Splines and the Notion of Geometric Continuity in Differential Geometry

John J. Jezioranski

Department of Computer Science
Computer Graphics Laboratory
University of Waterloo
Waterloo, Ontario, Canada
N2L 3G1

Technical Report CS-85-36
September 1985

Table of Contents

Acknowledgements

I am thankful to Professors Djoković, Gilbert, McKiernan, Shadwick, Webb and Zorzitto who were kind enough to answer many of my questions about mathematics. Their (possibly misinterpreted) answers to these questions are sprinkled throughout my thesis.

I thank the members of the Computer Graphics Laboratory for their patience in answering my questions about the lab computer. Especially, I want to thank Prof. Beatty, Peter Bumbulis and Prof. Field whose software was used to produce the figures in this thesis. (The figures were all done using splines!) Ian Allen kindly helped me through some problems with TROFF and its associated utilities. I would never have gotten VAXIMA to do what I wanted it to without the help of Prof. Char.

I am deeply indebted to my supervisor, Richard Bartels, who (1) helped get me admitted to graduate school, (2) knows that it is better to encourage graduate students for what they manage to do right than to criticize what they do wrong, (3) got me interested in this thesis topic, (4) met weekly with me to discuss my work and (5) painstakingly read some almost unreadable earlier drafts of this thesis.

ABSTRACT

This thesis is written with the computer graphics community as the intended audience. The main goal of the thesis is to translate the notion of a "kink-free" curve or surface into something precise. The graphics community has used the constructs of tangent line/plane and osculating circle/sphere from classical differential geometry to determine the "order" of a kink for the first two orders of differentiability. The fact that people cannot see the difference between twice-differentiable curves and more than twice-differentiable curves means that the definition of a kink of order higher than two will necessarily be less intuitive than it is for orders one and two. In the computer graphics literature there have been recent proposals for a constructive process for curves and surfaces which attempt to capture the notion of higher-order kinks. The standard texts in differential geometry do not appear to have studied the (local) notion of higher order kinks per se, although they provide a wealth of tools for studying local properties of curves and surfaces. We do not claim to have achieved a solid theory for these higher-order kinks. Rather, this thesis is a researcher's guide to possible avenues of generalization of kinks of orders one and two to higher order.

This thesis reaches into mathematics in two directions. Mainly, it reaches into differential geometry and singularity theory in several attempts to find the conditions to impose at spline joints in order to make the spline "kink-free". Higher-dimensional spline theory is currently an active area of research and is clearly deriving much of its foundation from combinatorial topology. We reach briefly into this recent literature and point to what we believe is an analogous and previously researched problem in differential topology. It is hoped that these great differential theories may be used in the discovery of higher-dimensional, higher-order "kink-free" splines.

Introduction

At the outset, the object of this thesis was to (1) construct a basis (the β-splines) for a linear space of geometrically continuous splines useful in computer graphics and (2) to consider the context and generalizations of these ideas in differential geometry and differential topology. The object of the thesis became more modest when the author found that the notion of geometric continuity was not as easy to find in the differential geometry literature as he had anticipated.

Intuitively, the curve and surface traces that are geometrically continuous of order k are those that don't have visible kinks of order k (although we allow them to intersect themselves). We generally equate the appearance of a kink at a point with the lack of a tangent at that point. A tangent is a geometric object that is related to first order differentiability. Consequently, we may say a curve (resp. surface) trace that has no tangent line (resp. tangent plane) at some point has a kink of order one there. The main goal of this thesis is to make this observation more precise and to find a natural extension to higher orders of differentiability.

Chapter 0 presents notation and should only be looked at if the reader finds the notation in any of chapters 1 through 7 confusing. Chapter 1 reviews some calculus and differential geometry that is fundamental to the notion of geometric continuity. Chapter 2 studies the nature of curves without using calculus and studies the meaning of geometric continuity of order 0. Chapter 3 is a sampling of concepts from the differential geometry of curves. The sampling includes material beyond that which is useful in support of our development in later chapters. It is the result of attempts to find a solid differential geometric foundation for geometric continuity of orders higher than one and two and is meant to contain ideas for further research in this direction. In chapter 4, we apply geometric continuity, insofar as chapter 3 has expressed it, to splines. In chapter 5, we introduce a vector space of splines every member of which is geometrically continuous. (Actually we introduce a vector space for any non-negative integer k such that all of the splines in the vector space are geometrically continuous of order k.) We also introduce a basis for this vector space that is useful for computer graphics. By "useful for computer graphics" we mean that (1) there is a basis that can be constructed by efficient and stable numerical techniques, (2) the elements of the basis have compact support, so that any curve constructed using a linear combination of basis elements will only change in a limited region if a single coefficient is adjusted, and (3) the basis forms a partition of unity, so that they can reasonably be used for control-vertex construction methods. The computation of this basis is detailed in appendix 2. We look briefly, in chapter 5, at (1) surfaces that are members of the tensor product of one such spline space with another and at (2) the collection of tensor products of members of the partition of unity bases as a basis for the tensor product space. In chapter 6, we consider polyhedral B-splines as a possible starting point for geometrically continuous surfaces (and higher dimensional geometrically continuous objects) more general than tensor products. Polyhedral B-splines are multivariate objects that were created in analogy to univariate B-splines. Work has been done trying to find collections of polyhedral B-splines that are a partition of unity and are a basis for a useful vector space. Chapter 7 is a last look back over the recommendations of this thesis for further research into the geometry of curves and surfaces (and objects more general than curves and surfaces) having geometric continuity of order greater than 2.

0. Notation and Background

This chapter is a log of notational conventions, definitions and results from various areas of mathematics that are peripheral to the thesis. Definitions that are central to this thesis will be presented in other chapters.

0.1 Sets and Functions

The definitions in this section will be used in virtually every chapter. We will use (1) lower case Roman letters for functions and members of sets, (2) upper case Roman letters for functions and sets and (3) upper case Greek letters for sets of sets. In general, a small-Roman-letter function is scalar-valued or vector-valued; a large-Roman-letter function is matrix-valued.

Definition 0.1.1: $\mathbb{Z}$ denotes the integers, $\mathbb{Z}^+$ denotes the positive integers, $\mathbb{Z}^-$ denotes the negative integers, $\mathbb{N}$ denotes the natural numbers $\{0,1,\ldots\}$, $\mathbb{R}$ denotes the real numbers, $\mathbb{R}^+$ denotes the positive real numbers and $\mathbb{R}^-$ denotes the negative real numbers.

Definition 0.1.2: If I is a bounded interval of $\mathbb{R}$, then the left endpoint of I is denoted left(I) and the right endpoint is denoted right(I). A *division* of I is an ordered collection of points $(t_i)_{i=0}^n$ such that left(I) $= t_0 < \ldots < t_n =$ right(I).

Definition 0.1.3: $f : A \to B : a \mapsto b$ (or $f : A \to B$, $f : a \mapsto b$ or $b = f(a)$) denotes a function that maps the set A into the set B and the element $a \in A$ into $b \in B$. For example, the function that maps each real number into its square could be denoted $\mathbb{R} \to \mathbb{R} : x \mapsto x^2$ (or $\mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ or $y = x^2$). This notation will imply that A is the *domain* of f; that is, $f(a)$ is defined for all $a \in A$. The *trace* (or *range* or *codomain*) of f is the set $f(A) :=$

$\{f(a):a \in A\}$. (The symbol $:=$ stands for "is defined as".) If B is a set then a *parameterization* $(A;f)$ of B is a set A together with a surjective (see below) function $f:A \rightarrow B$. We say that B is parameterized by $(A;f)$ (or just f).

N.B.: We use the term trace because it is the term in use in curve and surface theory. A curve trace is a collection of points that is the *range* of a special kind of *function* called a curve. In other words, a curve trace is a set that can be parameterized by a curve. Curves are defined precisely in chapter 2.

A function $f:A \rightarrow B:a \mapsto b$ is called *injective* (or 1:1 or *one to one* $\Leftrightarrow$ every $b \in f(A)$ has exactly one preimage (i.e. one member of A such that $f(a) = b$). (We will use $\Leftrightarrow$ to stand for "if and only if".) f is called *surjective* (or *onto*) $\Leftrightarrow f(A) = B$. f is *bijective* (or one to one and onto) $\Leftrightarrow$ it is surjective and injective. The inverse of f is denoted by f^{-1}. If $D \subset B$ then $f^{-1}(D) := \{a \in A : f(a) \in D\}$. (Here, f^{-1} is considered as a set valued function, so that f needn't be bijective for it to be well defined. However, if f is bijective, then f^{-1} will be a (bijective) function.) If $C \subset A$ then the function $C \rightarrow B:a \mapsto f(a)$ is called the *restriction* of f to C and is denoted $f|_C$.

Definition 0.1.4: If $U \subset V$ then the restriction to U of the identity map on V is called the *inclusion map*.

Definition 0.1.5: If $n \in \mathbb{N}$ then we denote by ndIRvs the collection of all n dimensional vector spaces over $\mathbb{R}$ and by fdIRvs the collection of all finite dimensional vector spaces over $\mathbb{R}$.

0.2 Derivatives

The notion of a derivative is central to this thesis and is used throughout it (except in the chapter on topological properties of curves (chapter 2)). Our main discussion of derivatives is in chapters 1 and 3. To define a derivative, all one uses is addition, scalar multiplication, a norm and limits. This is precisely what a Banach space has. The fact that limits make sense in a Banach space derives from its property of completeness:

Definition 0.2.1: If X is a normed linear space and $||\cdot||$ is a norm on X then a *Cauchy sequence* in X is a sequence $(x_1,x_2,...)$ (or $(x_i)_{i=1}^{\infty}$) with the following property. For any $\epsilon \in \mathbb{R}^+$, $\exists\ N \in \mathbb{Z}^+$ such that $d(x_m,x_n) = ||x_m - x_n|| < \epsilon\ \forall m, n > N$. A *convergent sequence* in X is a sequence $(x_i)_{i=1}^{\infty}$ with the following property. $\exists a \in X$ such that for any $\epsilon \in \mathbb{R}^+$ $\exists\ N \in \mathbb{Z}^+$ such that $d(x_m,a) < \epsilon\ \forall m > N$. X is said to be *complete* with respect to $||\cdot||$ if every Cauchy sequence is convergent. A *Banach Space* $(X,||\cdot||)$, is a normed linear space, X together with a norm $||\cdot||$ in which every Cauchy sequence converges. (There is no canonical calculus on objects with less structure - see (Keller, 1974) .)

In this thesis, we want to be able to discuss continuity and the topological structure it uses in a precise way. We also want to be able to discuss differentiability and the Banach space structure it uses in a precise way. The reader not familiar with Banach spaces may imagine $\mathbb{R}^n$ every time a Banach space is mentioned and $\mathbb{R}\times..\times\mathbb{R}$ every time a direct sum (or Cartesian product or direct product) of n Banach spaces is mentioned.

Proposition 0.2.2: If $(X_i, ||\cdot||_i)$, $i=1,..,n$ are Banach spaces then the direct sum of them, $\prod_{i=1}^{n} X_i$, with norm $||(x_1,..,x_n)||$ defined by $||x_1||_1+..+||x_n||_n$ is also a Banach space.

Proof: p. 152 of (Maurin, 1976) □

Definition 0.2.3: A *Euclidean space* is a finite-dimensional real inner-product space. If $n \in \mathbb{N}$ then the collection of all n dimensional Euclidean spaces is denoted by E^n. ($E^0 = \{\{0\}\}$.)

Any finite-dimensional normed linear space is a Banach space. Thus, any Euclidean space, with the norm induced by the inner product, is a Banach space. (For example $\mathbb{R}^n$ with the usual vector dot product as inner product is a Euclidean space.)

0.3 Topology

We present here some notions from topology; they are used mainly in chapter two. Proofs of the results stated in this section can be found for example in (Dugundji, 1970) .

Definition 0.3.1: A *topological space* (X, Δ) is a set X, together with a subset Δ of the power set of X (i.e. the collection of all subsets of X) such that

(1) $X, \varnothing \in \Delta$

(2) If I is a set suitable for indexing the members of Δ and $A_i \in \Delta \; \forall i \in I$ then $\bigcup_{i \in I} A_i \in \Delta$

(3) If J is a set suitable for indexing a finite subset of the members of Δ and $B_j \in \Delta$ $\forall j \in J$ then $\bigcap_{j \in J} B_j \in \Delta$

Δ is called the *topology* and the members of Δ are called the *open sets*. A *closed set* is a set whose complement is open. If $x \in X$ then a *neighborhood* (or nbd) of x is any set that contains an open set whose membership includes x. If $A \subset X$ then a neighborhood of A is any set that contains an open set that contains A.

Definition 0.3.2: If (X, Δ) and (Y, Ω) are topological spaces and $f: X \to Y$ then f is *continuous at* x $\Leftrightarrow$ for each neighborhood N of $f(x)$, $f^{-1}(N)$ is a neighborhood of x. If $U \subset X$ then f is *continuous on* U $\Leftrightarrow$ f is continuous at each $x \in U$. f is *continuous* $\Leftrightarrow$ f is continuous on X. f is a *homeomorphism* $\Leftrightarrow$ f is bijective and both f and f^{-1} are continuous. f is *open* $\Leftrightarrow$ f takes open sets onto open sets. f is *closed* $\Leftrightarrow$ f takes closed sets onto closed sets.

Proposition 0.3.3: If (X, Δ) and (Y, Ω) are topological spaces and $f: X \to Y$ then the following are equivalent:

(1) f is *continuous*,

(2) $\forall$ open sets V in Y, $f^{-1}(V)$ is open in X.

(3) $\forall$ closed sets C in Y, $f^{-1}(C)$ is closed in X.

Definition 0.3.4: If (X, Δ) is a topological space and $U \subset X$ then $\{ S : S = U \cap A$ and $A \in \Delta \}$ is called the *relative topology* from (X, Δ) on U.

Proposition 0.3.5: If X is a set and (Y, Δ) is a topological space and $f: X \to Y$, then $\{f^{-1}(A): A \in \Delta\}$ is a topology on X. Call this the *topology from* Y *induced by* f.

Example 0.3.6: The relative topology on a subset of a topological space is the topology from the topological space induced by the inclusion map.

Definition 0.3.7: If (X, Δ) is a topological space then $\Xi \subset \Delta$ is a *base* for Δ (and Δ is *generated by* Ξ) $\Leftrightarrow$ each $A \in \Delta$ can be written as a union of elements of Ξ.

Definition 0.3.8: If (X, Δ) is a topological space and $A \subset B \subset X$ and Ω denotes the relative topology on B from (X, Δ), then the *closure* of A with respect to B, $\bar{A}_B$, is the smallest (in the sense of containment) closed set in B (i.e. the smallest set whose complement is in Ω) which contains A. $\bar{A}$ denotes $\bar{A}_X$. int A denotes the largest open set contained by A. (X, Δ) is said to be *regular closed* $\Leftrightarrow$ $\overline{\text{int} A} = A \ \forall \ A$ closed in X. (X, Δ) is said to be *regular open* $\Leftrightarrow$ int $\bar{A} = A$) $\forall \ A$ open in X.

Definition 0.3.9: A topological space is said to be *Hausdorff* if any two distinct points have nonintersecting neighborhoods. A topological space is said to be *normal* if any two disjoint closed sets are contained in disjoint neighborhoods.

Note 0.3.10: The word "closed" in the definition of normal is important:

(1) Any two disjoint open sets are disjoint neighborhoods of themselves, so replacing the "closed" by "open" would be saying nothing at all.

