B. Dimsdale

Convex Cubic Splines

Abstract: Control curves used in industry are usually planar cubic splines, continuous and single-valued for a specific independent variable. If large slopes are specified at certain points, the spline coefficients, which are computed in order to preserve second derivative continuity of the spline, invariably lead to wildly oscillatory curves. Normally the large slopes occur at the ends of the curve, and this paper deals with such cases, developing methods which have been used successfully to combat the problems encountered. The end point where the difficulty occurs is ignored for the moment and a curve (spline or not) is designed for the remaining points. Then, with the point and slope adjacent to the difficult point known, the design is completed with one or more spline segments so that the resulting curve, which connects the difficult point to the next point, is convex (or concave), and has at least first derivative continuity. For the large finite slope case, two methods are described. The first method constructs the desired curve as a sequence of spline segments with finite slope everywhere, whereas the second expands the interval so that a single spline segment with infinite end slope passes through the difficult point with the required slope. The case of infinite end slope is also treated. In this and the preceding cases, second derivative continuity can be preserved at the juncture under certain conditions.

Introduction

In the aerospace, automotive, and shipbuilding industries (among others) plane curves are used as control curves by designers. These curves are continuous and single-valued for some specified independent variable convenient for the purpose of design. Since the data on which the design is based usually include a set of points through which the curve must pass, splines are most convenient. Cubic splines (parametric or otherwise) are normally used for a number of reasons.

Not infrequently other design requirements lead to specification of slopes at certain of the assigned points. If these slopes are of moderate size, there is generally no great problem. However, if they are very large or infinite, the process of fitting a spline to the data may result in wild oscillations of the curve, a highly undesirable feature.

This phenomenon is particularly prevalent when first and second derivative continuity of the spline is required. On the other hand, this continuity requirement satisfies some part of the designers' desire for a "smooth" curve. It is not generally possible to maintain such continuity and simultaneously eliminate the undesirable oscillations, though there are circumstances in which this can be done. Other considerations involved in designing smooth curves may involve moving points, changing slopes, and other matters; however, none of these falls within the scope of this paper. Let it suffice to say that control curve design is an iterative process involving negotiations among designers, engineers and others engaged in the process.

In practice, large and infinite slopes generally occur at the ends of the spline as, for example, in the design of aircraft fuselages. This suggests a practical and, as it turns out, very useful solution to the problem. If this is the case, the point (or points if large slopes occur at both ends) are ignored and a spline over the remaining points is devised to the satisfaction of the designer. This provides the point(s) and slope(s) adjacent to the point(s) at which the difficulty exists. Next, the design is completed with one or more spline segments in such a way that the result is an inflection-free (convex or concave) curve connecting the point(s) with large slope to the adjacent point(s). This paper is concerned only with construction of a convex or concave curve consisting of one or more cubic spline segments and connecting a point at which a large or infinite slope is required to another point of given slope. Since the only data required at the second point are its coordinates, slope, and perhaps second derivative, the problem dealt with in this paper is completely independent of any other features of the remaining curve.

It will be observed that the four possible cases—large (infinite) positive slope at the initial point, large (infinite) positive slope at the terminal point, large (infinite) negative slope at the initial point, and large (infinite) negative slope at the terminal point—can all be reduced to the first case by simple transformations. In this first case, to which the development addresses itself, the curve is convex.

For the case in which the design constraint involves a large but finite end slope, two methods are developed. One of these involves the use of a set of spline segments all of which have finite slopes, while the other uses a single spline segment with infinite slope, such that the segment passes through the point of large slope and has the required slope at that point. The first method results in slope continuity but requires insertion of additional points, at which second derivative continuity is violated. It has the advantage that the problem can always be solved provided that the slope at the point of connection is moderate. The term moderate is defined under the problem statement for the first method. The second method has the advantage that no new points need be introduced in the domain over which the final curve is to be designed. It also has slope and second derivative continuity in the interior of its interval of definition, which includes the original end point. Under certain circumstances second derivative continuity can be maintained at the connecting point. Its disadvantages are that it complicates subsequent calculations required for numerical control of machine tools, and certain engineering calculations. In addition, no solution exists for certain cases covered by the first method. In practice, no practical examples have arisen to date for which the second method provides no solution.

If the design constraint specifies an infinite end slope, it is a simple matter to specify an inflection-free parametric cubic segment. In this case a solution always exists, provided the next slope is finite.

In all cases a free parameter is available to the designer for adjusting the shape of the curve being devised.

Necessary and sufficient conditions for an inflection-free cubic segment

Consider a cubic defined by $P_1(x_1, y_1)$, m_1 and $P_2(x_2, y_2)$, m_2, where m_1 and m_2 are the associated finite slopes (Fig. 1). (Since the cubics dealt with are completely defined by two points and their associated slopes, a set of data P_1, m_1, P_2, m_2 may also be called a cubic or a cubic segment.) The equations of the cubic segment are

$$x = x_2 + (x_1 - x_2)u,$$

and

$$y = au^3 + bu^2 + cu + d,$$

where $0 \le u \le 1$. The cubic coefficients are defined as follows:

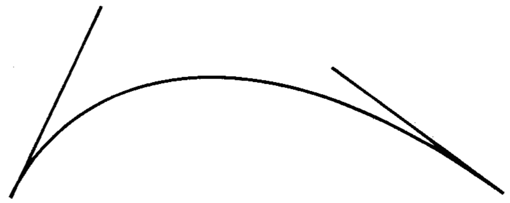


Figure 1 A cubic can be defined by the points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ where the associated slopes are m_1 and m_2, respectively.

