2
Curve and Surface Terms and Properties

Minimum variation curves, networks, and surfaces are designed to solve many of the problems in CAGD. In this chapter we introduce terms and identify properties necessary for an understanding of these problems and necessary to differentiate and evaluate the many solutions found in the literature.

Curves and surfaces for both design and approximation are characterized by those properties and attributes that affect their utility when applied to a particular problem. We will focus on the properties of curves and will include a discussion of surfaces only when necessary to make important distinctions between the properties of curves and surfaces. In the sections that follow we identify and define the terms associated with these characteristics:

2.1 Geometric Characterization of Curves

In describing the geometric character of a curve we consider four quantities: *position, tangent* $\hat{t}$*, curvature* κ, and *torsion* τ. For the purposes of this discussion, we will assume that our curve is described using a differentiable vector valued function,

$$\vec{C}(u) = \{x(u), y(u), z(u)\} \qquad u \in [a, b], \tag{2.1}$$

where u is an arbitrary curve parameter with the restriction that $\|\vec{C}'(u)\| \neq 0$ for all u, this is called a *regular parameterization.* We will also assume an alternate parameterization resulting from the inversion of $s(u)$ in equation (2.2). This results in what is called an *arc length parameterization.*

$$s(u) = \int_a^u \|\vec{C}'(u)\| du \tag{2.2}$$

Under this parameterization $\|\vec{C}'(s)\| = 1$ everywhere. We will refer to $\vec{C}(u)$ when discussing a curve with arbitrary regular parameterization, and refer to $\vec{C}(s)$ when discussing an arc length parameterized curve. Finally, we refer to a curve that is actually the graph of a function as $y = f(x)$.

To simplify the following discussion, we describe a sliding orthonormal coordinate system defined at each point on the curve. This trihedral frame is named the *Frenet frame* [52, p19]. The frame is composed of the curve's *tangent* $\hat{t}$*, normal* $\hat{n}$*,* and *binormal* $\hat{b}$, each of which is defined as follows. The tangent $\hat{t}(u)$ at a point $\vec{C}(u)$ is the direction of the curve at $\vec{C}(u)$ (Fig. 2.1). In terms of a parametric curve, the tangent is in the direction of the first derivative of the curve,

$$\hat{t}(s) = \vec{C}'(s)$$
$$\hat{t}(u) = \frac{\vec{C}'(u)}{\|\vec{C}'(u)\|}. \tag{2.3}$$

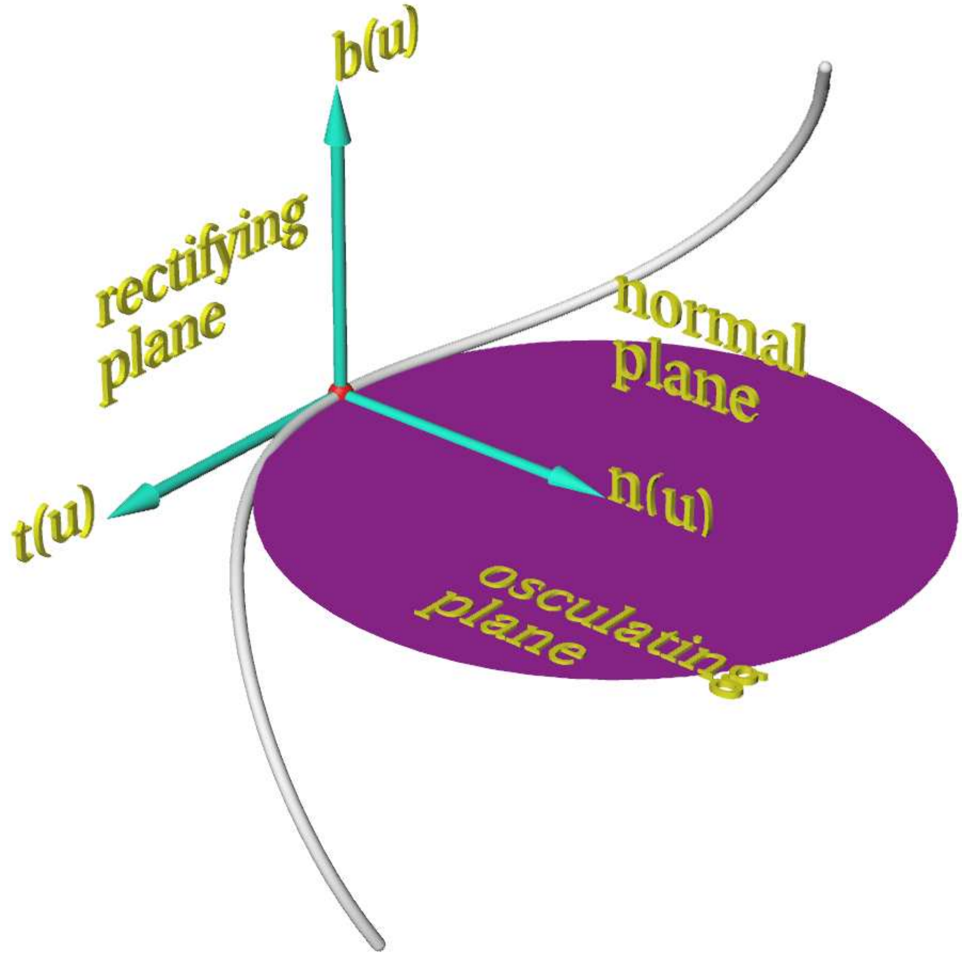


Figure 2.1. The Frenet Frame.

The frame is made up of the curve tangent, normal, and binormal vectors. These vectors, in turn, define a set of three planes: the osculating plane, the normal plane, and the rectifying plane.

The binormal to the curve at a point $\vec{C}(u)$ is the perpendicular to the plane the curve lies in at $\vec{C}(u)$ (Fig. 2.1). The binormal is perpendicular to the first and second derivatives,

$$\hat{b}(s) = \vec{C}'(s) \times \vec{C}''(s)$$

$$\hat{b}(u) = \frac{\vec{C}'(u) \times \vec{C}''(u)}{\left\| \vec{C}'(u) \times \vec{C}''(u) \right\|}. \qquad \textbf{(2.4)}$$

The normal to the curve is perpendicular to the tangent and binormal,

$$\hat{n}(s) = \hat{b}(s) \times \hat{t}(s)$$

$$\hat{n}(u) = \hat{b}(u) \times \hat{t}(u). \tag{2.5}$$

The coordinate frame we have defined spans three planes: the *osculating, normal,* and *rectifying* planes. The osculating plane is spanned by the tangent and normal, and is named for the circle it contains that “kisses” the curve (Fig. 2.1). The normal plane is spanned by the normal and binormal. The rectifying plane is spanned by the tangent and binormal, and is named for the fact that as it moves along a curve it sweeps out a rectifying developable surface. When this surface is “rolled out” flat on a plane, the curve that was used to generate it forms a straight line.