(2) If a topological space has the property that any two disjoint sets are contained in disjoint neighborhoods then, $\forall x \in X$, $X \backslash \{a\}$ and $\{a\}$ are disjoint sets having no disjoint supersets. Thus, omitting the word "closed" would imply that $\{a\}$ is open $\forall a \in X$, which

would mean that the topology on X is the power set of X.

Definition 0.3.11: If X is a set and $\{B_i\}_{i \in I} = \Gamma$ is a collection of subsets of X then Γ is a *cover* of $A \subset X \Leftrightarrow \bigcup_{i \in I} B_i = A$. A *subcover* of a cover Γ of A is a subset of Γ which is also a cover of A. If (X, Δ) is a topological space and $A \subset X$ and $\{B_i\}_{i \in I}$ is a collection of subsets of X, then Γ is an *open cover* of $A \Leftrightarrow \Gamma$ is a cover of A and $\Gamma \subset \Delta$. A topological space X is called *compact* $\Leftrightarrow$ it is Hausdorff* and every open cover of X contains a finite subcover. A subset A of a topological space (X, Δ) is called compact $\Leftrightarrow A$ is compact with the relative topology from (X, Δ) on A.

Theorem 0.3.12: If (X, Δ) is a Hausdorff topological space, then any compact set is regular closed.

Proposition 0.3.13: A continuous map takes compact sets onto compact sets.

Proposition 0.3.14: Closed subsets of a compact topological space are compact.

Proposition 0.3.15: Compact subsets of a Hausdorff topological space are closed.

A corollary of these three propositions is the following

Corollary 0.3.16: If X is a compact topological space, Y is a Hausdorff topological space and $f : X \to Y$ is continuous, then f is a closed map.

A further corollary is:

Corollary 0.3.17: If X is a compact topological space, Y is a Hausdorff topological space and $f : X \to Y$ is continuous and injective, then f is a homeomorphism onto its image.

---

* Be warned that many books do not make the requirement that the space be Hausdorff in the definition of compactness.

Definition 0.3.18: If (X_i, Δ_i) are topological spaces for $i = 1,..,n \in \mathbb{Z}^+$ then, denoting by Ξ the topology generated by $\{\prod_{i=1}^{n} A_i : A_i \in \Delta_i \text{ for } i=1,..,n\}$, $(\prod_{i=1}^{n} X_i, \Xi)$ is the *topological product space* of $((X_i, \Delta_i))_{i=1}^{n}$ and Ξ is called the *product topology*.

Definition 0.3.19: We will say that a topological space (resp. a map from a topological space into a set) has a property *semi-locally* if every point of the topological space has a neighborhood on which the property (resp. the map has the property) holds. We will say that a topological space (resp. a map from a topological space into a set) has a property *locally* if, for each point x of the topological space, every neighborhood of x contains a neighborhood of x on which the property holds (resp. the map has the property).

Proposition 0.3.20: If (X, Δ) is semi-locally compact (and thus Hausdorff) then (X, Δ) is locally compact.

Example 0.3.21: Thus, semi-local compactness and local compactness are equivalent for $\mathbb{R}^n$. For $\mathbb{R}^n$, semi-local and local properties are almost invariably equivalent.

Example 0.3.22: If (X, Δ) is a topological space and Y is a set and $f: X \to Y$ then f is semi-locally injective $\Leftrightarrow \forall x \in X, \exists U \in \Delta$ s.t. $f|_U$ is injective.

Example 0.3.23: If (X, Δ), (Y, Ω) are topological spaces and $f: X \to Y$ then f is a semi-local homeomorphism $\Leftrightarrow$ for each $x \in X$, $\exists U$ such that $x \in U \subset X$, and $f|_U : U \to f(U)$ is a homeomorphism.

Example 0.3.24: If I is an interval of $\mathbb{R}$ with left endpoint a and right endpoint b and $n \in \mathbb{Z}^+$ and $f: I \to \mathbb{R}^n$ then f is semi-locally injective $\Leftrightarrow \forall t \in I \setminus \{a, b\}$, $\exists \delta$ s.t. $\delta \in \mathbb{R}^+$, $[t-\delta, t+\delta] \subset I$ and $f|_{[t-\delta, t+\delta]}$ is injective, and further:
1) if $a \in I$, then $\exists \delta$ s.t. $\delta \in \mathbb{R}^+$, $[a, a+\delta] \subset I$ and $f|_{[a, a+\delta]}$ is injective;
2) if $b \in I$, then $\exists \delta$ s.t. $\delta \in \mathbb{R}^+$, $[b-\delta, b] \subset I$ and $f|_{[b-\delta, b]}$ is injective.

Proposition 0.3.25: If I is an interval of $\mathbb{R}$ and $f: I \to \mathbb{R}$ then the following are equivalent:

(1) f is semi-locally injective and continuous;

(2) f is injective and continuous;

(3) f is strictly monotone and continuous;

(4) $f: I \to f(I)$ is a homeomorphism.

1. Some Differential Calculus

In this chapter, we present intuitively the main ideas involved with geometric continuity. These ideas are presented in more detail in chapters 2 and 3. Some authors refer to a curve or surface as a function and others refer to it as a collection of points. We will use the convention that a curve trace is a collection of points that is the *range* of a special kind of *function* called a curve. In other words, a curve trace is a set that can be parameterized by a curve. Curves are defined precisely in chapter 2. Geometric continuity is associated with curve (and surface) traces so we will be concerned with whether the trace of the curve or surface *can* be parameterized in such a way that this parameterization has certain properties we will discuss.

1.1 Differentiability and Tangent Planes

A surface trace is geometrically continuous of order one if it has a continuously changing tangent plane. We will explain geometric continuity of other orders later, but first we will discuss conditions for a surface trace to be geometrically continuous of order one. In this section, we put aside the question of whether the tangent plane changes continuously and just discuss conditions for having a tangent plane at a given point. We use only calculus and intuition in this section. In the next section, we introduce differential geometry and identify the intuitive aspect of the present section as being geometric.

Definition 1.1.1: If U is an open subset of $\mathbb{R}^2$ and $f: U \to \mathbb{R}$ and $(a,b) \in U$ then f is *differentiable* at $(a,b) \Leftrightarrow \exists A, B$ s.t.

$$\lim_{(\mathrm{d}x,\mathrm{d}y)\to(0,0)} \frac{f(a+\mathrm{d}x, b+\mathrm{d}y) - f(a,b) - (A\,\mathrm{d}x + B\,\mathrm{d}y)}{((\mathrm{d}x)^2+(\mathrm{d}y)^2)^{1/2}} = 0.$$

The tangent plane to f at (a,b) is $(\mathrm{d}x, \mathrm{d}y) \mapsto f(a,b) + A\,\mathrm{d}x + B\,\mathrm{d}y$. The *derivative* (or differential or Fréchet derivative or strong derivative) of f at (a,b) is

$$df(a,b)=f'(a,b):(dx,dy) \mapsto A\,dx+B\,dy.$$

Recall that a real valued function of two real variables is (Fréchet) differentiable at a point $\Leftrightarrow$ it has a (non-vertical) tangent plane at that point.

Another kind of derivative introduced in advanced calculus is the *directional derivative* (or Gateaux derivative or weak derivative). Differentiability at a point implies directional differentiability at that point, but not vice versa. In fact, a function can have directional derivatives in every direction and not even be continuous there. For example,

$$(x,y) \mapsto \begin{cases} 0 & \text{if } x=y=0 \\ \frac{xy^2}{x^2+y^2} & \text{otherwise} \end{cases}$$

has directional derivatives in all directions at (0,0) but is not continuous there.

Now we recall the definition of Fréchet differentiability for vector valued functions of vector variables in the following:

Definition 1.1.2: If $c \in U \subset \mathbb{R}^m$ and $h: U \to \mathbb{R}^n$ then h is Fréchet differentiable at $c \Leftrightarrow \exists$ linear $L: \mathbb{R}^m \to \mathbb{R}^n$ s.t.

$$\lim_{\|dr\| \to 0} \frac{\|h(c+dr)-h(c)-L(dr)\|}{\|dr\|} = 0.$$

The definition of the differentiability of a function $h: U \to F$, where $U \subset E$ and E, F are Banach spaces, is the same — just replace $\mathbb{R}^m$ by E and $\mathbb{R}^n$ by F in the definition.

Now, instead of considering maps from open subsets U of $\mathbb{R}^2$ into $\mathbb{R}$ as we have been doing, consider maps h from open subsets U of $\mathbb{R}^2$ into $\mathbb{R}^3$. For the first definition of Fréchet differentiability, the property of being Fréchet differentiable (Def. 1.1.1) was sufficient for having a tangent plane. This is not the case for this second, more general definition of Fréchet differentiability (Def. 1.1.2), as is shown by the examples at the end of this section. In the remainder of this section, we explain how Fréchet differentiability, as given in Def. 1.1.1, defines the existence of a tangent plane as a special case of the following condition: If f (1) is differentiable at (a,b) and (2) has a rank-2 Jacobian at (a,b) then the trace of f has a tangent plane at $f(a,b)$.* A function satisfying these two properties is said to be *regular* at (a,b). It is this condition that is the appropriate guarantee for the existence of a tangent plane when we are dealing with vector-valued functions and Def. 1.1.2 is used.

What could make the difference between having a tangent plane in the case of Fréchet differentiable maps from U to $\mathbb{R}$ and not necessarily having a tangent plane for Fréchet differentiable maps from U to $\mathbb{R}^3$? To answer this, we must first notice that when $f:U \rightarrow \mathbb{R}$, the tangent plane is associated not with f but with the graph of f (the graph of f is the map $(x,y) \mapsto (x,y,f(x,y))$). This is a map from U to $\mathbb{R}^3$. Moreover, the Jacobian of this map is $\left[\frac{I_2}{J_f(a)}\right]$, which has rank 2, where I_2 is the 2×2 identity matrix and $J_f(a)$ is the Jacobian of f at a. In general, if $f:U \rightarrow \mathbb{R}^n$, where U is an open subset of $\mathbb{R}^m$ and f is differentiable at a, then the graph of f, $x \mapsto (x,f(x))$, has a tangent m-space at a because its Jacobian is $\left[\frac{I_m}{J_f(a)}\right]$, which has rank m. Thus graphs that are differentiable at a are members of the more general class of functions $h:U \rightarrow \mathbb{R}^p$, where U is a subset of $\mathbb{R}^m$, $m \leq p$, whose Jacobian $J_h(a)$ has full rank.

Thus, without loss of generality, we can forget about maps like those from $\mathbb{R}^2$ into $\mathbb{R}$ and consider only this more general class, at least in the context of curves, surfaces and their generalizations.

Note 1.1.3: Recall that a linear map of finite dimensional vector spaces is injective (resp. surjective) $\Leftrightarrow$ the matrix representing it has full column rank (resp. full row rank). Here $m<p$ and the matrix $J_h(a)$ representing $h'(a)$ is $p\times m$ so $h'(a)$ is injective $\Leftrightarrow$ $J_h(a)$ has rank m.

* On the other hand, a differentiable function f whose Jacobian has less than full rank at (a,b) might still have a tangent plane at $f(a,b)$. That is, the condition is sufficient but not necessary.

Definition 1.1.4: If $h'(a)$ has full rank then a is called a *regular point* of h and h is said to be *regular* at a. There are two cases for a regular point. (1) If $h'(a)$ is injective (thus, $m \leq p$ and rank $h'(a) = m$) then h is is said to be an *immersion* at a.
(2) If $h'(a)$ is surjective (thus $m \geq p$ and rank $h'(a) = p$) then h is is said to be an *submersion* at a.

Definition 1.1.5: If $b \in \mathbb{R}^p$ then b is a regular value of $h \Leftrightarrow$ every member of $h^{-1}(b)$ is a regular point of h.

Definition 1.1.6: A function f that maps an open set $U \subset \mathbb{R}^m$ into $\mathbb{R}^n$ is C^k-regular $\Leftrightarrow$ it is C^k and its Jacobian has full rank on all U.

Definition 1.1.7: A function f that maps an open set $U \subset \mathbb{R}^n$ into an (open) set $f(U)$ of $\mathbb{R}^n$ is a C^k*-diffeomorphism* $\Leftrightarrow$ is bijective, f is C^k and f^{-1} is C^k.

Note 1.1.8: We could have allowed the dimension of the domain and range to differ in the defintion and then proven that there is no C^k bijection that has a C^k inverse. See page 12 of (Guillemin, 1974) .

Theorem 1.1.9: (Inverse Function Theorem) If f is a C^k map of an open set $U \subset \mathbb{R}^n$ into an open set $f(U) \subset \mathbb{R}^n$ and the Jacobian of f is regular at a then f is a C^k-diffeomorphism in a neighborhood of a.

Note 1.1.10: Regular, immersion, submersion are all equivalent in the case that the dimension of the domain and range are the same. It is also equivalent to say that the derivative is an isomorphism.

We conjecture that C^k-regularity is a powerful condition with geometric significance, at least when it is applied to parameterized curves and surfaces. As examples of how deceptive continuous differentiability alone is, below we parameterize $y = |x|$ as a C^k

function for $k=1,..,\infty$. (We also parameterize $y=x$ as a function that is C^{k-1} but not C^k for each even k.) The derivative of the graph $y=|x|$ agrees with our intuition — it does not have even one derivative at (0,0). On the other hand, it seems that one parametric derivative is as good as infinitely many of them. Parametric differentiability alone seems to be devoid of geometric content. Define for each $k \in \mathbb{Z}^+$,

$$f_k : t \mapsto \begin{cases} (t^k, t^k) \text{ if } t \geq 0 \\ ((-1)^{k+1}t^k, (-1)^{k+1}t^k) \text{ if } t<0 \end{cases}$$

and

$$g_k : t \mapsto \begin{cases} (t^k, t^k) \text{ if } t \geq 0 \\ ((-1)^{k+1}t^k, (-1)^k t^k) \text{ if } t<0 \end{cases}$$

f_k (resp. g_k) has $y=x$ (resp. $y=|x|$) as its graph $\forall k \in \mathbb{Z}^+$. For each $k \in \mathbb{Z}^+$, g_k is C^{k-1} but not C^k. f_k is (1) analytic for k odd, (2) C^{k-1} but not C^k for k even. Thus, we have managed to parameterize a nice graph in a nasty fashion (for even orders) and a nasty graph in a nice fashion (for all orders). The following is an infinitely often continuously differentiable parameterization of a nasty graph.

$$h : t \mapsto \begin{cases} (e^{-\frac{1}{t^2}}, e^{-\frac{1}{t^2}}) \text{ if } t>0 \\ (0,0) \text{ if } t=0 \\ (-e^{-\frac{1}{t^2}}, e^{-\frac{1}{t^2}}) \text{ if } t<0 \end{cases}$$

is a C^∞ parameterization of $y = |x| < 1$.