$$a = a_1(x_1 - x_2),$$

$$b = b_1(x_1 - x_2),$$

$$c = m_2(x_1 - x_2),$$

and

$$d = y_2.$$

Further,

$$a_1 = -2R + m_1 + m_2,$$

$$b_1 = 3R - m_1 - 2m_2,$$

and

$$R = (y_1 - y_2)/(x_1 - x_2).$$

The values of u for the inflection point are

$$u_I = -b_1/3a_1$$
$$= (3R - m_1 - 2m_2)/(6R - 3m_1 - 3m_2).$$

If $a_1 = 0$ $(m_1 + m_2 = 2R)$ and $b_1 \ne 0$ $(m_1 + 2m_2 \ne 3R)$, then u_I will have an infinite value $(\pm\infty)$. However, for finite values of u_I (≤ 0 or ≥ 1), the following conditions must be met:

$$2m_1 + m_2 \le 3R \le m_1 + 2m_2 \qquad \text{for } m_2 > m_1,$$

or

$$m_1 + 2m_2 \le 3R \le 2m_1 + m_2 \qquad \text{for } m_2 < m_1. \tag{1}$$

In the special cases where u_I equals either zero ($3R = m_1 + m_2$) or one ($3R = 2m_1 + m_2$), a flat spot will exist at one end of the segment, which is generally considered undesirable in such curves. For this reason the inequalities (1) are strengthened to

$$(1 - \gamma)m_1 + \gamma m_2 \le R \le \gamma m_1 + (1 - \gamma)m_2 \qquad \text{for } m_2 > m_1,$$

and

$$\gamma m_1 + (1 - \gamma)m_2 \le R \le (1 - \gamma)m_1 + \gamma m_2 \qquad \text{for } m_2 < m_1, \tag{2}$$

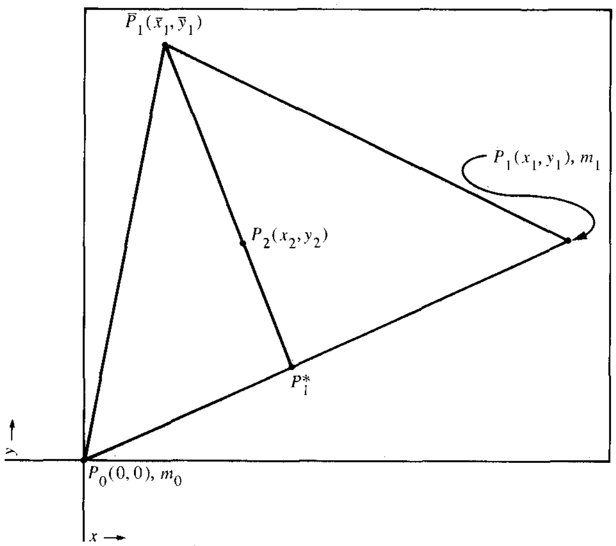


Figure 2 In the primary algorithm, the set of points generated describes a convex polygon and the point P_2 must be contained within the triangle defined by $P_0\bar{P}_1P_1$.

where $1/3 \leq \gamma \leq 1/2$. These conditions are now sufficient for the absence of an inflection point but are not necessary unless $\gamma = 1/3$.

Problem statement for the first method

In a plane with rectangular coordinates x, y, suppose that there are given two points P_0 $(0, 0)$ and P_1 (x_1, y_1), with slopes m_0 and m_1. In case P_0 is not at the origin, simply translate the origin to P_0; this does not alter slopes. Let

$$m_2 = y_1/x_1,$$

and assume that

$$0 < m_0,$$

$$m_1 < m_2 < m_0,$$

and

$$x_1 > 0. \tag{3}$$

These are the large and moderate conditions referred to in the introduction, for the particular case of large *positive initial* slope. With these conditions the objective is then to define a sequence of points $P_1, P_2, \cdots, P_N, P_0$ and associated slopes such that 1) successive pairs of points and slopes define cubics having no inflection point interior to the interval of definition, and 2) the collection of cubics has C_1 continuity and is convex (or concave).

Four cases involving large slope can occur: positive or negative, and initial or terminal. The large *positive initial* slope case is the only case that need be considered, since the other three cases can be reduced to this one. For example, for large *negative initial* slope m_0, where

$$0 > m_0,$$

$$m_1 > m_2 > m_0,$$

and

$$x_0 < x_1, \tag{4}$$

simply let $x' = x$ and $y' = -y$. For large *negative terminal* slope m_1, where

$$0 > m_1,$$

$$m_0 > m_2 > m_1,$$

and

$$x_0 < x_1, \tag{5}$$

let $x' = -x$, $y' = y$ and replace u with $1 - u$. Finally, for large *positive terminal* slope m_1, where

$$0 < m_1,$$

$$m_0 < m_2 < m_1,$$

and

$$x_0 < x_1, \tag{6}$$

let $x' = -x$, $y' = -y$ and replace u with $1 - u$.

The algorithm developed for the finite slope case consists of two parts, the primary algorithm and the sequence termination. The former is a repetitive procedure for generating a sequence of points, while the latter takes account of the fact that the sequence of points may not include the end point. For example, some point may have too large a slope.

Primary algorithm

It is assumed that the initial point is at the origin, an assumption requiring, at most, a translation of the origin. It is clear that the set of points generated must describe a convex polygon. In Fig. 2, let the slopes of the lines $P_0\bar{P}_1$ and $\bar{P}_1P_1$ be m_0 and m_1, respectively, so that

$$\bar{x}_1 = (y_1 - m_1x_1)/(m_0 - m_1),$$

and

$$\bar{y}_1 = m_0\bar{x}_1.$$

Since P_2 must be in the triangle defined by $P_0\bar{P}_1P_1$, let

$$P_1^* = \beta P_0 + (1 - \beta)P_1,$$

and

$$P_2 = \alpha\bar{P}_1 + (1 - \alpha)P_1^*,$$

where α, $\beta \in (0, 1)$. Next assign to P_2 the slope $m_2 = y_1/x_1$. It is now necessary to determine values of α and β (if they exist) such that the cubic segment P_2, m_2,

P_1, m_1 is convex. If this is so, those same values can be used to determine another point P_3 in the triangle defined by $P_0\bar{P}_2P_2$, etc.

