A Geometric Characterization of Parametric Cubic Curves

MAUREEN C. STONE
Xerox PARC
and
TONY D. DEROSE
University of Washington

In this paper, we analyze planar parametric cubic curves to determine conditions for loops, cusps, or inflection points. By expressing the curve to be analyzed as a linear combination of control points, it can be transformed such that three of the control points are mapped to specific locations on the plane. We call this image curve the *canonical curve*. Affine maps do not affect inflection points, cusps, or loops, so the analysis can be applied to the canonical curve instead of the original one. Since the first three points are fixed, the canonical curve is completely characterized by the position of its fourth point. The analysis therefore reduces to observing which region of the canonical plane the fourth point occupies. We demonstrate that for all parametric cubes expressed in this form, the boundaries of these regions are conics and straight lines. Special cases include Bézier curves, B-splines, and Beta-splines. Such a characterization forms the basis for an easy and efficient solution to this problem.

Categories and Subject Descriptors: G.1.1 [**Numerical Analysis**]: Interpolational—*spline and piecework polynomial interpolations*; I.3.5 [**Computer Graphics**]: Computational Geometry and Object Modelling—*curve, surface, solid, and object representations*

General Terms: Algorithms, Design

Additional Key Words and Phrases: Bezier curves, characterization, spline curves

1. INTRODUCTION

This paper describes a simple geometric method for determining whether a parametric cubic curve such as a Bézier curve, or a segment of a B-spline, has any loops, cusps, or inflection points. We call this determination the *characterization* of the curve, and these properties the *characteristics* of the curve. A curve with a *loop* is self-intersecting, one with a *cusp* has a point where the unit tangent

This work was supported in part by the National Science Foundation under grant numbers DMC-8602141 and DMC-8802949, the Xerox Corporation, and the Digital Equipment Corporation.
Authors' addresses: M. C. Stone, Electronic Documents Laboratory, Xerox PARC, 3333 Coyote Hill Rd.; Palo Alto, CA 94304; T. D. DeRose, Department of Computer Science, FR-35, University of Washington, Seattle, Wash. 98195.

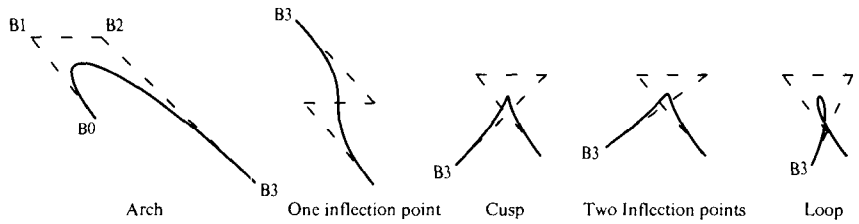

Fig. 1. Cubic Bézier curves showing a loop, cusp, and inflection points. Only the point $\mathbf{B}_3$ is moving.

vector is discontinuous, and one with an *inflection point* has a point where the curvature vanishes.

Characterizing cubic curves has wide-ranging applications. For instance, in numerically controlled milling operations, many of the algorithms rely on the fact that the trace of the curve is smooth—an assumption that is violated if a cusp is present. Inflection points often indicate unwanted oscillations in applications such as automobile body design and aerodynamics, and a surface that has a cross-section curve possessing a loop cannot be manufactured.

In most applications, parametric cubic curves are expressed as linear combinations of control points and basis functions. Because the characteristics of the curve do not change under affine transformations (the transformations including rotation, scaling, translation, and skewing), we can map the curve onto a canonical form so that the coordinates of three of the control points are fixed. As the fourth point moves about the plane, the curve may take on a loop, a cusp, or zero to two inflection points, depending only on the position of the moving points, as shown in Figure 1. The plane can therefore be partitioned into regions labeled according to the characterization of the curve segment when the fourth point is in each region. We show that the areas of the plane that define loops, inflection points, and cusps are all bounded by straight lines and conic curves. The simplicity of this form makes the characterization efficient to compute.

Previous work in this area has been done by Wang [12], who produced algorithms based on algebraic properties of the coefficients of the parametric polynomial and included some geometric tests using the B-spline control polygon. Su and Liu [11] have presented a specific geometric solution for the Bézier representation, and Forrest [6] has studied rational cubic curves. This paper presents a general method for making a geometric characterization of a (nonrational) parametric cubic that can be applied to any representation that is a linear combination of control points and basis functions. In Sections 3 and 4, we present the construction of a characterization diagram for Bézier curves. This diagram is simpler than the diagram of Su and Liu and can be more efficiently implemented. In Section 5, certain degenerate cases are addressed. Armed with the intuition gained in Sections 3 and 4, in Section 6 we describe a simple general procedure for constructing characterization diagrams for cubic curves of any type. This is done by showing that *all* characterization diagrams can be obtained from a two-dimensional slice through a common three-dimensional space.

The specific case for Bézier curves is presented before the general method strictly for purposes of making the material more accessible to less sophisticated

readers who wish to gain a better understanding of Bézier curves. More proficient readers may therefore wish to read Section 6 before reading Sections 3 and 4.

2. BACKGROUND AND PREVIOUS WORK

The general representation of a parametric cubic polynomial consists of a function $\mathbf{Q}(t) = (X(t), Y(t))$, where $X(t)$ and $Y(t)$ are each cubic polynomials with derivatives $\mathbf{Q}'(t) = (X'(t), Y'(t))$ and $\mathbf{Q}''(t) = (X''(t), Y''(t))$. Such a representation generally defines a curve of infinite length. To define a curve of finite length it is necessary to restrict the parameter t typically to the interval $[0, 1]$. In this paper, we use the terms *cubic curve*, *parametric cubic*, or *untrimmed curve* to refer to the infinite curve, and *segment*, *cubic segment*, or *trimmed curve* to refer to a bounded curve.

The *power basis representation* for $\mathbf{Q}(t)$ is defined as

$$\mathbf{Q}(t) = \sum_{i=0}^{3} \alpha_i t^i$$

This is a special case of a more general form where the curve is represented as a linear combination of control points and basis functions:

$$\mathbf{Q}(t) = \sum_{i=0}^{3} \mathbf{P}_i d_i(t), \tag{1}$$

where the $\mathbf{P}_i$ are the control points, and the $d_i(t)$ are the basis functions. This equation is often represented in matrix form [1].