The definitions of curvature and torsion are central to a discussion of a curve’s shape. Given the functions $\kappa(s)$ and $\tau(s)$ a unique curve shape is defined. With respect to the coordinate frame we just defined, curvature and torsion describe the rate at which the trihedral frame rotates. These quantities are thus measures of how quickly the curve is turning or bending. Curvature is an instantaneous measure of how much the curve is bending in the osculating plane away from the tangent direction. Torsion is an instantaneous measure of how much the curve is bending away from or out of the osculating plane, n.b. a curve with $\tau(s) = 0$ is planar. The *Frenet-Serret* formulas are a direct result of these definitions [52, p.19]:

$$\hat{t}'(s) = \kappa(s)\hat{n}(s)$$

$$\hat{b}'(s) = \tau(s)\hat{n}(s)$$

$$\hat{n}'(s) = -\kappa(s)\hat{t}(s) - \tau(s)\hat{b}(s). \tag{2.6}$$

Curvature is equal to the reciprocal of the *radius of curvature*. The radius of curvature is the radius of the osculating circle (Fig. 2.1). The center of the circle may also be established by finding the intersection of curve normal directions as they converge on the point $\vec{C}(u)$. The formulas for curvature are:

$$\kappa(s)\hat{n}(s) = \vec{C}''(s)$$

$$\kappa(u)\hat{n}(u) = \frac{(\vec{C}'(u) \times \vec{C}''(u)) \times \vec{C}'(u)}{\left\| \vec{C}'(u) \right\|^4}. \tag{2.7}$$

Note that when a curve is arc length parameterized, the second derivative is equal to curvature, because of this, the second derivative $\vec{C}''(u)$ is often used as an approximation to curvature. Similarly, the formulas for torsion are

$$\tau(s) = -\left\| \vec{C}'(s) \times \vec{C}'''(s) \right\|$$

$$\tau(u) = -\frac{\det\,[\vec{C}'(u), \vec{C}''(u), \vec{C}'''(u)]}{\left\| \vec{C}'(u) \times \vec{C}''(u) \right\|^2} . \qquad \textbf{(2.8)}$$

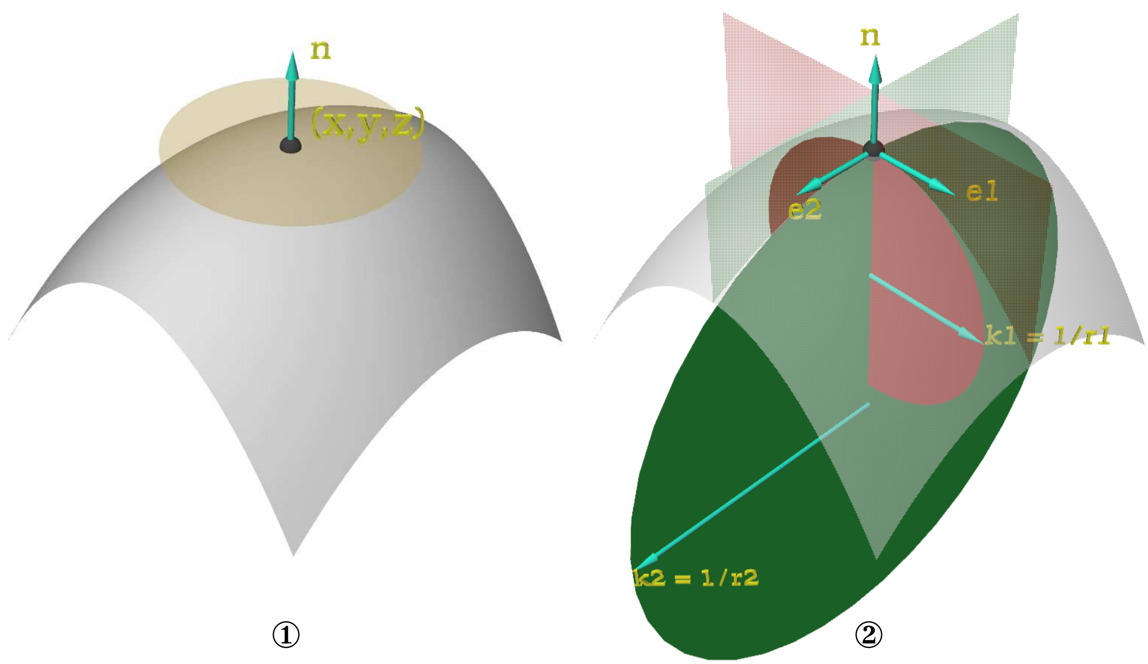


Figure 2.2. Surface Geometry.

A surface is locally characterized by ① position, surface normal, and ② principal directions and principal curvatures.

2.2 Geometric Characterization of Surfaces

In describing the geometric character of a surface we consider three quantities: *position*, *surface normal* $\hat{n}$ (a vector perpendicular to the tangent plane of the surface), and the *principal directions* $\hat{e}_1, \hat{e}_2$ and *principal curvatures* $\hat{e}_1, \hat{e}_2$, (Fig. 2.2). We assume that our surface is described using a differentiable vector valued function of two variables,

$$\vec{S}(u, v) = \{x(u, v), y(u, v), z(u, v)\} \qquad \{u, v\} \in [a, b]\,, \tag{2.9}$$

where u and v are arbitrary surface parameters with the restriction that $\left\|\vec{S}_u(u, v) \times \vec{S}_v(u, v)\right\| \neq 0$ for all u and v. In other words, the partial derivatives of $\vec{S}(u, v)$ must neither become colinear nor vanish. As in the discussion of curves, this is called a regular parameterization. Surfaces have no simple canonical parameterization analogous to a curve's arc length parameterization.

The normal to a surface is computed from the cross product of the first partial derivatives of the surface,

$$\hat{n} = \frac{\vec{S}_u(u, v) \times \vec{S}_v(u, v)}{\left\|\vec{S}_u(u, v) \times \vec{S}_v(u, v)\right\|}.$$

In order to describe the principal directions and principal curvatures of a surface, we must first define *normal curvature.* The *normal curvature* at a point on a surface in a direction specified by a surface tangent vector is determined from the intersection curve of the surface with the plane spanned by the surface normal and the given tangent vector. The principal directions, $\hat{e}_1$ and $\hat{e}_2$, and the principal curvatures, κ_1 and κ_2, at a point on a surface are the directions and magnitudes of the minimum and maximum of all possible normal curvatures at that point. The principal directions and curvatures are computed from the first and second fundamental forms from differential geometry [52],

$$E = \vec{S}_u \cdot \vec{S}_u \; F = \vec{S}_u \cdot \vec{S}_v \; G = \vec{S}_v \cdot \vec{S}_v .$$

and

$$e = \hat{n} \cdot \vec{S}_{uu} f = \hat{n} \cdot \vec{S}_{uv} g = \hat{n} \cdot \vec{S}_{vv} .$$

respectively. Specifically, the principal curvatures are the eigenvalues of the curvature tensor. The expression for the curvature tensor is

$$\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}, \tag{2.10}$$

where

$$
\begin{aligned}
a_{11} &= \frac{fF - eG}{EG - F^2} & a_{21} &= \frac{eF - fE}{EG - F^2} \\
a_{12} &= \frac{gF - fG}{EG - F^2} & a_{22} &= \frac{fF - gE}{EG - F^2}
\end{aligned}
.
$$

The principal directions are the eigenvectors of (2.10) relative to the basis $\{\vec{S}_u, \vec{S}_v\}$.

2.3 Specification

A key aspect of a curve is how it is defined, i.e., what information the designer must provide to fully define the curve's shape. Two major classes of curves are *interpolating curves* and *approximating curves.* Interpolating curves are defined by a sequence of positions that the curve must pass through (Fig. 2.3①). Approximating curves do not necessarily pass through their defining points, however, the curves *do* reflect the shape formed by the sequence of defining *control points* (Fig. 2.3②).

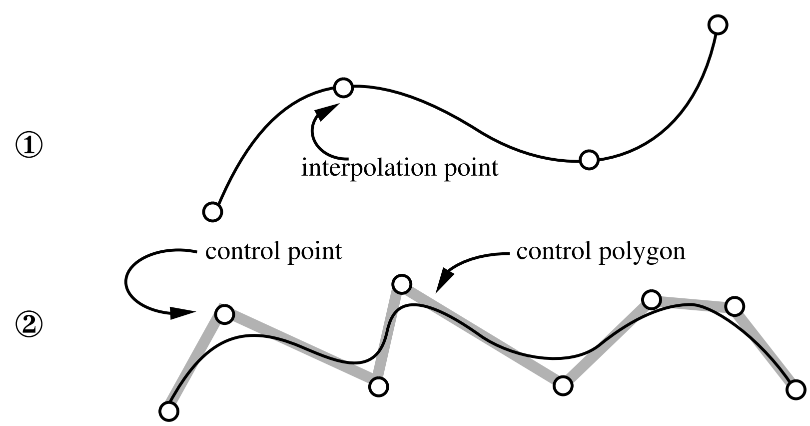


Figure 2.3. Interpolating and Approximating Curve.

