Handout #27 Original Handout #14 Tuesday, 9 November 1993

Original Lecture #9: 28 October 1993

Topics: Continuity Aspects of Spline Curves

Scribe: Joao L. Comba*

1 The General Idea of Splines

So far we've been analyzing the behavior of Bézier curves as isolated curves. To conclude this discussion let us summarize some of the nice properties of such curves:

- Affine Invariance: Bézier curves are invariant under affine maps. This property is a direct consequence of the *de Casteljau* algorithm, since points on the curve are defined using a sequence of linear interpolations (which are preserved in affine maps).
- Invariance under Affine Parameter Transformations: There is no need to define the curve in the interval [0,1], in fact we can use any arbitrary interval $a \le t \le b$ of the real line.
- Convex Hull: For $t \in [0,1]$, the point F(t) lies in the convex hull of the control polygon.
- Endpoint Interpolation: The Bézier curve passes through the first and the last point of the control polygon.

However, the Bézier representation of a complex shape will probably require a high-order polynomial (in fact a Bézier curve defined by n points will have degree n-1). In practice, high-order polynomials are not used, therefore such shapes are not suitably represented by Bézier curves.

A different approach, called splines, is used instead. The main idea is to have not just one curve but a concatenation of different curves linked together through special points called *joints*. Figure 1 illustrates the simple idea of a spline curve.

In order to transform this idea into a useful representation of complex shapes we must put some conditions at the junctions. More specifically, we are interested in the level of continuity that these junctions satisfy.

2 Continuity at Joints

Suppose that we are given two Bézier curves. What constraints do we have to put on their Bézier points — or their polar forms — if we want to guarantee that the two segments join smoothly, say to k^{th} order?

^{*}Based on the 1992 handout 19 (lecture 8)

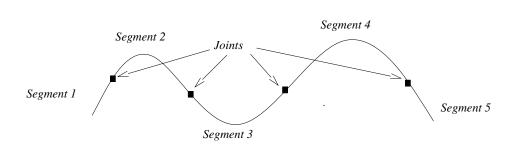


Figure 1: General Idea of Splines

2.1 Parametric vs. Geometric Continuity

Actually, there are two different notions of what it means to be smooth to k^{th} order. The first is called *parametric* continuity, and is written C^k ; the other is called either *geometric* or *visual* continuity, and is written G^k or VC^k . Parametric continuity means smoothness both of the curve and of its parameterization. That is, if a curve F is parametrically smooth, we can make a movie in which a car's position at time t is F(t) and the motion of the car in that movie will look smooth. Geometric continuity means simply the smoothness of the track that the car leaves in the snow after it passes by. For example, C^1 continuity means continuity of the tangent vector, while G^1 continuity continuity of slope; C^2 continuity means continuity of the acceleration vector, while G^2 continuity means continuity of the curvature. Parametric continuity is more expensive to arrange, in the sense that it uses up more degrees of freedom, but it is mathematically much simpler to deal with. In CS348a, we will focus on parametric continuity.

Figure 2 shows two examples to illustrate the different types of continuity. In case (a) we have both parametric and geometric continuity, but in case (b) we only have geometric continuity, because the parameter has a discontinuity of speed (from 0-1 to 1-4).

2.2 Parametric Continuity and Polar Forms

Parametric continuity and polar forms work together very neatly. Suppose F([p..q]) and G([q..r]) are two cubic segments. The level of parametric continuity at the joint at q between F and G turns out to be simply the number of polar arguments that can differ from q without destroying agreement between the resulting polar values of F and G.