Although planar parametric cubic curves are quite flexible, they are somewhat constrained in the diversity of their characterizations. A number of previous papers have examined these issues in the context of computer-aided geometric design and graphics (cf. [5], [7], [11] and [12]). Forrest [6] has gone somewhat further and examined rational cubic curves. For our purposes, the most relevant works are [11] and [12]. In particular, these authors show that loops, cusps, and inflection points are mutually exclusive. Thus, if a nondegenerate parametric cubic curve has a cusp, then it cannot have a loop or an inflection point; if it has a loop, then it cannot have either a cusp or an inflection point, and so forth. Moreover, a nondegenerate parametric cubic curve cannot have more than one cusp, one loop, or two inflection points. Further, they show that the presence of a loop, a cusp, or inflection points can all be determined by examining

$$F(t) = \det(\mathbf{Q}'(t), \mathbf{Q}''(t)) = X'(t)Y''(t) - X''(t)Y'(t),$$

which is a function that is proportional to the signed curvature of the curve at the point $\mathbf{Q}(t)$. (The signed curvature is defined to have a magnitude equal to the curvature of the curve, with the sign being chosen as positive if the cross product of $\mathbf{Q}'(t)$ and $\mathbf{Q}''(t)$ points in the z direction.)

The vanishing of $F(t)$ indicates the presence of inflection points since at such points the first and second derivative vectors are linearly dependent, which causes the curvature to vanish. Since $X(t)$ and $Y(t)$ are cubic functions, one would expect that $F(t)$ is a cubic function. However, using the power basis representation of $\mathbf{Q}(t)$, it is not difficult to show that the cubic term is zero,

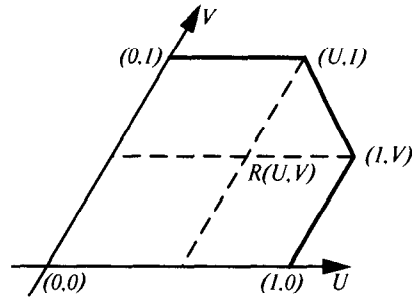

Fig. 2. The coordinate system of Su and Liu showing the construction of the characteristic point R.

implying that $F(t)$ is a quadratic of the form

$$F(t) = At^2 + Bt + C$$

where

$$A = 3 \det(\alpha_2, \alpha_3), \qquad B = 3 \det(\alpha_1, \alpha_3), \qquad C = \det(\alpha_1, \alpha_2). \tag{2}$$

Solutions for values of t to the equation $F(t) = 0$, therefore, indicate the presence of inflection points. The discriminant

$$\Delta = B^2 - 4AC$$

therefore becomes an important quantity. Wang, Su, and Liu [11], [12] show that the characterization of the curve can be entirely determined from A, B, and C as follows: If $A = 0$, then there is exactly one inflection point. Otherwise, if $\Delta > 0$, there are exactly two inflection points, if $\Delta < 0$ there is a loop, and if $\Delta = 0$ there is a cusp.

Su and Liu adapt the above results to the Bézier representation of the curve. Their approach is much the same as ours in that they choose a coordinate system, or equivalently an affine map, to fix six of the eight degrees of freedom in the equation of the curve (each control point is a position in the plane, and, therefore, embodies two degrees of freedom). Su and Liu map $\mathbf{B}_0$ to (1, 0), $\mathbf{B}_3$ to (0, 1), and the intersections of the lines $\mathbf{B}_0\mathbf{B}_1$ and $\mathbf{B}_2\mathbf{B}_3$ to (1, 1) as shown in Figure 2. This mapping effectively fixes the endpoints and the directions of the tangent vectors at $t = 0$ and $t = 1$; the remaining two parameters (U, V) are the lengths of the tangent vectors at the ends. The characterization of a cubic curve is, therefore, completely determined by the characteristic point $R = (U, V)$. The plane can be divided into regions for the segment of the curve corresponding to $0 \leq t \leq 1$ as shown in Figure 3.

There are a number of difficulties with the method of Su and Liu that we address in this paper. First, their method suffers from certain degeneracies that can be important in applications. For example, curves in which the first and last legs of the control polygon are parallel cannot be characterized by their construction since the point (1, 1) (and hence, (0, 0)) are not well defined in this case. Depending on the relative lengths of the legs, such curves can either have a loop, a cusp, or inflection points. Another (although less severe) degeneracy occurs when $\mathbf{B}_0$ and $\mathbf{B}_3$ are coincident. Second, the diagram of Figure 3 does not provide much intuition concerning the behavior of Bézier curves since the characteristic

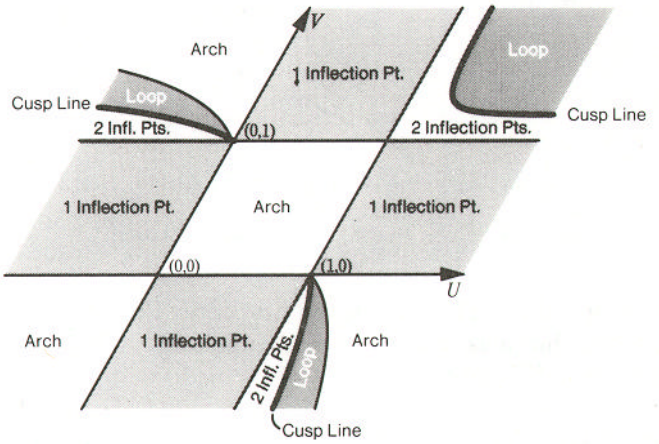

Fig. 3. Characterization diagram of Su and Liu for the trimmed curve. The label "Arch" indicates curves with no cusps, loops, or inflection points.

point must be constructed from the control points, and since the diagram has a fairly large number of disconnected regions. The diagrams produced by our method do not require the construction of an additional characteristic point. Furthermore, the particular diagram we construct using Bézier control points has fewer regions, has no ambiguous degeneracies, and is easier to interpret.