Two problems are encountered. One is the question of finiteness of the sequence. That is, is a point P_N with slope m_N ever attained such that P_0, m_0, P_N, m_N defines a convex cubic segment? The second problem is the possibility that even if the sequence is finite, m_{N-1} may be too small and m_N too large for sequence completion. This problem will be resolved later. Assume for the moment that β is fixed. Since

$$(x_2, y_2) = \alpha(\bar{x}_1, \bar{y}_1) + (1 - \alpha)(1 - \beta)(x_1, y_1),$$

it follows that

$$\begin{aligned}(y_1 - y_2)/x_1 \\ &= [\beta(1 - \alpha)m_2(m_0 - m_1) + \alpha m_1(m_0 - m_2)]/(m_0 - m_1),\end{aligned}$$

and

$$\begin{aligned}(x_1 - x_2)/x_1 = [\beta(1 - \alpha)(m_0 - m_1) \\ + \alpha(m_0 - m_2)]/(m_0 - m_1) > 0. \qquad (7)\end{aligned}$$

Since α and β are in the interval (0, 1) and $m_0 > m_2 > m_1$, it is now not difficult to show that with

$$\bar{m} = (y_1 - y_2)/(x_1 - x_2),$$

$$m_1 < \bar{m} < m_0.$$

Let

$$M = (m_0 - m_2)/(m_0 - m_1);$$

then

$$0 < M < 1,$$

and

$$\bar{m} = [\beta(1 - \alpha)m_2 + \alpha m_1 M]/[\beta(1 - \alpha) + \alpha M].$$

In order for the cubic to be convex in (x_2, x_1), it is necessary and sufficient that $m_2 + 2m_1 \leq 3\bar{m} \leq 2m_2 + m_1$. Since the denominator of $\bar{m}$ is positive, it follows that

$$\begin{aligned}(m_2 + 2m_1)[\beta(1 - \alpha) + \alpha M] \leq 3\beta(1 - \alpha)m_2 + 3\alpha m_1 M \\ \leq (2m_2 + m_1)[\beta(1 - \alpha) + \alpha M]. \qquad (8)\end{aligned}$$

The left inequality reduces to

$$(m_2 - m_1)[2\beta(1 - \alpha) - \alpha M] \geqq 0.$$

That is, since $m_2 > m_1$,

$$2\beta(1 - \alpha) - \alpha M \geq 0$$

and

$$\alpha \leq \beta/(\beta + M/2).$$

Similarly, from the right inequality (8) it follows that

$$\alpha \geq \beta/(\beta + 2M).$$

Thus, the necessary and sufficient condition for the cubic to be convex in (x_2, x_1) is

$$\alpha = \beta/(\beta + sM),$$

where $1/2 \leq s \leq 2$. Note that for $s = 2$, there is an inflection point at $x = x_2$, while for $s = 1/2$, an inflection point exists at $x = x_1$. For

$$\gamma/(1 - \gamma) \leq s \leq (1 - \gamma)/\gamma,$$

the conditions (2) for $m_2 > m_1$ are met.

Since $\alpha = \beta/(\beta + sM)$, it follows that

$$\bar{m} = (sm_2 + m_1)/(s + 1).$$

Also, from Eqs. (6),

$$x_2 = \left[\frac{s(1 - \beta)(m_0 - m_2) + \beta(m_2 - m_1)}{s(m_0 - m_2) + \beta(m_0 - m_1)}\right] x_1,$$

$$y_2 = \left[\frac{s(1 - \beta)m_2(m_0 - m_2) + \beta m_0(m_2 - m_1)}{s(m_0 - m_2) + \beta(m_0 - m_1)}\right] x_1,$$

and

$$m_2 = y_1/x_1.$$

Moreover,

$$(y_2/x_2) - m_2 = \frac{\beta(m_2 - m_1)(m_0 - m_2)}{\beta(m_2 - m_1) + s(1 - \beta)(m_0 - m_2)}. \qquad (9)$$

With s and β fixed, a sequence x_i, y_i, m_i can be computed as follows:

$$m_i = y_{i-1}/x_{i-1},$$

$$x_i = \left[\frac{s(1 - \beta)(m_0 - m_i) + \beta(m_i - m_{i-1})}{s(m_0 - m_i) + \beta(m_0 - m_{i-1})}\right] x_{i-1},$$

and

$$y_i = \left[\frac{s(1 - \beta)(m_0 - m_i) + \beta m_0(m_i - m_{i-1})}{s(m_0 - m_i) + \beta(m_0 - m_{i-1})}\right] x_{i-1}. \qquad (10)$$

For each interval P_{i+1}, P_i, a useful canonical form for the cubic is

$$x(u) = x_{i+1} + (x_i - x_{i+1})u,$$

and

$$\begin{aligned}y(u) = y_{i+1}\alpha_0(u) + y_i\alpha_1(u) + m_{i+1}(x_i - x_{i+1})\alpha_2(u) \\ + m_i(x_i - x_{i+1})\alpha_3(u),\end{aligned}$$

where

$$\alpha_0(u) = 2u^3 - 3u^2 + 1,$$

$$\alpha_1(u) = -2u^3 + 3u^2,$$

$$\alpha_2(u) = u^3 - 2u^2 + u,$$

and

$$\alpha_3(u) = u^3 - u^2.$$

Equation (9) becomes

$$m_{i+1} - m_i = \frac{\beta(m_i - m_{i-1})(m_0 - m_i)}{\beta(m_i - m_{i-1}) + s(1-\beta)(m_0 - m_i)}. \quad (11)$$

It has, incidentally, been shown that m_i for $i > 0$ is a monotone increasing sequence and is bounded by m_0. In the next section it is shown that a finite sequence suffices if and only if $\beta > s/(1+s)$.

It is convenient for a designer to fix s permanently, say at $s = 2/3$, and to vary β, thus modifying the shape of the convex curve. For uniformity of the parameter domain from the designer's point of view, a new parameter

$$\delta = (1+s)\beta - s$$

leads to the domain

$$0 < \delta < 1,$$

whatever value is assigned to s.

It is interesting to note that the algorithm allows the value $\delta = \beta = 1$. In this case, however, the point P_2 is on the line $P_0\bar{P}_1$ and C_1 continuity is not preserved.

It is also of interest to note the position of the inflection point relative to the interval of definition of the cubic. Let x_I be the abscissa of the inflection point and

$$u_I = (x_I - x_i)/(x_{i-1} - x_i).$$

It turns out that

$$u_I = (s-2)/(3s-3),$$

independent of β,

$$u_I > 1 \qquad \text{if } s < 1,$$

$$u_I < 0 \qquad \text{if } 1 < s < 2,$$

and

$$u_I = \infty \qquad \text{if } s = 1.$$

In the last case, the curve is either a parabola or a straight line.

Finiteness of the sequence

In order for the computed sequence of points to terminate, it is necessary, for some N, that

$$\gamma m_0 + (1-\gamma)m_N \le m_{N+1}. \quad (12)$$

Otherwise, there exists no N such that P_0 and P_N can be connected by a convex cubic. This problem is addressed by the following theorem.