1.2 Geometry and Differential Geometry

It is difficult to take the definitions of orders one and two in the literature and describe why geometric continuity of order two is an extension of the notion of geometric continuity of order one. We give a general definition of geometric continuity in this thesis which is basically that of (Bartels, 1984) . The goal of this thesis, that is, the description of this definition at order k as the application of a principle that can truly be called geometric of order k, has not been achieved. In the sense of this goal, geometric continuity is lacking formal mathematical depth. As a weak replacement for this, we present in this section four notes in which we try to provoke thoughts on geometric continuity in an almost philosophical manner. We use algebraic curves and equivalence classes of Taylor polynomials which have been used to great advantage in many branches of geometry. The first note is an intuitive approach to the notions of tangent line (osculating line) and osculating circle from classical differential geometry. These are the basis for the original presentations of geometric continuity (Barsky, 1981) . See also (Kahmann, 1983) . We will say that a curve trace is *geometrically differentiable* of order one (resp. two)* at one of its points if it has a tangent line (resp. osculating circle) there. Intuitively, a curve trace is *geometrically continuous* of order one if it has a tangent line at each of its points and the tangent line changes direction continuously as we move along the trace. Intuitively, a curve trace is *geometrically continuous* of order two if it has an osculating circle at each of its points and the center of the osculating circle moves continuously as we move along the trace. It is the *trace* that is referred to as having the property of geometric continuity or differentiability — *not* the parameterization. This line of presentation shows how the transition is made from Geometry to Differential Geometry and shows what is geometric about geometric differentiability. The problem with this approach to geometric continuity is that the author cannot think of how to generalize the notion to come up with a definition of geometric continuity of order higher than 2. The third note outlines another approach (with less geometric feel to it) to geometric

* The author is not implying that this derivative is more fundamentally geometric than other derivatives (like exterior, covariant and Lie derivatives) used in differential geometry. The name just implies a relationship with the name geometric continuity.

continuity that does generalize to higher orders of geometric continuity. Described in the fourth note is the Monge-Taylor map, which the author read about after completing the research for this thesis.

Note 1.2.1: Two distinct points determine a straight line uniquely in $\mathbb{R}^k$ with $k \geq 1$. Three distinct points determine a circle uniquely in $\mathbb{R}^k$ for $k \geq 2$ (where a straight line is also considered a circle). The question is, do four distinct points determine uniquely some curve from a class, WC[4], (which stands for "*W*hat *c*urve has these properties?") of curves in $\mathbb{R}^k$ for $k \geq 3$ that interpolates the 4 points and satisfies the natural extension of the properties satisfied by lines and circles, namely: (1) Circles are members of WC[4]. (2) WC[4] is a subset of the algebraic curves of degree 3. (3) The trace of any member of WC[4] is invariant under translations and rotations of the coordinate system. It is the invariance property that makes them geometric — lines and circles are ruler and compass constructions — *there is no need to talk about coordinate systems to conceptualize them.*

A curve trace can be parameterized to have a first derivative with respect to arc length at a point x $\Leftrightarrow$ a line through two points on the curve approaches a unique line as the two points approach x, independently of how they approach x. The first derivative with respect to arc length is a unit vector along the line.

A curve trace can be parameterized so as to have a second derivative with respect to to arc length at a point x $\Leftrightarrow$ a circle through three points on the curve trace approaches a unique circle as the three points approach x, independently of how they approach x. See (Faux, 1979) for an outline of the proof in 3 dimensions. The second derivative with respect to arc length is the vector from x to the center of this circle. Thus, the derivative with respect to arc length is itself pointing to these algebraic curves and saying that they are close friends. The extension to surfaces and higher dimensional immersions is easy for the first two derivatives. Three affinely independent points determine uniquely a plane in $\mathbb{R}^k$ with $k \geq 2$. Four distinct points such that at least three are affinely independent determine uniquely a sphere in $\mathbb{R}^k$ for $k \geq 3$ (where a plane is also considered a sphere). Again, the question is: do five distinct points such that at least three are affinely independent determine uniquely a member of some class WS[5] of algebraic surfaces in $\mathbb{R}^k$ for $k \geq 4$ (where a sphere is also considered a member of WS[5])?

Note 1.2.2: We require at least three of the points that are are approaching each other in the above discussion to be affinely independent because, otherwise, there is no unique way to pass a plane or sphere through them. What kind of additional information does the (1) non-vanishing of the Jacobian at a point give us over (2) continuous differentiabilty in a neighborhood of the point? The Jacobian of a function can have full rank at a point without its trace having a tangent, even if the function is a graph. The Jacobian must be continuous on a neighborhood of the point as well to guarantee the existence of a tangent. Continuity of the Jacobian on a neighborhood of the point without full rank at the point is not sufficient either. Rank deficiency in the Jacobian of a parametrization just means that the first derivative of the parameterization is not enough information to determine whether the curve trace has a tangent line. In some cases, when the Jacobian is rank deficient, higher derivatives can be used to determine whether or not a curve trace has a tangent line. See pages 40, 41 of (Goetz, 1970) .

Note 1.2.3: Because the author is unable even to determine whether or not WC[4] exists, he must retreat to less geometric objects. Suppose we drop the invariance requirement and consider polynomials. Polynomials have an invariance property as well. The trace of a polynomial of degree $\leq k$, determined by a coordinate system and the requirement that it interpolate $k+1$ given points, is invariant under translations of the coordinate system and scaling of its axes. To obtain an object that is geometric in the sense of not referring to one particular coordinate system, take the object to be the collection, over all coordinate systems with axes of three linearly independent directions (Perpendicular directions would do just as well as linearly independent ones.), of polynomials $p=(p_1,p_2,p_3)$ of degree k which interpolate the $k+1$ points. That is, p_i interpolates the ith component of the points. Consider, as we did before, the passage to the limit of the points approaching a given point and, in the process, the passage from geometry to differential geometry. Suppose we let $k+1$ points $\{t_i\}_{i=1}^{k+1}$ approach a point a in the domain of a "nice enough" curve f. Consider the polynomial p such that $p(t_i)=f(t_i)$ for $i=1,..,k+1$. Assuming f is injective in some neighborhood of a, when we let the t_i's approach a, p becomes the degree-k Taylor polynomial in the coordinate system of the curve at the point. (This is a special case of osculatory interpolation - see theorem 4.2 of (Conte, 1980) .) Doing the same in each linearly related coordinate system and each "nice enough" curve parametrizing the trace, we get a collection of degree k Taylor polynomials that (1) is geometric in the sense that it has no preference for any one coordinate system, and (2) gives a fundamental measure of differentiability of order k at the point. For a curve in $\mathbb{R}^n$, we pick a C^k parameterization $(f_1,..,f_n):t \mapsto (x_1,..,x_n)$ and take as the geometric object of order k the collection of Taylor polynomials over all coordinate systems with $(x_1,..,x_n)$ axes with linearly independent directions. This collection of Taylor polynomials is essentially the k-jet of $f:I \to \mathbb{R}^n$, where I and $\mathbb{R}^n$ are considered as C^k manifolds, except that coordinate systems must be replaced by charts which are related by diffeomorphisms, not just by nonsingular linear transformations.

Note 1.2.4: Suppose f is a C^k-regular function from an open subset U of the plane into $\mathbb{R}^3$ and $(a,b) \in U$. The tangent plane at (a,b) is spanned by the first partials of f at (a,b). Define three axes (a trihedron), all passing through $f(a,b)$, one in each of the directions of the first partials at (a,b) and one orthogonal to the tangent plane. In a small enough neighborhood of $f(a,b)$ in $\mathbb{R}^3$, the surface trace can be expressed as the graph of a function g using the axes of the first partials as independent axes and the axis orthogonal to these two as the dependent axis. The map that takes (a,b) into the degree k Taylor polynomial of g, and similarly for each other $(a,b) \in U$, is called the Monge-Taylor map for f. Although the parameterization f is used to constuct the trihedra on the surface trace, the construction of the Taylor polynomials from these trihedra does not use the parameterization. Presumably, such a C^k-varying collection of trihedra could have been constructed without a C^k-regular parameterization and the variation of the Taylor coefficients with position on the surface could be studied without using a parameterization. A C^k-regular parameterization provides the means construct such trihedra and, through the Monge-Taylor map, the means to study how the surface trace connects these trihedra together. In chapter 3, we show how this construction can be extended to cases where the codomain has 2 or more dimensions more than te domain.

1.3 Geometric Differentiability

Here we concentrate on geometric differentiability. As with the Fréchet derivative, it is important to distinguish between this geometric differentiability at a point and the continuous geometric differentiability on a set. In accordance with the literature, we will refer to "continuous geometric differentiability in the set U" as "geometric continuity in the set U" and denote it $G^k(U)$. The C in C^k is for (k times) continuous (differentiability *in a set*) and has no significance at a point. Likewise, G^k is meaningless at a point. Continuity becomes a consideration when we have geometric differentiability at each of a set of points and we want the geometric derivative to change continuously throughout the set. There has been no use so far for associating unique geometric derivatives or "geometric continuitives" with each point of a geometrically continuous curve. The only issue of importance is whether a curve segment has the property of being geometrically continuous of some order. We have sketched what geometric differentiability is but we have not yet indicated what the geometric derivatives themselves are. Likely candidates are (1) derivatives with respect to arc length (which give rise to many of the quantitative geometric measures used in differential geometry), (2) tangent lines and osculating circles (for the first two, anyway) and (3) jets.

In the following figures, the solid lines are the trace of f, tr(f), while the dotted lines just join the pieces together and serve to highlight the jumps between the pieces. Each piece of the trace touches the two bounding dashed curves.

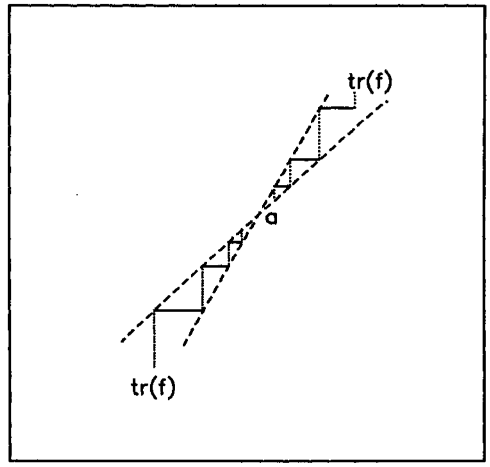

Figure 3: tr(f) has no geometric derivative at a

The functions in the figures are not curves because they are not continuous. In chapter one, we attempt to motivate why geometrically continuous functions of order zero should be defined as curves. After these examples, geometric differentiability will only be applied to curves (or surfaces or their generalizations). In the first figure, the two bounding curves have different tangent lines at a so the trace of f is not geometrically differentiable at a.

We illustrate the difference between pointwise geometric differentiability and geometric continuity by finding a point a and function f such that the trace of f is once geometrically differentiable at a but is not zero times geometrically continuous in any neighborhood of a.

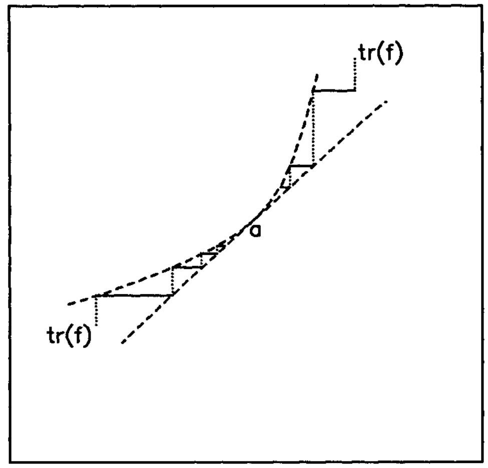

Figure 4: tr(f) is once geometrically differentiable at a but not zero times geometrically continuous in any neighborhood of a

Figure 4 shows the graph of f. Likewise, figure 5 is the graph of f for tr(f) twice geometrically differentiable at a, but not zero times geometrically continuous in any neighborhood of a. The two curve traces in figure 4 bounding tr(f) are meant to be tangent at a. The two curves bounding tr(f) in figure 5 are meant to have the same tangent line and osculating circle at a. The trace of f, tr(f), in figure 4 is once but not twice geometrically differentiable at a because the two bounding curves have different osculating circles at a.

The author is indebted to Prof. Djoković who provided figure 4 in answer to the question "Can you give an example of a curve that has a derivative at a point but no derivative in any neighborhood of that point?".

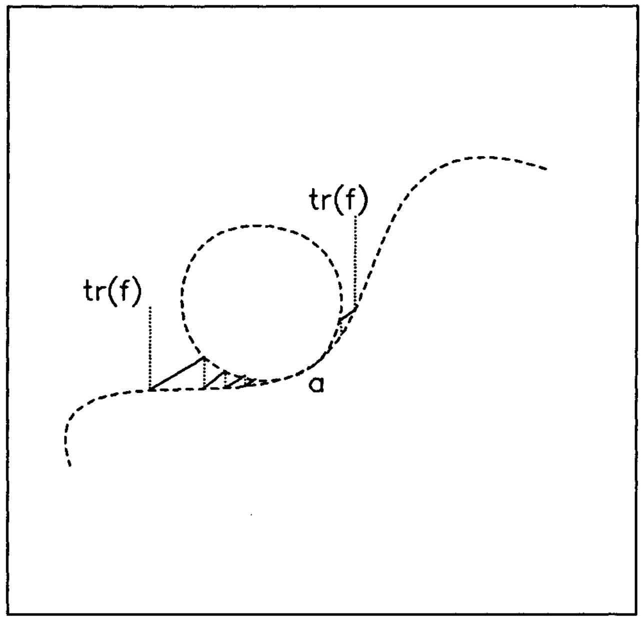

Figure 5: tr(f) is twice geometrically differentiable at a but not zero times geometrically continuous in any neighborhood of a

In the next chapter we see how these derivatives at distinct points can be glued together in such a way that the curve doesn't have visual kinks.

2. Topological Properties of Curves

We would probably be getting too philosophical if we asked, "What is it that we like about curve traces (resp. surface traces) better than other point sets that can only be expressed as other functions of one (resp. two) variables?". But this question seems to be the natural precursor to the question, "What is it about geometric continuity that makes it a geometric or natural kind of continuous differentiability?". Functions of one variable that are geometrically continuous of order 1 or more are neccessarily curves as defined below. We argue that it is natural to say that a function of one variable is geometrically continuous of order 0 if it is a curve. The hypothesis of this chapter is that we like the same thing (geometric continuity) in a derivative that we like in a curve (resp. surface) — we like it to be locally homeomorphic with (the corresponding derivative on) the line (resp. plane). The problem then is to come up with a notion (definition) of what a curve trace is and a notion of what a curve trace with nice derivative traces is. The definition should (1) avoid pathologies and (2) give rise to interesting theorems.

2.1 Definition of a Curve

The most common definition of a curve is from topology.

Definition 2.1.1: (First definition of a curve) A *curve* is a continuous map from [0,1] into a topological space. (A closed curve is a continuous map from the unit circle in the plane into a topological space.)

An example of the trace of such a curve is $[0,1]\times[0,1]$. It is the trace of the 2 dimensional Peano curve. See p. 105 of (Dugundji, 1970) . We do not want surfaces to be special cases of curves, so we do not adopt this definition. Another definition of a curve is given in (Goetz, 1970) (whose exposition we follow closely in this chapter).

Definition 2.1.2: If $n \in \mathbb{Z}^+$ and $A \subset \mathbb{R}^n$ then A *is a simple arc* $\Leftrightarrow \exists a, b, f$ s.t. $a, b \in \mathbb{R}$ and $a<b$ and $f:[a,b] \rightarrow A$ is continuous and bijective. (where continuity is with respect to the relative topologies from the ordinary topologies on $\mathbb{R}$, $\mathbb{R}^n$, respectively).

Example 2.1.3: A circle is not a simple arc. (The two endpoints of the interval have to map to separate points for the map to be injective.)