3. GEOMETRIC CHARACTERIZATION

In this section and the next, we describe the construction of a characterization diagram for Bézier curves. In Section 6 a much more general method is developed, and it is shown how all characterization diagrams are related.

The prescription for generating the characterization diagram is to select a basis, in this case the Bézier basis, and then choose a coordinate system, or equivalently an affine map, that fixes the coordinates of three of the four control points. The coefficients of the parametric equation 1 can then be expressed as linear combinations of the control points, three of which are constant; this is called the *canonical form* of the curve. As the fourth point varies, the curve may take on a loop, a cusp, or zero to two inflection points, depending on the position of the moving point. The plane can, therefore, be partitioned into regions labeled according to the characterization of the curve segment when the fourth point is in each region.

Note that it may not be possible to produce the canonical form. We defer discussing such degenerate cases until Sections 5 and 6.

The Bézier representation for a planar parametric cubic Q(t) is defined by four control points in the plane, $\mathbf{B}_0, \mathbf{B}_1, \mathbf{B}_2$, and $\mathbf{B}_3$. This representation is linearly related to the basic definition such that

$$\mathbf{Q}(t) = \sum_{i=0} \mathbf{B}_i \binom{3}{i} t^i (1-t)^{3-i}. \tag{3}$$

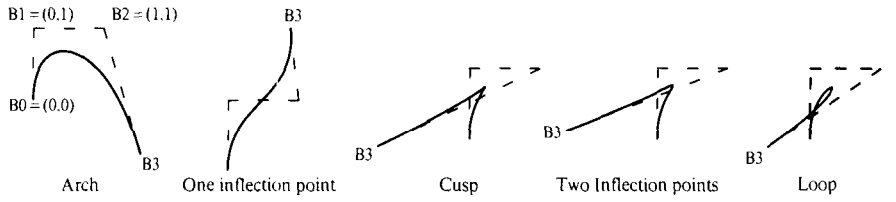

Fig. 4. Cubic Bézier curves in canonical form.

We choose a canonical form that maps the first three Bézier control points such that they fall along the sides of an isosceles triangle, $\mathbf{B}_0 = (0, 0)$, $\mathbf{B}_1 = (0, 1)$, and $\mathbf{B}_2 = (1, 1)$ as shown in Figure 4.

We then substitute the coordinates of B_0, B_1, and B_2 into Equation 3 to determine the $X(t)$ and $Y(t)$ component functions for the canonical curve:

$$X(t) = (B_{3x} - 3)t^3 + 3t^2 \tag{4}$$

$$Y(t) = B_{3y}t^3 - 3t^2 + 3t. \tag{5}$$

These equations can be used to compute the characterizing values A, B, and C defined in Equation 2:

$$A = 9(B_{3x} + B_{3y} - 3) \tag{6}$$

$$B = -9(B_{3x} - 3) \tag{7}$$

$$C = -9 \tag{8}$$

$$\Delta = B_{3x}^2 - 2B_{3x} + 4B_{3y} - 3 \tag{9}$$

Let us look first at the regions of the plane that characterize the untrimmed cubic curve. We then trim the curve and the diagram for the segment corresponding to $t \in [0, 1]$.

The untrimmed curve has a cusp if and only if $\Delta = 0$. For this canonical form, this equation describes a parabola, symmetric around $x = 1$, passing through the points $(-1, 0)$, $(1, 1)$, and $(3, 0)$. We call this curve *the cusp curve* since, if $\mathbf{B}_3$ is positioned somewhere along this curve, the parametric cubic has a cusp. The cusp curve divides the plane into two regions, one corresponding to loops and one corresponding to two inflection points. The equation $\Delta < 0$ is true in the region below the parabola, indicating that if $\mathbf{B}_3$ falls in this region the cubic has a loop. Therefore, if $\mathbf{B}_3$ falls in the region above the parabola, $\Delta > 0$ and the curve has two inflection points.

There are two other characteristics to discuss for the curve, both cases where the curve degenerates to a more constrained form. The first is the case in which the cubic terms of Equations 4 and 5 simultaneously vanish, implying that the cubic curve reduces to a quadratic curve, that is, a parabola. This occurs when $\mathbf{B}_3$ is located at $(3, 0)$. For this reason, we call this point the *parabolic point.*

The remaining case occurs when the parametric cubic becomes the graph of a cubic function, that is, one of the parametric equations can be reduced to a linear function. When this occurs, the degree of $F(t)$ drops from two to one, implying at most one root, and hence, at most one inflection point. This situation is

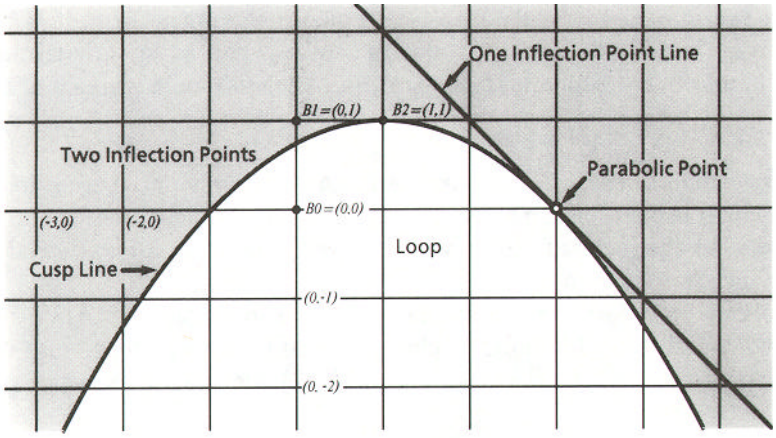

Fig. 5. The characterization diagram for the untrimmed Bézier cubic in canonical form.

indicated by the condition $A = 0$. On our diagram, this characteristic can therefore be described by the line:

$$x + y - 3 = 0 .$$

We call this line the *one inflection point line*. It is tangent to the cusp curve at the parabolic point (3, 0).

These results are summarized in Figure 5. For a Bézier cubic in canonical form, the position of B_3 on this diagram can be used to determine if the curve has a loop, a cusp, one or two inflection points, or is a parabola. If the parametric cubic is actually a straight line, all four control points are collinear and cannot be put into canonical form for this diagram.