Theorem If $\beta > s(1-\beta)$, $m_2 = y_1/x_1 < \gamma m_0 + (1-\gamma)m_1$ for any fixed γ such that $0 < \gamma < 1$, then for some *finite* N,

$$x_{N+1} = y_N/x_N,$$

and Eq. (12) holds. If $\beta \le s(1-\beta)$, then always

$$m_{N+1} < \gamma m_0 + (1-\gamma)m_N.$$

Proof Define

$$H_i = (m_{i+1} - m_i)/(m_0 - m_i).$$

By hypothesis, $H_1 < \gamma$; also $H_i > 0$. If $\beta \le s(1-\beta)$, then

$$\beta(m_i - m_{i+1}) + s(1-\beta)(m_0 - m_i) \ge \beta(m_0 - m_{i-1}).$$

From Eq. (11),

$$(m_{i+1} - m_i)/(m_0 - m_1) \le (m_i - m_{i-1})/(m_0 - m_{i-1}),$$

or

$$H_i \le H_{i-1} \qquad \text{for all } i \ge 1.$$

Hence

$$H_i < \gamma,$$

and

$$m_{i+1} < \gamma m_0 + (1-\gamma)m_i.$$

That is, an infinite sequence of cubics defined by the primary algorithm does not meet the specified requirements.

Now consider

$$\beta = s(1-\beta) + \xi,$$

where $\xi > 0$, and the identity

$$m_{i+1} - m_i = \frac{(m_i - m_{i-1})(m_0 - m_i)}{m_0 - m_{i-1}} + \frac{\xi(m_0 - m_i)^2(m_i - m_{i-1})}{(m_0 - m_{i-1})[\beta(m_0 - m_{i-1}) - \xi(m_0 - m_i)]}.$$

Now,

$$\beta(m_0 - m_{i-1}) - \xi(m_0 - m_i)$$
$$= \beta(m_i - m_{i-1}) + s(1-\beta)(m_0 - m_i) < \beta(m_0 - m_{i-1}),$$

and

$$\frac{m_{i+1} - m_i}{m_0 - m_i} > \frac{m_i - m_{i-1}}{m_0 - m_{i-1}} + \frac{\xi(m_0 - m_i)(m_i - m_{i-1})}{\beta(m_0 - m_{i-1})(m_0 - m_{i-1})};$$

thus,

$$H_i > H_{i-1}[1 + \xi(m_0 - m_i)/\beta(m_0 - m_{i-1})] > H_{i-1}.$$

But

$$(m_0 - m_1)/(m_0 - m_{i-1}) = 1 - H_{i-1},$$

hence

$$H_i > H_{i-1}[1 + \xi/\beta - (\xi/\beta)H_{i-1}]$$
$$= (1 + \xi/\beta)H_{i-1} - (\xi/\beta)H_{i-1}^2.$$

The hypothesis that $H_i < \gamma < 1$ for all i yields

$$H_i > [1 + (\xi/\beta)]\,H_i - (\xi\gamma/\beta)H_i,$$

more simply written as

$$H_i > cH_{i-1},$$

where $c = 1 + \xi(1 - \gamma)/\beta$ and $c > 1$. Thus, the hypothesis $H_i < \gamma$ for all i leads to the contradiction that H_i increases without limit. Therefore, there exists an N such that

$$H_N \geq \gamma,$$

and since H_i is a monotone increasing in this case,

$$H_i > \gamma \qquad \text{for all } i > N.$$

In order to complete the algorithm it is necessary to investigate certain problems arising in terminating the sequence.

Sequence termination

It is now assumed that $\beta > s/(1 + s)$. The objective of the primary algorithm is to produce a sequence of points $P_2, \cdots, P_N$ such that P_0 and P_N can be connected by a convex cubic segment having slope m_0 at P_0, m_N at P_N. This requires that

$$\gamma m_0 + (1 - \gamma)m_{N-1} \leq m_N \leq (1 - \gamma)m_0 + \gamma m_{N-1};$$

that is,

$$\gamma \leq H_{N-1} \leq 1 - \gamma.$$

If $H_1 > 1 - \gamma$, the primary algorithm is inapplicable, and this case will be resolved by another technique. If $\gamma \leq H_1 \leq 1 - \gamma$, the points P_0, P_1 and slopes m_0, m_1 already define a convex cubic and there is no problem. Lastly, if $H_1 < \gamma$, the previous theorem guarantees the existence of a minimum $N \geq 2$ such that $H_{N-1} < \gamma$ and $H_N \geq \gamma$. If also $H_N \leq (1 - \gamma)$, then P_0, m_0, P_{N+1}, m_{N+1} confine a convex cubic as required, and the sequence of points and slopes derived by the primary algorithm is P_2, m_2, $\cdots$, P_{N+1}, m_{N+1}. If however $H_N > 1 - \gamma$, the primary algorithm is no longer applicable. That is,

$$m_0 > m_{N+1} > (1 - \gamma)m_0 + \gamma m_N,$$

and the slope m_{N+1} is too large to allow a convex cubic segment connecting P_0 and P_{N+1}.

For this case define

$$m'_{N+1} = (1 - \gamma)m_0 + \gamma m_N, \; H'_{N+1} = (m'_{N+1} - m_N)/(m_0 - m_N) = 1 - \gamma > \gamma,$$

and

$$m_N < m'_{N+1} < m_{N+1}.$$

The cubic segment P_N, m_N, P_{N+1}, m_{N+1} has continuous and monotonic increasing slope on its interval. Hence there exists a unique point P'_{N+1} with slope m'_{N+1}. When P_{N+1}, m_{N+1} is replaced by P'_{N+1}, m'_{N+1}, the problem is resolved. It remains only to develop the necessary algorithm. In the notation of the section on the problem statement of the first method,

$$x = x_{N+1} + (x_N - x_{N+1})u,$$

$$y = au^3 + bu^2 + cu + y_{N+1},$$

where

$$a = -2(y_N - y_{N+1}) - (m_{N+1} + m_N)(x_N - x_{N+1}),$$

$$b = 3(y_N - y_{N+1}) - (2m_{N+1} + m_N)(x_N - x_{N+1}),$$

and

$$c = m_{N+1}(x_N - x_{N+1}).$$

It is required that

$$dy/dx = m'_N;$$

that is,

$$f(u) = 3au^2 + 2bu + c - m'_N(x_N - x_{N+1}) = 0.$$

Compute the solution u of the quadratic equation where $0 < u < 1$. Note that

$$f(0) = (m_{N+1} - m'_N)(x_N - x_{N+1}) < 0,$$

and

$$f(1) = (m_N - m'_N)(x_N - x_{N+1}) > 0,$$

so that there is indeed a unique value of u.