An interpolating curve passes through its defining points. An approximating curve passes near or through its defining *control points*. The collection of line segments connecting the control points is referred to as the *control polygon.*

Curves shapes based on user provided positions represent the simplest form of specification. Implicit in the specification of a sequence of points is their ordering; a curve designer may also be required to specify the parametric spacing of the points along the curve, their *parameterization*. In addition to interpolated points, curve definitions may

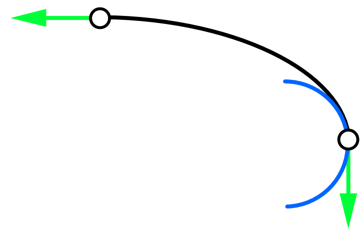


Figure 2.4. An Interpolating Curve.
A curve specified by its end positions, tangent directions, and one curvature value.

include higher order geometric specifications such as *tangent* direction, *curvature*, and *torsion* (Fig. 2.4). Curves may also be defined by positions augmented with *parametric derivatives*, with respect to the independent curve parameter. Such specifications concern the method used to represent the curve rather than the intrinsic geometry of the curve. Finally, curves may have associated *shape handle*s, parameters which are neither geometric nor parametric in nature, but control the shape of the curve in a predictable fashion. *Tension* [193] is a typical shape handle, an increase in tension causes the curve to decrease in fullness, as though the curve were being pulled more tightly to its defining control polygon (Fig. 2.5).

2.4 Computational Form

Curves may be represented and computed using a variety of techniques and forms. The three basic forms are *implicit, parametric,* and *explicit*. For example, a unit circle centered at the origin may be defined in each of these forms.

1. Implicit:

   $$x^2 + y^2 - 1 = 0$$

   The implicit form is particularly useful for point classification (determining whether a point is inside, on, or outside a closed curve). In general, there is no direct method for computing a sequence of points that lie on an implicit curve.

2. Explicit:

   $$y = \sqrt{1 - x^2}$$
   $$y = -\sqrt{1 - x^2}$$
   $$-1 \le x \le 1$$

   The explicit form, the simplest of curve forms, is useful for curves representing functions. In the case of the circle, two separate equations are required because it is multi-valued. The explicit form is commonly used in the approximation of functional data.

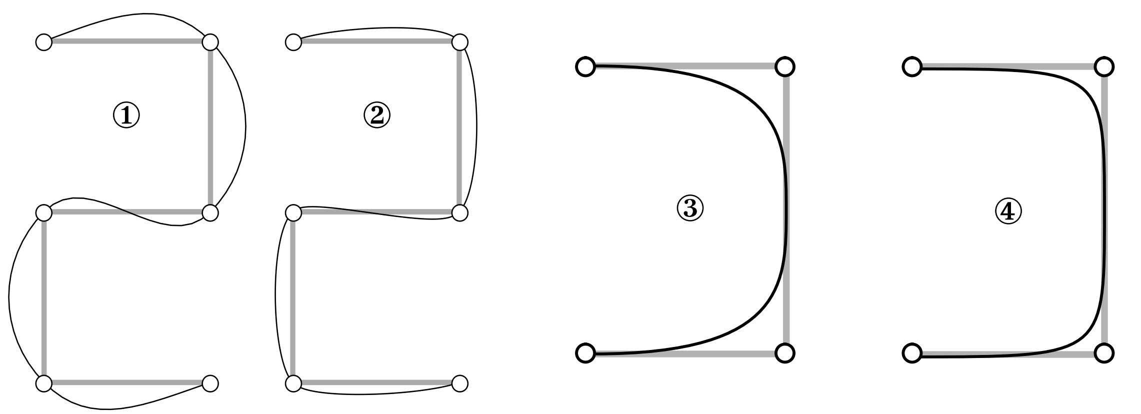


Figure 2.5. Increasing Tension in a Curve.

Tension is varied from a low value (①, ③) to a high value (②, ④) causing the curve to follow its defining control polygon more closely. (①,②) interpolating curve, (③, ④) approximating curve.

3. Parametric:

$$x = \sin u \qquad\qquad x = \frac{1-u^2}{1+u^2}$$
$$y = \cos u \qquad \text{or} \qquad y = \frac{2u}{1+u^2}$$
$$-\pi \le u \le \pi \qquad\qquad -\infty \le u \le \infty$$

Finally, the parametric form, sometimes referred to as *vector-valued,* provides a simple mechanism for the representation of curves. Points on the curve are computed as a vector valued function of an independent parameter. Because of their flexibility and the ease with which sequences of points on a curve may be calculated, the parametric form has emerged as the favored approach for CAD applications.

Curves may be further classified by the type of functions which they employ. Parametric curves are defined using *polynomial*, *rational polynomial*, *trigonometric*, and *exponential* functions. Polynomial and rational polynomial functions are preferred because of the ease with which they may be computed, only requiring the elementary operations on real numbers, addition and multiplication for plain polynomials, and division for rational polynomials. Further, polynomials are easily differentiated, and at least theoretically, may always be symbolically integrated.

Finally, curve generation techniques may be classified based on their ability to interpolate or approximate *multi-valued* data. A curve that only supports *function* interpolation and approximation is referred to as *single-valued* (Fig. 2.6). *Function interpolants* are *single-valued,* while *curve interpolants* can be *multi-valued*.

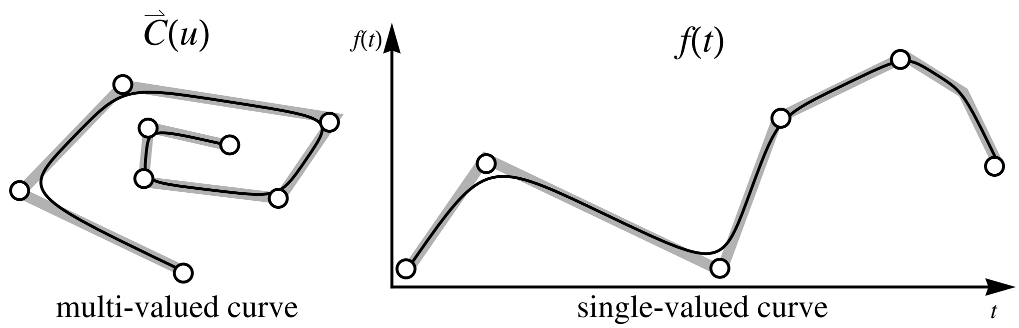


Figure 2.6. Multi- and single-valued curves.

2.5 Continuity

Because of the limited descriptive power of a single parametric polynomial, multiple parametric polynomials may have to be pieced together to form a more complicated curve. The resulting curve is referred to as *piecewise polynomial*. In order to provide guarantees concerning the smoothness of such a composite curve, the notion of *continuity* has been introduced. Continuity refers to the smoothness of the joints between adjacent polynomial pieces. Consider the piecewise polynomial curve that is globally parameterized from 0 to 3 (Fig. 2.7). The curve is made up of three polynomial pieces or *segments*. In order for the curve to be considered to be first order parameter continuous (C^1), the first derivatives of the segments must be equal at the joints:

$$\vec{c}'_0(1) = \vec{c}'_1(0) \qquad \text{and} \qquad \vec{c}'_1(1) = \vec{c}'_2(0). \tag{2.11}$$

In order for the curve to be second order parameter continuous (C^2), both the first and second derivatives have to be equal:

$$\begin{aligned} &\vec{c}'_0(1) = \vec{c}'_1(0) \qquad \vec{c}'_1(1) = \vec{c}'_2(0) \\ &\vec{c}''_0(1) = \vec{c}''_1(0) \qquad \vec{c}''_1(1) = \vec{c}''_2(0). \end{aligned} \tag{2.12}$$

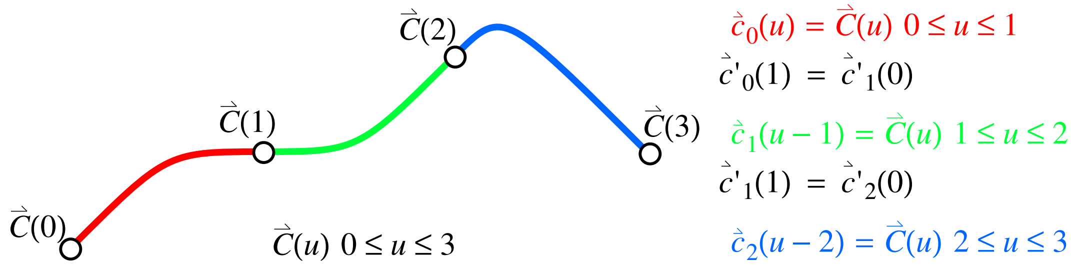


Figure 2.7. A C^1 Piecewise Parametric Curve.
Curve segments join so that the first derivatives are equal at the segment-segment joints.