4. THE REGIONS FOR THE CUBIC SEGMENT

The characterization diagram for the entire cubic curve was determined without restricting the domain parameter t. The diagram for the cubic segment is obtained by trimming away the portions of the regions that correspond to values of t outside the interval [0, 1]. We do this by individually trimming the cusp curve, the loop region, and the inflection point region.

To trim these regions, we use a number of geometric arguments which are specific to our choice of the Bézier representation. However, we show in Section 6 how these procedures are actually more general.

One particular geometric property we use that may not be familiar to some readers is the *variation diminishing* property [3,4]. Simply stated, it guarantees that every line intersects the control polygon in at least as many points as it intersects the curve itself.

4.1 Trimming the Cusp Curve

The portion of the cusp curve corresponding to $t \in [0, 1]$ is obtained by using the variation diminishing property. Since a cusp is a degenerate loop, a line passing through the cusp can intersect the curve three times, whereas a convex

polygon can only be cut twice by a line. For our canonical form, if $\mathbf{B}_3$ is on the cusp curve, the control polygon is always convex when $x > 1$. The trimmed cusp curve, therefore, corresponds to the portion of the parabola where $x \leq 1$.

4.2 Trimming the Loop Region

The loop region is trimmed by considering only the region where a loop (i.e., a double point) occurs for some value t in [0, 1]. The boundaries of the new region correspond to the locus of points where a double point occurs either at $t = 0$ or $t = 1$ as indicated in Figure 6.

The region boundary corresponding to a double point at $t = 0$ can be found by noting that such a double point is characterized by the existence of a value of t on the canonical curve $\mathbf{Q}(t)$ in the interval [0, 1] such that

$$\mathbf{Q}(t) - \mathbf{B}_0 = 0.$$

If we solve this equation for B_3, we get a parametric representation for the trimming curve which can then be reduced to algebraic form. We can also use a result from [12] which shows that this curve can be expressed as

$$B^2 - 3AC = 0$$

The result is a portion of the parabola described by

$$x^2 - 3x + 3y = 0, \qquad X < 0,\ Y < 0.$$

The region boundary corresponding to a double point at $t = 1$ can be determined in a similar fashion. In this case, we seek the existence of a value of t in [0, 1] such that $\mathbf{Q}(t) - \mathbf{B}_3 = 0$. The corresponding result from [12] is

$$(B + 2A)^2 - 3A(A + B + C).$$

The resulting curve is the ellipse:

$$x^2 + y^2 + xy - 3x = 0, \qquad 0 \leq X < 1,\ 0 \leq Y < 1.$$

4.3 Trimming the Inflection Point Region

An untrimmed cubic generally has either a loop or two inflection points (ignoring for the moment the cusp curve and the one inflection point line), whereas a cubic segment can have zero, one, or two inflection points, depending on whether the roots of $F(t)$ fall on the interval [0, 1]. We, therefore, expect to partition the inflection point region into three cases. Figure 7 is used to analyze the untrimmed inflection point region.

Region 0. *Zero Inflection Points.* If $\mathbf{B}_3$ lies in this region, then the control polygon is convex, so a line can intersect the control polygon in at most two points. If the curve segment had an inflection point, it would be possible to intersect the curve in three points, thereby violating the variation diminishing property. Region 0, therefore, corresponds to curves where both inflection points lie outside $t \in [0, 1]$.

Region 1. *One Inflection Point.* If $\mathbf{B}_3$ lies in this region, the signed curvature at $\mathbf{B}_0$ is negative, and the signed curvature at $\mathbf{B}_3$ is positive, implying that there are an odd number of roots in the interior of the curve segment. Since the curve

Fig. 6. Two curves showing loops appearing at (a) $t = 0$ and (b) $t = 1$.

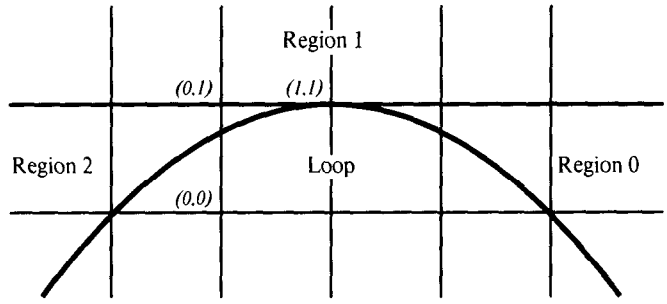

Fig. 7. The untrimmed diagram with the inflection point area broken into regions 0, 1, and 2, corresponding to the number of inflection points in the interval [0, 1].

can have a maximum of two inflection points, Region 1 corresponds to curves where exactly one inflection point occurs inside the curve segment.

Region 2. Two Inflection Points. Region 2 of Figure 7 corresponds to curves where exactly two inflection points occur inside the curve segment. If $\mathbf{B}_3$ lies in this region, then the signed curvature is negative at $\mathbf{B}_0$ and positive at $\mathbf{B}_3$, implying an even number of roots of $F(t)$ (possibly zero) in the interval $t \in [0, 1]$. However, it can be shown [10, 12] that at least one of the roots occurs in the interval [0, 1] for all points in Region 2, implying that there are at least two roots corresponding to two inflection points in the interval.

Figure 8 shows Figure 5 trimmed for the cubic segment defined by $t \in [0, 1]$. A region labeled "Arch" has been added to indicate curve segments that have no loops, cusps, or inflection points. The loop region has been trimmed to an irregular shape, and the inflection point region has been partitioned into regions of 0, 1, and 2 inflection points. The one inflection point line has disappeared. Referring to Figure 5, we can see that it crosses the regions that correspond to zero or one inflection point in the untrimmed diagram. Therefore, it is no longer necessary to distinguish it.