For the case $H_i > 1 - \gamma$, there is no cubic available, as in the previous case. There are two possible approaches to this problem. The first, apparently practical, approach assumes that m_1 is not small compared to m_0, and simply ignores the problem. The other approach is to use a pair of Bernstein-Bezier cubics [1, 2] (see Fig. 3). Briefly, given P_0, m_1, P_1, m_1, the point $\bar{P}_1$ is defined. Let P_{01}, P_{02} trisect the line $P_0\bar{P}_1$ and P_{11}, P_{12} trisect the line $\bar{P}_1P_1$. Choose a point Q_0 somewhere between P_{01}, P_{02} and a point Q_4 somewhere between P_{11}, P_{12}. The frame P_0, P_{01}, Q_1, Q_2 defines a convex cubic segment over the interval $(x_0, \bar{x}_1)$. The curve is convex since the frame is convex. It is a true cubic since the abscissae are equally spaced. Its end points and slopes are the same as the end points and slopes of the frame. The frame Q_2, Q_3, P_{12}, P_1 achieves the same results for the interval $(\bar{x}_1, x_1)$. Moreover, continuity and C_1 continuity at Q_2 are preserved.

The same technique could be used in the previous case where m_{N+1} is too large.

Problem statement for the infinite slope case

As has been described elsewhere [3–5], a convenient choice of parametrization over an interval (x_0, x_1) has

$$x = x_0 + (x_1 - x_0)g(u),$$

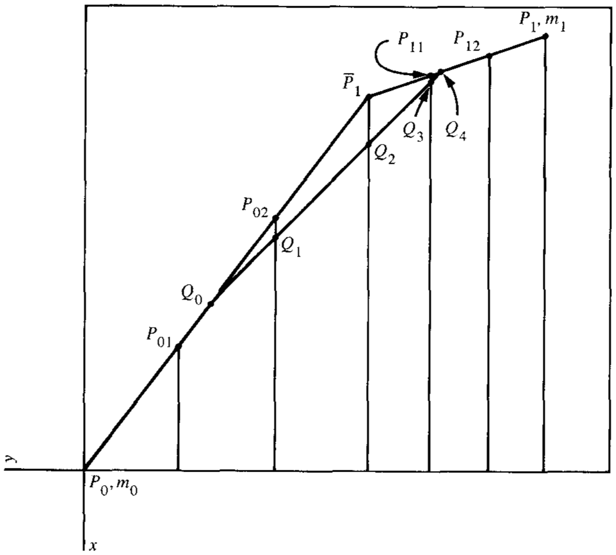


Figure 3 The frames P_0, P_{01}, Q_1, Q_2 and Q_2, Q_3, P_{11}, P_1 define two convex cubic segments over the intervals $(x_0, \bar{x}_1)$ and $(\bar{x}_1, x_1)$ in the Bernstein-Bezier approach.

where

$$g(u) = 2u^2 - u^3,$$

or

$$g(u) = u^2$$

for infinite slope dy/dx at $x = x_0$. For the infinite slope at $x = x_1$,

$$g(u) = (1 - u)^2(1 + u),$$

or

$$g(u) = (1 - u)^2,$$

respectively. The variable y is, of course, a cubic function of u.

Thus, as before, $P_0(x_0, y_0)$, m_0 and $P_1(x_1, y_1)$, m_1 are given and the constraints $m_0 > 0$, $m_1 < m_2 < m_0$, and $x_0 < x_1$ remain the same. It is not useful in the present case to translate P_0 to the origin. The goal here is to find a spline segment on (x^*, x_1) passing through P^*, P_0, P_1, with slopes as indicated in Fig. 4. As for the other three cases, the same transformations (3)–(5) as before are required to reduce them to the present case.

Calculation of P*

Bernstein-Bezier cubics are used to solve the problem by constructing the frame P_1, P_2, P_3, P^* (see Fig. 4). Here $P^*\bar{P}$ is a vertical line, and $\bar{P}P_1$ has slope m_1;

$$P_2 = (1 - s)P_1 + s\bar{P},$$

and

$$P_3 = (1 - r)P^* + r\bar{P},$$

where $r, s \in (0, 1)$. Let $X = x_1 - x^*$ and $Y = y_1 - y^*$. Then

$$\bar{P} = (x^*, y_1 - m_1X),$$

$$P_2 = (x_1 - sX, y_1 - sm_1X),$$

and

$$P_3 = (x^*, y^* + rY - rm_1X).$$

The Bernstein-Bezier cubic is

$$P(u) = (P_1 - 3P_2 + 3P_3 - P^*)u^3 + 3(P_2 - 2P_3 + P^*)u^2 + 3(P_3 - P^*)u + P^*,$$

where $u \in (0, 1)$. The x component of $P(u)$ is

$$x(u) = (3s - 2)Xu^3 + 3(1 - s)Xu^2 + x^*,$$

and this is precisely $x^* + Xg(u)$, where $g(u) = 2u^2 - u^3$ when $s = 1/3$, and $g(u) = u^2$ when $s = 2/3$.

Now define

$$F_0 = (3s - 2)u^3 + 3(1 - s)u^2 - 1 = (u - 1)[(3s - 2)u^2 + u + 1],$$

$$F_1 = (3r - 2)u^3 + 3(1 - 2r)u^2 + 3ru - 1 = (u - 1)^2[(3r - 2)u - 1],$$

$$F_2 = (s - r)u^3 + (2r - s)u^2 - ru = u(u - 1)[(s - r)u + r],$$

$$F_3 = (3r - 2)u^2 + 2(1 - 2r)u + r = (u - 1)[(3r - 2)u - r],$$

and

$$F_4 = m_1[3(s - r)u^2 + 2(2r - s)u - r] - m_0[(3s - 2)u^2 + 2(1 - s)u]. \tag{13}$$

The y component of u is given by

$$y(u) = Y[(3r - 2)u^3 + 3(1 - 2r)u^2 + 3ru] + 3m_1X[(s - r)u^3 + (2r - s)u^2 - ru] + y^*,$$

or

$$y(u) = y^*\alpha_0(u) + y_1\alpha_1(u) + 3r[(y_1 - y^*) - m_1(x_1 - x^*)]\alpha_2(u) + 3sm_1(x_1 - x^*)\alpha_3(u). \tag{14}$$

Now suppose that u has a value such that $P(u) = P_0$. That is,

$$x(u) = x_0,$$

$$y(u) = y_0,$$

and also

$$\dot{y}(u)/\dot{x}(u) = m_0,$$

where the dot notation signifies differentiation with respect to u.

