under the assumption that the $G_i(\bar{u})$ sum to one without further scaling.

Thus the amount of re-computation required by the change of a shape parameter is independent of the number of control vertices defining the curve, and the change in shape is local.

4.8. Positivity

In order to demonstrate a convex hull property for the discretely-shaped Beta-splines we need to establish that the basis functions are non-negative. The argument we will give[5] relies on an assumption that $\beta_1, \beta_2 \geq 0$. G^2 continuity prohibits negative values of β_1, for in such a case the unit tangent vector changes direction. Negative values of β_2 do not violate G^2 continuity, but can result in basis functions which change sign.

For the time being we shall assume that $G_i(\bar{u})$ goes positive at $\bar{u}_i$; that is, $G_i(\bar{u}_i+\varepsilon) > 0$ for all sufficiently small $\varepsilon > 0$. If this is not the case then we consider $-G_i(\bar{u})$ below. Thus we are actually about to show that $G_i(\bar{u})$ does not change sign on $[\bar{u}_i,\bar{u}_{i+4})$. When this preliminary effort is complete we shall see that, in fact, $G_i(\bar{u})$ must be non-negative.

Writing out the segment polynomials for $G_i(\bar{u})$ we have

$$G_i(\bar{u}) = \begin{cases} s_{i,-1}(\bar{u}) & \text{for } \bar{u}_i \leq \bar{u} < \bar{u}_{i+1} \\ s_{i,-2}(\bar{u}) & \text{for } \bar{u}_{i+1} \leq \bar{u} < \bar{u}_{i+2} \\ s_{i,-3}(\bar{u}) & \text{for } \bar{u}_{i+2} \leq \bar{u} < \bar{u}_{i+3} \\ s_{i,-4}(\bar{u}) & \text{for } \bar{u}_{i+3} \leq \bar{u} < \bar{u}_{i+4}\ . \end{cases}$$

Since $G_i(\bar{u})$ is a linear combination of

[5] The argument given here expands upon one suggested to us by Larry Schumaker, for which the authors express their thanks.

$$(\bar{u}-\bar{u}_j)_+^3 \quad \text{for} \quad j = i, i+1, i+2, i+3, i+4$$

$$(\bar{u}-\bar{u}_j)_+^2 \quad \text{for} \quad j = i+1, i+2, i+3, i+4$$

$$(\bar{u}-\bar{u}_j)_+^1 \quad \text{for} \quad j = i+1, i+2, i+3, i+4$$

it follows that

$$s_{i,-1}(\bar{u}) = c_i(\bar{u}-\bar{u}_i)^3$$

for some constant $c_i > 0$. This establishes that $s_{i,-1}(\bar{u})$ cannot have a zero within the interval $(\bar{u}_i, \bar{u}_{i+1})$.

Since $G_i(\bar{u})$ is a linear combination of G^2 functions, it must also be G^2. This means that

$$\begin{aligned}
\beta_{1,i+4}^2 s_{i,-4}^{(2)}(\bar{u}_{i+4}) + \beta_{2,i+4} s_{i,-4}^{(1)}(\bar{u}_{i+4}) &= 0 \\
\beta_{1,i+4} s_{i,-4}^{(1)}(\bar{u}_{i+4}) &= 0 \\
s_{i,-4}(\bar{u}_{i+4}) &= 0
\end{aligned}$$

since $G_i(\bar{u})$ has been constructed to be identically zero from the right at $\bar{u}_{i+4}$. These equations imply that

$$s_{i,-4}(\bar{u}_{i+4}) = s_{i,-4}^{(1)}(\bar{u}_{i+4}) = s_{i,-4}^{(2)}(\bar{u}_{i+4}) = 0 \ .$$

Since $s_{i,-4}(\bar{u})$ is a cubic polynomial, this implies (by Taylor's theorem) that $s_{i,-4}(\bar{u})$ can be expressed as

$$s_{i,-4}(\bar{u}) = d_{i+3}(\bar{u}-\bar{u}_{i+4})^3$$

for some constant d_{i+3} (whose sign is yet to be discovered). As in the case of $s_{i,-1}(\bar{u})$, this guarantees that $s_{i,-4}(\bar{u})$ cannot have a zero within the interval $(\bar{u}_{i+3}, \bar{u}_{i+4})$. Our goal is to establish that $s_{i,-2}(\bar{u})$, $s_{i,-3}(\bar{u})$, and $s_{i,-4}(\bar{u})$ are each positive on their respective segment intervals.

We will consider these segment polynomials from left to right, or from right to left, as circumstances dictate, studying each transition from one to its neighbour across the joining knot. Of importance in establishing the number of zeros a polynomial can have within an interval is the pattern of signs which its derivatives have at the endpoints of the interval:

Theorem (Budin-Fourier):

If $p(\bar{u})$ is a polynomial of exact order k (its $k-1^{st}$ derivative is a nonzero constant), then $Z_{(a,b)}(p)$, the number of zeros of p (counting multiplicities) on the interval $a < \bar{u} < b$ satisfies

$$Z_{(a,b)}(p) \leq k - 1$$

$$- S^+[p(a), -p^{(1)}(a), \ldots, (-1)^{k-1}p^{(k-1)}(a)]$$

$$- S^+[p(b), p^{(1)}(b), \ldots, p^{(k-1)}(b)]$$

where $S^+[v_0, \ldots, v_{k-1}]$ stands for the count of sign changes which is obtained for the sequence of numbers $v_0, \ldots, v_{k-1}$ by regarding each zero value as $+1$ or -1 as needed to maximize the count.

We have, for instance,

$$\begin{array}{ll}
s_{i,-1}(\bar{u}_i) = 0 & s_{i,-1}(\bar{u}_{i+1}) > 0 \\
s^{(1)}_{i,-1}(\bar{u}_i) = 0 & s^{(1)}_{i,-1}(\bar{u}_{i+1}) > 0 \\
s^{(1)}_{i,-1}(\bar{u}_i) = 0 & s^{(1)}_{i,-1}(\bar{u}_{i+1}) > 0 \\
s^{(3)}_{i,-1}(\bar{u}_i) > 0 & s^{(3)}_{i,-1}(\bar{u}_{i+1}) > 0 \\
\hline
\bar{u}_i & \bar{u}_{i+1}
\end{array}$$

or, in brief,

$$\begin{array}{ll}
0 & + \\
0 & + \\
0 & + \\
+ & + \\
\hline
\bar{u}_i & \bar{u}_{i+1}
\end{array}$$

for which

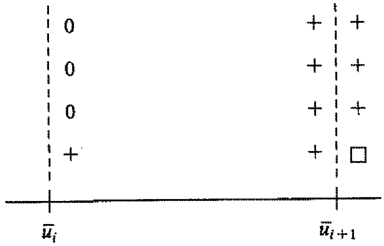
$$Z_{(\bar{u}_i, \bar{u}_{i+1})}(s_{i,-1}) \leq 4 - 1 - 3 - 0 = 0 \ .$$

The Budin-Fourier theorem is proved and discussed in [Schumaker1981a].

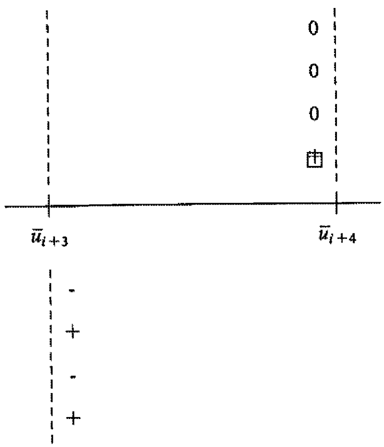
The G^2 conditions let us infer how the value and the first two derivatives of a segment polynomial influence the value and first two derivatives of its neighbouring polynomial. For instance,

$$\begin{aligned}
s_{i,-2}(\bar{u}_{i+1}) &= s_{i,-1}(\bar{u}_{i+1}) > 0 \\
s^{(1)}_{i,-2}(\bar{u}_{i+1}) &= \beta_{1,i+1} s^{(1)}_{i,-1}(\bar{u}_{i+1}) > 0 \\
s^{(2)}_{i,-2}(\bar{u}_{i+1}) &= \beta^2_{1,i+1} s^{(2)}_{i,-1}(\bar{u}_{i+1}) + \beta_{2,i+1} s^{(1)}_{i,-1}(\bar{u}_{i+1}) > 0
\end{aligned}$$

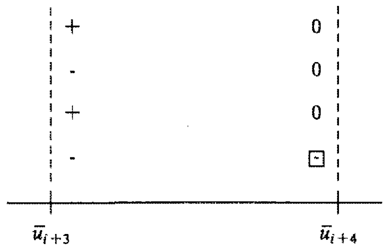
which may be depicted as



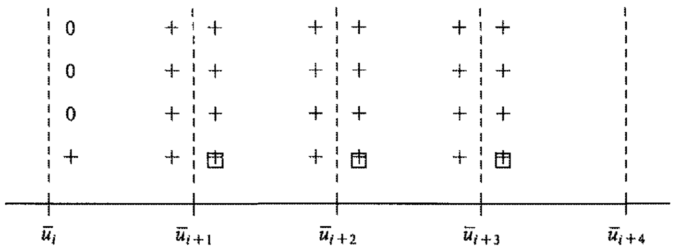
where the small box indicates that $s^{(3)}_{i,-2}(\bar{u}_{i+1})$ is unknown. We will proceed on a case–by–case basis to determine what sign configurations are legal over the segment intervals. Where signs are unknown, we will examine each of the possibilities in turn. Where obvious impossibilities show up, we will backtrack and alter earlier choices until all avenues have been explored. In all cases, since the sign of $s_{i,-4}(\bar{u})$ is unknown on the interval $(\bar{u}_{i+3}, \bar{u}_{i+4})$, we must investigate both of the possibilities



and

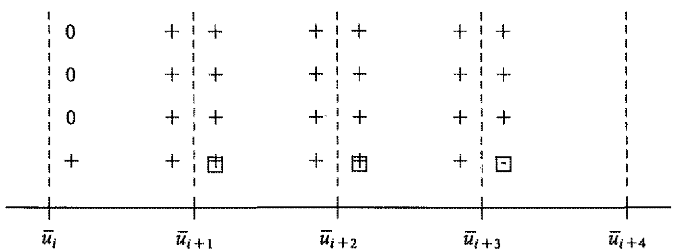


Clearly the first possibility, consistent with choosing + from left to right at every opportunity, is



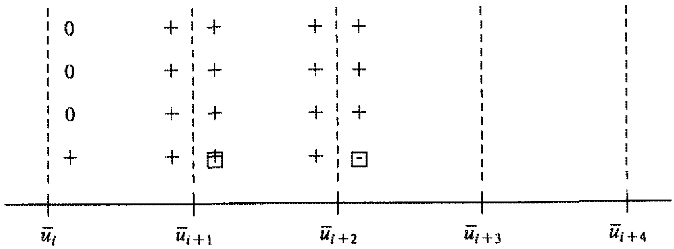
This yields a sign pattern on the right of $\bar{u}_{i+3}$ which is inconsistent with either option for $s_{i,-4}(\bar{u})$. (Note that the choices of + for the third derivatives may be regarded as including the possibility of a zero value.)