Some intuition for this diagram can be found by looking at the curve sequence in Figure 9. The sequence begins with a curve with $\mathbf{B}_3$ in the "Arch" region. As $\mathbf{B}_3$ crosses the $Y = 1$ line, it acquires an inflection point at $t = 1$. As $\mathbf{B}_3$ moves above the $Y = 1$ line, the inflection point moves into the curve. Moving $\mathbf{B}_3$ counterclockwise until it is once more on the $Y = 1$ line, the curve acquires a

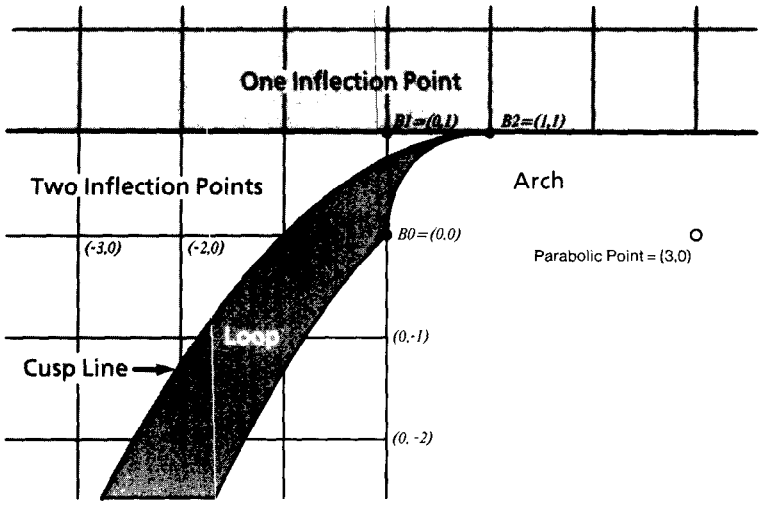

Fig. 8. The diagram shown in Figure 5 trimmed for the cubic segment defined by $t \in [0, 1]$.

second inflection point at $t = 1$. As $\mathbf{B}_3$ moves down toward the cusp line, the inflection points move together toward the point of high curvature, becoming coincident when $\mathbf{B}_3$ is on the cusp curve. Moving $\mathbf{B}_3$ below the cusp line, the curve develops a loop at the point of high curvature. This loop gets bigger as $\mathbf{B}_3$ moves to the right, opening back to an arched curve once $\mathbf{B}_3$ passes out of the loop region.

5. DEGENERATE FORMS

The use of our canonical form relies on the ability to map the two internal points plus one endpoint onto the isosceles triangle (0, 0), (0, 1), (1, 1). There are several cases in which this mapping is degenerate because the control points are collinear or coincident. The canonical form defines $\mathbf{B}_0 = (0, 0)$, $\mathbf{B}_1 = (0, 1)$, and $\mathbf{B}_2 = (1, 1)$ with $\mathbf{B}_3$ moving. If these three points are collinear but not collinear with $\mathbf{B}_3$, then the mapping $\mathbf{B}_3 = (0, 0)$, $\mathbf{B}_2 = (0, 1)$, and $\mathbf{B}_1 = (1, 1)$ with $\mathbf{B}_0$ moving produces identical results since the shape of a Bézier curve is unaffected if its control points are reversed. If all four control points are collinear but disjoint, the curve can be drawn as a straight line. However, the motion of t along this line varies with the order of the points along the line as shown in Figure 10. The curve is only parametrically a straight line if the four points are equally spaced.

If $\mathbf{B}_1 = \mathbf{B}_2$, the curve is a full cubic having an inflection point at each end. If $\mathbf{B}_1 = \mathbf{B}_2$ and this point equals an endpoint, the trace of the curve is a line, and the parameterization varies cubically with $\mathbf{Q}(t)$ moving monotonically along the curve. The curve has a zero length tangent vector and zero curvature at the end with the coincident control points. If $\mathbf{B}_0 = \mathbf{B}_1$ and $\mathbf{B}_2 = \mathbf{B}_3$, the trace of the curve is a line, the parameterization is still cubic with $\mathbf{Q}(t)$ moving monotonically along the curve, and the curve has a tangent vector of zero length at each end. Finally, if all four points are coincident, the curve collapses to a single point.

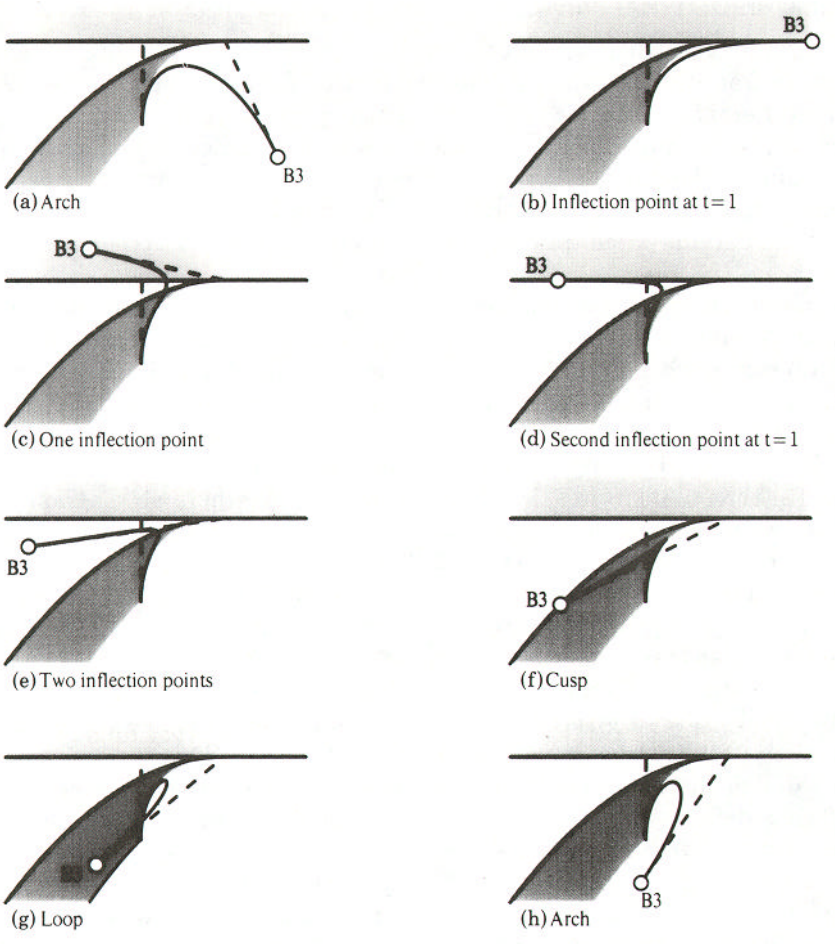

Fig. 9. A sequence of figures showing $\mathbf{B}_3$ moving around the trimmed diagram.