Definition 2.1.4: (Second definition of a curve) If $n \in \mathbb{Z}^+$ and $C \subset \mathbb{R}^n$ then C *is a curve trace* $\Leftrightarrow \exists I, f$ such that I is an (any) interval of $\mathbb{R}$ and $f:I \rightarrow C$ is continuous, surjective and locally injective. Such an f is *a curve having C as its trace* (or just *a curve* or *a parameterization of C*).

The fact that there is no injective continuous function from [0,1] onto [0,1] $\times$ [0,1] is proven in (Hobson, 1957) . The fact that that there is no locally injective continuous function from [0,1] onto [0,1] $\times$ [0,1] was proven for the author by Professor Gilbert. His proof uses the result in Hobson and a proposition that we will state below. We state the proposition (because of its own interest) but omit the rest of the proof. This only shows that the definition doesn't let this particular pathology through. However, our hypothesis is that the definition doesn't let any pathologies through. A curve in $\mathbb{R}^n$ is a continuous, semi-local injection $\Leftrightarrow$ it is a semi-local homeomorphism $\Leftrightarrow$ it is a local homeomorphism. This follows from the following result: if f is an injective, continuous function from a compact space X into a Hausdorff space Y, then X and $f(X)$ are homeomorphic. An outline of a proof of this result is given in chapter 0. In the case that the curve doesn't intersect itself, the trace of the curve is a one dimensional topological manifold. See (Boothby, 1975) for more on manifolds. For example, a circle with the relative topology from the plane is a one dimensional topological manifold. If the curve intersects itself, then its trace (with the relative topology from the plane) might not be a topological manifold. Locally, it is a topological manifold, i.e. the curve can be restricted to a subinterval such that the trace of this restriction is a topological manifold.

Definition 2.1.5: An *immersion* (or *topological immersion*) is a locally injective continuous map from one topological manifold to another.

Thus, our notion of a curve in $\mathbb{R}^n$ is an immersion of an interval of $\mathbb{R}$ into $\mathbb{R}^n$. Our notion of a surface trace in $\mathbb{R}^n$, $n \geq 2$ is an immersion of a special kind of set in $\mathbb{R}^2$ into $\mathbb{R}^n$. We do not want to immerse a line segment for example — this would be a curve. The general m-trace is an immersion of a special subset of $\mathbb{R}^m$ into $\mathbb{R}^n$. Take the special subsets to be, say, connected sets with a nonempty interior. We probably also want our intervals (and special subsets) to be compact as well to avoid curves like the line with irrational slope on the torus, which is dense in the torus but is C^∞-regular. Alternatively, we could require curves to be rectifiable.

An m-trace that is geometrically continuous of order k is a C^k-immersion of the same kind of subset of $\mathbb{R}^m$ into $\mathbb{R}^n$. In chapter 1, we presented geometrically differentiable functions of one variable that were not curves. Functions of one variable that are geometrically continuous of order greater than one are always curves. We argue as follows that geometric continuity of order 0 is precisely the condition that the function be a curve, that is, the condition that the function be locally injective and continuous. A C^k immersion is a function whose first k derivatives are continuous and locally injective (See chapter 3 for more on this.), so, for us, a more convenient name for a curve than a topological immersion is a C^0 immersion. Then, appling the rule that a G^k function is a C^k immersion for $k=0$ as well, we have come informally to the conclusion that a G^0 function of one variable is precisely a curve.

Proposition 2.1.6: If X is a set and $c, d \in \mathbb{R}$, $c<d$ and $f:[c,d] \to X$ is locally injective, then $\exists n$, $(t_i)_{i=0}^n$ s.t. $c=t_0<\ldots<t_n=d$ and $f|_{[t_i,t_{i+1}]}$ is injective for $i=1,\ldots,n$.

Proof: Let Θ denote the relative topology on $[c,d]$ from the ordinary topology on $\mathbb{R}$. f is locally injective so each $a \in [c,d]$ has a neighborhood A_a s.t. $f|_{A_a}$ is injective so for each $a \in [c,d]$ $\exists \epsilon_a \in \mathbb{R}^+$ s.t. $f|_{(a-\epsilon_a, a+\epsilon_a) \cap [c,d]}$ is injective. Let $\Xi :=$

$\{(a-\epsilon_a, a+\epsilon_a)\bigcap [c,d]\}_{a \in [c,d]}$ then Ξ is a Θ-open cover of $[c,d]$ and $[c,d]$ is compact so Ξ contains a finite cover $\{(a_i-\epsilon_{a_i}, a_i+\epsilon_{a_i})\bigcap [c,d]\}_{i=1}^{n}$. Choose the finite cover so that $a_1<\ldots<a_n$ and $(\,(a_i-\epsilon_{a_i}, a_i+\epsilon_{a_i})\bigcap [c,d]\,)\ \Delta\ (\,(a_{i+1}-\epsilon_{a_{i+1}}, a_{i+1}+\epsilon_{a_{i+1}})\bigcap [c,d]\,) \neq \emptyset$ for $i = 1,\ldots,n-1$. (Δ denotes the symmetric difference, $A \Delta B := (A \bigcup B) \setminus (A \bigcap B)$.)

Thus if $t_0 := c$, $t_i := \dfrac{a_i+\epsilon_{a_i}+a_{i+1}-\epsilon_{a_{i+1}}}{2}$ for $i = 1,\ldots,n-1$ and $t_n := d$ then $f\,|_{[t_i, t_{i+1}]}$ is injective for $i=0,\ldots,n$-1. □

Corollary 2.1.7: If I is an interval of $\mathbb{R}$ with left endpoint c and right endpoint d and (X,Γ) is a topological space and $f: I \rightarrow X$ is continuous, locally injective then $\exists n$, $(t_i)_{i=0}^{n}$ s.t. $c=t_0<\ldots<t_n=d$ and $f\,|_{I\bigcap[t_i, t_{i+1}]}$ is injective for $i=0,\ldots,n$-1.

Proof: Extend f to $\tilde{f}: \bar{I} \rightarrow X$ by continuity. Take n, t_i, $i=0,\ldots,n$ determined by applying the previous proposition to $\tilde{f}$. $\tilde{f}$ agrees with f on I so $f\,|_{I\bigcap[t_i, t_{i+1}]} = \tilde{f}\,|_{I\bigcap[t_i, t_{i+1}]}$, which is injective for $i=0,\ldots,n$-1 because $\tilde{f}\,|_{[t_i, t_{i+1}]}$ is injective for $i=0,\ldots,n$-1. □

Example 2.1.8:
$f(t) := (\cos(3t)\cos t, \cos(3t)\sin t)$ is a curve despite the fact that it intersects itself (i.e. is not injective). The origin corresponds in a locally injective fashion to $t = \frac{\pi}{6}, \frac{\pi}{2}, \frac{5\pi}{6}$.

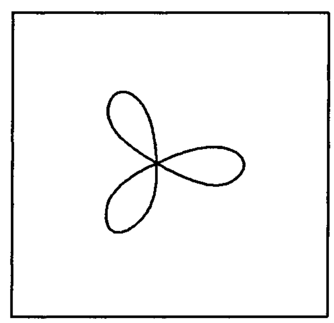

Figure 1: Cloverleaf

Note 2.1.9: Goetz says on p.28 that "The intuition behind our definition of a curve is that a curve [trace] could be drawn with one stroke without stopping and returning along the line which led to the point we stopped at."

2.2 Oriented Curves

We add orientation to curves in this section.*

Example 2.2.1: Suppose I is an interval of $\mathbb{R}$ and $f: I \to$ "⅄". The trace of f has 3 endpoints so at least one of them must correspond under f to an interior point α of I. Hence f cannot be locally injective at α.

Note 2.2.2: If we want to allow curves to stop and trace back over themselves (as in the previous example) then we just have to change the definition of curve so that it is piecewise locally injective rather than locally injective.

Definition 2.2.3: A surface is a locally injective continuous map from a connected set with a nonempty interior in the plane into $\mathbb{R}^n$, $n \geq 2$.

Definition 2.2.4: If I (resp. J) is an interval of $\mathbb{R}$ with left endpoint a (resp. α) and right endpoint b (resp. β), $n \in \mathbb{Z}^+$ and $f: I \to \mathbb{R}^n$, $g: J \to \mathbb{R}^n$ are curves then *f and g are equivalent* (resp. *opposite*) $\Leftrightarrow \exists t$ s.t. $t: J \to I$ is continuous, increasing (resp. decreasing) and surjective and $f(t(x)) = g(x)\ \forall x \in J$.

Note 2.2.5: Two curves are equivalent $\Leftrightarrow$ they trace out the same set of points in the same order as we proceed from left to right through their respective intervals.

If a curve intersects itself then it is possible to have 2 parametric representations of C that are neither equivalent nor opposite.

Definition 2.2.6: A curve trace A is *a simple closed curve trace* $\Leftrightarrow \forall x \in A$, $A \setminus \{x\}$ is a curve trace. A *curve trace intersects itself* $\Leftrightarrow$ it contains a simple closed curve trace.

---

* Tensor product splines, discussed later in this thesis, need not be orientable. For example, a Möbius band can be constructed just by joining a few patches end to end.

Note 2.2.7: A curve trace A is a simple closed curve trace $\Leftrightarrow \forall B \subset A$ s.t. B is a curve trace, $A \backslash B$ is a curve trace. (There may be some B s.t. $A \backslash B$ is a curve trace while A is not a simple closed curve trace, as in the following example.)

Example 2.2.8: The limaçon i.e. the trace of $r = a(1+2\cos\theta)$ $(a>0)$ (see figure 2) is not a simple closed curve trace because removing the point A leaves a set that is not a curve trace.

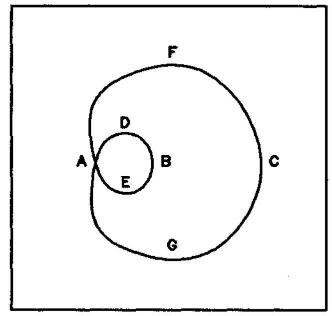

Figure 2: Limaçon

Proposition 2.2.9: Any two curves f, g having the same trace are either equivalent or opposite if the trace doesn't intersect itself.

Corollary 2.2.10: A curve trace is a simple arc $\Leftrightarrow$ it doesn't intersect itself.

Note 2.2.11: Equivalence and opposition of curves in $\mathbb{R}^n$ are equivalence relations.

Definition 2.2.12: The equivalence classes under equivalence of curves are *oriented curves*. If $f : I \rightarrow C$ is a parametric representation of C then denote by $[f]$ the parametric representation containing f.

Definition 2.2.13: If $f : I \rightarrow C$, $g : J \rightarrow C$ are elements of the same oriented curve, say $f(t(\tau)) = g(\tau)$ $\forall \tau \in J$ then $((a,f) \sim (b,g) \Leftrightarrow t(b) = a)$ is an equivalence relation. The equivalence classes are called *points of the oriented curve*. Denote by $[a,f]$ or $[t;I;f]$ the equivalence class containing (a,f).

Since two curves are equivalent $\Leftrightarrow$ they trace out the same set of points in the same order, we can think of a oriented curve invariantly as a curve trace together with an order of traversal of the trace.

Example -1.1.1: The points $[\frac{\pi}{6},f]$, $[\frac{\pi}{2},f]$ $[\frac{5\pi}{6},f]$ of example 0 are distinct points of $[f]$ although $f(\frac{\pi}{6}) = f(\frac{\pi}{2}) = f(\frac{5\pi}{6})$.

Note -1.1.2: We gave examples of continuous curves with traces that filled space and hypothesized that this was not possible for locally injective continuous curves. We now ask whether the analogous thing can happen to the derivative. If we think of the range of the derivative as being pairs, each consisting of a point on the curve and its derivative with respect to arc length there,* is it possible for the trace of the derivative of a continuously differentiable function to fill space? The derivative of a function that is regular as well as continuous is certainly locally injective as the following argument of Professor Djoković shows. A curve that is continuous and has a full rank Jacobian at a has a full rank Jacobian in some neighborhood of a. Thus, the only way for the curve trace to pass through this point $f(a)$ again in this neighborhood is through a closed curve with no kinks. Doing this in every neighborhood of a means that the derivative must spin continuously back to itself in every neighborhood of a, but this means that the derivative at a does not exist. As an example of a curve that is continuously differentiable but not locally injectively continuously differentiable, consider

* In the terminology of differential geometry, these are points of the tangent bundle.

$$t \mapsto \begin{cases} t^2 \sin^2(\frac{1}{t})(\cos t, \sin t) & \text{if } t \neq 0 \\ (0,0) & \text{if } t = 0 \end{cases}$$

This may not fill space, but it has bad enough behavior near (0,0) that we don't want to think of this as a G^1 curve.

3. Curves with Differentiability

We continue our discussion of curves, but with differential structure in addition to topological structure. Some differential structures do not give us the topological structure on the trace that we demanded in the last chapter. Continuous regular differentiability gives us this topological structure not only on the trace, but on the traces of the derivatives as well. (See the discussion at the end of chapter 2.)

3.1 Arc Length and Singularities

Proposition 3.1.1: If E is a Euclidean space and I is an interval of $\mathbb{R}$ and $n \in \mathbb{Z}^+$ and $z \in I$ and $f : I \to E$ is C^n and $\overset{(1)}{f}(t) \neq 0 \ \forall t \in I$ and $s_z^f(u) := \int_z^u \|\overset{(1)}{f}(t)\| dt \ \forall u \in I$ then s_z^f is C^n, where $\|\cdot\|$ denotes the Euclidean norm and $\overset{(k)}{f}$ denotes the kth derivative of f.

Proof: $\overset{(1)}{s_z^f}(u) = ((\overset{(1)}{f}(u) | \overset{(1)}{f}(u)))^{1/2}$; which has n derivatives — see p.171 of (Maurin, 1976) □

We will also write $s((I;f);\cdot)$ or s for s_z^f. Suppose we allow $\overset{(1)}{f}(t) = 0$ in the above definition of s. s remains constant for $t_1 \leq .. \leq t_2 \Leftrightarrow f$ remains constant for $t_1 \leq .. \leq t_2$. (because s is continuous and nondecreasing and $s((I,f);t_2) - s((I,f);t_1) = \int_{t_1}^{t_2} \|\overset{(1)}{f}(u)\| du$). If $\overset{(1)}{f}(u) = 0$ (Since $\overset{(1)}{s}(u) = \|\overset{(1)}{f}(u)\|$, this implies $\overset{(1)}{s}(u) = 0$.) at isolated points then s, considered as a function of t, is strictly increasing but the inverse of s is not continuously differentiable (the derivative is $\pm\infty$ at these points). Also, s is not differentiable of order greater than one, even if f is. If s is twice differentiable,

$$\overset{(2)}{s}(u)=\frac{(\overset{(1)}{f}(u)\,|\,\overset{(2)}{f}(u))}{\|\overset{(1)}{f}(u)\|}.$$

Definition 3.1.2: If $m, n, p \in \mathbb{Z}^+$ and C is a curve trace in $\mathbb{R}^n$ then C is *piecewise C^m-regular* $\Leftrightarrow \exists f, I, \{t_i\}_{i=0}^{p}$ s.t. I is an interval of $\mathbb{R}$, $\{t_i\}_{i=0}^{p}$ is a division of I and f is a parameterization of C and the restriction of f to the interval (t_i, t_{i+1}), $f\,|_{(t_i, t_{i+1})}$, is C^m-regular for $i=0,\ldots,p$-1. Such an f is itself said to be a *piecewise C^m – regular curve.* If $\overset{(m)}{f}(t_i)$ does not exist or $\overset{(1)}{f}(t_i) = 0$ then t_i *is a singularity of order m of f.* If for every curve $g: J \rightarrow C$ in $[f]$, the point $u_i \in J$ corresponding to t_i is a singularity of g then (t_i, f) *is an essential singularity of order m of* $[f]$. Any point of $[f]$ that is not an essential singularity of order m of $[f]$ is a *nonsingular point of order m of* $[f]$. Any point of I that is not a singular point of order m of f is a *nonsingular point of order m of f.*

Example 3.1.3: The limaçon in figure 2 traversed in the orders ADBEAGCFA, ADBEAFCGA are both parametric curves but in the first case the second incidence of A in the traversal is an essential singularity whereas the second traversal has no essential singularities.