The next possibility is



which is also inconsistent with either option for $s_{i,-4}(\bar{u})$.

The next set of choices from left to right causes us to consider

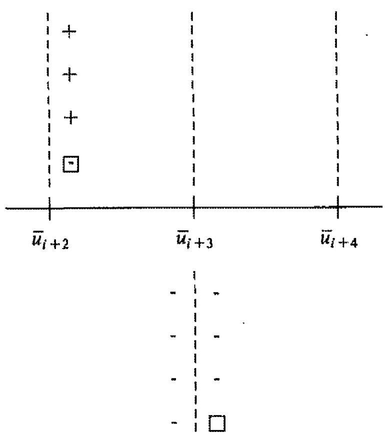


By Taylor's expansion, $s_{i,-3}(\bar{u})$ is given by

$$\begin{aligned} s_{i,-3}(\bar{u}) \;=\; & s_{i,-3}(\bar{u}_{i+2}) \\ & + s_{i,-3}^{(1)}(\bar{u}_{i+2})(\bar{u}-\bar{u}_{i+2}) \\ & + s_{i,-3}^{(2)}(\bar{u}_{i+2})(\bar{u}-\bar{u}_{i+2})^2 \\ & + s_{i,-3}^{(3)}(\bar{u}_{i+2})(\bar{u}-\bar{u}_{i+2})^3 \;. \end{aligned}$$

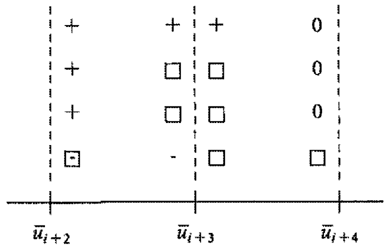
Since $s_{i,-3}(\bar{u}_{i+2})$, $s_{i,-3}^{(1)}(\bar{u}_{i+2})$, and $s_{i,-3}^{(2)}(\bar{u}_{i+2})$ are assumed to be positive, with $s_{i,-3}^{(3)}(\bar{u}_{i+2})$ assumed negative, it is easy to see that $s_{i,-3}(\bar{u})$ has precisely one zero to the right of $\bar{u}_{i+2}$. The question of interest is whether that zero falls in the interval $(\bar{u}_{i+2}, \bar{u}_{i+3}]$ or not.

In case it does, we have by inspection

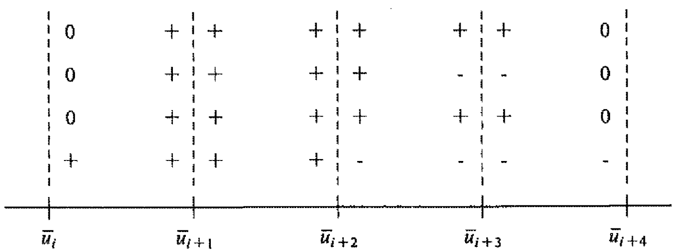


where the first minus sign in each column under $\bar{u}_{i+3}$ may be regarded as including the possibility of a zero value. Clearly, this is incompatible with $s_{i,-4}(\bar{u})$.

In case the zero of $s_{i,-3}(\bar{u})$ does not occur in the segment interval, we have

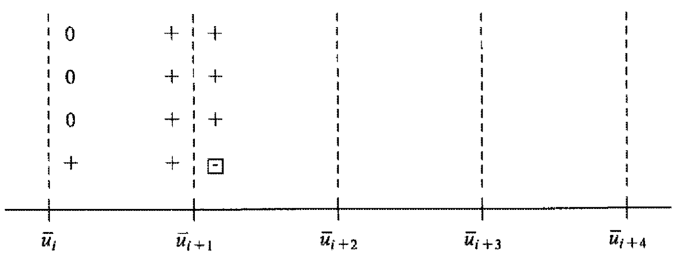


Note that this situation is compatible only with the sign pattern $(0,0,0,-)$ for $s_{i,-4}(\bar{u}_{i+4})$ and not with the pattern $(0,0,0,+)$. In the compatible case we would have



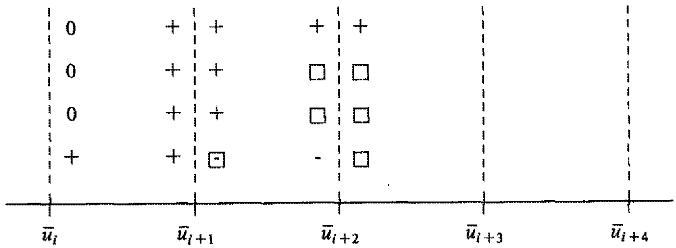
The sign pattern which this reveals for $(\bar{u}_{i+2},\bar{u}_{i+3})$ suggests that $s_{i,-3}(\bar{u})$ can have no more than $4-1-5=-2$ zeros in this interval, which is nonsense. This means that the sign pattern in the interval is incompatible with a cubic.

The next possibility is

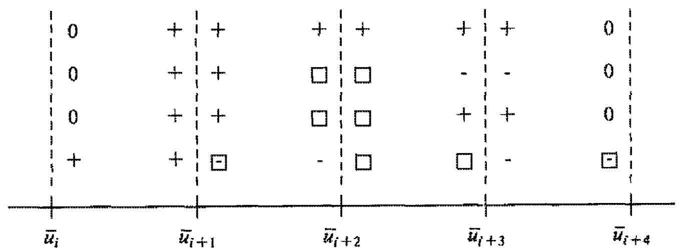


As was established in a similar case above, $s_{i,-2}(\bar{u})$ has precisely one zero to the right of $\bar{u}_{i+1}$, which may or may not occur in the interval $(\bar{u}_{i+1},\bar{u}_{i+2})$.

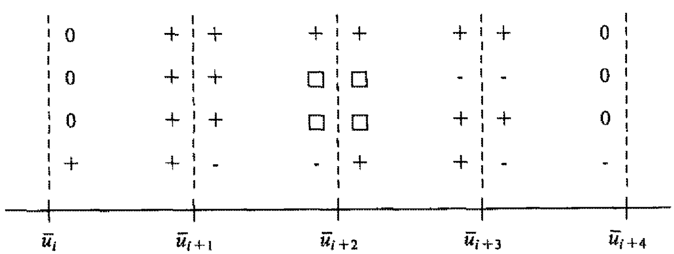
In the case of no zero in this interval, we have



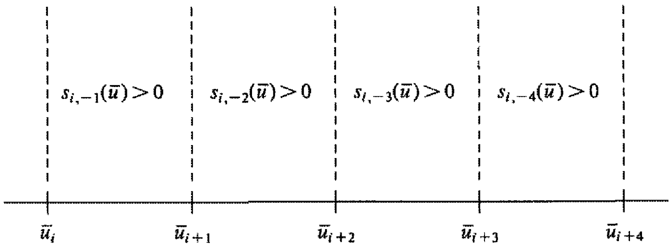
In this case, we begin by considering the first of the two possibilities for $s_{i,-4}(\bar{u})$, which yields



If a zero is to occur, then it can only take place in the interval $(\bar{u}_{i+2}, \bar{u}_{i+3})$. This requires that $s_{i,-3}^{(3)}(\bar{u}) > 0$, since from Taylor's Theorem we know that a sign pattern of $(+,-,+,-)$ at $\bar{u}_{i+3}$ would force $s_{i,-3}(\bar{u})$ to go monotonically to $+\infty$ as $\bar{u}$ goes goes to $-\infty$. Hence in this case $s_{i,-3}(\bar{u})$ must have the sign pattern $(+,\square,\square,+)$ at $\bar{u}_{i+2}$ and the pattern $(+,-,+,+)$ at $\bar{u}_{i+3}$. The Budin-Fourier theorem would already imply, on this information alone, that $s_{i,-3}(\bar{u})$ could have no more than one zero within the interval in question. In fact we know that it cannot have one zero since we have assumed that $s_{i,-3}(\bar{u}_{i+2}) > 0$; hence it must have no zeros on $(\bar{u}_{i+2}, \bar{u}_{i+3})$. The sign pattern we may infer from the discussion of this case is the following:

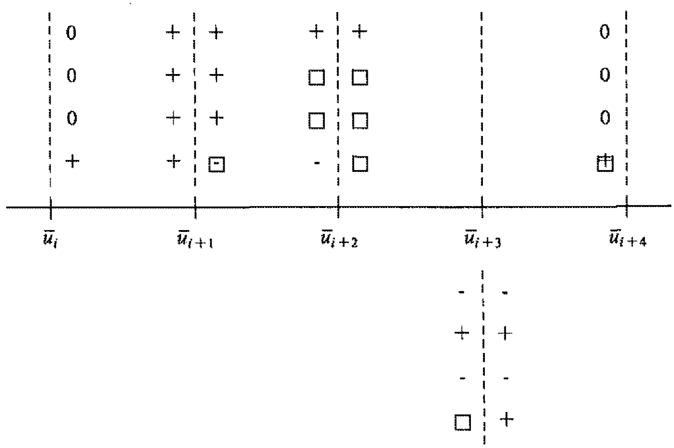


Since we have seen that no zeros exist on $[\bar{u}_i, \bar{u}_{i+4}]$ for this sign pattern, we have

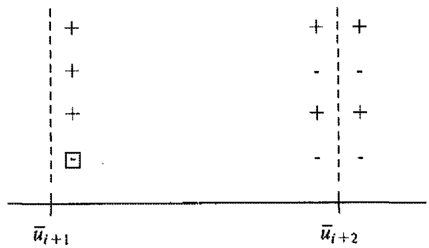


for this case, which is exactly the situation we desire.