From these last three equations, it follows that

$$XF_0 = x_0 - x_1,$$

$$YF_1 + 3m_1XF_2 = y_0 - y_1,$$

and

$$YF_3 + XF_4 = 0.$$

Hence,

$$X = (x_0 - x_1)/F_0,$$

$$Y = (y_0 - y_1)/F_1 - 3m_1(x_0 - x_1)F_2/F_0F_1,$$

and

$$F_1F_4 - 3m_1F_2F_3 + \bar{m}F_0F_3 = 0, \tag{15}$$

where

$$\bar{m} = (y_1 - y_0)/(x_1 - x_0).$$

If u exists satisfying the relationship among the Fs, then X and Y are also known, hence x^*, y^*. Note that F_0 and F_1 do not vanish for the allowed values of r and s; also $0 < u < 1$. Equation (15) reduces to

$$Br + C = 0 \qquad r = -C/B,$$

where

$$B = \bar{m}[(3s - 2)u^2 + u + 1](3u - 1)$$
$$+ m_1 [(1 - u)^2 - 3su^2]$$
$$- 3um_0[(3s - 2)u^2 + 2(1 - s)u],$$

and

$$C = -2\bar{m}u[(3s - 2)u^2 + u + 1] + m_1su(u + 2)$$
$$+ m_0u(1 + 2u)[(3s - 2)u + 2(1 - s)]. \tag{16}$$

Let

$$H = (\bar{m} - m_1)/(m_0 - m_1),$$

$$f_1(u) = 6H[(3s - 2)u^2 + u + 1]$$
$$- 3(1 + 2u)[(3s - 2)u + 2(1 - s)],$$

and

$$f_2(u) = 3H[(3s - 2)u^2 + u + 1]$$
$$-3u[(3s - 2)u + 2(1 - s)].$$

Then

$$r = uf_1(u)/[uf_1(u) - (1 - u)f_2(u)],$$

and for $0 < r < 1$, it follows that $f_1(u)f_2(u) < 0$.

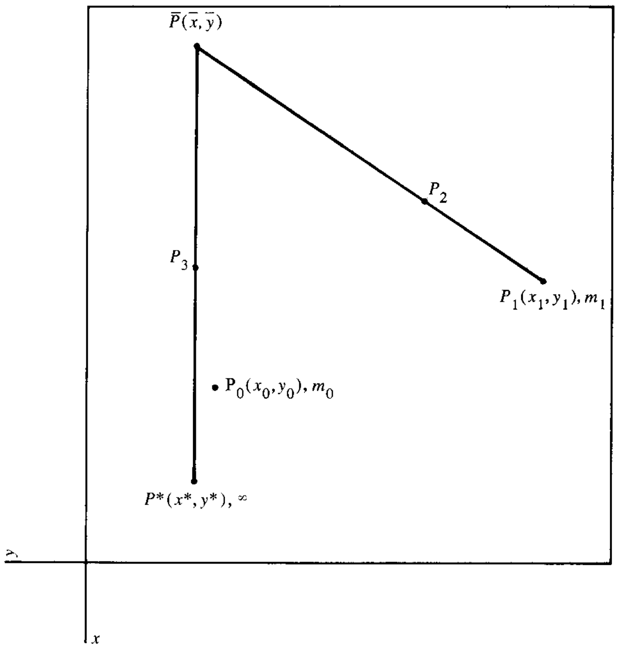


Figure 4 Bernstein-Bezier cubics are used to solve the problem of infinite slope by constructing the frame P_1, P_2, P_3, P^*, where $P^*\,\bar{P}$ is a vertical line.

For $s = 2/3$,

$$f_1(u) = (6H - 4)u + (6H - 2),$$

and

$$f_2(u) = (3H - 2)u + 3H.$$

Here, $f_1(u) = 2[f_2(u) - 1]$, whence $0 < f_2(u) < 1$. If $3H - 2 < 0$, then

$$(3H - 1)/(2 - 3H) < u < 3H/(2 - 3H),$$

and for $0 < H \le 1/3$,

$$0 < u < 3H/(2 - 3H). \tag{17}$$

For $1/3 < H < 1/2$,

$$(3H - 1)/(2 - 3H) < u < 1. \tag{18}$$

If $H \ge 1/2$, then $u \ge 1$, which is not allowed. Similarly, $3H - 2 \ge 0$ leads to an invalid solution. In this case the problem can only be dealt with if $0 < H < 1/2$. If a solution is necessary, the previous method must be used.

The parameter provided to the designer here is

$$\delta = u(2 - 3H)/3H \qquad \text{for } H \le 1/3,$$

and

$$\delta = [(2 - 3H)u - (3H - 1)]/(3 - 6H) \qquad \text{for } 1/3 < H < 1/2.$$

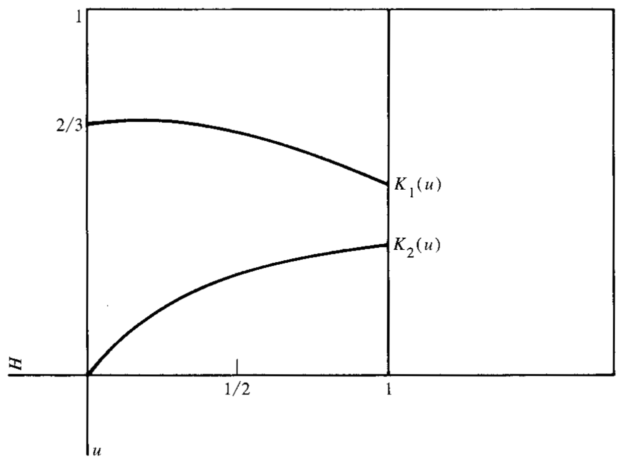


Figure 5 Convex functions $K_1(u)$ and $K_2(u)$. Note the five different ranges for H as defined in Eqs. (19).