Consider the case of C^0 continuity to start with. The segments $F([p\mathinner{.\,.} q])$ and $G([q\mathinner{.\,.} r])$ join with C^0 continuity at q precisely when F(q)=G(q), that is, when f(q,q,q)=g(q,q,q). Thus, the joint has C^0 continuity precisely when the polar forms f and g agree on the particular argument triple (q,q,q). That's the same thing as saying that f and g agree on all argument

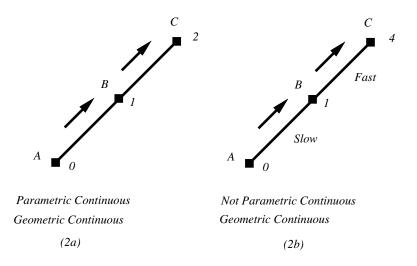


Figure 2: Parametric vs. Geometric Continuity

triples that have zero elements different from q.

At the other extreme, consider C^3 continuity. Since the curves F and G are cubic, the only way that they can meet with C^3 continuity at q is for them to be identical curves. In this case, we surely have f(u, v, w) = g(u, v, w) for any three polar arguments u, v, and w, so f and g agree on all argument triples. Thus, the number of polar arguments that can differ from q without destroying agreement of the polar values is 3.

Here's the statement and proof of the general case.

Theorem 1. Two n^{th} -degree curve segments F([p..q]) and G([q..r]) join with C^k continuity at the joint q — that is, they agree at q parametrically to k^{th} order — precisely when their n-polar forms agree on all sequences of polar arguments that include at most k values different from q.

Proof: Suppose first that the polar forms of F and G do agree on all sequences of n polar arguments that include at most k values different from \bar{q} . Then, in particular, f and g must agree on the sequences

$$(\underline{\bar{q},\ldots,\bar{q}},\underbrace{\delta,\ldots,\delta}_m)$$

for all m from 0 to k. From a previous lecture, we have

$$F^{(m)}(\bar{q}) = n(n-1)\cdots(n-q+1)f(\underbrace{\bar{q},\ldots,\bar{q}}_{n-m},\underbrace{\delta,\ldots,\delta}_{m}),$$

and similarly for $G^{(m)}(\bar{q})$. Since the factor out front is the same for F as for G, we conclude that the 0^{th} through k^{th} derivatives of F and G agree at q, which is what we mean when we say that F and G join at q with C^k continuity.

Conversely, suppose that the 0^{th} through k^{th} derivatives of F and G agree at q. Working backwards through the above, we conclude that the polar forms f and g must agree on the

argument sequences

$$(\underline{\bar{q},\ldots,\bar{q}},\underbrace{\delta,\ldots,\delta}_m)$$

for m from 0 to k. We want to show that f and g agree, in fact, on any argument sequence of the form

$$(\underline{\bar{q},\ldots,\bar{q}},\bar{u}_1,\ldots,\bar{u}_k),$$

where the times \bar{u}_1 through \bar{u}_k are arbitrary. Note that we have

$$\bar{u}_i = \bar{q} + (u_i - q)\delta$$
 for i from 1 to k .

To prove the equality

$$f(\underline{\bar{q},\ldots,\bar{q}},\bar{u}_1,\ldots,\bar{u}_k)=g(\underline{\bar{q},\ldots,\bar{q}},\bar{u}_1,\ldots,\bar{u}_k),$$

we substitute $\bar{q} + (u_i - q)\delta$ for \bar{u}_i for i from 1 to k and expand by multilinearity on both sides. The result on the left side will be a linear combination of polar values of the form

$$f(\underline{\bar{q},\ldots,\bar{q}},\underbrace{\delta,\ldots,\delta}_{m}),$$

where m is at most k. The result on the right will be the same linear combination of the polar values

$$g(\underline{\bar{q},\ldots,\bar{q}},\underbrace{\delta,\ldots,\delta}_{m}).$$

But we saw above that those polar values are the same for f and g.

To illustrate this discussion, let us enumerate the conditions of continuity for the case of cubics. The continuity constraints to the joint of two polynomials can be expressed in terms of polar forms as:

$$C^{0} \leftrightarrow f(q,q,q) = g(q,q,q)$$

$$C^{1} \leftrightarrow f(q,q,u) = g(q,q,u) \quad \forall u$$

$$C^{2} \leftrightarrow f(q,u,v) = g(q,u,v) \quad \forall u,v$$

$$C^{3} \leftrightarrow f(u,v,w) = g(u,v,w) \quad \forall u,v,w$$

To conclude, let us say that we will use the terminology C^{-1} to refer to the case where we have *no* continuity whatsoever.