Considering the second possibility for $s_{i,-4}(\bar{u})$ yields



We may remain compatible with the sign pattern at the right of $\bar{u}_{i+2}$ only if $s_{i,-3}(\bar{u})$ has a zero in the interval $(\bar{u}_{i+2}, \bar{u}_{i+3})$. The sign patterns are enough to ensure that only a single zero is possible, since $s_{i,-3}(\bar{u})$ must otherwise have three inflection points. When the Budin-Fourier Theorem is applied to $(\bar{u}_{i+2}, \bar{u}_{i+3})$ the sign pattern at the right of $\bar{u}_{i+2}$ already implies that the number of zeros in the interval could be no more than one, and this fixes the sign pattern at the right of $\bar{u}_{i+2}$ to be $(+,-,+,-)$. This gives us the following information about $(\bar{u}_{i+1}, \bar{u}_{i+2})$, the adjoining interval to the left:



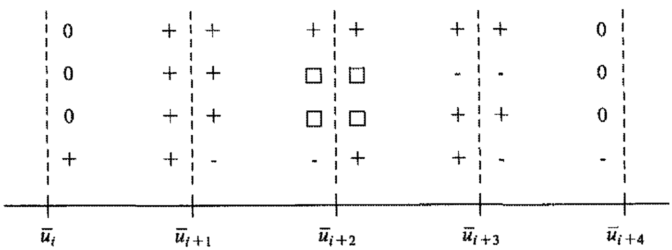
This shows that $s_{i,-2}^{(1)}(\bar{u})$ must have a zero in this interval. The Budin-Fourier theorem would give the number of zeros of $s_{i,-2}^{(1)}$, based upon this information alone, to be bounded by -2, which is impossible. The sign patterns which we have inferred for this interval are incompatible for a cubic.

All cases have now been covered. The final outcome is that only the case in which $s_{i,-4}(\bar{u}) > 0$ for $\bar{u}_{i+3} \le \bar{u} < \bar{u}_{i+4}$ is compatible with any of the possibilities from left to right. And in all of the compatible cases

$$G_i(\bar{u}) > 0$$

throughout the interval $(\bar{u}_i, \bar{u}_{i+4})$. This establishes the result we wanted.

Next we enquire a bit further into the sign structure of the compatible cases. We have seen that



in these cases. Consider the interval $(\bar{u}_{i+1}, \bar{u}_{i+2})$. Since the sign count at the left end is 2, the Budin-Fourier theorem would allow at most one sign change on the right. That is, only the following possibilities exist on the left side of $\bar{u}_{i+2}$

+		+		+
+	or	+	or	-
+		-		-
-		-		-

The first of these implies the top-to-bottom pattern $(+,+,+,+)$ on the right of $\bar{u}_{i+2}$, which is incompatible with the sign pattern at $\bar{u}_{i+3}$.

The second of these possibilities implies a pattern of $(+,+,\square,+)$ on the right of $\bar{u}_{i+2}$, and the sign-patterns in $(\bar{u}_{i+2},\bar{u}_{i+3})$ will be legal only if the undetermined sign is $-$.

The last possibility implies a pattern of $(+,-,-,+)$ on the right of $\bar{u}_{i+2}$, which is compatible with the pattern at $\bar{u}_{i+3}$.

This means that we are permitted either of the following situations at $\bar{u}_{i+2}$

$$\begin{array}{c|c} + & + \\ + & + \\ - & - \\ - & + \\ \hline \multicolumn{2}{c}{\bar{u}_{i+2}} \end{array}$$

or else

$$\begin{array}{c|c} + & + \\ - & - \\ - & + \\ - & + \\ \hline \multicolumn{2}{c}{\bar{u}_{i+2}} \end{array}$$

Finally we resolve the question posed at the beginning of this subsection, namely the question of whether $G_i(\bar{u})$ is non-negative or non-positive. An argument nearly identical to that used for $G_i(\bar{u})$ suffices to show that $\Delta_i^3 g_i(\bar{u})$ is also either entirely non-negative or entirely non-positive; since in fact we have $\Delta_i^3 g_i(\bar{u}_{i+4}) = 1$, $\Delta_i^3 g_i(\bar{u})$ must actually be non-negative. But since

- $\Delta_i^4 g_i(\bar{u}) = -[\Delta_i^3 g_{i+1}(\bar{u}) - \Delta_i^3 g_i(\bar{u})] = \Delta_i^3 g_i(\bar{u}) - \Delta_i^3 g_{i+1}(\bar{u})$ on $[\bar{u}_i,\bar{u}_{i+4})$, and
- $\Delta_i^3 g_{i+1}(\bar{u})$ is zero on $[\bar{u}_i,\bar{u}_{i+1})$, and
- $\Delta_i^3 g_i(\bar{u})$ is positive on $[\bar{u}_i,\bar{u}_{i+1})$,

it follows that $G_i(\bar{u})$ is positive on $[\bar{u}_i, \bar{u}_{i+1})$ Since $G_i(\bar{u})$ cannot change sign, it is therefore positive throughout $[\bar{u}_i, \bar{u}_{i+4})$.

4.9. A Closed Symbolic Representation

Analyzing the properties of the discretely-shaped Beta-splines, and the rendering of discretely-shaped Beta-spline curves, would be facilitated if we could obtain a compact symbolic representation of their basis functions. Unfortunately we have not, as yet, been able to do so for non-uniform knot sequences, and it seems likely that any such representation will prove to be quite complicated. However, we have been able to analyze two significant special cases: the quadratic Beta-splines over arbitrary knot sequences and the cubic Beta-splines over a uniform knot sequence.

5. Quadratic Discretely-Shaped Beta-Splines

The G^1 continuous quadratic Beta-splines are significantly simpler than the G^2 continuous cubic Beta-splines discussed above. For the quadratic Beta-splines the truncated power basis consists of

$$
\begin{aligned}
g_i(\bar{u}) &= (\bar{u} - \bar{u}_i)_+^2 + \sum_{j=i+1}^{m+3} a_{i,j} (\bar{u} - \bar{u}_j)_+^1 \\
&= (\bar{u} - \bar{u}_i)_+^2 + a_{i,i+1} (\bar{u} - \bar{u}_{i+1})_+^1 + \cdots + a_{i,m+3} (\bar{u} - \bar{u}_{m+3})_+^1
\end{aligned}
$$

and Algorithm I becomes

Algorithm II

for $i \leftarrow 0$ **step** 1 **until** $m+2$ **do**
    $Sa \leftarrow 0$
    for $j \leftarrow i+1$ **step** 1 **until** $m+3$ **do**
        $a_{i,j} \leftarrow (\beta_{1,j} - 1) \left[2(\bar{u}_j - \bar{u}_i)^2 + Sa \right]$
        $Sa \leftarrow Sa + a_{i,j}$
    endfor
endfor

Since we expect quadratic basis functions to be non-zero over three rather than four successive intervals, the differencing is arranged so as to cause the third difference to be zero for $\bar{u}_{i+3} \leq \bar{u} < \bar{u}_{i+4}$. If we fix i, the computational differencing proceeds as follows.

$$g_j(\bar{u}) = u^2 + A_{i,j}u + B_{i,j} \quad \text{for } j = i, i+1, i+2, i+3,$$

where

$$A_{i,i} = (a_{i,i+1} + a_{i,i+2} + a_{i,i+3}) - 2\bar{u}_i$$
$$B_{i,i} = -(a_{i,i+1}\bar{u}_{i+1} + a_{i,i+2}\bar{u}_{i+2} + a_{i,i+3}\bar{u}_{i+3}) - \bar{u}_i^2$$

$$A_{i,i+1} = (a_{i,i+2} + a_{i,i+3}) - 2\bar{u}_{i+1}$$
$$B_{i,i+1} = -(a_{i,i+2}\bar{u}_{i+2} + a_{i,i+3}\bar{u}_{i+3}) - \bar{u}_{i+1}^2$$

$$A_{i,i+2} = (a_{i,i+3}) - 2\bar{u}_{i+2}$$
$$B_{i,i+2} = -(a_{i,i+3}\bar{u}_{i+3}) - \bar{u}_{i+2}^2$$

$$A_{i,i+4} = -2\bar{u}_{i+3}$$
$$B_{i,i+4} = -\bar{u}_{i+3}^2 .$$

Just as for the cubic case, the first differences $\Delta_i^1 g_i(\bar{u})$, $\Delta_i^1 g_{i+1}(\bar{u})$ and $\Delta_i^1 g_{i+2}(\bar{u})$ are defined by

$$\begin{aligned} \Delta_i^1 g_j(\bar{u}) &= \frac{g_{j+1}(\bar{u}) - g_j(\bar{u})}{A_{i,j+1} - A_{i,j}} && \text{for all } \bar{u} \text{ and } j = i,\, i+1,\, i+2 \\ &= \bar{u} + \frac{B_{i,j+1} - B_{i,j}}{A_{i,j+1} - A_{i,j}} && \text{for } \bar{u}_{i+3} \le \bar{u} < \bar{u}_{i+4} \\ &= \bar{u} + C_{i,j} && \text{for } \bar{u}_{i+3} \le \bar{u} < \bar{u}_{i+4} . \end{aligned}$$