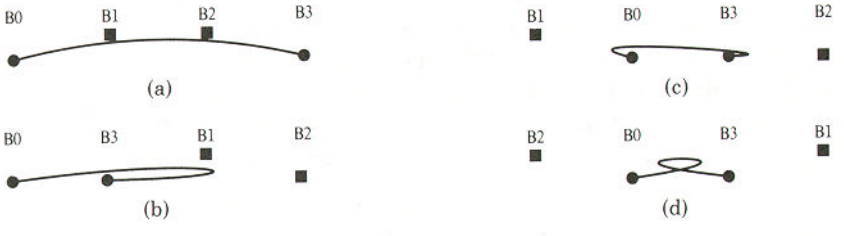

Fig. 10. Four cases for collinear cubics. Assume the control points converge to the horizontal line defined by the two endpoints. (a) Case (1), monotonic motion of t: (b) Case (2), one fold; t changes direction once; (c) Case (3), two folds; t changes direction twice; (d) Case (4), two folds; t changes direction three times.

This completes the description of our characterization diagram for Bézier curves. Further discussion of degenerate situations and a detailed description of how to implement this particular diagram can be found in [10]. The next section expands this result to include a large class of similar diagrams.

6. A MORE GENERAL VIEW

Su and Liu construct a characterization diagram that is quite different in structure from the characterization diagram of Figure 8. Yet another diagram with a different structure could be constructed by fixing $\mathbf{B}_0$, $\mathbf{B}_2$, $\mathbf{B}_3$ and letting $\mathbf{B}_1$ vary. The diagram obtained in this way would, as indicated by Figure 11, have a disconnected loop region. In addition, one could imagine many more variations, including the study of characterization diagrams for other curve types such as B-splines, Catmull–Rom splines, or Beta-splines. It is the purpose of this section to describe the precise relationship between all such diagrams. One of the principle results of this section, and indeed of the entire article, is that *all* region diagrams produced by fixing three control points and letting the remaining point vary, no matter what the curve type or which control point varies, can be obtained from a projective (that is, a rational linear) transformation of Figure 8.

Although characterization diagrams such as the ones in Figures 3 and 8 are two dimensional, the relationship between various diagrams only becomes apparent in three dimensions. We refer to the 3-space where all diagrams can be unified as *characterization space.* To see how characterization space is defined and used, we recall from Section 2 that the characterization of a curve $\mathbf{Q}$ can be completely determined from the function $F(t) = \det(\mathbf{Q}'(t), \mathbf{Q}''(t)) = At^2 + Bt + C$, or equivalently, from the numbers A, B, and C. Thus, we associate with every planar parametric cubic curve a point (A, B, C) in characterization space. The statement that the curve $\mathbf{Q}$ has a cusp if $B^2 - 4AC = 0$ is viewed geometrically as saying that the point in characterization space corresponding to $\mathbf{Q}$ is on the implicit surface defined by $B^2 - 4AC = 0$. This surface is a double cone through the origin called the *cusp cone.* Other surfaces of interest are the $t = 0$ *cone* defined by $B^2 - 3AC = 0$ indicating the presence of a loop at $t = 0$, the $t = 1$ *cone* defined by $(E + 2A)^2 - 3A(A + B + C) = 0$ indicating the presence of a loop at $t = 1$, and the *single inflection point plane* defined by $A = 0$ indicating the presence of only one inflection point on the untrimmed curve.

Based on these surfaces, regions of characterization space can be segmented into volumes corresponding to curve segments possessing loops, cusps, one or two inflection points, or none of the above. Since each of these surfaces is either a double cone or a plane, the resulting volumes are relatively simple.

Two-dimensional characterization diagrams are obtained by taking two-dimensional cuts through characterization space. For instance, the diagram of Figure 8 corresponds to taking a planar section through characterization space. More specifically, Equations 6, 7, and 8 show that as $\mathbf{B}_3$ varies, the point

$$(A, B, C) = (9(B_{3x} + B_{3y} - 3), -9(B_{3x} - 3), -9) \tag{10}$$

in characterization space traces out the plane $C = -9$; Figure 8 is, therefore, the result of intersecting the volumes in characterization space with the plane $C = -9$. Equivalently, the canonical plane of Figure 8 can be viewed as being embedded in characterization space, with the embedding given by eq. 10. The fact that the $C = -9$ plane is parallel to a side of the cusp cone explains why the cusp curve in Figure 8 is a parabola instead of some other conic section. Since the canonical plane never intersects the $C = 0$ plane, none of the curves where

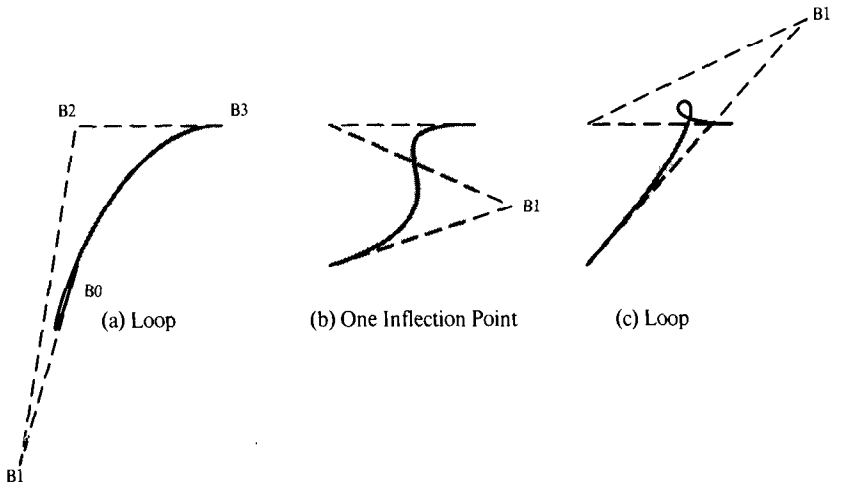

Fig. 11. $\mathbf{B}_0$, $\mathbf{B}_2$, and $\mathbf{B}_3$ are mapped to (0, 0), (0, 1), (1, 1). As $\mathbf{B}_1$ moves, we see that there are two disjoint loop regions. The gray region is the unit square.