3 Geometric Interpretation of the Parametric Continuity

Last section discussed mathematical constraints in terms of polar forms, necessary for having continuity at joints. In this section we will evaluate the different degrees of continuity in terms of the control points that define the spline.

Usually, we have C^0 in splines, that is, the curve segments at a joint must be connected. In other words, the point at which one curve segment ends is the same point where the next segment starts.

Let us now evaluate the conditions for higher order continuity.

3.1 C^1 Continuity

The constraints that we must put over the control points to obtain C^1 continuity are a little more elaborate than in the previous case. However, these conditions are easily derived if we write the constraints in terms of polar forms, as discussed in the previous section. Remember that to have C^1 continuity the following condition must be satisfied:

$$C^1 \leftrightarrow f(q,q,u) = g(q,q,u) \ \forall u$$

Let's consider the example of Figure 3, where we have two cubic Bézier segments. To satisfy the condition, the points f(1,1,u) and g(1,1,u) must be identical. One necessary condition for this to happen is that f(0,1,1), g(1,1,2) and the point $f(1,1,1) \equiv g(1,1,1)$ must be collinear. But this condition is not sufficient. In order to assure that g(0,1,1) is equal to $\underline{f(0,1,1)}$ (and by symmetry that $\underline{g(1,1,2)} \equiv f(1,1,2)$) the ratio of the lengths of the segments $\underline{g(1,1,1)g(1,1,2)}$ and $\underline{f(0,0,1)f(1,1,1)}$ must be the same as the ratio of the range of the parametrization in g and g. Once we know one of these ratios, we can put constraints on the other.

In the case of equally spaced parametrizations, it is easy to see that $f(1,1,1) \equiv g(1,1,1)$ must be the middle point of the segment $\overline{f(0,1,1)g(1,1,2)}$. This is the case in Figure 3. In general, to have C^1 continuity the point f(1,1,1) must be located on this segment at a distance from the points proportional to the ratio of their parametrization intervals.

To conclude this discussion, let us say that once we have C^1 we also have C^0 . In general, when we have C^m we will also have all other lower degrees continuities $C^k(0 \le k \le m)$. For the following cases, we will assume when discussing C^k continuity that we have C^{k-1} continuity.

3.2 C^2 Continuity

In a similar way, we can take a look at the case of C^2 continuity. In terms of polar forms the condition we must satisfy is:

$$C^2 \leftrightarrow f(q, u, v) = g(q, u, v) \ \forall u, v$$

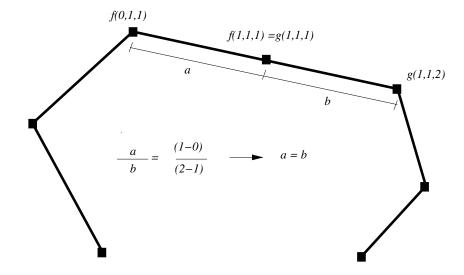


Figure 3: C^1 Continuity

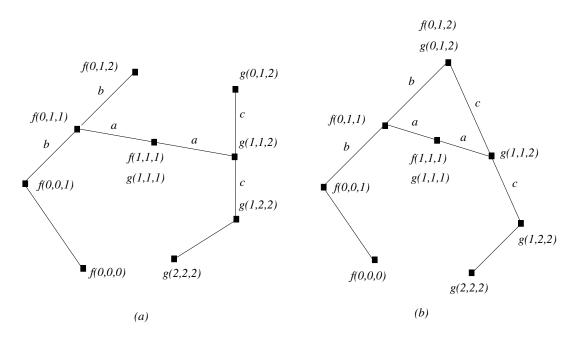


Figure 4: C^2 Continuity

In the case of figure 4 we would like to have C^2 at f(1,1,1). Therefore, the points f(0,1,2) and g(0,1,2) must be the same. By construction, we add these points to the figure and see that in case (a) we do not have C^2 continuity, because these points are different, but in case (b) these two points are coincident, so we have C^2 .