We can now cancel the linear term by forming the two functions $\Delta_i^2 g_i(\bar{u})$ and $\Delta_i^2 g_{i+1}(\bar{u})$ as

$$\begin{aligned} \Delta_i^2 g_j(\bar{u}) &= \frac{\Delta_i^1 g_{j+1}(\bar{u}) - \Delta_i^1 g_j(\bar{u})}{C_{i,j+1} - C_{i,j}} && \text{for all } \bar{u} \text{ and } j = i,\, i+1 \\ &= 1 && \text{for } \bar{u}_{i+3} \le \bar{u} < \bar{u}_{i+4} . \end{aligned}$$

Finally we compute the function

$$\begin{aligned} \Delta_i^3 g_i(\bar{u}) &= -[\Delta_i^2 g_{i+1}(\bar{u}) - \Delta_i^2 g_i(\bar{u})] \\ &= 0 && \text{for all } \bar{u}_{i+3} \le \bar{u} < \bar{u}_{i+4} , \end{aligned}$$

with which we replace $g_i(\bar{u})$; $\Delta_i^3 g_i(\bar{u})$ is our quadratic discretely-shaped Beta-spline.

With the aid of Vaxima [Bogen1977a, Fateman1982a] we have been successful in obtaining reasonable symbolic representations of these differences and of the quadratic Beta-splines. In particular, we have

$$\Delta_i^1 g_i(\bar{u}) = -\frac{g_{i+1}(\bar{u}) - g_i(\bar{u})}{2\beta_{1,i+1}\beta_{1,i+2}\beta_{1,i+3}(\bar{u}_{i+1}-\bar{u}_i)}$$

$$\Delta_i^1 g_{i+1}(\bar{u}) = -\frac{g_{i+2}(\bar{u}) - g_{i+1}(\bar{u})}{2\beta_{1,i+2}\beta_{1,i+3}(\bar{u}_{i+2}-\bar{u}_{i+1})}$$

$$\Delta_i^1 g_{i+2}(\bar{u}) = -\frac{g_{i+3}(\bar{u}) - g_{i+2}(\bar{u})}{2\beta_{1,i+3}(\bar{u}_{i+3}-\bar{u}_{i+2})}$$

$$\Delta_i^2 g_i(\bar{u}) = -\frac{2\beta_{1,i+1}\beta_{1,i+2}\beta_{1,i+3}[\Delta_i^1 g_{i+1}(\bar{u}) - \Delta_i^1 g_i(\bar{u})]}{\beta_{1,i+1}\bar{u}_{i+2} + (1-\beta_{1,i+1})\bar{u}_{i+1} - \bar{u}_i}$$

$$\Delta_i^2 g_{i+1}(\bar{u}) = -\frac{2\beta_{1,i+2}\beta_{1,i+3}[\Delta_i^1 g_{i+2}(\bar{u}) - \Delta_i^1 g_{i+1}(\bar{u})]}{\beta_{1,i+2}\bar{u}_{i+3} + (1-\beta_{1,i+2})\bar{u}_{i+2} - \bar{u}_{i+1}}$$

$$\Delta_i^3 g_i(\bar{u}) = -[\Delta_i^2 g_{i+1}(\bar{u}) - \Delta_i^2 g_i(\bar{u})] \ .$$

Explicit formulae for the three basis segments comprising the non-zero portion of the i^{th} basis function $\Delta_i^3 g_i(\bar{u})$ are as follows (see Figure 5).

$$s_{i,-1} = \frac{(\bar{u}-\bar{u}_i)^2}{(\bar{u}_1-\bar{u}_0)\,[\beta_{1,i+1}\bar{u}_{i+2} + (1-\beta_{1,i+1})\bar{u}_{i+1} - \bar{u}_i]}$$

$$\begin{aligned} s_{i,-2} = {} & \frac{2(\beta_{1,i+1}-1)(\bar{u}_{i+1}-\bar{u}_i)(\bar{u}-\bar{u}_{i+1}) - (\bar{u}-\bar{u}_{i+1})^2 + (\bar{u}-\bar{u}_i)^2}{(\bar{u}_{i+1}-\bar{u}_i)\,[\beta_{1,i+1}\bar{u}_{i+2} + (1-\beta_{1,i+1})\bar{u}_{i+1} - \bar{u}_i]} \\ & - \frac{\beta_{1,i+1}(\bar{u}-\bar{u}_{i+1})^2}{(\bar{u}_{i+2}-\bar{u}_{i+1})\,[\beta_{1,i+1}\bar{u}_{i+2} - (\beta_{1,i+1}-1)\bar{u}_{i+1} - \bar{u}_i]} \\ & - \frac{(\bar{u}-\bar{u}_{i+1})^2}{(\bar{u}_{i+2}-\bar{u}_{i+1})\,[\beta_{1,i+2}\bar{u}_{i+3} - (\beta_{1,i+2}-1)\bar{u}_{i+2} - \bar{u}_{i+1}]} \end{aligned}$$

$$s_{i,-3} = \frac{\beta_{1,i+2}(\bar{u}-\bar{u}_{i+3})^2}{\beta_{1,i+1}(\bar{u}_{i+3}-\bar{u}_{i+2})(\bar{u}_{i+3}-\bar{u}_{i+2})(\bar{u}_{i+2}-\bar{u}_{i+1})} \ .$$

By definition these functions have compact support – $\Delta_i^3 g_i(\bar{u})$ is non-zero only on $(\bar{u}_i, \bar{u}_{i+3})$. Using Vaxima, one can verify by direct evaluation of the three basis functions which are non-zero on a given interval that they sum to one.

It is interesting to see how easily we can show directly that they are non-negative for $\beta_1 > 0$. Consider $s_{i,-1}(\bar{u}_i)$. First we note that $s_{i,-1}(\bar{u}_i)$ is 0. Moreover, $(\bar{u} - \bar{u}_i)^2$ is positive so long as $\bar{u} > \bar{u}_i$ and $(\bar{u}_1 - \bar{u}_0)$ is positive by assumption. Let us rewrite the remaining term $[\beta_{1,i+1}\bar{u}_{i+2} + (1-\beta_{1,i+1})\bar{u}_{i+1} - \bar{u}_i]$ as $\beta_{1,i+1}(\bar{u}_{i+2} - \bar{u}_{i+1}) + (\bar{u}_{i+1} - \bar{u}_i)$. Clearly this expression is linear in $\beta_{1,i+1}$, has positive slope, and has a zero at $-(\bar{u}_{i+1}-\bar{u}_i)/(\bar{u}_{i+2}-\bar{u}_{i+1}) < 0$. Hence this expression will be positive for all $\beta_{1,i+1} \geq 0$. Consequently $s_{i,-1}(\bar{u})$ will be positive for $\bar{u}_i < \bar{u} < \bar{u}_{i+1}$.

It is similarly easy to see that $s_{i,-3}(\bar{u})$ is positive so long as $\bar{u}_{i+2} \leq \bar{u} < \bar{u}_{i+3}$ and $\beta_{1,i+2} > 0$, which suffices.

The complexity of $s_{i,i-2}(\bar{u})$ makes a direct analysis painful. However, it is easy to see that the first derivative of $s_{i,i-1}(\bar{u})$ at $\bar{u}_{i+1}$ is positive, and that the first derivative of $s_{i,-3}(\bar{u})$ at $\bar{u}_{i+2}$ is negative. Since $s_{i,-2}(\bar{u})$ is a quadratic and positive at both $\bar{u}_{i+1}$ and $\bar{u}_{i+2}$, it cannot then be negative for $\bar{u}_{i+2} < \bar{u} < \bar{u}_{i+3}$.

Thus the quadratic Beta-splines are non-negative. Because we already know that they sum to one, we can conclude that the i^{th} curve segment of a quadratic Beta-spline curve lies within the convex hull of $\mathbf{V}_{i-3}$, $\mathbf{V}_{i-2}$ and $\mathbf{V}_{i-1}$ since these are the control vertices which are weighted by non-zero basis functions on $[\bar{u}_{i-1}, \bar{u}_i)$.