Similarly, to achieve *n*th order parameter continuity (C^n), the first n derivatives of the curve segments must be continuous at the segment-segment joints. For a complete discussion of curve and surface continuity see [47].

A less restrictive form of continuity is that of *geometric continuity,* denoted G^n for *n*th order geometric continuity. It stipulates that curve segments can be made to meet with parametric continuity under *some* suitably chosen parameterization, i.e., it must be possible to reparameterize adjacent segments such that they meet with parametric continuity. Examining the case of G^1 continuity, unit tangent vector continuity is necessary and sufficient for first order geometric continuity to be achieved. For G^2 continuity, the curve segments must have the same curvature at the junction in addition to tangent continuity. A result of geometric continuity is the availability of extra degrees of freedom which are useful for controlling curve shape.

Consider a curve made up of cubic Hermite segments. A cubic Hermite segment [59] is specified by the positions and first parametric derivatives at its end points, $\vec{c}_i(0), \vec{c}_i(1), \vec{c}'_i(0), \vec{c}'_i(1)$. To form a C^1 continuous curve, the positions and first derivatives must be explicitly shared at segment-segment joints. To form a G^1 continuous curve it is only necessary that the segments share tangent direction, $\hat{t}_i(u)$, at these joints (Fig. 2.8). To accomplish this and to take advantage of all the available degrees of freedom, we redefine the cubic Hermite segment to be defined by the positions, tangent directions, and first derivative magnitudes at its end points,

$$\vec{c}_i(0), \vec{c}_i(1), \hat{t}_i(0), \hat{t}_i(1), m_i(0), m_i(1). \tag{2.13}$$

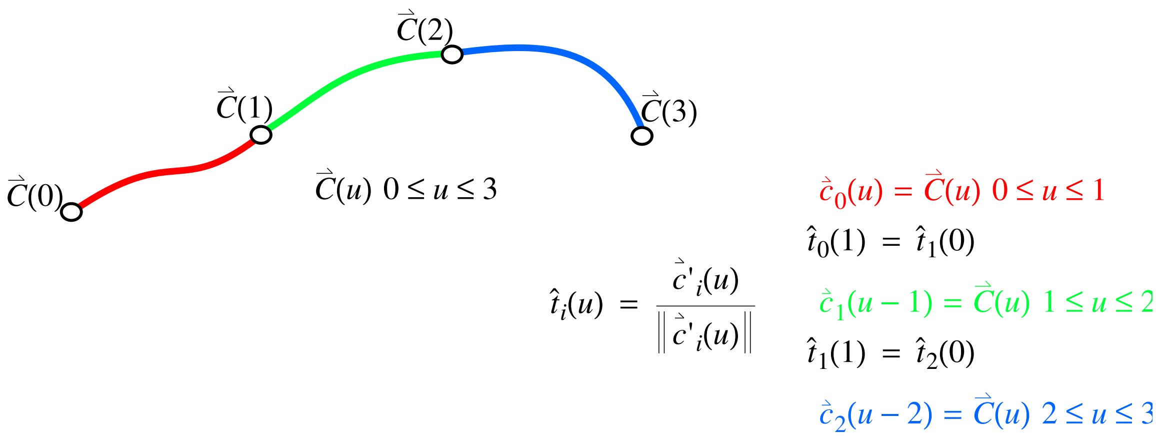


Figure 2.8. A G^1 Piecewise Parametric Curve.

Curve segments join such that the tangent directions are equal at the segment-segment joints.

To map to the traditional Hermite form, we scale the tangent direction by the associated magnitude producing the appropriate derivative,

$$\vec{c}\,'_i(0) = m_i(0)\hat{t}_i(0) \qquad \vec{c}\,'_i(1) = m_i(1)\hat{t}_i(1). \tag{2.14}$$

The use of geometric continuity has produced two extra degrees of freedom per curve segment, $m_i(0), m_i(1)$. These degrees of freedom may be used as shape handles to further control the shape of the curve without losing geometric continuity.

As a more concrete measure of the descriptive power of curves exhibiting geometric versus parametric continuity, we examined the relative degree of the parametrically continuous curve required to exactly reproduce a geometrically continuous curve. We found that a degree dk, C^k curve is required to reproduce the shape of a degree d, G^k curve. Consider a G^k curve composed of curve segments $\vec{C}_{0\ldots n}(u)$, we may assume without loss of generality that each segment is parameterized with $u \in [0, 1]$. Starting at the first joint, between $\vec{C}_0(u)$ and $\vec{C}_1(u)$, we fix the parameterization of the first segment and reparameterize the second (producing $\vec{C}'_1(u)$) such that their derivatives match at the joint where they meet:

$$\vec{C}_0^{(1)}(1) = \vec{C}'^{(1)}_1(0)$$
$$\vec{C}_0^{(2)}(1) = \vec{C}'^{(2)}_1(0)$$
$$\ldots$$
$$\vec{C}_0^{(k)}(1) = \vec{C}'^{(k)}_1(0)$$

The required reparameterization is of the form

$$u = \sum_{i=1}^{k} \beta_i u'^i; \qquad \textbf{(2.15)}$$

u is replaced by a degree k polynomial in u', the parameter of the reparameterized curve. Since $\vec{C}_1(u)$ is a degree d polynomial, replacing u with a degree k polynomial, creates a degree dk polynomial, $\vec{C}'_1(u')$. The coefficients of (2.15) are calculated by making the substitution and differentiating $\vec{C}_1(u')$;

$$\vec{C}_0^{(1)}(1) = \beta_1 \vec{C}_1^{(1)}(0)$$
$$\vec{C}_0^{(2)}(1) = \beta_1^2 \vec{C}_1^{(2)}(0) + 2\beta_2 \vec{C}_1^{(1)}(0)$$
$$\vec{C}_0^{(3)}(1) = \beta_1^3 \vec{C}_1^{(3)}(0) + 6\beta_1\beta_2 \vec{C}_1^{(2)}(0) + 6\beta_3 \vec{C}_1^{(1)}(0) \qquad \textbf{(2.16)}$$

$$\ldots$$

Using this system of equations, the β_i are easily found. Since $\vec{C}'_1(u')$ is reparameterized, the range of u' is now $0 \ldots x$, where $1 = \sum_{i=1}^{k} \beta_i x^i$. Continuity between segments $\vec{C}'_1(u')$ and $\vec{C}_2(u)$ is similarly established

$$\vec{C}'^{(1)}_1(x) = \beta_1 \vec{C}_2^{(1)}(0)$$
$$\vec{C}'^{(2)}_1(x) = \beta_1^2 \vec{C}_2^{(2)}(0) + 2\beta_2 \vec{C}_2^{(1)}(0)$$
$$\vec{C}'^{(3)}_1(x) = \beta_1^3 \vec{C}_2^{(3)}(0) + 6\beta_1\beta_2 \vec{C}_2^{(2)}(0) + 6\beta_3 \vec{C}_2^{(1)}(0) \qquad \textbf{(2.17)}$$

$$\ldots$$

Continuing this process creates a series of curve segments that meet with C^k parametric continuity and which are each parameterized from $0 \ldots x_i$. A single unified parameterization can be established by further substitution without affecting the degree of the result.