Note 3.1.4: One cannot see an orientation on a curve, but one can usually see whether there is an orientation that makes the curve look nice. Despite the fact that the limaçon has some orientations with kinks, it looks smooth precisely because *there is* a nice orientation.

Theorem 3.1.5: If f is a piecewise C^m-regular curve then all members of $[f]$ are piecewise C^m-regular

Proof: page 24 of (Bruce, 1984) □

Definition 3.1.6: A oriented curve that contains a piecewise C^m-regular curve is a *piecewise* C^m-regular oriented curve.

Definition 3.1.7: If $n \in \mathbb{Z}^+$ and C is a curve trace in $\mathbb{R}^n$ then C is *rectifiable* $\Leftrightarrow \exists I, f$ s.t. I is an interval of $\mathbb{R}$ and $f : I \rightarrow C$ is a parametric representation of C and $l := \sup\{ \sum_{k=1}^{p} \|f(t_k) - f(t_{k-1})\|;\ (t_k)_{k=0}^{p}$ is a division of $\bar{I}$ $\}$ (i.e. the supremum with respect to all divisions of $[a,b]$) exists and is finite. l is the *length* of f.

Example 3.1.8: If a circle of radius r is parameterized to be traversed k times then the length of the parameterization is $2k\pi r$. Thus we can parameterize a circle to have any length $\geq 2\pi r$.

Theorem 3.1.9: If $(X, \|\cdot\|)$ is a Banach space and $F : X \times X \rightarrow \mathbb{R}$ is continuous and positive homogeneous in the second variable (i.e. $F(x, cy) = c\ F(x,y)\ \forall c \in \mathbb{R}^+$) then

$$L : (I; f) \mapsto \int_{\mathrm{left}(I)}^{\mathrm{right}(I)} F(f(t), \frac{\mathrm{d}f}{\mathrm{d}t}(t))\mathrm{d}t$$

is an invariant of oriented curves (i.e. is invariant under equivalent reparameterizations i.e. if $(I;f)$, $(J;g)$ are members of the same oriented curve then $L((I;f)) = L((J;g))$.

Proof: p. 368 of (Maurin, 1976) □

Corollary 3.1.10:

$$s((I;f),t) := \int_{\mathrm{left}(I)}^{t} \|\frac{\mathrm{d}f}{\mathrm{d}u}(u)\|\mathrm{d}u$$

is an invariant of oriented curves i.e. if $[t; I; f] = (y; J; g)$ then

$$\int_{\mathrm{left}(I)}^{t} \|\frac{\mathrm{d}f}{\mathrm{d}u}(u)\|\mathrm{d}u = \int_{\mathrm{left}(J)}^{y} \|\frac{\mathrm{d}g}{\mathrm{d}x}(u)\|\mathrm{d}x$$

Proof: Take $F(x,y) = \|y\|$ in the previous theorem. □

Theorem 3.1.11: If $n \in \mathbb{Z}^+$ and I is a finite interval with left endpoint a and right endpoint b and $f: I \to \mathbb{R}^n$ is a piecewise C^1–regular curve then $f(I)$ is rectifiable and $\text{length}(f) = \int_a^b \|\overset{(1)}{f}(t)\| dt$

Proof: page 158 of (Bishop, 1964) □

Corollary 3.1.12: If f, g are members of the same parametric curve and f, g are piecewise C^1–regular then their lengths are equal. (This is the same as the corollary of the previous theorem.)

Definition 3.1.13: If I is an interval of $\mathbb{R}$ and $z \in I$ and $f: I \overset{\text{surjective}}{\to} C$ is a piecewise C^1–regular curve then $s_z^f(t)$ *is an arc-length parameterization of* C. (The arc-length coordinate is unique up to a choice of the origin.)

Note 3.1.14: Since f is piecewise C^1–regular, $\overset{(1)}{f}(u) \neq 0$ except possibly at the ends of the pieces so s_z^f is a strictly increasing function of t so s_z^f has a strictly increasing piecewise C^1–regular inverse t_z^f so $g(s) := f(t_z^f(s))$ is an equivalent parametric representation i.e. any parametric curve that contains a piecewise C^1–regular curve also contains an arc-length parameterization.

If I (resp. J) is an interval of $\mathbb{R}$ and $f: I \to C$, $g: J \to C$ are equivalent piecewise C^1–regular parametric representations of C and $z \in I$ and ς is the corresponding point of J (exists because f, g are equivalent) then $g(\tau) = f \circ t_z^f \circ s_\varsigma^g(\tau)$.

Theorem 3.1.15: If $n \in \mathbb{Z}^{\geq 0}$ and $f: I \to C$ is an arc-length parameterization of C and $[f]$ is piecewise C^n-regular and $a \in I$ then $[a, f]$ is an essential singularity of order n of $[f]$ $\Leftrightarrow$ a is a singularity of order n of f.

Proof: p. 35 of Goetz □

Definition 3.1.16: If $n \in \mathbb{Z}^{\geq 0}$ and I, J are intervals of $\mathbb{R}$ and $f: I \to C_1$, $g: J \to C_2$ are arc-length parameterizations and $a \in I$, $\alpha \in J$, $A \in C_1 \cap C_2$ and $f(a) = A = g(\alpha)$ then $[f], [g]$ *have contact of order* n *at* $[a, f], [\alpha, g]$ $\Leftrightarrow$ $\|f(a+h) - g(\alpha+h)\| \in o(h \mapsto h^n)$ but $\|f(a+h) - g(\alpha+h)\| \notin o(h \mapsto h^{n+1})$. ($o(q)$ is the collection of all functions p such that $\lim_{x \to 0} \frac{p(x)}{q(x)} = 0$.)

Theorem 3.1.17: If $n \in \mathbb{Z}^{\geq 0}$ and I, J are intervals of $\mathbb{R}$ and $f: I \to C_1$, $g: J \to C_2$ are C^{n+1}-regular parameterizations of C_1, C_2, respectively, and $a \in I$, $\alpha \in J$, $A \in C$ and $f(a) = A = g(\alpha)$ and a, α are not singular points of order n of f, g respectively then $[f], [g]$ have contact of order n at $[a, f], [\alpha, g]$ $\Leftrightarrow$ for any arc-length parameterizations h, k contained in $[f], [g]$, respectively, if we let $[b, h] := [a, f]$, $[\beta, k] := [\alpha, g]$ then $h^{(i)}(b) = k^{(i)}(\beta)$ for $i = 0, \ldots, n$ and $h^{(n+1)}(b) \neq k^{(n+1)}(\beta)$.

Proof: p. 38 of Goetz □

Note 3.1.18: Any continuously regularly differentiable parameterization can be used to generate the arc-length parameterization. (Continuous differentiability is not good enough.) On the other hand, the arc-length parameterization is itself a continuous regular parameterization. Thus, geometric continuity of a curve trace can be expressed either as the condition that a continuously regularly differentiable parameterization exist or that a continuously differentiable arc-length parameterization exist.

3.2 Submersions and Contact

There are other useful ways of describing a curve trace besides a parameterization. So far, we have been asking when a map from $U \subset \mathbb{R}^m$ into $\mathbb{R}^n$, $n > m$, is nice. When is a map from $\mathbb{R}^m$ into $\mathbb{R}^n$, $n < m$, nice? For example, if $F: \mathbb{R}^2 \to \mathbb{R}$, under what conditions is $F^{-1}(\{0\})$ a nice curve trace? These sets can be ugly - see page 56 of (Bruce, 1984) . One answer again is, when the Jacobian has full rank and all the derivatives are continuous. Thus we see that the notion of a regular point is very important in both cases. Despite the fact that curves are constructed as parameterizations ($n>m$ case - i.e. immersion) in computer graphics, the $m>n$ case (i.e. submersion) is useful for obtaining a better understanding of geometric continuity.

In the light of this news, let us take another look at immersions. An immersion is always a locally homeomorphic mapping - see pages 117, 324 of (Goetz, 1970) . For a C^k-regular curve in $\mathbb{R}^n$, we can add $n-1$ axes to the domain and a natural map induced by the curve will be a C^k diffeomorphism. For example, if $\gamma:I \rightarrow \mathbb{R}^2:t \mapsto (X(t),Y(t))$ is a C^k-regular curve, then $\phi:I^2 \rightarrow \mathbb{R}^2:(x,y) := (X(x),y+Y(x))$ is injective $\Leftrightarrow$ γ is injective and is a C^k diffeomorphism in some neighborhood of (a,b) if $\overset{(1)}{X}(a)\neq 0$. (This follows from the inverse function theorem.) Since γ is regular, if $\overset{(1)}{X}(a)=0$, then $\overset{(1)}{Y}(a)\neq 0$. Thus, at least one of ϕ and $\psi:I^2 \rightarrow \mathbb{R}^2:(y+X(x),Y(x))$ is a C^k diffeomorphism in some neighborhood of each $(a,b) \in I^2$. The generalization of this example (which comes from page 52 of (Bruce, 1984)) is the parameterized manifold described on page 53 of (Bruce, 1984) . Submersions and immersions are related through the result that any parameterized n dimensional C^k-manifold in $\mathbb{R}^{n+q}$ is locally the inverse image of a regular value of a C^k map from an open subset of $\mathbb{R}^{n+q}$ into $\mathbb{R}^q$. An outline of proof of this result is given on page 61 of (Bruce, 1984) . The idea is that a C^k-regular map from a subset of $\mathbb{R}^m$ into $\mathbb{R}^n$ extends naturally to a local diffeomorphism from a subset of $\mathbb{R}^{\max(m,n)}$ into $\mathbb{R}^{\max(m,n)}$ just by adding coordinate axes to the domain or range to make the dimensions of the domain and range the same. Generalizing more yet, submersions and immersions are special cases from the important class of functions f whose derivatives f' have constant rank in a neighborhood of the given point. See page 240 of (Maurin, 1976) . Perhaps geometric continuity can be defined for any maps which have constant positive rank in open sets.

In the rest of this section, we write down two propositions from (Bruce, 1984) that come from singularity theory. (Singularity theory is a part of differential topology - see (Guillemin, 1974) .) We do this partly because the Monge-Taylor map is studied in singularity theory, but mainly because singularity theory, especially as presented in Brice and Giblin's book, is the only field that the author has found that studies pathologies of curves in depth. Although the singularities are mostly not the pathologies that have motivated the introduction of geometric continuity into the literature, the methods used there may also be useful for studying geometric continuity. The two propositions (1) make the above statements about regular values more precise and (2) point toward the notion of transversality, which generalizes the notion of a regular value. In what follows, take manifolds of dimension n to be generalizations of nonintersecting curves (1 dimensional) and surfaces (2 dimensional). Manifolds may have kinks - see chapter 7. Everything is C^k unless otherwise mentioned.

Proposition 3.2.1: If $f:\mathbb{R}^m \rightarrow \mathbb{R}^n$, $n<m$, is a submersion at a then $\exists$ an open neighborhood U of a such that $f^{-1}(f(a))$ is a parameterized $n-m$ —manifold.* A further generalization of this result is

Proposition 3.2.2: If $M \subset \mathbb{R}^{m+p}$ is an m dimensional manifold and $f:M \rightarrow \mathbb{R}^{n+q}$ is transverse to an n dimensional manifold N in $\mathbb{R}^{n+q}$ then $f^{-1}(N)$ is a manifold.

Thus, although we have not defined transversality,** we can at least note that the second proposition is a generalization of the first and that the generalization involves transversality.

Now, we define contact between a C^k curve in $\mathbb{R}^n$ and a C^k hypersurface in $\mathbb{R}^n$.† Although we do not use the definition in this chapter, we state it here (1) to have it close to the other definition of contact and (2) because it is important in the classification of

* For the definition of a parameterized manifold, see page 53 of (Bruce, 1984) .

** The definition of transversality is on page 159 of (Bruce, 1984) .

† More generally, one can define contact between any two intersecting sets. See page 22 of (Pogorelov,)

critical points, which is a part of transversality theory and (3) it is closely related to definition of jets given in (Gardner,) .

Definition 3.2.3: If 0 is a regular value of F then γ and $F^{-1}(0)$ has k point *contact* (or k fold contact) at $a \Leftrightarrow$ the first $k-1$ derivatives of $g(t) := F(\gamma(t))$ are 0 at a but the kth is nonzero.

Example 3.2.4: The hyperplane through p perpendicular to u is the set of all x such that $F(x) = <x-p|u> = 0$, where $<.|.>$ denotes the usual inner product in $\mathbb{R}^n$. The hypersphere with center u that passes through p is the set of all x such that $F(x) = ||x-u||^2 - ||u-p||^2 = 0$.

In the plane, this definition of contact is equivalent to the definition of contact of parameterized curves. See page 33 of (Pogorelov,) .

As with the other definition, the order of contact is independent of reparameterizations of the curve. For curves in the plane, the order of contact with the line and circle can be used to say many things about the curve.

3.3 The Frenet Formulae

First, we state a nice result, that holds if the first k derivatives are linearly independent, and watch how it degenerates as dependencies are introduced.

Theorem 3.3.1:

(Frenet formulae) If (I,x) is a C^n arc-length parameterization of a curve trace in E^n and $\{x^{(1)}(s_0),...,x^{(n)}(s_0)\}$ is linearly independent, then $(e_1(s),...,e_n(s))$ obtained by Gram-Schmidt on $(x^{(1)}(s_0),...,x^{(n)}(s_0))$ satisfies:

(1) $x^{(1)}(s) = e_1(s)$

(2) $e_2(s) = \frac{x^{(2)}(s)}{||x^{(2)}(s)||}$

(3) $e_j^{(1)}(s) = q_j(s)e_{j+1}(s) - q_{j-1}(s)e_{j-1}(s)$ for $j = 1,\ldots,n$, where $e_0(s) = 0 = e_{n+1}(s)$ and $q_k(s) > 0$ for $k = 1,\ldots,n$-1.

Proof: p. 374 of Maurin □†

A C^n arc-length parameterization exists ⇔ there is a C^n-regular parameterization f of the same curve trace. Also, the span of the first n derivatives of any two C^n-regular parameterizations is the same. There is a C^k change of parameter (diffeomorphism) between them and the chain rule gives the k^{th} derivative of one parameterization as a linear combination of the first k derivatives of the other parameterization. We will show later in this section that linear independence of the first n derivatives at a point and continuity of the first n derivative maps imply that the first n derivatives are linearly independent on a neighborhood of the point. Thus, we could have phrased the above theorem using any C^n-regular parameterization. This is done on pages 11-13 of (Klingenberg, 1978) . He also shows that if $e(t)=(e_1(t),..,e_n(t))$ is the Frenet frame arising from one C^k regular parameterization and $b(u)=(b_1(u),..,b_n(u))$ is the Frenet frame obtained by a reparameterization $t=h(u)$ of the domain then $e=b \circ h$. If we build up a matrix $Q(s)$ of the curvatures on the first superdiagonal and the negative of the curvatures on the first subdiagonal then part (3) of the above theorem says that $e^{(1)}(s)=Q(s)e(s)$. For an arbitrary C^k-regular parameterization f the result is $e^{(1)}(t)=\|f^{(1)}(t)\|Q(s(t))e(t)$.