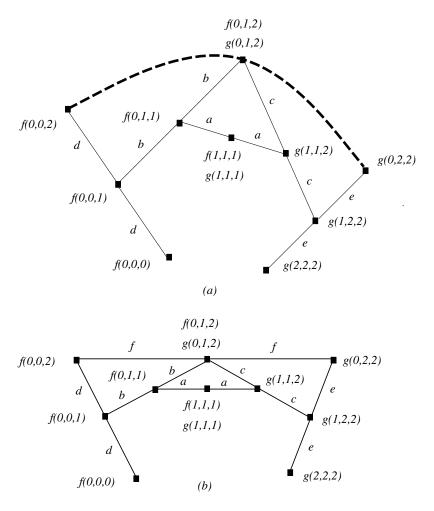


Figure 5: C^3 Continuity

3.3 C^3 Continuity

In a similar way, we can take a look at the case of C^3 continuity. In terms of polar forms the condition we must satisfy is:

$$C^3 \leftrightarrow f(u, v, w) = g(u, v, w) \ \forall u, v, w$$

In the case of figure 5 we would like to have C^3 at f(1,1,1). Not only must f(0,1,2) = g(0,1,2) but also f(0,0,2) = g(0,0,2). Therefore $f(0,1,2) \equiv g(0,1,2)$ must be interpolated on the line $\overline{f(0,0,2)g(0,0,2)}$. This does not happen in the construction of case (a), but happens in case (b). In other words, to assure C^3 continuity the control points must be located as if they were obtained by the *de Casteljau* algorithm of the combined curve.

4 Spline curves

Now that we can control the level of parametric continuity at a joint in terms of the polar forms of the joining curves, we are ready to consider assembling a sequence of curve segments into a spline curve.

4.1 Knots

Suppose that we choose to assemble our spline curve out of segments of degree at most n. What continuity should we ask for at the joints?

A joint with C^n continuity wouldn't be a joint at all; the entering and leaving segments would be adjacent segments of the same polynomial curve. The highest continuity that we can ask for at a joint that leaves any flexibility at all is C^{n-1} continuity. At a C^{n-1} joint, the derivative vectors from the 0^{th} up through the $(n-1)^{st}$ are smooth, but the n^{th} derivative jumps discontinuously from one value to another.

In addition to joints with C^{n-1} continuity, we can, if we like, allow joints with C^{n-m} continuity for various values of m greater than 1. At such a joint, the 0^{th} through $(n-m)^{th}$ derivatives are smooth, while the remaining m derivatives may all jump discontinuously. Letting m=n, for example, we can allow $C^{n-n}=C^0$ joints, where only the 0^{th} derivative, the position of the particle, is continuous. By convention, we even allow the case m=n+1, which leads to joints with C^{-1} continuity.

The parameter value that occurs at a joint is called a *knot*. If our spline curve has C^{n-1} continuity at the corresponding joint, the knot is called *simple*. Suppose that the joint has only C^{n-2} continuity. The standard convention is to say that the corresponding knot is a *double knot*, that is, it should be thought of as two separate, simple knots that have coalesced. In general, a knot of *multiplicity m* is the parameter value corresponding to a joint where the spline curve has C^{n-m} continuity, and should be thought of as a cluster of *m* simple knots that have coalesced.

Associating a multiplicity n with a knot in this way works out well because a multiple knot really does behave like a limiting case of a cluster of simple knots. For example, consider assembling line segments into an affine spline, that is, a polyline. A typical joint in such a spline is a vertex, where the position of the particle is continuous but the velocity undergoes a sudden jump. Such a joint has C^0 continuity. By our convention above, we also allow an affine spline to have joints with C^{-1} continuity, where even the position of the particle undergoes a jump. Consider such a C^{-1} joint. Say that our polyline L arrives at the point P at time r and then leaves from the point Q, with $P \neq Q$; that is, we have

$$\lim_{t \uparrow r} L(t) = P$$
, but $\lim_{t \downarrow r} L(t) = Q$.