Schoenberg [191] introduces a specific instance of a piecewise polynomial function, the *spline*. While his definition is of a scalar spline function used for approximation, we consider a vector valued spline curve:

$$U = (u_o, u_1, \ldots, u_n) \qquad u_i \in \Re, u_i \le u_{i+1}, i = 0, \ldots, n-1 \qquad \textbf{(2.18)}$$

U is called a *knot vector* and is composed of *knot values* u_i. A vector valued function $\vec{C}(u)$ of degree d is a spline function if it satisfies two conditions. First, the segments $\vec{C}_i(u)$ of which it is composed, are each parameterized $u \in [u_i, u_{i+1}]$, and are polynomials of degree d. Second, $\vec{C}(u)$ and its derivatives $\vec{C}'(u), \vec{C}''(u), \ldots, \vec{C}^{(d-1)}(u)$ are continuous, $\vec{C}(u) \in C^{d-1}$. The notion of a spline curve is used extensively in curve design and has become somewhat less strictly defined as a curve composed of $C\infty$ segments meeting with specified continuity, either parametric or geometric.

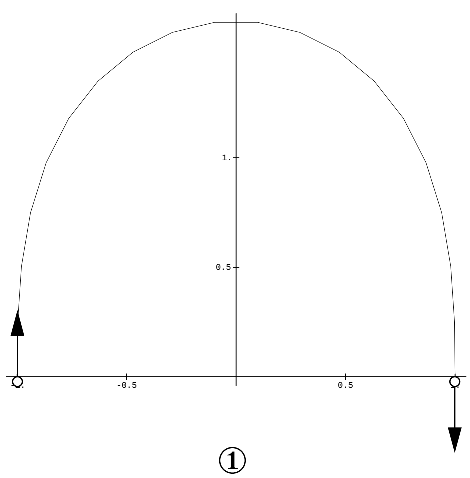
2.6 Order

Order refers to two distinct characteristics of a curve. First, when referring to a polynomial or rational polynomial representation, the order of a curve refers to the order of the polynomials employed. When speaking about rational polynomials, the orders of the numerator and denominator need not be equal and thus both are stated. Order is the number of degrees of freedom in a scalar-valued polynomial, thus it is one greater than degree; a quintic polynomial is sixth order. Second, order also refers to the accuracy of an interpolation technique. The *order of interpolation* is the maximum order of a polynomial that is exactly recreated from sampled data. The *order of convergence* refers to the rate at which an approximating curve converges on an analytic curve as the number of samples is increased.

2.7 Existence

The question of existence of a solution to a problem is widely studied. In the context of curve design, the question of existence is whether or not a curve generation technique or algorithm produces a desired result for all input specifications, for only some inputs, or for no inputs at all. Alternatively, the question may be whether the curve generated is useful, e.g. is the solution curve of finite length? Most work on formal proof of existence is in the area of solving differential equations. For most constructive curve generation techniques existence is easily proved, or it is possible to characterize the necessary and sufficient restrictions on inputs for solutions to exist. In some cases existence may be conditional on the representation used; a differentiable curve may exist that meets the curve specification, but that curve may not be in the space of curves covered by the representation used.

Existence proofs for curve generation techniques using iterative methods such as functional minimization require the derivation of characteristic differential equations and their subsequent analysis. Correctly characterizing a system based on functional minimization with all its associated constraints is extremely complex, and proof of existence is difficult or even impossible. In the absence of an existence proof, it may still be possible to heuristically identify configurations that lead to unusable solutions, e.g. a solution curve of infinite length. For example, a minimum energy curve (MEC, (3.1.1)) will run off to infinite length if the angular difference between supports is greater than π (Fig. 2.9). In order to guarantee that an MEC exists for a given interpolation problem, the arc length of the curve must be limited. The necessity of this restriction is pointed out by Birkhoff and deBoor [13] who observe that a "satisfactory" interpolant composed of circles of infinite radius may be constructed to solve any interpolation problem while consuming zero energy. Jerome [104] established the sufficiency of limiting the arc length of the solution curve.



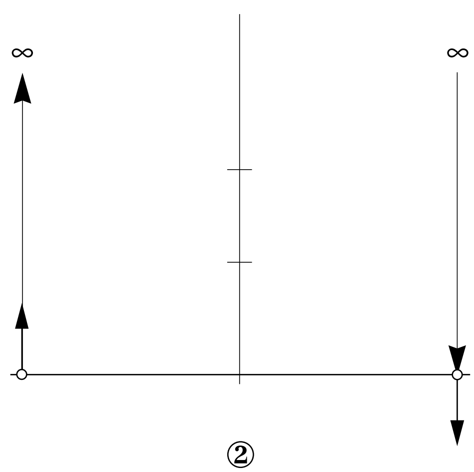


Figure 2.9. A Minimum Energy Curve.
① the curve resulting from clamping end points with tangent directions differing by π.
② the curve where the difference in tangent directions is ε greater than π.

2.8 Uniqueness

Once the existence of a solution to a curve generation problem has been established (empirically or otherwise), it may be desirable to find out if there is more than one solution to the problem as stated. Again, constructive curve generation techniques normally result in the calculation of a single curve. Curves based on functional minimization often have multiple solutions which represent multiple local minima in the energy landscape. It is often insufficient to characterize the unique solution as that curve for which the functional is globally minimized. Because in fact, in many cases, the global minimum is not the desired minimum, e.g. the global minimum of an MEC is a curve of infinite length. If a unique solution does not exist, it is normally necessary to modify the functional, to add extra conditions, to change the approach used in calculating the curve, or some combination of these. In the case of the MEC, the continuum method (described below) can be used to compute a locally unique solution. The continuum method is a technique where a continuous family of solutions is found and the desired solution is normally found at one end of the continuum. We start our continuum with an MEC limited in arc length to the accumulated chord length of the sequence of points connected one to the next. The continuum is formed by gradually relaxing the arc length constraint until a stable configuration or an upper arc length bound is reached. A stable configuration is a curve shape that is a local minimum of the original MEC functional, free of arc length constraints.

2.9 Sensitivity

The sensitivity of a curve generation technique refers to how a curve's shape varies with its defining data and even the algorithm used to compute its shape. First, is the derived solution sensitive to the starting point? This applies primarily to curves computed using iterative techniques such as minimization and is closely related to the issue of *uniqueness*. Second, if the specification of a curve is modified continuously, does the solution vary continuously? Figure 2.9 is an example of high sensitivity to small change, the specification of curve ① is modified by ε and curve ② results, a curve of infinite length. Finally, does a small change in specification lead to a small change in solution? During interactive design, it is important that a curve's shape respond in a reasonable and controllable fashion to a user's actions. A minor modification to a curve's specification should not result in a gross change in shape.

2.10 Shape Preservation

Shape preservation is that aspect of a curve technique which determines how faithfully it reproduces the shape of its defining data. For example, consider a curve defined by interpolation points that vary monotonically in some direction, e.g. non-decreasing x and y. If the curve resulting from monotonically varying points is also monotonic then the curve is said to preserve *monotonicity* (Fig. 2.10). Similarly, *convexity* preservation states that if the points defining a curve form a convex path then a convexity preserving curve must also remain convex (Fig. 2.11). Restated, a convexity preserving curve may have no more inflection points than its defining control polygon.