Definition 3.3.3: q_j in the above theorem is called the jth *curvature* of the oriented curve. In the case $n=3$ of space curves, q_1 is called the *curvature* and q_2 is called the

† A better study of curves would examine how the order of differentiability of the q_js affects the curve. Perhaps the confusion about degenerate curves, discussed later in this section, could be dispelled by studying their curvatures. On the other hand, if we make some assumptions about the q_js, we get the *fundamental theorem of curve theory*:

Theorem 3.3.2: If I is an interval of IR and q_i is a positive, C^n function on I for $i = 1,..,n$, then there is an oriented curve of class C^{n+1} in E^n, unique up to a Euclidean motion, whose curvatures are the q_j.

Proof: p. 378 of (Maurin, 1976) □

torsion of the oriented curve. We call the collection $(x(s),e_1(s),..,e_n(s))$ the *moving n-hedron.*

Note 3.3.4: Because geometric continuity is to be a concept that is designed to avoid degeneracies, we are taking the time to study as many degeneracies as we can think of to design geometric continuity to be a notion that avoids them. We got around some of these degeneracies by requiring that the derivative be continuous for some parameterization. We got around more by requiring that the derivative be non-zero as well. Because of the preceding theorem, we have cause to wonder about another possible source of degeneracy. The Frenet-Serret formulae, which are important in classical differential geometry, require that the derivatives with respect to arc-length be linearly independent. (This is independent of the parameterization.) The question is, does the linearly dependent case correspond to some kind of kink in the curve?

Theorem 3.3.5: If $(X,\|\cdot\|)$ is a Banach space and $\{v_i\}_{i=1}^n \subset X$ and $\exists\{w_{ji} : 1 \le i \le n, 1 \le j \le \infty\}$ s.t. for each j, $\{w_{ji}\}_{i=1}^n$ is linearly dependent and for each i, $\lim_{j\to\infty} w_{ji} = v_i$, then $\{v_i\}_{i=1}^n$ is linearly dependent.

Proof: The proof was given to the author by Prof. Zorzitto. For each j, $\exists\{a_{ji}\}\subset \mathrm{IR}$ s.t. $\sum_{i=1}^n a_{ji} w_{ji} = 0$ and $\sum_{i=1}^n |a_{ji}| \neq 0$. Let $b_{ji} := \frac{a_{ji}}{\sum_{i=1}^n |a_{ji}|}$. $(b_{j1})_{j=1}^\infty$ is a bounded (by ± 1) sequence, so by the Bolzano Weierstrass theorem, it has a convergent subsequence $(b_{j_k 1})_{k=1}^\infty$. Using this index sequence as input for $i=2$, repeat the argument: $(b_{j_k 2})_{k=1}^\infty$ is a bounded sequence so it has a convergent subsequence $(b_{j_{k_l} 2})_{l=1}^\infty$. After having done this n times, the nth index sequence J is such that $(b_{ji})_{j\in J}$ is convergent for each i. Let $b_i := \lim_{j \in \mathcal{J}} b_{ji}$ for $i = 1,..,n$. To show that $\sum_{i=1}^n b_i v_i = 0$, show that

$$\lim_{j\to\infty}\|\sum_{i=1}^{n} b_{ji} w_{ji} - \sum_{i=1}^{n} b_i v_i\| = 0.$$

To show this, show that each of the terms on the right side of

$$\|\sum_{i=1}^{n} b_{ji} w_{ji} - \sum_{i=1}^{n} b_i v_i\| \le \sum_{i=1}^{n}\|b_{ji} w_{ji} - b_i v_i\|$$

go to zero in the limit.

$$\|b_{ji} w_{ji} - b_i v_i\| \le \|b_{ji} w_{ji} - b_{ji} v_i\| + \|b_{ji} v_i - b_i v_i\|$$

$$\le |a_{ji}|\,\|w_{ji} - v_i\| + |a_{ji} - a_i|\,\|v_i\|$$

The $|a_{ji}|$ are bounded and $\|v_i\|$ is fixed, while $\|w_{ji} - v_i\|$ and $|a_{ji} - a_i|$ go to zero in the limit. □

Note 3.3.6: Another way of stating this result is, the collection of $(v_1,..,v_n) \in \prod_{i=1}^{n} X$ such that $\{v_1,..,v_n\}$ is linearly independent is open in $\prod_{i=1}^{n} X$, with, say, the norm $\|(x_1,..,x_n)\| := \sum_{i=1}^{n}\|x_i\|$. Alternatively, the collection of linearly dependent n-tuples is closed in $\prod_{i=1}^{n} X$. Given any collection of n linearly independent vectors, there are neighborhoods of each vector such that any selection of n vectors, one from each neighborhood, is linearly independent. Because the first n derivatives were assumed linearly independent and continuous at a point in the Frenet-Serret theorem, the above theorem implies that they are also linearly independent in a neighborhood of of that point. What does the interior, A, of the set of linearly dependent n-tuples look like? On A, given any collection of n linearly dependent vectors, there are neighborhoods of each vector such that any selection of n vectors, one from each neighborhood, is linearly dependent. We will desribe a result for the case that the first n derivatives are linearly dependent in a neighborhood of a point. The case where the first n derivatives are linearly dependent at a point, but not in any neighborhood of the point is discussed briefly on pages 69-71 of (Vaisman, 1984) for space curves and for the special case that (1) the first $p-1$ are zero and the pth is nonzero, (2) the $p+1$th through $p+q-1$th are linear multiples of the pth, (3) $\exists r$ such that the pth, $p+q+r$th, pth and $p+q$th derivatives at the point are linearly independent. In such a case, p is called the order of the singularity, q is called the class of the singularity and r is called the rank of the singularity. The problem with this construction is that it might not lead to a Frenet frame that changes continuously. See page 54 of (Goetz, 1970) .

Proposition 3.3.7: If $x:I \rightarrow E^n$ is a C^n arc-length parameterization of a curve and $\{x^{(1)}(s),$ $\ldots, x^{(k)}(s)\}$ is linearly independent, but $x^{(k+1)}(s) = \sum_{i=1}^{k} c_i(s) x^{(i)}(s)$ for $s \in [a-r, a+r]$, then $x([a-r, a+r])$ lies in the translate of a k dimensional subspace of E^n but not in the translate of any $k-1$ dimensional subspace.

Proof: page 376 of (Maurin, 1976) □

Note 3.3.8: The condition on the $k+1$th derivative in the proposition is implicitly a condition on all higher derivatives as well. By differentiating this expression, all derivatives of order higher than k are in the span of the first k derivatives $\forall\ s \in [a-r, a+r]$.

If $n=3$ and $k=2$, then the previous proposition says that the space curve settles into a plane for a little while. In the plane, we construct a Frenet bihedron instead of a Frenet trihedron. If the torsion settles nicely to zero as the curve trace settles into a plane, we can just add a third unit vector to the bihedron to get a trihedron that changes continuously at the points where the curve enters the plane. On the other hand, is it possible for the trihedron to spin as the curve enters the plane in such a way that the orientation of the trihedron approaches no limit? There are certainly decent curves that are degenerate. For example, the curve trace whose graph is $y=x^3$ has a Frenet bihedron that turns into a unihedron at $x=y=0$.

At any point on a curve in $\mathbb{R}^n$ with a Frenet n-frame, there is a neighborhood of the point on which the curve trace can be expressed exactly as $x \mapsto (x, Y_1(x), .., Y_n(x))$ where the independent coordinate x is the coordinate along the tangent line and the remaining $n-1$ coordinates are the coordinates along the remaining $n-1$ axes of the Frenet frame. Take the degree k Taylor polynomial of this map and according to page 173 of (Bruce, 1984) what we have done is to "...associate to each point $t \in I$ a polynomial of degree k which carries all the infinitesimal information about the curve γ at t up to order k". This map of a curve into this Taylor polynomial is called the Monge-Taylor map. (Bruce, 1984) does not worry about the degenerate case. They assume only the existence of a tangent line and construct the remaining perpendicular axes as follows for $n=3$. By Sard's theorem,* there exists a direction in $\mathbb{R}^3$ such that no tangent line to the (compact) curve is parallel to this direction. Orthogonally project a fixed vector in this direction on the plane normal to the tangent line at each point of the curve. Together with the unit tangent vector field, this gives two C^k unit vector fields on the curve. Taking the cross product of these two at each point gives us a C^k moving trihedron. In $\mathbb{R}^4$, we would be able to find two such directions by applying Sard's theorem twice. We described the Monge-Taylor map for a C^k-regular surface in $\mathbb{R}^3$ in chapter 1. We could find two axes orthogonal to the tangent plane for a surface in $\mathbb{R}^4$ by likewise applying Sard's theorem.

* Sard's theorem is the main technical result in transversality theory, which generalizes the notion of a regular value, which is fundamental to geometric continuity. Interestingly, Sard also introduced a general definition of a spline. See page 215 of (Groetsch, 1980) .

4. Spline Curves

4.1 Applying Geometric Continuity to Splines

In the previous chapter, we defined a collection of points as being geometrically continuous if it is the image of a special kind of parameterization. Finding such a parameterization can be difficult. In this chapter, we define a vector space of splines, such that the trace of every element in the spline space is geometrically continuous. In this spline space, we don't have to look hard for special parameterizations — they are all special! What has to be shown is that such a restriction leaves us with enough elements in the space for it to be useful. If we can demonstrate an efficient method for selecting elements from this space based on some useful kinds of input (say a function or collection of points that we want to approximate or interpolate) and demonstrate that the selected element can be made to fit the input requirements as closely as desired, then the vector space and procedure will have precise control over not only the degree of approximation but also over the geometric properties of the approximant. We will also be striving to construct a compact support basis for this spline space, so that as few as possible pieces of a spline curve or surface are changed by changing one of the coefficients in its expression as a linear combination of this basis. This gives us *local control* over elements of the spline space.

In the following we give a loose, working definition of a spline. A general definition of a spline can be found on page 216 of (Groetsch, 1980) . Specific cases of the definition as well as a more thorough discussion of splines can be found in (de Boor, 1976) .

Definition 4.1.1: A *spline curve (trace)* in $\mathbb{R}^n$ is (the trace of) a piecewise analytic function, each piece from an interval of $\mathbb{R}$ into $\mathbb{R}^n$. Moreover, the intervals (pieces of the trace) and their ends are ordered so that each piece has a *left* end that is associated with the *right* end of the *previous* piece (except the first piece) and a *right* end which is associated with the *left* end of the *next* piece (except the rightmost piece). Finally, this association satisfies some conditions that determine the type of the spline. The association between ends of intervals (pieces of the trace), is called a *knot* (*joint*).

G^k spline curve traces (G^k splines for short) satisfy the G^k condition between joints. Again one can think of G^k splines invariantly (without parameterizations) but in practice we will always refer to them through parameterizations.

In chapter 3, we gave several equivalent characterizations of geometric continuity. We now repeat these characterizations, but for the specific case of splines, i.e. the case of geometrically continuous pieces being joined together. More than just repeating the characterizations, we wish to change such statements of chapter 3 as "there exists a parameterization such that..." into a specific condition on the given parameterizations of the pieces. We do this for characterizations 2, 3 and 4 of chapter 3. The pieces will be taken to be parameterized by analytic functions and thus (because any collection of points which is the trace of an analytic curve is G^k $\forall k \in \mathbb{N}$) their traces will be geometrically continuous of all orders. Because the pieces are G^k, the only points of contention are the joints between the pieces.

Characterizations 1 and 2: Characterization 1 is Barsky's original characterization, described in chapter 1, of having an osculating line (order 1) or osculating circle (order 2) which "changes continuously" as we traverse the curve trace. To make this precise, we use Barsky's terminology. Geometric continuity of orders one and two are continuity of the unit tangent and curvature vectors, respectively. This characterization has not been extended to higher orders or to surfaces or higher dimensional manifolds. Suppose $f(t)$ is the curve representing the "left-hand" piece at a joint, say $t \le t_0$, and $g(u)$ is the curve representing the "right-hand" piece at the same joint, $u \ge u_0$, where $f(t_0)=g(u_0)$. Let $f_L^{(1)}(t_0)$ denote the derivative from the left of f at t_0, and $g_R^{(1)}(u_0)$ denote the derivative from the right of g at u_0. The conditions one obtains by applying the above characterizations of orders 1 and 2 to such parameterizations are as follows. (See (Barsky, 1981) for a derivation.):

$$g_R^{(1)}(u_0)=\beta_1 f_L^{(1)}(t_0)$$

and

$$g_R^{(2)}(u_0)=\beta_1^2 f_L^{(2)}(t_0)+\beta_2 f_L^{(1)}(t_0)$$

respectively, where β_1 and β_2 are real numbers. β_1 is taken to be positive because the $\beta_1<0$ case is the case of a cusp.

Note 4.1.2: We will take a diversion here to be sure in what follows that the trace has a tangent line and not a cusp. In the case of a cusp, the two pieces have tangent lines as the end point is approached and the two limiting tangent lines coincide, but the union of the two traces does not have a tangent line at the cusp. Because the parameterization is regular, the tangent always points in the direction in which the curve is being traced, in the sense that $f(t+h)=f(t)+hf^{(1)}(t)$. For points near the joint, the tangent points away from the joint (i.e. h in the above expression is negative if $f(t+h)$ is the joint and $f(t)$ is a point on the curve near the joint) if it corresponds to the left end of the parameter interval and toward the joint (h is positive) if it corresponds to the right end of the parameter interval. Thus there are the following cases. Suppose $f:[a,b]\rightarrow \mathbb{R}^n$, $g:[w,x]\rightarrow \mathbb{R}^n$ are C^1-regular. (1) If $f(a)=g(w)$ and their traces have the same one-sided tangent lines there then the union has a cusp of order one if β_1 is positive and a tangent line if β_1 is negative. (2) If $f(b)=g(x)$ and their traces have the same one-sided tangent lines there then the union has a cusp of order one if β_1 is positive and a tangent line if β_1 is negative. (3) If $f(b)=g(w)$ and their traces have the same one-sided tangent lines there then the union has a cusp of order one if β_1 is positive and a tangent line if β_1 is positive. Case (4) is just case (3) with f and g reversed.

From now on, we will take f to be defined and C^k-regular to the left of t_0 and g to be defined and C^k-regular to the right of u_0. We will always make the choice of sign that gives a tangent line instead of a cusp.