Since this joint has only C^{-1} continuity, the knot r is a double knot. Let $r_1 < r_2$ be two distinct times that are both quite close to r. We can approximate the polyline L arbitrarily closely by a polyline M with two C^0 joints, one at time r_1 with $M(r_1) = P$ and the other at time r_2 with $M(r_2) = Q$. Over the short time interval $[r_1 ... r_2]$, the spline M scoots at high speed from P to

Q, thus smoothing out the jump in position that occurs in L. In a similar way, one can view a knot of any multiplicity m as a cluster of m simple knots that have coalesced.

Another way to think about this issue of knot multiplicity is to say that a knot is the time at which one has the right to break a derivative. At a simple knot, we are only allowed to break a single derivative, the n^{th} . At a double knot, each of the two simple knots buried inside the double knot gives us the right to break one derivative, so overall we can break both the n^{th} and $(n-1)^{st}$ derivatives. And so forth. At an (n+1)-fold knot, we can break all of the derivatives including the 0^{th} derivative, the position; so the resulting joint has C^{-1} continuity. It doesn't make sense to talk about a knot whose multiplicity is higher than n+1.

4.2 Knot sequences

With these conventions about knot multiplicity, we can incorporate all of the smoothness information about a spline curve — the time of each joint and the level of smoothness at each joint — in a single sequence of numbers, as follows. We make a list of all of the knots in which each knot is repeated according to its multiplicity, and then we sort that list into non-decreasing order. The result is called the *knot sequence* of the spline.

For example, suppose that we have a cubic spline with the knot sequence

$$(\ldots, 1, 2, 4, 4, 5, \ldots).$$

The resulting spline curve F will have a segment F([1..2]), a segment F([2..4]), and a segment F([4..5]), as well as possibly further segments on both ends. Since the knot 2 has multiplicity 1, the joint at time 2 between F([1..2]) and F([2..4]) will have C^2 continuity. But the knot 4 is a double knot, so the joint between F([2..4]) and F([4..5]) will be only C^1 .

So far, we have been talking as if knot sequences were bi-infinite, so that there are no ends. In practice, however, we can deal only with finite sequences of knots, so we have to worry a little bit about what happens at the ends. The cleanest convention is to demand that the end knots of a finite knot sequence be knots of multiplicity n+1. For example, consider a finite cubic spline F with the knot sequence

$$(0,0,0,0,1,2,4,4,5,6,6,6,6)$$
.

The joint at time 0 between whatever comes before — which we don't know anything about — and the segment F([0..1]) will be a C^{-1} joint. But there is no relationship between the entering and leaving segments at a C^{-1} joint, so it won't hurt us any that we don't know what came before time 0. Similarly, the fact that 6 is a quadruple knot means that the joint between F([5..6]) and whatever comes after — which we also don't know anything about either — is a C^{-1} joint. In fact, we can often get by with a little less: with first and last knots that have multiplicity only n, instead of n+1. But going all the way to multiplicity n+1 is certainly safe.

One common type of knot sequence is one in which all of the knots are simple knots and they are equally spaced. That is, the knots form an arithmetic progression on the parameter

line. This common situation is called the *uniform case*, and a spline whose knot sequence is uniform is called a *uniform spline*. The geometry of uniform splines is particularly regular, as we will see. Note that all of the knots in a uniform spline are simple knots, so a uniform n^{th} -degree spline has C^{n-1} continuity at all of its joints.

References

- [1] Carl de Boor (1978), A Practical Guide to Splines, Springer-Verlag.
- [2] Lyle Ramshaw (1989), Blossoms are polar forms, Computer Aided Geometric Design **6**, 323–358.
- [3] Gerald Farin (1993), Curves and Surfaces for CAGD (Third Edition), Academic Press.