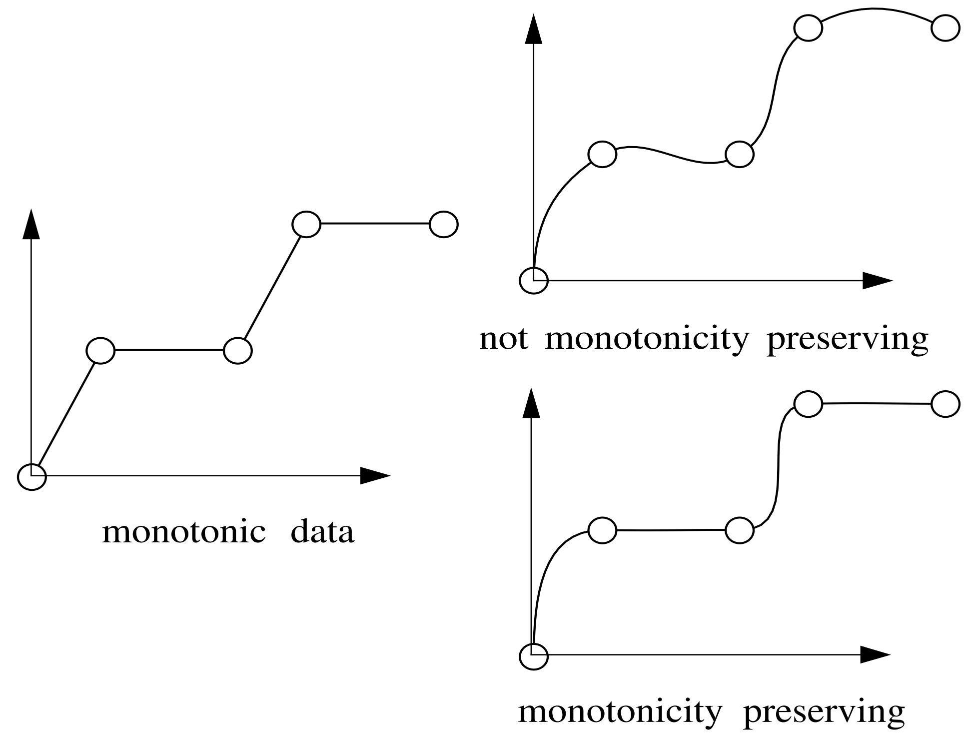


Figure 2.10. Monotonicity Preservation.
Two curves are fit to monotonically varying data points. The upper curve does not preserve monotonicity. The lower curve is non-decreasing preserving the monotonicity of the data.

An even stronger property than convexity preservation and monotonicity preservation is the *variation diminishing property*[191]. It states that a curve does not oscillate more often about any straight line than the piecewise linear interpolation of the data points. An

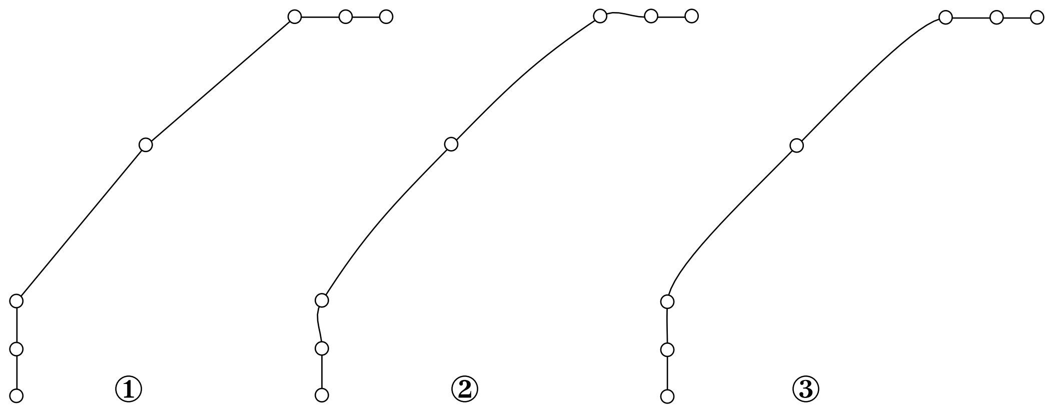


Figure 2.11. Convexity Preservation.

Two curves are fit to data points which are convex ①, i.e. the control polygon has no inflections. Curve ② has inflection points and thus is not convexity preserving. Curve ③ has no inflections, accurately reflecting the data and thus is convexity preserving.

alternative statement of the property is that no straight line should intersect the generated curve a greater number of times than it intersects the curve's control polygon. A variation diminishing curve is always smoother than the data used to define it (Fig. 2.12).

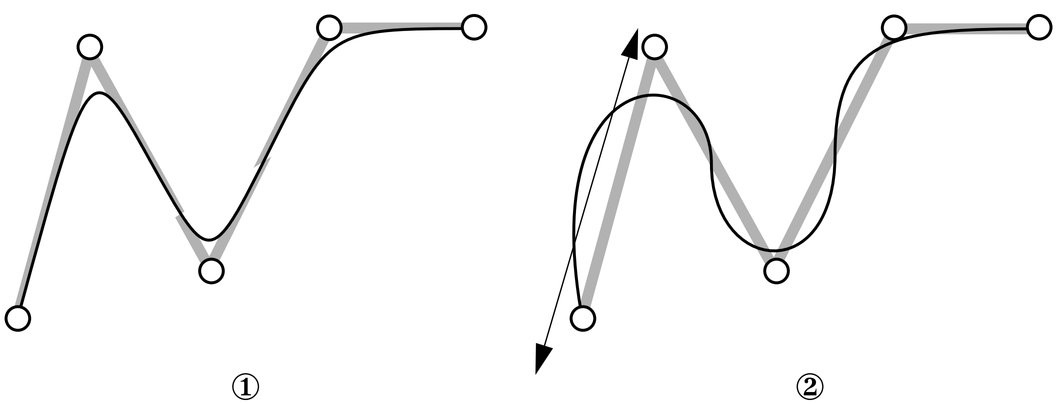


Figure 2.12. Variation Diminishing Curve.

Curve ① *smooths* the associated defining data, where curve ② does not. There is no line which intersects curve ① a larger number of times than it intersects the control polygon.

2.11 Convex Hull Property

The convex hull property states that a curve remains within the convex hull of its defining points (Fig. 2.13①). Note that this property also only applies to approximating curves because interpolating curves *must* go outside the control hull. In Figure 2.13② we see the broader definition of the convex hull property. Here a piecewise quadratic curve remains within the union of the convex hulls of the points taken three at a time. Similarly, if the curve were composed of cubic segments the curve would remain within the union of the convex hulls of the points taken four at a time.

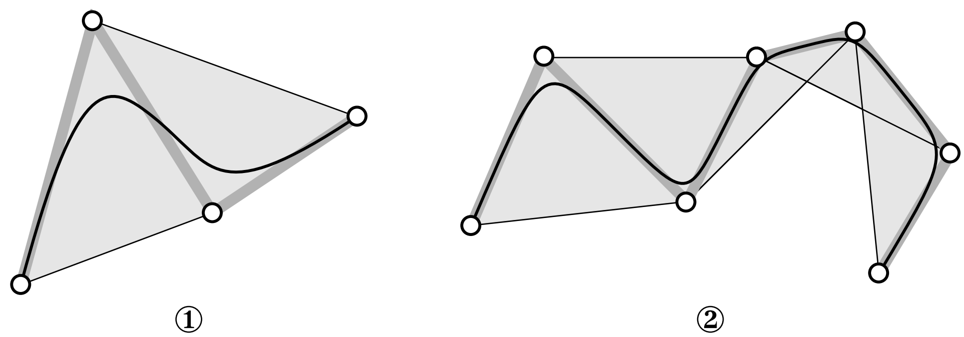


Figure 2.13. The Convex Hull Property.

① the curve remains within the convex hull of its defining control points. ② a piecewise quadratic curve remains within the union of the convex hulls computed three points at a time.