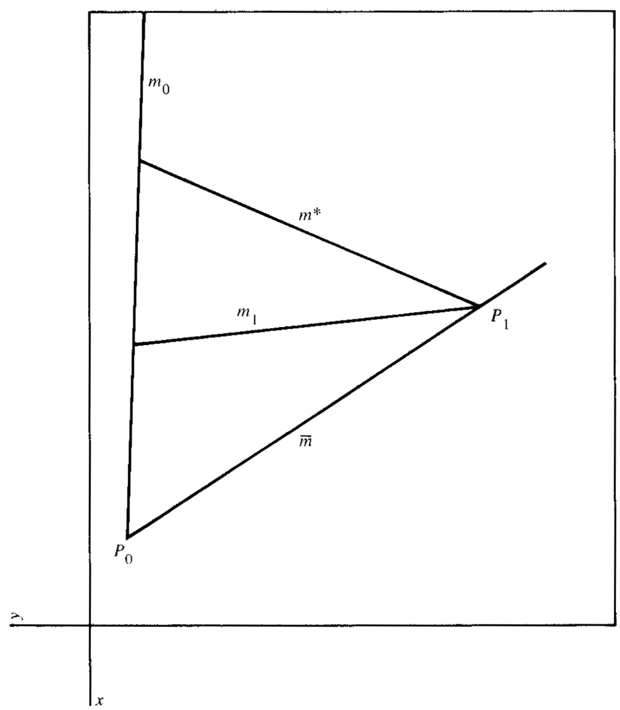


Figure 6 For the slope m_0, lines such as m_1 are acceptable as terminal slopes.

Thus he can choose any δ such that $0 < \delta < 1$. Then u, $f_1(u)$, $f_2(u)$, r, the F functions and finally x^*, y^* are computed in order.

For $s = 1/3$, the equations are

$$f_1(u) = 6(1 - H)u^2 + (6H - 5)u + (6H - 4),$$

and

$$f_2(u) = 3(1 - H)u^2 + (3H - 4)u + 3H.$$

Let

$$K_1(u) = (6u^2 - 5u - 4)/6(u^2 - u - 1),$$

and

$$K_2(u) = (3u^2 - 4u)/3(u^2 - u - 1).$$

Now

$$f_1(u) > 0 \Leftrightarrow H > K_1(u);$$

$$f_1(u) < 0 \Leftrightarrow H < K_1(u);$$

$$f_2(u) > 0 \Leftrightarrow H > K_2(u);$$

$$f_2(u) < 0 \Leftrightarrow H < K_2(u).$$

It is easily shown that both K_1 and K_2 are convex functions, that $K_1(0) = 2/3$ and $K_1(1) = 1/2$, that

$$\max K_1(u) = (25 - 2\sqrt{5})/30 \qquad \text{at } u = \sqrt{5} - 2,$$

that $K_2(0) = 0$, $K_2(1) = 1/3$, and that

$$\max K_2(u) = (10 - 2\sqrt{5})/15 \qquad \text{at } u = 3 - \sqrt{5}.$$

Thus

$$K_2(u) < H < K_1(u),$$

as in Fig. 5.

There are five different ranges for H.

1. $H \leq 1/3 \qquad 0 < u < u_{21}$;
2. $1/3 < H \leq (10 - 2\sqrt{5})/15$
   $u_{21} < u < u_{22}$ or $u_{22} < u < 1$;
3. $(10 - 2\sqrt{5})/15 < H < 1/2 \qquad 0 < u < 1$;
4. $1/2 \leq H < 2/3 \qquad 0 < u < u_{11}$;
5. $2/3 \leq H < (25 - 2\sqrt{5})/30 \qquad u_{11} < u < u_{12}$, (19)

where u_{21} and u_{22} are the smaller and larger roots, respectively, of $K_2(u) = H$ for $f_2(u) = 0$. Likewise, u_{11}, u_{12} are the smaller and larger roots for $f_1(u) = 0$. That is,

$$u_{11} = (5 - 6H - \sqrt{180H^2 - 300H + 121})/12(1 - H),$$

$$u_{12} = (5 - 6H + \sqrt{180H^2 - 300H + 121})/12(1 - H),$$

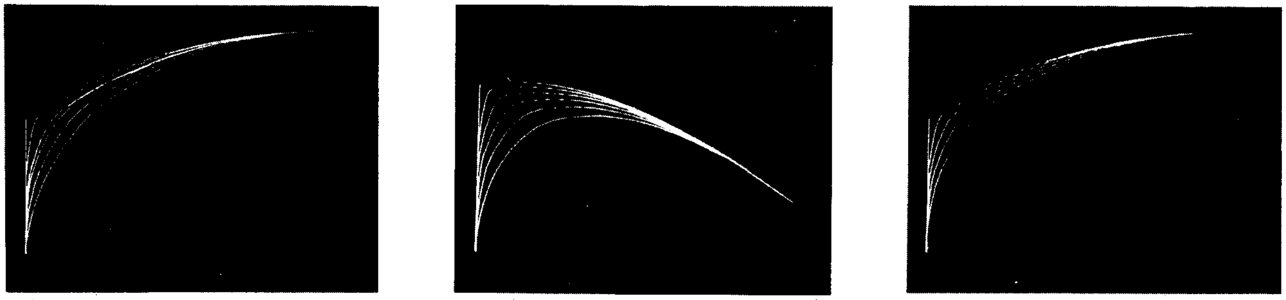
$$u_{21} = (4 - 3H - \sqrt{45H^2 - 60H + 16})/6(1 - H),$$

and

$$u_{22} = (4 - 3H + \sqrt{45H^2 - 60H + 16})/6(1 - H). \qquad (20)$$

This case, although more complex, allows a considerably wider range of H. Again, a design parameter δ between 0 and 1 can be introduced for the convenience of the designer. The second range of u for the second range of H above can be ignored.

$H = (\bar{m} - m_1)/(m_0 - m_1)$ is a monotonically decreasing function of m_1, and $H = 1$ for $m_1 = -\infty$. Hence, there is a largest m_1, here called m^*, such that this method does not



(a) (b) (c)

Figure 7 Finite slope results for the cubic segment $P_0(x_0, y_0)$, m_0; $P_1(x_1, y_1)$, m_1 for the δ values of 0.1, 0.3, 0.5, 0.7, 0.9 and 1 (bottom to top curves), where $s = 2/3$. In all cases P_0 is defined as the origin (0, 0). The chosen variables m_0, $P_1(x_1, y_1)$ and m_1 are as follows: (1) 10^3, (4, 4) and 0; (b) 10^2, (4, 1) and -1; and (c) 10^5, (4, 4) and 10^{-2}.