$C = 0$ can be characterized by the diagram of Figure 8. These curves are precisely the curves that cause the degeneracies described in Section 5.

The diagram one would obtain by fixing $\mathbf{B}_0$, $\mathbf{B}_2$, and $\mathbf{B}_3$ and letting $\mathbf{B}_1$ vary would correspond to a different planar cut through characterization space. In particular, direct calculation for the dependence of A, B, and C on $\mathbf{B}_2$ shows that the embedding for this case is given by

$$(A, B, C) = (9(7B_{1x} + 2B_{1y} + 1), -9(2B_{1x} + B_{1y}), -9B_{1x}),$$

which is a parametric description for the plane $A + 2B + 7C - 9 = 0$, meaning that the diagram can be described by slicing this plane through characterization space. This plane cuts both branches of the cusp cone (the curve of intersection being a hyperbola), explaining why the loop regions indicated in Figure 11 are disconnected.

The fact that the characterization diagram produced by moving $\mathbf{B}_3$ and the characterization diagram produced by moving $\mathbf{B}_2$ correspond to planar slices through characterization space is not a phenomenon that is peculiar to Bézier curves. The diagram obtained by varying one coefficient in *any* polynomial basis corresponds to a planar cut. This implies, for instance, that the diagram produced when moving a point in a B-spline, Catmull–Rom, or Beta-spline representation of a curve has regions bounded by lines and conic sections (assuming that the knots of the B-spline and the shape parameters of the Beta-spline are held fixed). Moreover, the determination of which planar section of characterization space is appropriate is quite simple. Suppose that all but one of the control points $\mathbf{P}_i$ in eq. 1 is held fixed. To determine which planar section through characterization space to use, compute A, B, and C as a function of the coordinates of a moving point, thereby defining the embedding into characterization space. As we show

below, A, B, and C are guaranteed to be linear functions of the moving point, implying that the diagram in fact corresponds to a planar section.

From Equation 2, each of A, B, and C are of the form $\det(\alpha_k, \alpha_l)$ where $k \neq l$. These functions are, therefore, clearly linear if all but one of $\alpha_0, \ldots, \alpha_3$ are held fixed. To show that A, B, and C are linear if all but one control point is held fixed, we recall that the representation of a polynomial curve in one basis can be converted into a representation in any other basis by matrix multiplication (cf. [1]). If m_{ij} denote the coefficients of this matrix, then

$$\det(\alpha_k, \alpha_l) = \det\left(\sum_r m_{kr}\mathbf{P}_r, \sum_s m_{ls}\mathbf{P}_s\right).$$

The fact that determinants are linear in each of their arguments can be used to rewrite this as

$$\det(\alpha_k, \alpha_l) = \sum_r \sum_s m_{kr}m_{ls} \det(\mathbf{P}_r, \mathbf{P}_s). \tag{11}$$

This shows that if all but one of the $\mathbf{P}_i$ are held fixed, then each of the terms in the double summation of eq. 11 is either a constant or a linear function, thereby showing that A, B, and C are linear functions.

The plane $A - B + 2C + 1$ yields a particularly revealing section through characterization space as shown in Figure 12. This plane intersects the cusp cone, the $t = 0$ cone, and the $t = 1$ cone in ellipses that show the special nature of these surfaces. The cusp curve meets each of the other two curves tangentially at two points, and all three curves meet tangentially at the parabolic point. The common tangent line at the parabolic point corresponds to $A = 0$, the tangent line at the other point of intersection between the cusp curve and the $t = 0$ trim curve corresponds to $C = 0$, and the tangent line at the other point of intersection between the cusp curve and the $t = 1$ trim curve corresponds to $A + B + C = 0$. Finally, the $\mathbf{B} = 0$ line connects the two points of intersection between the cusp curve and the $t = 0$ curve, and the $B + 2A = 0$ line connects the two points of intersection between the cusp curve and the $t = 1$ curve.

Remark. Mathematical purists familiar with projective geometry (cf. [7] and [8]) may have noticed that since the defining equations for the cusp cone, the $t = 0$ cone and the $t = 1$ cone are homogeneous polynomials of degree two, and characterization space is more naturally thought of as a projective 2-space rather than as a linear 3-space. That is, for characterization purposes, a curve $\mathbf{Q}(t)$ can be placed in correspondence with the point in projective 2-space having homogeneous coordinates $[A, B, C]$ where A, B, and C are given by $\det(\mathbf{Q}'(t), \mathbf{Q}''(t)) = At^2 + Bt + C$.

In this context, planar sections through 3-space are interpreted as choosing an affine subspace by sending one of the projective lines to infinity. For the diagram of Figure 8, for instance, the line that has been sent to infinity is the $C = 0$ line. As was demonstrated earlier, curves in correspondence with the projective points of this line could not be characterized by the diagram. This is not only a problem with our canonical form, it is a problem with *every* canonical form obtained by fixing three control points and moving one. This deficiency is a topological consequence of the fact that these methods correspond to affine subsets—they

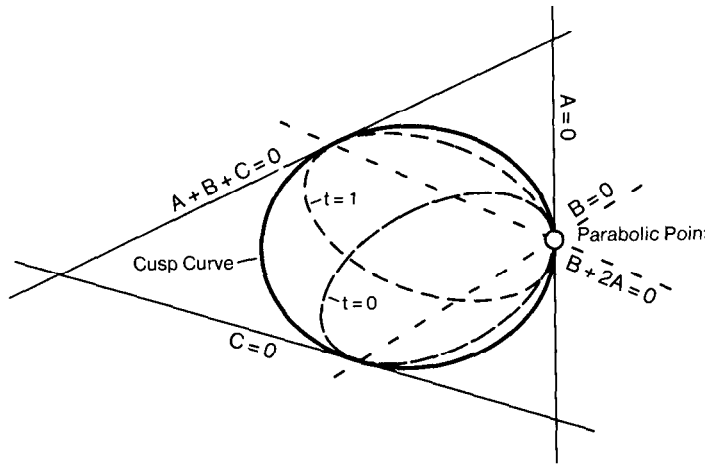

Fig. 12. A planar section through characterization space showing the critical curves, lines, and tangencies.

must therefore send some line to infinity and lose the ability to characterize the associated curves. The projective setting also makes it clear that every such diagram can be created via a projective transformation of Figure 12. This follows from the fact that every pair of affine subsets is related by a projective transformation in much the same way that every pair of bases are related by a linear transformation.