Since a unit tangent vector is (except for sign) the first derivative with respect to arc length, and the curvature vector is the second derivative with respect to arc length, requiring the one sided derivatives with respect to arc length on either side of a knot to be equal is equivalent to Barsky's conditions (that the one-sided limits of the unit tangent and curvature vectors are equal). This is a special case of the characterization of geometric continuity we gave in chapter 3: namely, that derivatives up to order k with respect to arc length exist and are continuous. These equivalent descriptions of geometric continuity are invariant. Equating first and second derivatives with respect to arc length (and taking the sign that gives a tangent line instead of a cusp) gives:

$$\frac{f_L^{(1)}(t_0)}{\|f_L^{(1)}(t_0)\|}=\frac{g_R^{(1)}(u_0)}{\|g_R^{(1)}(u_0)\|}$$

$$=\frac{\dfrac{g_R^{(2)}(u_0)}{\|g_R^{(1)}(u_0)\|^2}-\dfrac{f_L^{(2)}(t_0)}{\|f_L^{(1)}(t_0)\|^2}}{\left(\dfrac{f_L^{(1)}(t_0)}{\|f_L^{(1)}(t_0)\|}\,\middle|\,\dfrac{g_R^{(2)}(u_0)}{\|g_R^{(1)}(u_0)\|^2}-\dfrac{f_L^{(2)}(t_0)}{\|f_L^{(1)}(t_0)\|^2}\right)}$$

where $\|\cdot\|$ and $(\cdot|\cdot)$ denote the Euclidean norm and inner product, respectively. Solving for $g_R^{(1)}$ and $g_R^{(2)}$ in the above equations give us Barsky's equations:

$$g_R^{(1)}(u_0)=\beta_1 f_L^{(1)}(t_0) \quad \text{where } \beta_1=\frac{\|g_R^{(1)}(u_0)\|}{\|f_L^{(1)}(t_0)\|} \quad \text{and}$$

$$g_R^{(2)}(u_0)=\beta_1^2 f_L^{(2)}(t_0)+\beta_2 f_L^{(1)}(t_0) \quad \text{where } \beta_2=\frac{(f_L^{(1)}(t_0)\,|\,g_R^{(2)}(u_0)-\beta_1^2 f_L^{(2)}(t_0))}{\|f_L^{(1)}(t_0)\|^2}$$

The difference is that equating one-sided derivatives with respect to arc length generalizes to orders higher than 2 whereas we know of nobody who has generalized along the lines of thought of continuous unit tangent vector and continuous curvature vector. Moreover, the condition one obtains by equating one-sided arc length derivatives looks like chain rule in the sense of the following

Characterization 3: Characterization 3 is a manifestation of the condition that the curve trace have a C^k-regular parameterization. If $f:[a,b]\to f([a,b])\subset\mathbb{R}^n$ and $g:[x,y]\to g([x,y])\subset\mathbb{R}^n$ are C^k bijections and $f(b)=g(x)$ and $f_L^{(1)}(b)\neq 0\neq g_R^{(1)}(x)$ then the trace is $G^k \Leftrightarrow \exists\ c,\ h$ s.t. $h:[b,c]\to[x,y]$ is a C^k bijection and $h(b)=x$ and $\tilde{f}^{(i)}(b)$ exists for $i=0,..,k$ where

$$\tilde{f}(t):=\begin{cases} f(t) & \text{if } a\le t\le b \\ g(h(t)) & \text{if } b\le t\le c \end{cases}$$

Note that the existence of $\tilde{f}^{(i)}(b)$ (i.e., the requirement that $\tilde{f}_L^{(i)}(b)$, $\tilde{f}_R^{(i)}(b)$ exist and are equal) is equivalent to the requirement $f_L^{(i)}(b)=(g\circ h)_R^{(i)}(b)$. This is basically the chain rule argument given in (Bartels, 1984) . See also (Barsky, 1984) and (Ramshaw, 1984) . The right side is expanded using the chain rule and, taking $\beta_i :=$ $h_R^{(i)}(b)$, we will get the same formula for any k as we get from characterization 2.

If there exists an h such that the first k derivatives of $\tilde{f}$ can be made continuous, this means that the trace of the degree k Taylor polynomial of the parameterization f of one side can be made to coincide with the trace of the degree k Taylor polynomial of the reparameterization, $g\circ h$, of the parameterization g of the other side. What about the converse — does the trace of the degree k Taylor polynomials coinciding for one pair of C^k-regular parameterizations imply that the union of the two traces can be parameterized by a C^k-regular curve $\tilde{f}$? We prove that this is so in characterization 4.

Characterization 4: Characterization 4 is a manifestation of the condition that the curve trace can be written as an immersion which has a k-jet. Suppose that

$$\sum_{i=1}^{k}\frac{g_R^{(i)}(u_0)}{i!}(u-u_0)^i =: \hat{g}(u)$$

and

$$\sum_{i=1}^{k}\frac{f_L^{(i)}(t_0)}{i!}(t-t_0)^i =: \hat{f}(t)$$

have the same trace and that $g_R^{(1)}(u_0)\neq 0\neq f_L^{(1)}(t_0)$. It does not make sense to refer to g as a reparameterization of f in a neighborhood of the joint because the traces of f and g do not overlap. However, the traces of their Taylor polynomials do, so we can now refer to one Taylor polynomial as being a reparameterization of the other. $\hat{f}^{(1)}(t_0)\neq 0$ because $f_L^{(1)}(t_0)\neq 0$ (and $\hat{g}^{(1)}(u_0)\neq 0$ because $g_R^{(1)}(u_0)\neq 0$). Thus, $\hat{f}^{-1}$ exists on some $\mathbb{R}^n$-neighborhood of $\hat{f}(t_0)$. We are using the notation for the inverse of a function loosely here. We mean that $\hat{f} \circ \hat{f}^{-1}$ is the identity on some neighborhood of t_0 and $\hat{f}^{-1}|_C \circ \hat{f}$ is the identity on some C-neighborhood of $f(t_0)$, where C is the portion of the trace of $\hat{f}$ that falls within the domain of $\hat{f}^{-1}$. Let $h := \hat{f}^{-1} \circ \hat{g}$ then

$$\sum_{i=1}^{k}\frac{g_R^{(i)}(u_0)}{i!}(u-u_0)^i = \sum_{i=1}^{k}\frac{f_L^{(i)}(t_0)}{i!}(h(u)-t_0)^i$$

Since two polynomials that agree on a neighborhood of a point agree everywhere, h can be extended to be defined for all u. Since $h(u_0) = t_0$,

$$\sum_{i=1}^{k}\frac{g_R^{(i)}(u_0)}{i!}(u-u_0)^i = \sum_{i=1}^{k}\frac{f_L^{(i)}(t_0)}{i!}(h^{(1)}(u_0)(u-u_0)+\frac{h_L^{(2)}(u_0)}{2!}(u-u_0)^2$$

$$+\ldots+\frac{h_L^{(k)}(u_0)}{k!}(u-u_0)^k+o((u-u_0)^{k+1})^i$$

$$=\sum_{i=1}^{k}\frac{f_L^{(i)}(t_0)}{i!}((h^{(1)}(u_0)^i(u-u_0)^i+\ldots)$$

Equating coefficients of $(u-u_0)^i$ for $i=1,..,n$ gives the chain rules of orders 1 through n. This is a consequence of the observation that if $q,r:\mathbb{R}\to\mathbb{R}^n$, $s:\mathbb{R}\to\mathbb{R}$ are analytic and $q=r \circ s$ then equating coefficients of $(u-u_0)^k$ in

$$\sum_{k=0}^{\infty}\frac{q^{(k)}(u_0)}{k!}(u-u_0)^k = \sum_{l=0}^{\infty}\frac{r^{(l)}(s(u_0))}{l!}(\sum_{m=0}^{\infty}\frac{s^{(m)}(u_0)}{m!}(u-u_0)^m-s(u_0))^l$$

gives the chain rule of order k. One also has to note that the chain rule of order k is also obtained if each of the three power series is truncated at order k.

Is there a restriction on the β_i's in view of the fact that they arise from the chain rule or in view of the expressions from characterization 2? For β_1, any positive real number is permissible. Negative values correspond to cusps and zero is not possible because β_1 is the scalar factor relating the one-sided derivatives of the two pieces, which are both nonzero because the parameterization is regular. We will find in constructing the β-splines that they have the partition of unity property for only some values of the β_i's. (The partition of unity property is the property that each β-spline is everywhere non-negative and the sum of all the β-splines is 1.) Is this because these values of the β_i's cannot be realized through the chain rule? We do not know.

Characterization 5: The Monge Taylor map exists to degree k at the joint. Some practical methods for determining when this happens are given on pages 175, 176 of (Bruce, 1984)

.

4.2 A Historical Perspective of Geometric Continuity

We have given several characterizations of geometric differentiabilty of arbitrary order, although they are hardly as intuitively geometric as the characterizations of first and second geometric differentiability given by contact with once (resp. twice) geometrically differentiable measuring sticks — lines (resp. circles). Tangent lines and osculating circles are not new to differential geometry but Barsky has taken a fresh (at least, it was not emphasized in books that the author has read) and very interesting view towards them: contact of order two of a curve trace with a line constitutes a geometric kind of first derivative and contact of order three with a circle constitutes a geometric kind of second derivative. His work begs the question 'What is a geometric, third-derivative measuring-stick equivalent of line and circle?'.

There were two major discoveries, in the author's opinion, that led to the idea of geometric continuity as presented in this thesis, and to the resulting β-splines. The first was Barsky's PhD thesis, which (1) made explicit the idea the *order* of differentiability in a nonparametric (or geometric) sense, (2) gave a solid description of first and second

geometric continuity for curves and (3) derived the β-splines of degree ≤ 3.*
The second was the observation that Barsky's conditions (to be discussed in this chapter) look like the chain rules of order 0, 1, 2. ((Bartels, 1983) and (Ramshaw, 1984)) The origin of this thesis is really the quest to give geometric meaning to this chain rule argument. In fact, a parallel thesis by one of Barsky's graduate students, T. DeRose, embarked in the same direction and the same synopsis of the chain rule argument is given in a resulting technical report (Barsky, 1984) (and in (Ramshaw, 1984)) as is given here. That is, it is a special case of the requirement that k derivatives with respect to arc length (equivalently, a C^k-regular parameterization) exist.

(Goodman, 1985) proposed an extension of the algebraic conditions defining β_1 and β_2 to formal algebraic conditions for β's associated with higher-order derivatives. These conditions were proposed as a generalization of β-splines not on geometric principles but because they permitted the theory which Goodman had developed for cubics to be extended to higher degrees. A linear map on a vector space of polynomials, say $\mathbb{P}^k$, can be regarded as a linear map of polynomial coefficients or, equivalently, as a linear map of the derivatives of the polynomial. Because the conditions at each knot for Goodman's splines, C^k-splines and G^k-splines all have the form that a linear combination of the one-sided derivatives from the left of the knot is equal to a linear combination of the derivatives from the right of the knot, at the knot, they are all g-splines (page 316 of (Schumaker, 1981)).

The author feels that there are pointers in chapter 3 for further research (especially in singularity theory) but acknowledges failure, so far, in establishing anything like the substantial theory of geometric continuity that there is for orders one and two.

---

* Actually, others besides Barsky in Computer Aided Geometric Design have published this notion. For example, (Farin, 1983) calls it visual C^k-continuity.

5. Beta Splines and Tensor Products of Them

In this chapter, we construct a vector space of geometrically continuous polynomial splines and a useful basis for it, the β-splines. Then, we consider tensor products of copies of this vector space and look briefly at surfaces obtained this way.

5.1 A Vector Space of Geometrically Continuous Polynomial Splines

Definition 5.1.1: Let $\hat{\mathbb{P}}_{k,(t_i)_{i=0}^m}$ denote the vector space of piecewise polynomial functions with knot vector $(t_i)_{i=0}^m$ whose pieces all have degree $\leq k$ and are $k-1$ times geometrically continuous.*

(This is isomorphic to a subspace of the the direct sum of m copies of $\mathbb{P}_k$ by the map which takes a polynomial segment into the polynomial with the same coefficients .) Let $\hat{\mathbb{P}}^n_{k,(t_i)_{i=0}^m}$ be the direct sum of n copies of $\hat{\mathbb{P}}_{k,(t_i)_{i=0}^m}$. If $\{b_i\}_{i=1}^m$ is a basis for $\hat{\mathbb{P}}_{k,(t_i)_{i=0}^m}$, then

$$(1,0,..,0)b_1,..,(1,0,..,0)b_m$$

$$\cdots$$

$$(0,..,0,1)b_1,..,(0,..,0,1)b_m$$

is a basis for $\mathbb{P}^n_{k,(t_i)_{i=0}^m}$. Members of $\mathbb{P}^n_{k,(t_i)_{i=0}^m}$ have the form $f(u) = \sum_{i=1}^m v_i b_i(u)$. where $v_i \in \mathbb{R}^n$ for $i=1,..,m$.

Next we show that the chain rule argument applied to a spline curve f in $\mathbb{R}^n$ just reduces to the chain rule on each of the basis functions b_i (independent of n). See also (Barsky, 1984) .

* As yet, nobody has worked with spline spaces where the curve is G^{k_i} in a neighborhood of t_i (i.e. geometrically continuous splines with confluence).

Proposition 5.1.2: Let D_L (resp. D_R) denote the derivative from the left (resp. right). Let f_i denote the piece of f on $t_{i-1} \leq t \leq t_i$ and b_{ij} denote the piece of b_i on $t_{j-1} \leq t \leq t_j$. Let $c_{kr}(t_i)$ be defined by

$$D_R^k \,|_{t_i} f = \sum_{r=0}^{k} c_{kr}(t_i) D_L^r \,|_{t_i} f$$

(i.e., by applying the chain rule argument to f) then

$$D_R^k \,|_{t_i} b_i(t) = \sum_{r=0}^{k} c_{kr}(t_i) D_L^r \,|_{t_i} b_i(t).$$

Proof: Substituting the expression for f in the "chain rule" gives

$$\sum_{i=1}^{m} v_i D_R^k \,|_{t_j} b_i(t) = \sum_{i=1}^{m} v_i \sum_{r=0}^{k} c_{kr}(t_i) D_L^r \,|_{t_i} b_i(t).$$

Since the v_i are arbitrary,

$$i\, D_R^k \,|_{t_i} b_i(t) = \sum_{r=0}^{k} c_{kr}(t_i) D_L^r \,|_{t_i} b_i(t) \quad \square$$

Because this condition is independent of n, we don't have to recompute the b_i's for each n. We can do it once and for all. Figure 6 is the graph of a uniform cubic β-spline with $\beta_1 = 4$ and $\beta_2 = 0$, while figure 7 is the graph of a uniform cubic β-spline with $\beta_1 = 1/4$ and $\beta_2 = 0$ (see appendix 1). The first reaction to one of these things is to ask, "How could such an awful looking thing give a curve with a continuous unit tangent vector?". One must remember that the graph of a β-spline (i.e. $u \mapsto (u, b(u))$) with $\beta_1 \neq 1$ is not a G^1 curve (because $u \mapsto u$ does not have the required first derivative jump). It is not the graph of the β-splines that one should look at. They are always taken in combinations with points in $\mathbb{R}^n$ as coefficients. The point is that the same jump is being introduced in each co-ordinate and the jump is controlled so that the jump in parameter space does not show up as a visual jump on the curve. We started by worrying that parametric continuous differentiability might not be good enough to guarantee geometric continuity. But, it turns out that for polynomial splines, parametric continuous differentiability does guarantee geometric continuity. In fact, the space of continuously differentiable splines (with no confluence) is a proper subspace of the space of geometrically continuous splines. The β_i's give additional degrees of freedom.