6. Uniform Cubic Discretely-Shaped Beta-Splines

Now let us return to the cubic Beta-splines, and suppose that the underlying knot sequence is uniform, so that $\bar{u}_i - \bar{u}_{i-1} = 1$ for all i. If, for an arbitrary set of positive $\beta_{1,i}$'s and $\beta_{2,i}$'s, we compute the four segments which are non-zero on the interval $[\bar{u}_{i+3}, \bar{u}_{i+4})$ (see Figure 6) and sum them using Vaxima, we find that they sum to one. As for the quadratic Beta-splines, it is interesting to see how we can show directly that they are non-negative. Two of the basis segments are trivial. We find that

$$s_{i+3,-1}(\bar{u}) = \frac{1}{\delta 1}\left[\, 2(\beta_{2,i+5} + 2\beta_{1,i+5}^2 + 2\beta_{1,i+5})u^3 \,\right]$$

where

$$\begin{aligned}\delta 1 = {} & \beta_{2,i+4}\beta_{2,i+5} + 2\beta_{1,i+4}^2\beta_{2,i+5} + 4\beta_{1,i+4}\beta_{2,i+5} + 2\beta_{2,i+5} \\ & + 2\beta_{1,i+5}^3\beta_{2,i+4} + 4\beta_{1,i+5}^2\beta_{2,i+4} + 2\beta_{1,i+5}\beta_{2,i+4} + 4\beta_{1,i+4}^2\beta_{1,i+5}^3 \\ & + 4\beta_{1,i+4}\beta_{1,i+5}^3 + 8\beta_{1,i+4}^2\beta_{1,i+5}^2 + 12\beta_{1,i+4}\beta_{1,i+5}^2 + 4\beta_{1,i+5}^2 \\ & + 4\beta_{1,i+4}^2\beta_{1,i+5} + 8\beta_{1,i+4}\beta_{1,i+5} + 4\beta_{1,i+5}\end{aligned}$$

and

$$s_{i+3,-4}(\overline{u}) = \frac{1}{\delta 4}\left[\, 2\beta_{1,i+3}^3(\beta_{2,i+2} + 2\beta_{1,i+2}^2 + 2\beta_{1,i+2})(1-u)^3 \,\right]$$

where

$$\begin{aligned}\delta 4 = \; & \beta_{2,i+2}\beta_{2,i+3} + 2\beta_{1,i+2}^2\beta_{2,i+3} + 4\beta_{1,i+2}\beta_{2,i+3} + 2\beta_{2,i+3} \\ & + 2\beta_{1,i+3}^3\beta_{2,i+2} + 4\beta_{1,i+3}^2\beta_{2,i+2} + 2\beta_{1,i+3}\beta_{2,i+2} + 4\beta_{1,i+2}^2\beta_{1,i+3}^3 \\ & + 4\beta_{1,i+2}\beta_{1,i+3}^3 + 8\beta_{1,i+2}^2\beta_{1,i+3}^2 + 12\beta_{1,i+2}\beta_{1,i+3}^2 + 4\beta_{1,i+3}^2 \\ & + 4\beta_{1,i+2}^2\beta_{1,i+3} + 8\beta_{1,i+2}\beta_{1,i+3} + 4\beta_{1,i+3} \;.\end{aligned}$$

It is easy to see that these two basis segments will be positive since all of the β's are positive and $0 \leq u < 1$. The denominators $\delta 1$ and $\delta 4$ can, of course, be factored further, but we have left them in this form for simplicity.

The remaining two segments require more effort. The segment $s_{i+3,-2}(\overline{u})$ may be written as

$$s_{i+3,-2}(u) = \left(c_0 + c_1 u + c_2 u^2 \right) - \left(c_3 u^3 \right) . \tag{26}$$

The following argument establishes that $s_{i+3,-2}(u)$ is positive on (0,1):

- c_1, c_2, c_3, $s_{i+3,-2}(0) = c_0$ and $s_{i+3,-2}(1) = c_0 + c_1 + c_2 - c_3 \equiv d_3$ are all sums of products of positive values, like $\delta 1$ and $\delta 4$, and are therefore themselves positive;
- hence we may represent c_3 as $c_0 + c_1 + c_2 - d_3$;
- since $0 < u < 1$, we have $1 > u > u^2 > u^3$;
- therefore

$$\begin{array}{lcl} c_0 & > & c_0 u^3 \\ c_1 u & > & c_1 u^3 \\ c_2 u^2 & > & c_2 u^3 \\ 0 & > & -d_3 u^3 \;; \end{array}$$

- therefore

$$c_0 + c_1 u + c_2 u^2 + 0 > \left(c_0 + c_1 + c_2 - d_3 \right) u^3 = c_3 u^3 \;;$$

- therefore (26) is positive on (0,1), as desired.

An exactly analogous argument suffices for the right middle segment if it is written in the form

$$s_{i+3,-3}(u) = \left(c_0 + c_1(1-u) + c_2(1-u)^2 \right) - \left(c_3(1-u)^3 \right) . \tag{26}$$

A variety of important properties follow from the fact that the uniform cubic

Beta-splines are non-negative and sum to one:

- the i^{th} segment $\mathbf{Q}_i(\bar{u})$ lies within the convex hull of $\mathbf{V}_{i-4}$, $\mathbf{V}_{i-3}$, $\mathbf{V}_{i-2}$ and $\mathbf{V}_{i-1}$;
- if $\mathbf{V}_{i-4} = \mathbf{V}_{i-3} = \mathbf{V}_{i-2}$ then this point will be interpolated, and the curve segment defined by these three points and $\mathbf{V}_{i-1}$ will be a straight line;
- if $\mathbf{V}_{i-4} = \mathbf{V}_{i-3}$ then the first point on the curve segment defined by these two points, together with $\mathbf{V}_{i-2}$ and $\mathbf{V}_{i-1}$ must lie on the line segment joining $\mathbf{V}_{i-3}$ and $\mathbf{V}_{i-2}$ and the curvature there will be zero.

(If one assumes that the $G_i(\bar{u})$ do not need further scaling in order to sum to one, or if one computes the scale factors which produce a partition of unity, then these results apply also to discretely-shaped Beta-splines over a non-uniform knot sequence.)

It is possible to verify, with the aid of Vaxima, that as $\beta_{2,i+2}$ is made arbitrarily large $\mathbf{Q}(\bar{u}_{i+2})$ converges to $\mathbf{V}_i$. This behaviour, which the uniformly-shaped and continuously-shaped Beta-splines display as well, naturally associates the joint at $\bar{u}_{i+2}$ with the control vertex $\mathbf{V}_i$, and so we sometimes speak loosely of the "β_2 value associated with $\mathbf{V}_i$" when referring to $\beta_{2,i+2}$ (and similarly for β_1).

If $\beta_{1,i} = \beta_1$ and $\beta_{2,i} = \beta_2$ for all i, we then obtain the uniformly-shaped Beta-spline for β_1 and β_2.

In many applications the ability to manipulate β_2 may be sufficient, and we therefore list the basis segments on the interval $[\bar{u}_{i+3}, \bar{u}_{i+4})$ for the special case in which the knots are spaced one unit apart and the β_1 values all have the value one:

$$s_{i+3,-1}(u) = \frac{2(\beta_{2,i+5} + 4)u^3}{\delta 1}$$

$$\begin{aligned} s_{i+2,-2}(u) = & -\frac{2(\beta_{2,i+4} + 4)}{\delta 1\, \delta 2} \Big[\beta_{2,i+3}\beta_{2,i+4}\beta_{2,i+5} + 8\beta_{2,i+3}\beta_{2,i+5} \\ & + 8\beta_{2,i+3}\beta_{2,i+4} + 3\beta_{2,i+4}\beta_{2,i+5} + 44\beta_{2,i+3} \\ & + 24\beta_{2,i+4} + 28\beta_{2,i+5} + 144 \Big] u^3 \\ & + \frac{(\beta_{2,i+4} + 4)}{\delta 2} \Big[3(\beta_{2,i+3} + 2)u^2 + 6u + 2 \Big] \end{aligned}$$

$$s_{i+1,-3}(u) = -\frac{2(\beta_{2,i+3} + 4)}{\delta 2\, \delta 3} \Big[\beta_{2,i+2}\beta_{2,i+3}\beta_{2,i+4} + 8\beta_{2,i+2}\beta_{2,i+4}$$

$$+ 3\beta_{2,i+2}\beta_{2,i+3} + 8\beta_{2,i+3}\beta_{2,i+4} + 28\beta_{2,i+2}$$

$$+ 24\beta_{2,i+3} + 44\beta_{2,i+4} + 144 \Big] (1-u)^3$$

$$+ \frac{(\beta_{2,i+3}+4)}{\delta 3} \Big[3(\beta_{2,i+4}+2)(1-u)^2 + 6(1-u) + 2 \Big]$$

$$s_{i,-4}(u) = \frac{2(\beta_{2,i+2}+4)(1-u)^3}{\delta 3}$$

where

$$\delta 1 = (\beta_{2,i+4}\beta_{2,i+5} + 8\beta_{2,i+4} + 8\beta_{2,i+5} + 48)$$

$$\delta 2 = (\beta_{2,i+3}\beta_{2,i+4} + 8\beta_{2,i+3} + 8\beta_{2,i+4} + 48)$$

$$\delta 3 = (\beta_{2,i+2}\beta_{2,i+3} + 8\beta_{2,i+2} + 8\beta_{2,i+3} + 48) \ .$$

By inspection it is clear that so long as $\beta_{1,i}, \beta_{2,i} \geq 0$ the above representation for discretely-shaped Beta-splines over a uniform knot sequence are necessarily well-defined – the denominators cannot vanish, even though the differencing representation of the discretely-shaped Beta-splines admits of this possibility.

7. Examples

Figures 7-16 illustrate various of the properties discussed above. It is illuminating to see how changes in the basis functions shown in Figures 8, 10 and 14 produce the curves shown in Figures 7, 9, 11 and 13. One sees also how the curves lie within the convex hull of their corresponding control vertices; Figures 17 and 18 illustrate the failure of a curve to lie within the convex hull of its control points when a β_2 value is negative.

Figure 19 demonstrates the similar, though not identical, tension-like effects produced by manipulating β_1 and β_2. Figure 20 is produced by varying several shape parameters simultaneously.

Figures 21 and 22 illustrate the locality provided by the discretely-shaped Beta-splines.

8. Conclusions

The discretely-shaped Beta-splines which we have defined offer an attractive alternative to the continuously-shaped Beta-splines defined in [Barsky1983b]. Since they rely on the controlled introduction of discontinuities of the first and second parametric derivatives at joints, we may think of them as providing an intuitive means of utilizing a portion of the B-splines over knot sequences containing multiple vertices, much as the continuously-shaped Beta-splines provide an intuitive means of utilizing certain relatively high degree polynomials.

For uniformly spaced knots we have seen that the gross behaviour of discretely-shaped and continuously shaped Beta-spline curves is similar as the shape parameters are varied, and as multiple vertices are introduced. Since the discretely-shaped Beta-splines are composed of purely cubic polynomial segments, however, they can be efficiently rendered by forward differencing.