2.12 Linearity

Consider a technique for creating curves that takes as input a series of points $\vec{p}_i$ and produces a curve $\vec{C}_p(u)$ that is swept out by varying the curve parameter *u*. The technique used to create $\vec{C}_{p_i}(u)$ is considered to be linear if the following identity holds for all inputs:

$$\vec{C}_{\vec{p}_i + \vec{q}_i}(u) = \vec{C}_{\vec{p}}(u) + \vec{C}_{\vec{q}_i}(u).$$

2.13 Symmetry (w.r.t. ordering)

Curves defined by sequences of points may be sensitive to the order in which the points are processed. A curve fitting technique is *symmetric* if the order in which the points are processed has no effect on the shape of the resulting curve, i.e. a curved shape defined by points $\vec{p}_0 \ldots \vec{p}_n$ should be equivalent to the shape defined by points $\vec{p}_n \ldots \vec{p}_0$ [24].

2.14 Invariance Under Transformation

Invariance under transformation means that a curve should not change shape under a change of the coordinate system in which the data is described. Furthermore, under certain transformations of the data, the resulting shape may change in a simple predictable manner. Considering a transformation T applied to the data defining a curve and to the curve itself if these produce the same curve:

$$\mathrm{T} \cdot \vec{C}(p_i, t) = \vec{C}(\mathrm{T} \cdot p_i, t), \qquad \textbf{(2.19)}$$

then the curve is said to be invariant under that transformation. There are three types of transformational invariance of particular interest: *invariance under similarity transformations, invariance under affine transformations,* and *invariance under affine parameter transformations.* First, let us describe a hierarchy of transformations in terms of the geometric measures they preserve. Rigid body motions preserve both angular and length measures. Similarity transformations augment rigid body motions with uniform changes of scale; angular measures and ratios of lengths are preserved. Finally, affine transformations preserve parallel lines, and the ratios of lengths of parallel lines.

We start by considering function interpolation and approximation. The curve fit to data from a function should not change shape if the units of measure are changed, i.e. it should not change shape if scaling is performed along the axes (nonuniform scaling.) It is generally not required that function interpolators exhibit transformational invariance under anything more than nonuniform scaling.

Curve modeling techniques used in CAD applications should exhibit invariance under similarity transformations. It should be possible to apply modeling transformations (e.g. rotation, scaling, etc.) to a curve without the curve changing shape. There are curve methods that are invariant under affine transformations, such invariance is not a prerequisite for CAD utility, in fact it is not always desirable behavior (Fig. 2.14). In Figure 2.14① affine invariance and a nonuniform scaling results in exaggerated overshoot. In Figure 2.14② a similar nonuniform scaling results in an oval curve where a circular curve is appropriate.

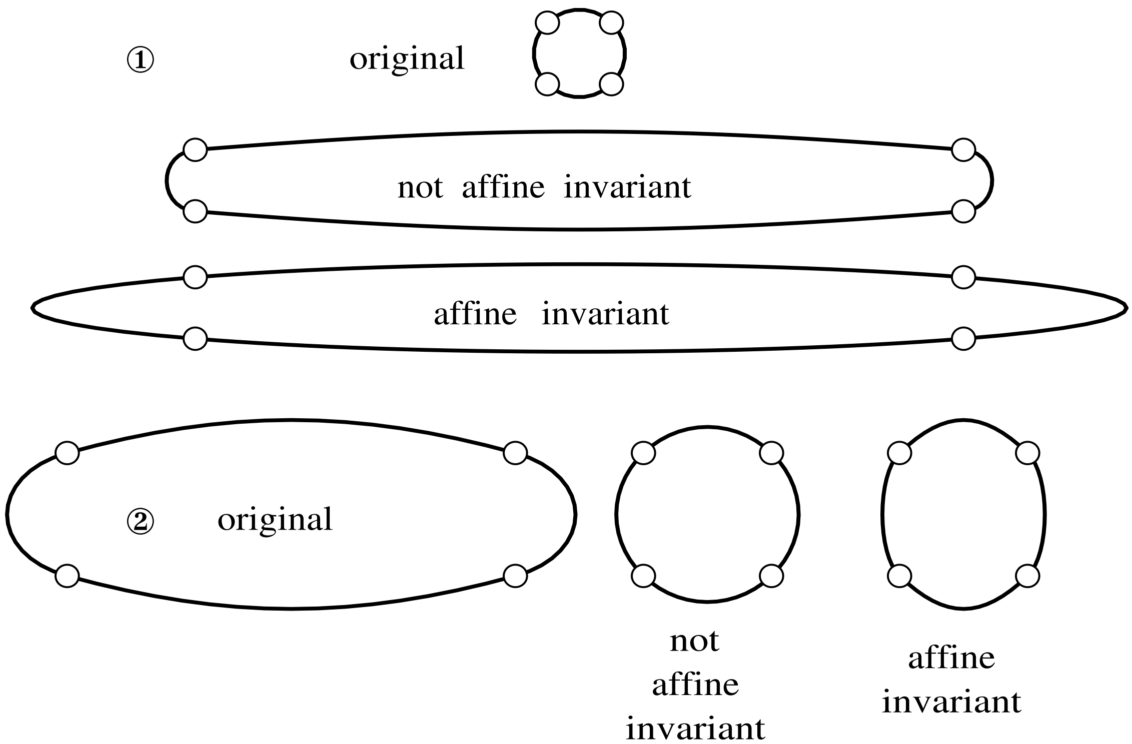


Figure 2.14. The Effect of Affine Invariance.
① demonstrates overshoot as a result of affine invariance. ② demonstrates compression as a consequence of affine invariance.

2.15 Locality

Consider a curve defined to interpolate a series of points, if the movement of a single point changes the shape of the entire curve then the technique used to generate the curve exhibits no locality whatsoever and is considered *global*. *Locality* refers to the extent that one part of a specification is limited to affect only a small, well defined portion of a curve. For example, if a single point to be interpolated is moved, how does the resulting curve change shape? If a curve's shape only changes in the neighborhood of the relocated point, then the technique exhibits *local* behavior (Fig. 2.15). For an interactive design process it is very important to have control over the locality of modification. A designer may become frustrated if a modification at one end of a curve resulted in an unexpected change

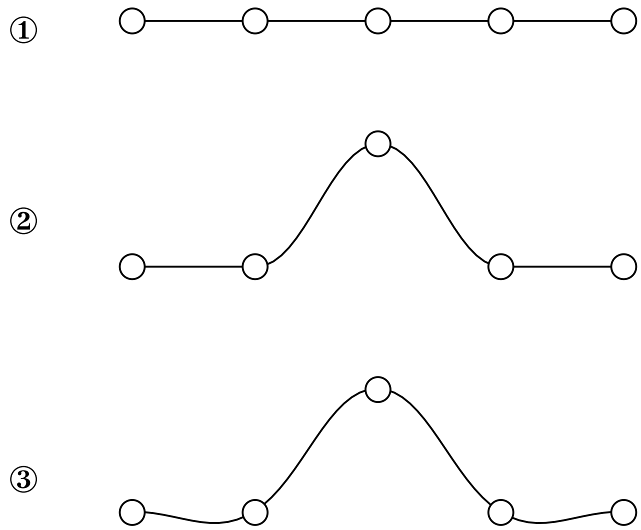


Figure 2.15. Local vs. Global control.

The central point of a curve ① is moved. The curve resulting from the technique exhibiting local control ② only changes shape in those spans adjacent to the relocated point. The curve resulting from a global technique ③ changes throughout in response to the modification.

at the other end of the curve. *Pseudo-local* control is exhibited by techniques where relocation of a control point causes a formally global change in the curve definition, but the *apparent* change is quickly attenuated as the curve leaves the neighborhood of the modification.