Interestingly, the characterization diagram of Su and Liu (Figure 3) does not correspond to a planar cut through characterization space, but it does correspond to a *bilinear* one. Wang shows that with a suitable choice of coordinates,

$$A = 3(UV - U - V)$$
$$B = 3U(2 - V)$$
$$C = U(V - 3),$$

where U and V are as indicated in Figure 2. The embedding of Su and Liu's (U, V) plane into characterization space is therefore given by

$$(A, B, C) = (3(UV - U - V), 3U(2 - V), U(V - 3)),$$

which describes a bilinear surface that maps the corners of the unit square in (U, V) space into the points $(0, 0, 0)$, $(-2, 4, -3)$, $(-2, 0, 0)$, and $(-2, 2, 2)$. The nonlinear nature of this surface explains why the characterization diagram of Su and Liu (Figure 3) is not a projective transformation of the characterization diagram of Figure 8.

7. CONCLUSIONS

In this paper, we have developed a simple geometric method for characterizing cubic curves based on the positions of the control points. The method is based on the idea that three of the four points can be fixed and the shape determined

by analyzing the position of the fourth point. Depending on the position of the fourth point, the curve can possess a loop, a cusp, or up to two inflection points. The presence of these features is determined by partitioning the plane into regions based on the position of the fourth point. A specific example is given for Bézier curves. The resulting characterization diagram is seen to be simpler and more instructive than the previous diagram of [11].

Of course, there is a large choice of which point to move and which representation of the curve to use, each leading to a different characterization diagram. It is therefore important to understand at a higher level how these diagrams are related. We have answered this question by showing that all characterization diagrams result from exacting two-dimensional slices from a common, universal three-dimensional space that we call characterization space. It was further shown that the diagrams obtained by fixing all but one coefficient in any polynomial basis corresponds to a simple planar section through characterization space. Figure 8 is an example of such a planar slice, corresponding to a plane parallel to one of the axes through characterization space. Finally, it was shown that the increased complexity of the diagram of Su and Liu (Figure 3) was a result of the fact that it corresponds to a nonplanar slice through characterization space.

ACKNOWLEDGMENTS

Much of the intuition for this work was achieved by playing interactively with a Bézier cubic segment in the Gargoyle interactive illustration system at Xerox PARC [2, 9]. Furthermore, we were able to programatically construct Gargoyle objects for each of the boundary curves and interactively test the diagram for the trimmed curve. We also used Gargoyle to explore Su and Liu's diagram. We believe that without this interaction we could not have progressed to our results as quickly and as confidently. Gargoyle was also used to produce all of the figures. Our motivation to work on this problem at this time was originally inspired by Jules Bloomenthal's need for a method to define tree branches without loops in them. Jules was also able to provide us with a glimpse of the three cones in characterization space and verify their shape and relationship to one another. Finally, this work has been a pleasant and evenly balanced collaboration, and the order of the authors names should be considered reverse-alphabetical.

REFERENCES

1. BARTELS, R. H., BEATTY, J. C., AND BARSKY, B. A. *An Introduction to Splines For Use in Computer Graphics and Geometric Modeling.* Morgan-Kaufmann, Palo Alto, Calif. 1987.
2. BIER, E., AND STONE, M. Snap-dragging. *Comput. Graph. 20*, 4 (Aug. 1986), 233–240.
3. BOEHM, W., FARIN, G., AND KAHMANN, J. A survey of curve and surface methods in CAGD. *Comput. Aided Geom. Des. 1*, 1 (July 1984), 1–60.
4. FARIN, G. *Curves and Surfaces for Computer Aided Geometric Design.* Academic Press, Orlando, Fla. 1987.
5. FORREST, A. R. Shape classification of the non-rational twisted cubic curve in terms of Bézier polygons. CAD Group Document No. 52, University of Cambridge, Cambridge, England, Dec. 1970.
6. FORREST, A. R. The twisted cubic curve: A computer-aided geometric design approach. *Comput. Aided Des. 12*, 4 (July 1980), 165–172.
7. PATTERSON, R. R. Parametric cubics as algebraic curves. *Comput. Aided Geom. Des. 5* (1988), 139–159.

8. PENNA, M. A., AND PATTERSON, R. R. *Projective Geometry and Its Applications to Computer Graphics*. Prentice-Hall, Englewood Cliffs, NJ, 1986.
9. PIER, K., BIER, E., AND STONE, M. An Introduction to Gargoyle: An interactive illustration tool. In *Document Manipulation-Typography, Proceedings of the International Conference on Electronic Publishing, Document Manipulation-Typography (EP88)* (Nice, France, Apr. 20–22, 1988). Cambridge University Press, 1988, pp. 223–238.
10. STONE, M. C., AND DEROSE, T. D. Characterizing cubic Bézier curves. Tech. Rep. EDL-88-8, Xerox Palo Alto Research Center, Palo Alto, Calif., Dec. 1988.
11. SU, B., AND LIU, D. An affine invarient and its application in computational geometry. *Scientia Sinica (Series A) 24*, 3 (Mar. 1983), 259–267.
12. WANG, C. Y. Shape classification of the parametric cubic curve and parametric B-spline cubic curve. *Comput. Aided Des. 13*, 4 (1981), 199–206.

Received June 1988: revised and accepted February 1989