We have not yet obtained a closed-form symbolic representation for the discretely-shaped Beta-splines over non-uniform knot sequences, although we are still attempting to do so. Other questions which we are exploring include:

- a proof that the differencing construction we have presented directly yields basis functions which partition unity, without the need for further scaling;
- a characterization of the values of the shape parameters for which the differencing fails because of a zero division (for non-uniform knot sequences), and a suitable extension of our definition for $G_i(\bar{u})$ to cover these cases;
- a representation of the $G_i(\bar{u})$ in terms of B-splines, and consequently a general subdivision scheme for the Beta-splines;
- an optimal factorization of the closed form representation of the uniform cubic discretely-shaped Beta-splines;
- a comparison of the cost of rendering curves defined using B-splines, uniformly-shaped Beta-splines and discretely-shaped Beta-splines.

Goodman has independently investigated the existence and properties of local basis functions for a variety of spline which contains the G^2 cubic splines as a special case [Goodman1983a]. Without providing computational descriptions, his results show that there exist basis functions with support on four successive intervals, that these basis functions are unique up to a scale factor, have no zeros, and sum to one. He shows that his development can generate the uniformly-shaped cubic Beta-splines (over a uniform knot sequence). In an even more abstract setting, chapter 11 of [Schumaker1981a] provides a general context for analyzing classes of piecewise functions which includes the Beta-splines as a very special case. His results can be used to provide the dimensionality of the space of Beta-splines, as well as conditions under which we may expect a local basis to exist.

9. Pictures

Pictures

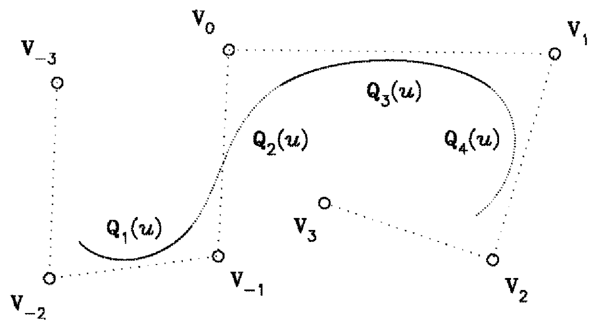


Figure 1. An example of a curve defined by a sequence of *control vertices*, represented here by small circles, near which the curve passes. The lightly dotted line connecting the control vertices forms the *control polygon*, and indicates the order in which the control vertices are to be approximated. The solid and heavily dotted curves represent distinct curve *segments*. Each is a single parametric cubic. The point at which two successive segments meet is called a *joint*. The value of the parameter $\bar{u}$ which corresponds to a joint is called a *knot*. This particular curve was constructed from uniform cubic B-splines.

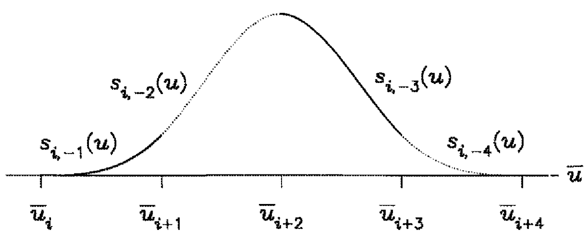


Figure 2. The uniform cubic B-spline $B_i(\bar{u})$ centered at $\bar{u}_{i+2}$. It is zero for $\bar{u} \leq \bar{u}_i$ and for $\bar{u} \geq \bar{u}_{i+4}$. The nonzero portion of $B_i(\bar{u})$ is comprised of the four cubic polynomial segments $s_{i,-1}(\bar{u})$, $s_{i,-2}(\bar{u})$, $s_{i,-3}(\bar{u})$ and $s_{i,-4}(\bar{u})$.

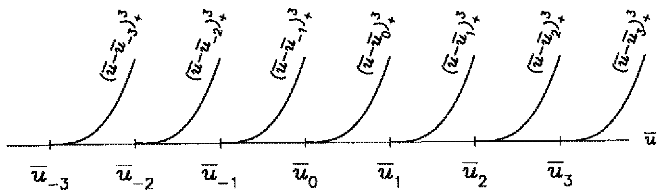


Figure 3. The "one-sided cubic spline basis functions." All of these functions grow unboundedly large as we move from left to right. Since the x or y coordinate of a curve will lie in some modestly bounded range, it will be necessary to cancel large positive values of some of these functions by using large negative values of others, a likely source of numerical error (if not of overflow).

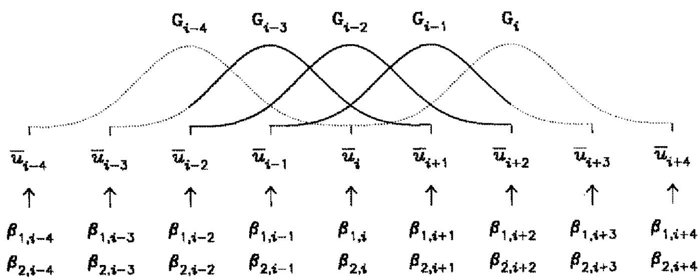


Figure 4. The basis segments which are effected by the change in β_1 or β_2 at $\bar{u}_i$.

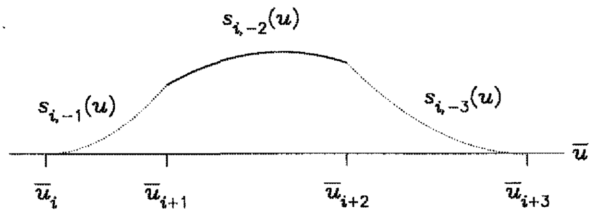


Figure 5. A quadratic Beta-spline. Notice that it consists of only three nonzero segments.

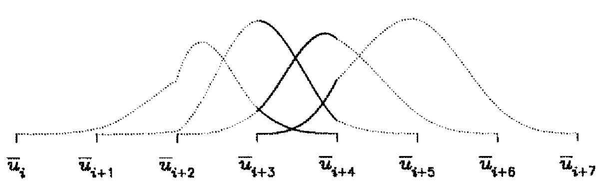


Figure 6. The four basis functions $G_i(\bar{u})$, $G_{i+1}(\bar{u})$, $G_{i+2}(\bar{u})$ and $G_{i+3}(\bar{u})$ which are nonzero on the particular segment $[\bar{u}_{i+3}, \bar{u}_{i+4})$.

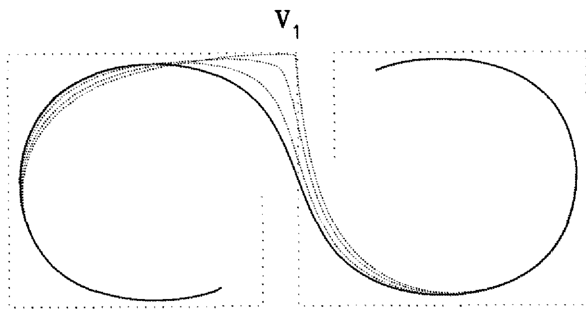


Figure 7. The solid line is a uniform cubic B-spline curve (β_1 and β_2 have the values 1 and 0 at every joint). The dotted curves result when the value of β_2 at the joint nearest V_1 is set to 2, 10 and 100, respectively. Increasing values of β_2 draw the joint in question towards V_1. For clarity the control polygon is shown, but not the control vertices.

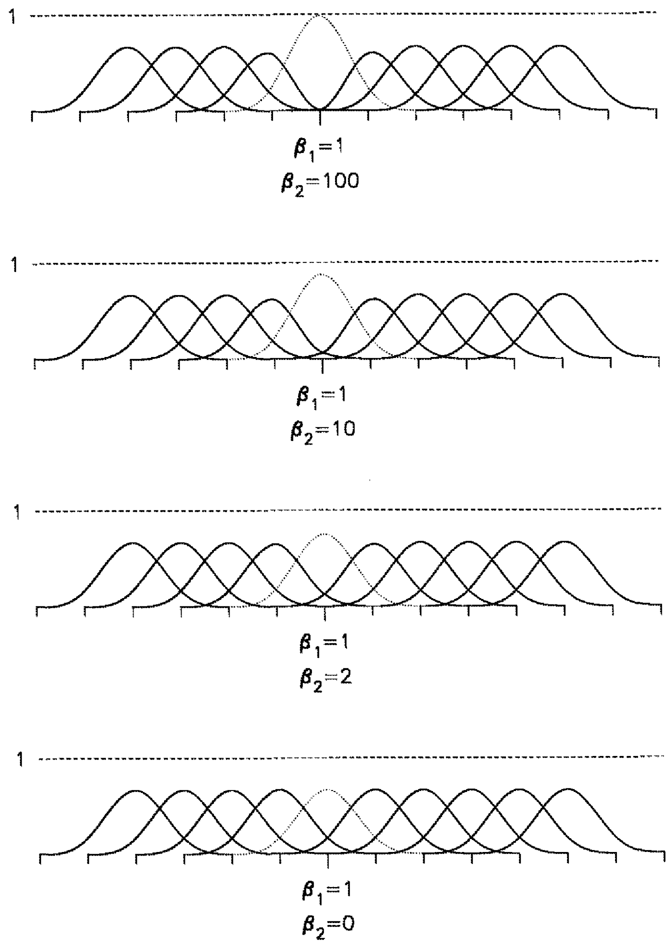


Figure 8. The basis functions corresponding to the four curves of Figure 7. β_1 and β_2 have the values 1 and 0 at all knots except the one explicitly labeled.