On the other hand, it is also desirable to allow larger scale modifications to a curve's shape; e.g. a curve with fine detail might require global bending. We refer to those techniques which provide explicit control over the locality of changes as having *variable locality.*

2.16 Consistency

The concept of *consistency* refers to a characteristic of interpolation techniques. In both *function* and *curve interpolation* a consistent technique is one that preserves the shape of a curve when an interpolation problem is augmented with additional constraints or data points that match the geometry of the curve. In Figure 2.16① we see an original curve, defined by six positional constraints. In Figure 2.16②,③ we see two possible results of adding an additional positional constraint, i.e., a point lying exactly on the curve. In ②, an

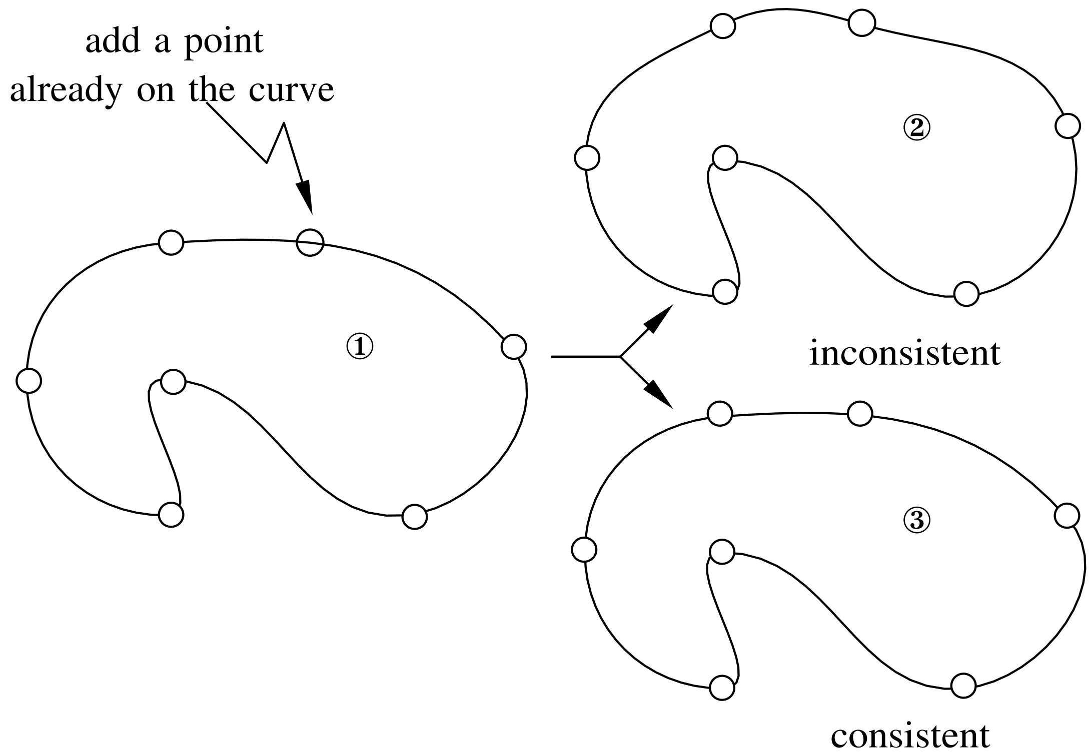


Figure 2.16. Consistency.

Consistency is the property that a curve's shape remains unchanged when a compatible constraint is added to an interpolation problem.

inconsistent technique produces a curve of modified shape. In ③, precisely the same curve is produced. The concept of consistency is most often achieved by techniques which, by some measure, produce the *best* curve possible under the given constraints. If a new best curve were produced by the addition of a new point on the curve then the previous "best" curve was apparently not the best—a contradiction!

2.17 Versatility

In section 2.4 we described the various computational forms that are used to describe curves. Another aspect of curves and the techniques used to generate them is their inherent *versatility.* As we have seen, some curve techniques are only capable of modeling single-valued curves, while others may be multi-valued. In addition, some techniques may only be applicable to planar curves, while others may model space curves with piecewise planar segments, while still others may be true space curves. For example, a quadratic curve is by definition a planar curve. The versatility of a curve technique is also determined by what types of curves can be represented exactly. For example, circular and elliptical arcs can only be *approximated* by polynomial curves, whereas rational polynomial curves may exactly model conic sections.

2.18 Visual Appearance - Fairness

The inherent subjectivity of assessing the appearance of a curve makes the definition of pleasing appearance and fairness difficult. The difficulty of arriving at a definition is compounded by the fact that it is application specific. Despite, or perhaps because of this, a number of definitions have been put forth.

A curve's shape is completely described by the curve's curvature and torsion as they vary along its length. For plane curves this may be reduced to curvature because torsion is zero. From this observation it should not be surprising that most quantitative measures of curve quality are stated in terms of curvature and sometimes torsion. The earliest discussion of fairness this author has found predates CAD and curve design by many years. In *Aesthetic Measure* [15], George Birkhoff provides quantitative analysis of various art forms: polygonal forms, ornaments and tilings, vases, music and poetry. It is in the section concerning vases that he describes the "Requirements for Regularity of Contour". This discussion of the curvature of the outline of a vase remains a preeminent definition of fairness:

> "The following further consideration of the requirements for regularity of contour is to be regarded as tentative. It is clear that such requirements can never be given any very satisfactory formulation.
>
> Consider a convex curve made up of arcs of circles of different radii, tangent to one another at their common end points. Evidently the impression produced is not that of a single unified curve, especially if the radii are alternately larger and smaller.
>
> A first obvious requirement, therefore, is that the curvature varies gradually (that is, continuously) along the curve and oscillates as few times as possible in view of the prescribed characteristic points and tangents. In particular the curvature should not oscillate more than once on any arc of the contour not containing a point of inflection. By inspection of various vase forms like those shown later, it is found that this condition is satisfied in practice.
>
> A second requirement is that the maximum rate of change of curvature be as small as possible along the contour. This condition eliminates both unnecessarily large curvatures along the contour and unnecessarily rapid changes in curvature.
>
> Although the strict application of these two conditions would be very difficult, still they may be regarded as substantially satisfied when it is not feasible to modify the curve so as to diminish either the greatest curvature or the rapidity of change of curvature. The curves of contour actually employed will be found not to permit of such modification.

The eye can follow with ease curves meeting these two requirements, just because of the small curvature and its small rate of change."

As mentioned above there are many definitions of fairness; some other definitions of fairness are paraphrased below:

1. A curve is fair if its curvature plot is continuous and consists of only a few monotone pieces [59].
2. "A frequency analysis of the radius of curvature plotted against arc length might give some measure of fairness, the lower the dominant frequency, the fairer the curve [M. Sabin from 70].
3. A fair curve has minimum strain energy [180].
4. A curve is "fair" if it can be drawn with a small number of french curve segments [56].
5. A curve's curvature plot must be almost piecewise linear, with only a small number of segments. Continuity of curvature is an obvious additional requirement [56].
6. The curve should be convexity preserving [202].
7. A curve is characterized as fair if its curvature plot is continuous, has the appropriate sign (if the convexity of the curve is prescribed), and it is as close as possible to a piecewise monotone function with as few monotone pieces as possible [187].
8. Fairness is measured as the integral of the square of the second derivative of the curve [146].
9. In many design applications a gentle, gradual development of curvature along a curve is much desired and is often used as a subjective measure of curve fairness [135].
10. A curve is fair if its curvature plot consists of relatively few monotone pieces [57].
11. The properties desired of a fair curve are smoothness, shape preservation, absence of extraneous inflection points and the like, but a curve which satisfies all of these criteria as well as the original constrains may still fail to be fair.... Fairness measures must depend only on the geometric invariants of the curve and be independent of the curve's parametrization. A curve's shape should minimize either the variation of the radius of curvature or the variation of curvature [182].
12. A plane curve is called *fair* if the following three conditions are satisfied: (1) The curve is G^2; (2) There are no unwanted inflection points on the curve; (3) The curvature of the curve varies in an even manner. Expanding on (3): (i) The number of extreme points of the curvature should be as small as possible. (ii) The curvature of the curve between two adjacent extreme points should vary almost linearly [27].

These definitions, in particular Birkhoff's, point to the approaches we have taken. Birkhoff suggests minimizing the maximum curvature, and minimizing the variation of curvature overall.