(a) (b) (c)

Figure 8 Infinite slope results for the same cubics given in Fig. 7, where the values of δ are 0.1, 0.3, 0.5, 0.7 and 0.9 for $s = 1/3$.

lead to a solution (see Fig. 6). If H^* is the smallest value of H (1/2 or $(25 - 2\sqrt{5})/30$), which is not acceptable, then

$$m^* = (m - H^* m_0)/(1 - H^*).$$

A solution exists for any $m_1 > m^*$ and $m_1 < m_0$.

Second derivative continuity can be attained at the connecting point for a certain interval of values of a specified second derivative. This is developed in Reference [6].

Infinite initial slope

If the slope at P_0 is taken now to be infinite, with r and s as before, then the Bernstein-Bezier polynomial is obtained when P^* is replaced by P_0. Then

$$x(u) = x_0 + (x_1 - x_0)g(u),$$

$$g(u) = (3s - 2)u^3 + 3(1 - s)u^2,$$

$$\begin{aligned} y(u) = y_0 &+ [(3r - 2)(y_1 - y_0) + (3s - 3r)m_1(x_1 - x_0)]u^3 \\ &+ [3(1 - 2r)(y_1 - y_0) + (6r - 3s)m_1(x_1 - x_0)]u^2 \\ &+ [3r(y_1 - y_0) - 3rm_1(x_1 - x_0)]u, \end{aligned}$$

and

$$\begin{aligned} y(u) &= y_0\alpha_0(u) + y_1\alpha_1(u) \\ &+ 3r[(y_1 - y_0) - m_1(x_1 - x_0)]\alpha_2(u) \\ &+ 3sm_1(x_1 - x_0)\alpha_3(u). \end{aligned} \tag{21}$$

At $u = 1$,

$$d^2y/dx^2 = 2(m_2 - m_1)(r - 1)/3(x_1 - x_0)s^2.$$

Let μ be the initial second derivative for the next interval. If

$$2(m_1 - m_2)/3(x_1 - x_0)s^2 \le \mu \le 0,$$

then, if

$$r = 1 + 3\mu(x_1 - x_0)s^2/2(m_2 - m_1),$$

and

$$0 < r < 1, \tag{22}$$

C_2 continuity is maintained for $x = x_1$.

Summary

The purpose of this section is to summarize the results of the previous development in an algorithmic form.

Again, these methods deal only with large or infinite end slopes and the four possible cases have been reduced to one, large or infinite positive initial slope, as described previously. Once the transformed problem is solved and the canonical form obtained, the reverse transformations solve the original problem. Note that $\alpha_0(1-u) = \alpha_1(u)$, $\alpha_1(1-u) = \alpha_0(u)$, $\alpha_2(1-u) = -\alpha_3(u)$, $\alpha_3(1-u) = -\alpha_2(u)$ and, for the infinite slope case, if $g(u) = 2u^2 - u^3$, then $g(1-u) = (1-u)^2(1+u)$ and conversely; if $g(u) = u^2$, then $g(1-u) = (1-u)^2$ and conversely.

For the first method, values of γ and s must be chosen with $1/3 < \gamma < 1/2$, $\gamma/(1-\gamma) \leq s \leq (1-\gamma)/\gamma$ if flat spots are to be avoided. For the examples of the next section, $\gamma = 0.4$ and $s = 2/3$. One must next verify the shape conditions (3). If these conditions are met, iterate using Eqs. (10), with any fixed β such that $1 > \beta > s/(1-s)$, or the equivalent $\delta = (1-s)\beta - s$ such that $1 > \delta > 0$. Simultaneously compute $H_i = (m_{i+1} - m_i)/(m_0 - m_i)$, and terminate the iteration according to the specifications given in the section on sequence termination.

For the second method again verify the slope conditions (3). Compute $\bar{m} = (y_1 - y_0)/(x_i - x_0)$, $H = (\bar{m} - m_1)/(m_0 - m_1)$. Now, depending upon which infinite slope form is used [$g(u) = 2u^2 - u^3$ or $g(u) = u^2$], check the ranges (17) and (18) or (19) and (20). If a valid H range exists, select an arbitrary u from the associated u range. If not, revert to the first method. For example, in the first case, if $H = 0.25$, then any u such that $0 < u < 0.6$ is usable. Compute $r = -C/B$ from Eqs. (16), then F_0, F_1, F_2 from (13), X and Y from (15); $x^* = x_1 - X$; $y^* = y_1 - Y$ and the canonical form from Eq. (14).

For infinite end slope the canonical form is given in Eq. (21), where r is arbitrarily chosen with $0 < r < 1$. If a second derivative μ is specified at the connecting point and if (22) is satisfied, then this choice of r produces the required continuity.

Photographs of curves produced from typical finite and infinite slope results for the cubic $P_0(x_0, y_0)$, m_0; $P_1(x_1, y_1)$, m_1 are given in Figs. 7 and 8. These curves were generated on an oscilloscope connected to a computer, which had been programmed to implement the mathematics described in this paper and to display the appropriate cubic segments.

References

1. J. A. Adams, "Geometric Concepts for Computer Graphics," *Report AD-750 753*, U.S. Naval Academy, Annapolis, Maryland, 1972.
2. P. E. Bezier, "Example of an Existing System in the Motor Industry: The Unisurf System," *Proceedings of the Royal Society of London*, 1971, p. 207.
3. B. Dimsdale, "On Multiconic Surfaces," *Technical Report G320-2661*, IBM Scientific Center, Los Angeles, California, 1974.
4. A. Inselberg, "Cubic Splines with Infinite Derivatives at Some Knots," *IBM J. Res. Develop.* **20**, 430 (1976).
5. B. Dimsdale, "On Multiconic Surfaces II, Spline Elements," to be published as LASC report.
6. B. Dimsdale, "Convex Cubic Splines II," *Technical Reports G320.2691 and G320.2692*, IBM Scientific Center, Los Angeles, California, 1977.

Received May 16, 1977; revised August 15, 1977

The author recently retired from the IBM Data Processing Division's Los Angeles Scientific Center, Los Angeles, California 90045.