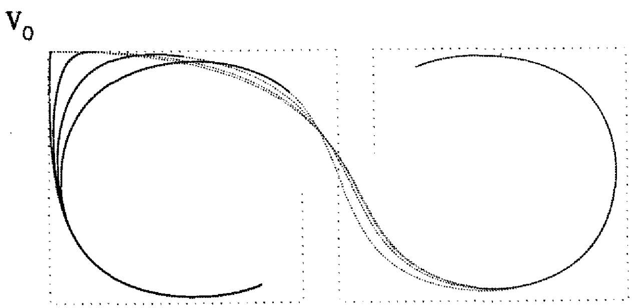


Figure 9. Starting from the same uniform cubic B-spline curve as appears in Figure 7, we successively increase β_1 at the joint between the solid and dotted portions of the curve, so that it has the values 1, 4, 16 and 256. As β_1 is increased the joint is pulled towards $\mathbf{V}_0$.

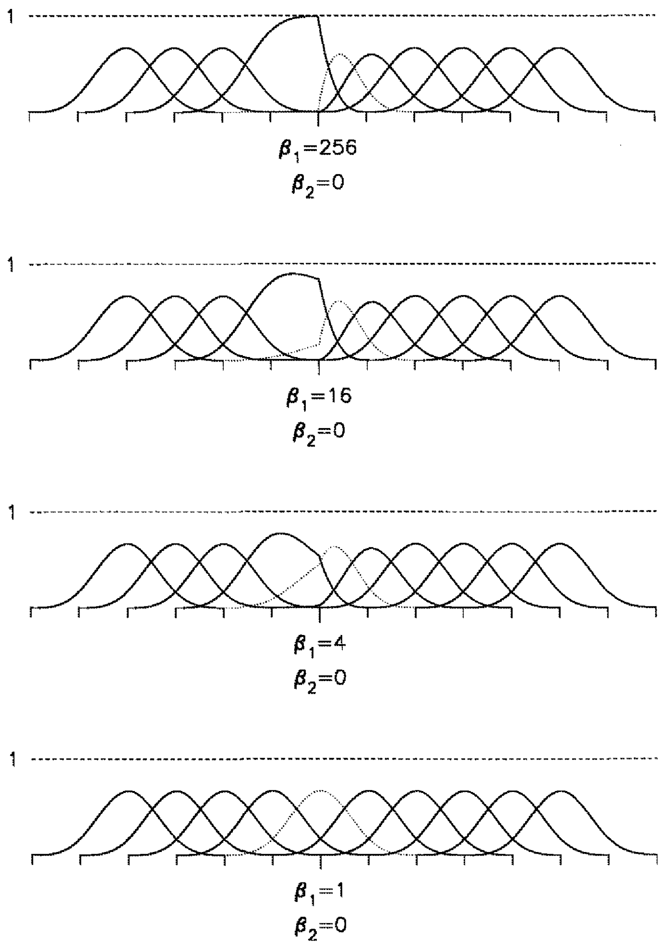


Figure 10. The basis functions corresponding to the four curves of Figure 9. β_1 and β_2 have the values 1 and 0 at all knots except the one explicitly labeled.

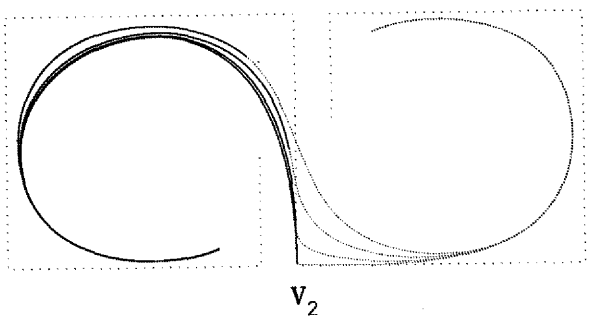


Figure 11. Symmetric behaviour occurs if we set β_1 to the values 1, 1/4, 1/16 and 1/256, respectively, with $\beta_2 = 0$. This time the joint is pulled towards V_2.

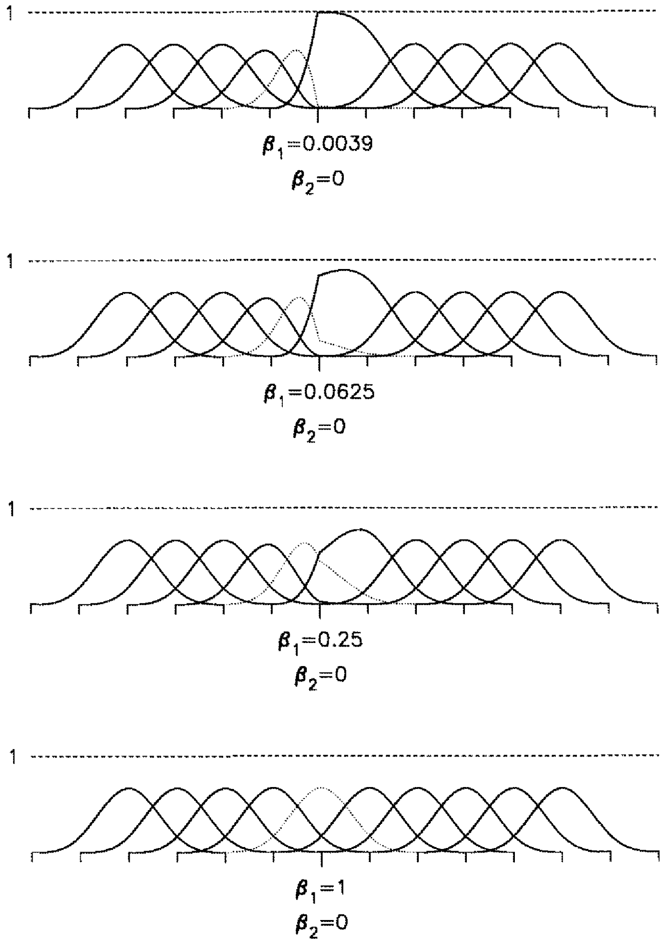


Figure 12. The basis functions corresponding to the four curves of Figure 11. β_1 and β_2 have the values 1 and 0 at all knots except the one explicitly labeled.

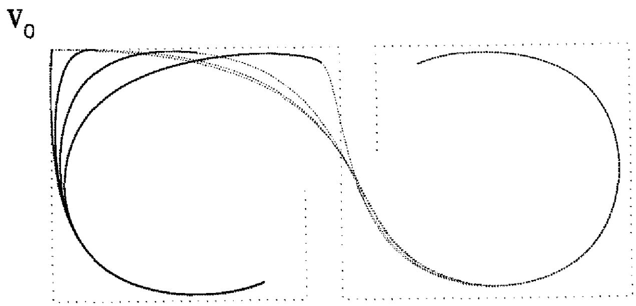


Figure 13. The β_1 values here are the same as in figure 9 except that the value of β_2 at the joint in question is 10 in each case instead of 0. Again the joint is pulled towards V_0. Recall that increasing β_2 at that joint has the effect of pulling the curve towards V_1.

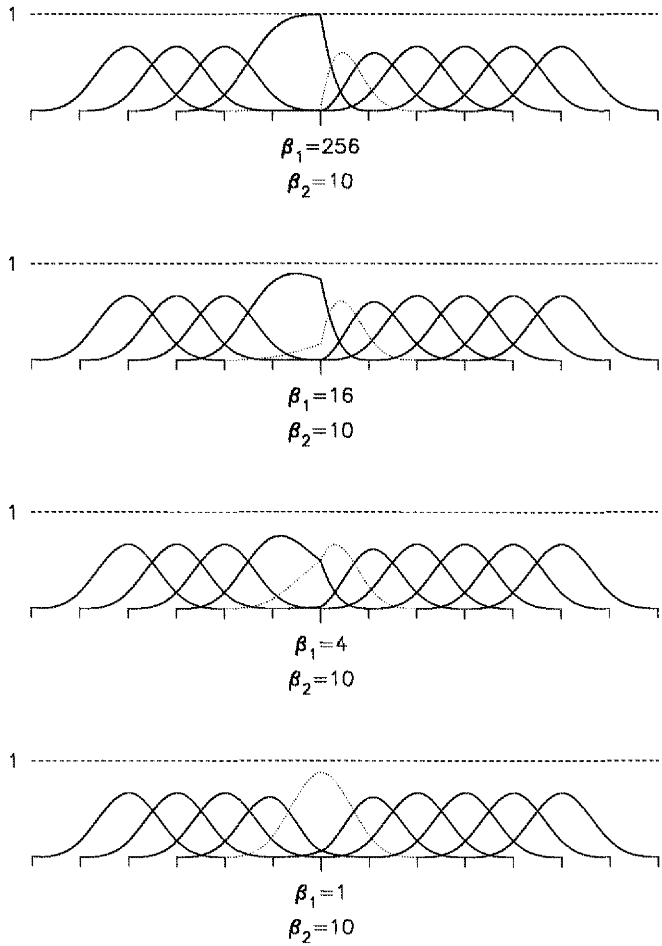


Figure 14. The basis functions corresponding to the four curves of Figure 13. β_1 and β_2 have the values 1 and 0 at all knots except the one explicitly labeled.

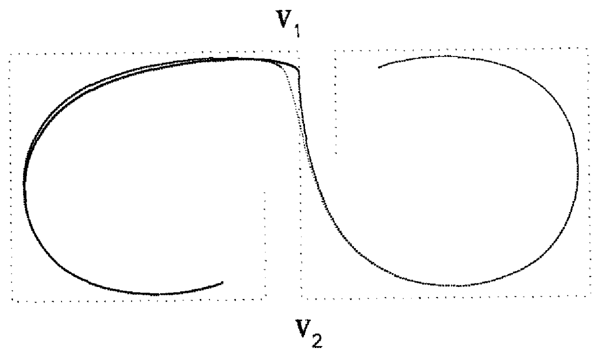


Figure 15. The β_1 values here are the same as in figure 11 except that the value of β_2 at the joint in question is 10 in each case instead of 0. Note that in this case the joint does not converge to V_2. Tensing the curve toward V_1 by setting a high value on β_2 at the joint has inhibited the convergence.

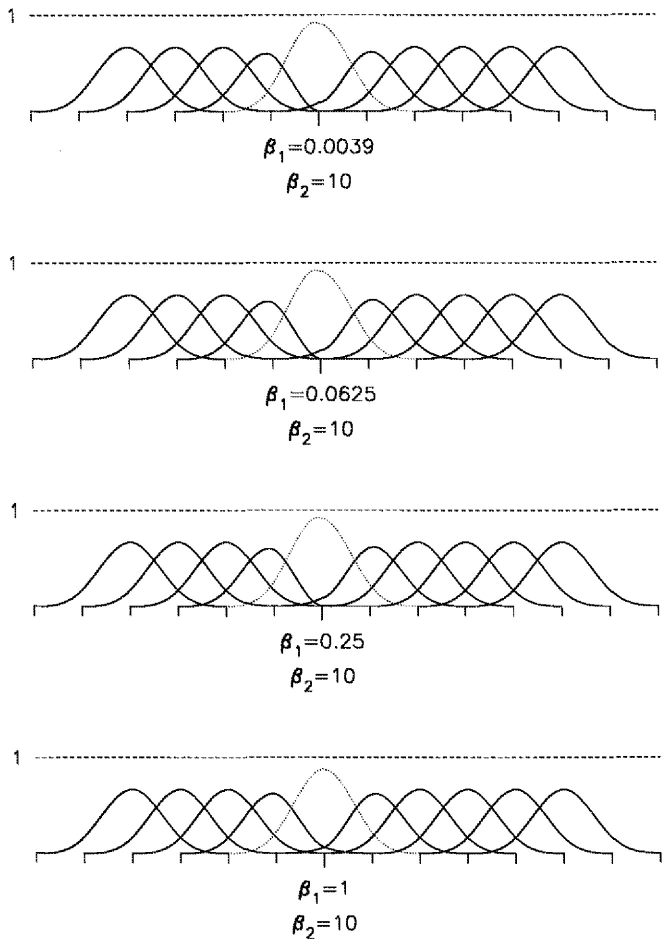


Figure 16. The basis functions corresponding to the four curves of Figure 15. β_1 and β_2 have the values 1 and 0 at all knots except the one explicitly labeled.

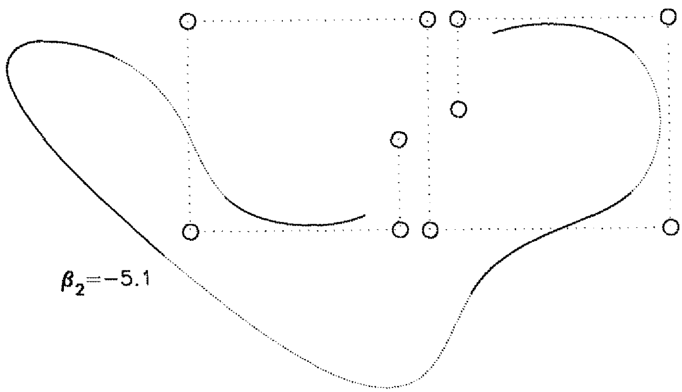


Figure 17. For negative values of β_2 the curve may pass outside the convex hull. β_1 has the value 1 and β_2 the value 0 at every joint except the one explicitly indicated.

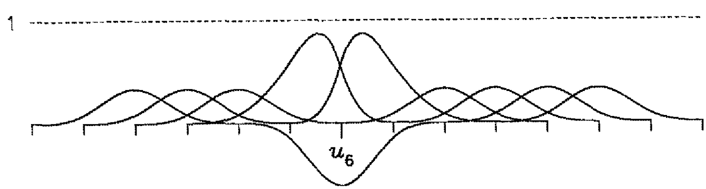


Figure 18. These are the (unscaled) Beta-splines with which the curve of Figure 17 is defined. Notice the negative basis function centered over the knot at which $\beta_2 = -5.1$. This is not a violation of the convex hull property established in the text, which holds only for positive values of β_1 and β_2.

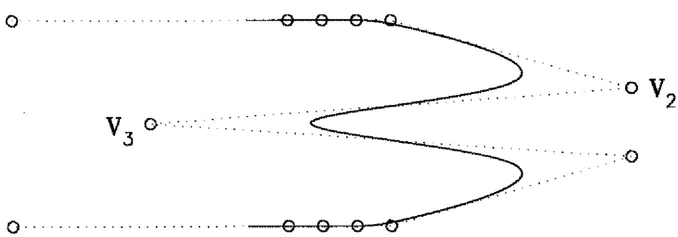


Figure 19a. A uniform discretely-shaped Beta-spline curve. Actually this is a C^2 spline curve since β_1 and β_2 have the values 1 and 0 throughout the curve, which should be compared with the curves below.

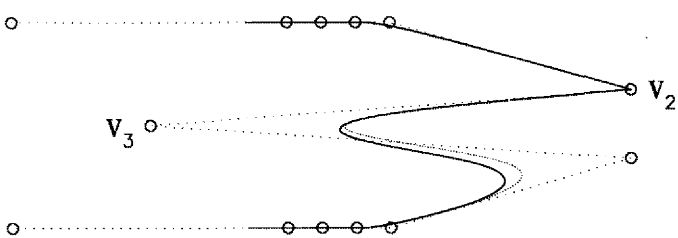


Figure 19b. The solid curve here is obtained from the curve of Figure 19a by increasing β_1 at V_3 from 1 to 10,000. The dotted curve is obtained from Figure 19a by instead increasing β_2 at V_2 from 0 to 10,000. In both cases a further increase in the shape parameter produces no observable change in the figure.

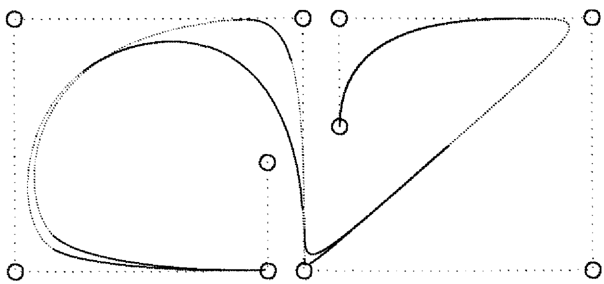


Figure 22. In this case we have changed the knot spacing for the third segment.

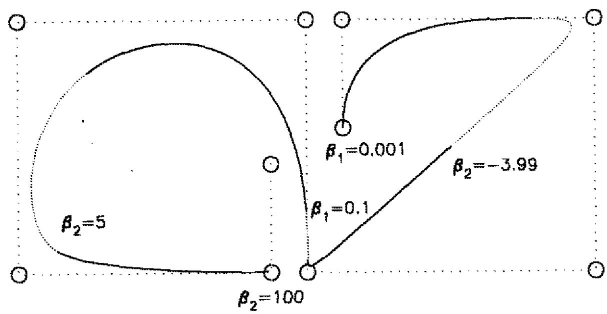


Figure 20. Some compound variations in the shape parameters. β_1 and β_2 have the values 1 and 0 except as shown.

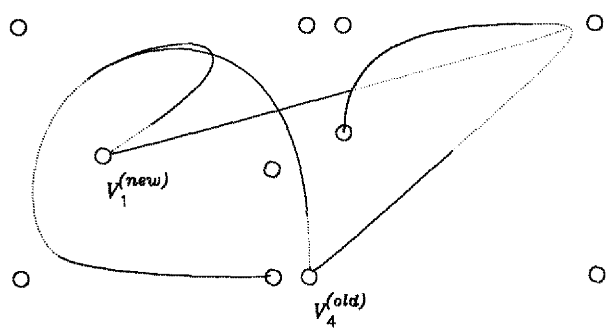


Figure 21. Here we see the effect produced by moving one of the vertices defining the curve of Figure 20. Notice that only four curve segments are altered. (The control polygon has been omitted here to enhance visibility of the curves.)

10. References

[Barsky1981a] 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.

[Barsky1982a] Brian A. Barsky and John C. Beatty (1982), Varying the Betas in Beta-Splines, CR-82-49, Department of Computer Science, University of Waterloo, Waterloo, Ontario, 42 pages [also available as TR CSD-82-112 from the Computer Science Division of the University of California at Berkeley].

[Barsky1983a] Brian A. Barsky (1983), The Beta-spline: A Curve and Surface Representation for Computer Graphics and Computer Aided Geometric Design, Computer Science Division, University of California, Berkeley, California 94720 [submitted for publication].

[Barsky1983b] Brian A. Barsky and John C. Beatty (1983), Local Control of Bias and Tension in Beta-splines, *Transactions on Graphics* **2**(2), April, 27-52 [also available in the proceedings of the 1983 Siggraph Conference, which are the July 1983 issue of the Computer Graphics Quarterly, 17:3 pp 193-218].

[Bartels1983a] Richard H. Bartels (1983), Splines in Interactive Computer Graphics, *Proceedings of the 1983 Dundee Conference on Numerical Analysis*, 26 pages [invited paper].

[Bartels1983b] Richard H. Bartels, John C. Beatty, and Brian A. Barsky (1983), An Introduction to the Use of Splines in Computer Graphics, CS-83-09, Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1, 182 pages.

[Bogen1977a] Richard Bogen, Jeffrey Golden, Michael Genesereth, and Alexander Doohovskoy (1977), *MACSYMA Reference Manual*, Massachusetts Institute of Technology [version nine].

[Boor1978a] Carl de Boor (1978), *A Practical Guide to Splines*, Applied Mathematical Sciences Volume 27, Springer-Verlag.

[Fateman1982a] Richard J. Fateman (1982), Addendum to the MACSYMA Reference Manual for the VAX, University of California, Berkeley.

[Goodman1983a] T. N. T. Goodman (1983), Properties of Beta-Splines, Department of Mathematical Sciences, University of Dundee, Dundee DD1 4HN, Scotland, 22.

[Schumaker1981a] Larry L. Schumaker (1981), *Spline Functions: Basic Theory*, John Wiley & Sons.