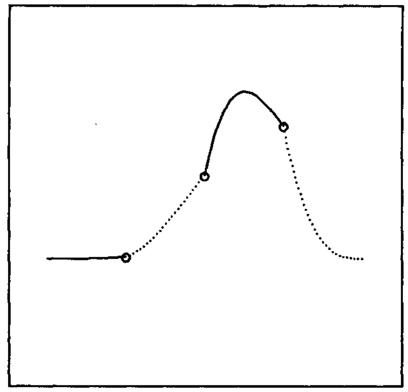

Figure 6: $\beta_1 = 4$, $\beta_2 = 0$ (at each knot)

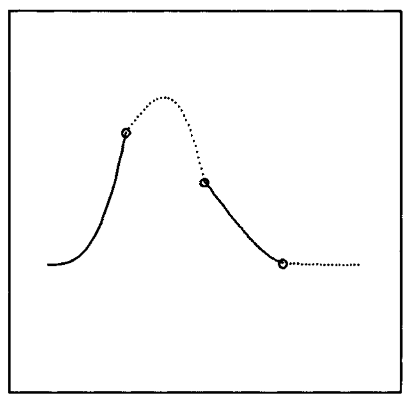

Figure 7: $\beta_1 = .25$, $\beta_2 = 0$ (at each knot)

Are C^k splines useful in some applications where other β-splines would not be? As another example to indicate that the answer might be no, consider the differential equation

$$\overset{(1)}{x} = f(x(t))$$

$$x(t)=(x_1(t),x_2(t))$$

C^k splines are widely used for obtaining approximate solutions to differential equations. They can be integrated easily to yield amazingly good results even for functional differential equations near unstable limit cycles. What we really want is trajectories with a continuously changing tangent line. G^k splines give us this and they provide a larger class of potential solution curves.

We describe several methods for obtaining β-splines. The actual programs to do this are in appendix 1. Wherever the methods overlap, they produce the same results. Program 1 in the appendix generates and solves the linear system obtained by applying the chain rule argument with the same choice of β_i's at every knot and with unit knot spacing, (Bartels, 1983) . The denominator common to each of the β-spline pieces is the determinant of the coefficient matrix of the linear system that was solved in the program. The zeroes of this determinant correspond to choices of the β's for which the β-splines are no longer a basis for the spline space. For cubic β-splines with uniform knot spacing and the same β's at all knots and $\beta_1 = 1$, the determinant is zero at $\beta_2 = -12$. At $\beta_2 = -4$, the first derivative of the graph of the β-spline goes from having 1 zero to having 3 zeroes (The graph of the β-spline goes from looking like the back of a dromedary to looking like the back of a bactrian) and at $\beta_2 = -8$, the β-spline goes negative (and continues to be negative for some portions of the curve for all $\beta_2 \leq -8$). The subprogram "usgspline" of program 1 is itself a program which has the property that the basis functions it produces sum to 1. What aspect of the constraints is giving rise to the property that these basis functions sum to 1? Besides summation to 1, the positivity of the basis functions for a wide range of the β's is a curious artifact of the construction. Finding out what these ranges are would be useful. As with B-splines, the positivity of β-splines can be investigated by considering collections of them as Tchebycheff systems. See (Rice, 1964) and (Schumaker, 1981) for more on this.

In program 2 of appendix 1, the knot spacing is still uniform, but now the β's can be different at different knots. Here, making the splines add up to 1 at the endpoint does not make them add up to 1 on the whole interval.

With non-uniform knot spacing, the question is: what do we replace the condition above of summation to 1 at a point with? In program 3 of appendix 1, we first construct the one-sided basis of (Bartels, 1983) . To construct a β-spline f_{ii} starting at knot uu_{ii}, start by requiring that $f_{ii}(u) = 0 \ \forall \ u \ \in [uu_{n+1}, uu_{n+2}) =: I$. The piece of f_{ii} on I is a non-trivial linear combination of the onesided functions starting at $uu_{ii}, .., uu_{ii+n+1}$ (The first $n+1$ of these one-sided functions will be linearly independent except for some interesting choices of the β's.) because $n+2$ pieces of degree $\leq n$ cannot possibly linearly independent. We have $n+2$ coefficients to determine, but requiring the linear combination to be 0 gives us $n+1$ conditions, one for each power, so we have one condition left over. If we require $f_{jj}, .., f_{jj+n}$ to add up to 1 on $uu_{jj+n} \leq u \leq uu_{jj+n+1}$ then there are $n+1$ conditions to satisfy, but there are $n+1$ functions, each with 1 free condition, so we now have the same number of conditions as unknowns: $(n+1)(n+2)$. VAXIMA could handle nothing more than the linear case.

In program 4, we give the construction of the β-splines given in (Bartels, 1983) . Here the β-splines are constructed by analogy with the divided difference definition of B-splines. For, β-splines however, the construction is not exactly divided differencing. Surprisingly, these 'divided difference' β-splines appear to add up to 1. The output for the quadratic case is given in appendix 1. It looks from the quadratic case that the β's are local parameters of the β-splines insofar as changing the β's at one of the knots affects only those β-splines which are non-zero there.

The bases dual to the truncated powers and B-splines (pages 102 and 116 of (de Boor, 1978) , respectively) are used to some advantage by de Boor, so it is probably worth trying to do the same for the geometrically continuous truncated powers and the β-splines.

The coefficients $\{v_i\}_{i=1}^{p}$ (called the control vertices) of B-splines could have arisen from any application (interpolation, approximation, whatever). But, because they have compact support, once one has expressed a G^k polynomial spline in terms of β-splines, a single v_i or $\beta_j(r)$ can be varied without altering any but the B-splines that are non-negative for this parameter. This is called *local control* and is important in computer aided design. Two procedures useful for B-splines in computer graphics are: (1) taking the coefficients as input, and (2) given a function g, taking the coefficients $v_i = g(t_i^*)$, where $t_i^* = \frac{t_{i+1}+..+t_{i+k-1}}{k-1}$. The second choice gives rise to splines with the property that any straight line crosses g at least as many times as it crosses the spline. See p. 160, 161 of (de Boor, 1978) . This is called the "shape preserving" or "variation diminishing" property. It would be interesting to see if the same choice for β-splines also gives rise to a variation diminishing spline.

5.2 Tensor Product Splines

The tensor product of the original spline space with itself is a vector space. One builds up surfaces in $\mathbb{R}^n$ by taking control vertex combinations as we did for curves: $S(u,v)=\Sigma w_{jk} b_k(u) b_j(v)$ i.e. by building up direct sums of the tensor product.

One can show (just by checking ratios of the three 2-by-2 subdeterminants of the Jacobian) that such tensor product surfaces have a unit normal (hence, a tangent plane) that changes continuously accross patches of the surface. The argument for Bézier patches in (Kahmann, 1983) should be extendible to an argument about continuity of curvature over patch boundaries for tensor products of β-splines. Tensor product patches can be pathological in the sense that they are not what we have defined as surfaces. If, however, we take enough control points that a becomes an interior point of the union of the resulting patches, then there is also a tangent plane at a. This is because parallelism is transitive and the unit normals of the surface approach eachother in pairs at patch boundaries, so all the unit normals approach (up to sign) a common value. In figure 2, which is a poor rendering of something which is supposed to look like a sphere, we see that the lonely point in figure 1 has become an interior point when more control points (and thus patches) were added. The control points for figure 1 (proceeding from left to right, then top to bottom) are:

-1	1	0	-1	-1	0	1	-1	0	1	1	0
0	1	1	0	-1	1	0	-1	-1	0	1	-1
1	1	0	1	-1	0	-1	-1	0	-1	1	0
0	1	-1	0	-1	-1	0	-1	1	0	1	1

The 16 control points for figure 1 are obtained by taking the vertices of a square (the first row) and rotating them 3 times by $\frac{\pi}{2}$. The "sphere" is obtained by repeating the first 3 entries of each row and appending them to the end of the row (to get 7 columns per row) and then repeating the first 3 rows and appending these to the 4 for the single patch (to get a total of 7 rows of 7 columns), in the same way that the cylinder and doughnut are obtained in (Bartels, 1983)

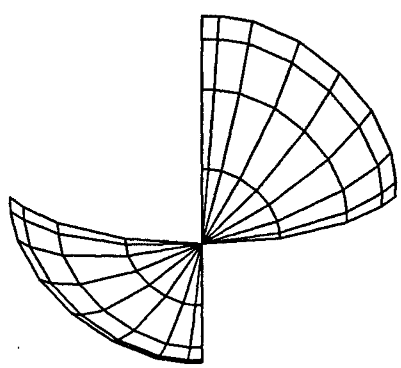

Figure 8: A Patch of a "Sphere"

Just as a curve can be be given invariantly (and uniquely, up to orientation) as the vector field of unit tangent vectors tangent to it, so a (nice enough) surface trace can be invariantly (and uniquely up to orientation) given as the vector field of unit normal vectors. The maps $u \mapsto \frac{\frac{\mathrm{d}g}{\mathrm{d}u}}{\|\frac{\mathrm{d}g}{\mathrm{d}u}\|}$ and $(u,v) \mapsto \frac{\frac{\partial f}{\partial u} \times \frac{\partial f}{\partial v}}{\|\frac{\partial f}{\partial u} \times \frac{\partial f}{\partial v}\|}$ (Gauss map) are fundamental to many of the constructs in classical Differential Geometry (see chapter 3 of (do Carmo, 1976)). For surfaces in higher dimensions, there is no such unique vector field. This is an indication that we will have to be content to look at equivalence classes of objects rather than uniquely determined objects.

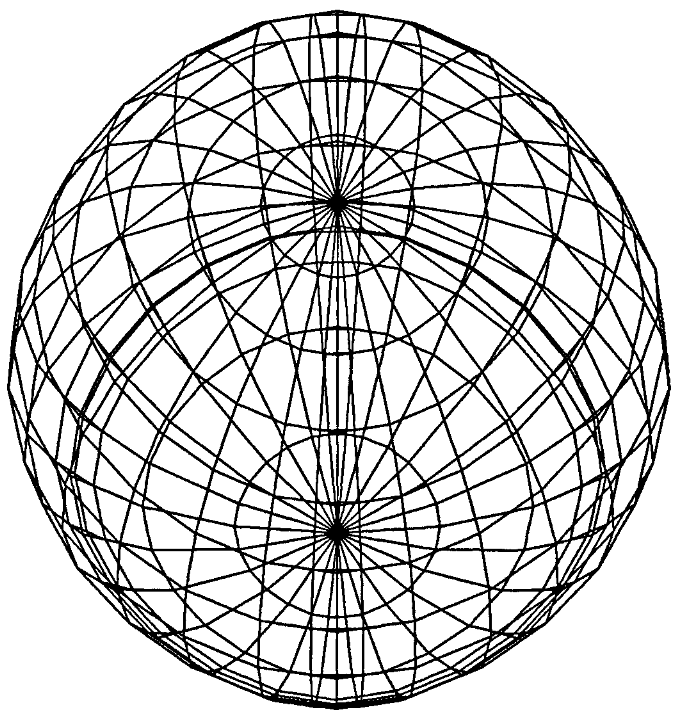

Figure 9: "Sphere"

The problem of matching up pieces of splines at boundaries is more complicated for surfaces than for curves. The problem of matching up patches has been researched more for polyhedral splines than for splines which preserve geometric properties.

6. Polyhedral B-Splines

If, in the β-splines, we take $\beta_1=1$ and $\beta_k=0\ \forall k>1$ (this is the case that the parameterization is C^k), we get the B-splines. In fact, B-splines are also defined with confluence (where we let the degree of continuous differentiability vary from knot to knot). These univariate B-splines have been generalized through the Hermite-Gennochi formula to a multivariate analogue called polyhedral B-splines.* Over the last eight years there has been active research by de Boor, Dahmen, Höllig and Micchelli on polyhedral B-splines. (Recently, others have started studying them as well.)

In 1976, de Boor introduced the simplicial B-spline in the literature, attributing their motivation to Schoenberg, who did a huge amount of work on univariate B-splines and gave them their name. B is for basis of $\mathbb{P}_{k,\xi,\nu}$, defined as follows. $\mathbb{P}_{k,\xi,\nu}$ is the spline space of piecewise polynomials of degree $k-1$ on a knot vector ξ such that each member is C^{ν_i-1} at ξ_i.) Schoenberg claims that Laplace must have known about B-splines because of their convolution properties. In his PhD thesis and the resulting paper (Kergin, 1980) , through trying to introduce a multivariate divided difference (a whole class of them satisfying properties like Leibniz rules have since been introduced), Kergin introduced a natural kind of interpolation in $\mathbb{R}^n$. Both of these results have since been generalized. In 1982, the polyhedral B-spline was introduced in the literature by de Boor and Höllig (de Boor, 1982) . Three kinds of convex polyhedra B have been studied so far for the purposes of interpolation and approximation: simplices (the original ones), cones (also discussed in terms of multivariate truncated powers) and cubes (box splines - these are the easiest to deal with because a linear independent collection with good approximation power can be obtained just by taking translates of one of them). Variants of Kergin's interpolation scheme have also been discussed in the literature.

* We stay away from the terminology "multivariate B-splines" because this may refer to either simplicial B-splines or polyhedral B-splines.

On p.59 of (de Boor, 1982) , de Boor says that Kergin's work "in a way gave impetus to all the material yet to be discussed in these lectures" (which was polyhedral B-splines and approximation from the span of polyhedral B-splines). The hardest proof in Kergin's paper is an application of Stoke's theorem. de Boor and Höllig state that the recurrence relations for polyhedral B-splines are a simple consequence of Stoke's theorem. Stoke's theorem is a differential geometric result — in fact Kergin used differential forms and exterior derivatives in his proofs.

If $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ and $X : \mathbb{R}^n \overset{\text{affine}}{\rightarrow} \mathbb{R}^s$ then define $M_\omega(x\,|X)$ by

$$\int_{\mathbb{R}^s} f(x) M_\omega(x\,|X)\mathrm{d}x = \int_{\mathbb{R}^n} \omega(u) f(X(u))\mathrm{d}u \quad \forall f \in C_0(\mathbb{R}^s).$$

$M_\omega(x\,|X)$ is, in general, a distribution but, for the special case that ω is the characteristic function of a convex polyhedron (See page 35 of (Dahmen, 1983) for a precise description of the condition on ω.), that makes $M_\omega(x\,|X)$ correspond to a polyhedral B-spline, it can be identified with a function. (A useful list of the distributions that result for various choices of ω is given in (Dahmen, 1983) .) A large amount of the research on this subject so far has been spent on trying to determine linearly independent collections of polyhedral B-splines that are useful for approximation and interpolation. The difficulty in finding bases and their approximation power is closely related to how messy the problem of fitting these B-splines together differentiably can be. Now, the integration in the above formula is over a convex polyhedron, and polyhedra are the underlying sets of simplicial complexes. The study of fitting together complexes that are differentiably immersed or imbedded in a manifold and of approximating maps from simplicial complexes into manifolds is a part of Differential Topology (chapter 2 of (Munkres, 1966)). The integrals in the above equation would make sense on any oriented Riemannian manifold because any such manifold has a well determined volume element (that allows us to integrate functions). Triangulations correspond to diffeomorphisms in differential topology (see page 79 of (Munkres, 1966)), so if the imbeddings of the complexes are triangulations as well, the sense of approximation in differential topology may possibly be useful in studying the approximation of functions with polyhedral B-splines.

The questions of interest here are; (1) can multivariate β-splines be constructed in a way analogous to the polyhedral B-splines and used in approximation theory and (2) can the approximation of maps in differential topology be made to preserve geometric properties of the original map? Just as B-splines are special cases of β-splines in the univariate case, we are hoping that multivariate β-splines will turn out to be related to polyhedral B-splines.

Tensor products of univariate B-splines, besides being computationally very efficient, are also being found to have good approximation properties in the box spline investigations of de Boor and Höllig. But one should not conclude the other kinds of polyhedral B-splines will not be useful in computer graphics. Box splines arising from two dimensional grids are being found to give unified descriptions of many finite elements. As another example, the de Casteljau-recursions for Bézier polynomials (important in computer aided design) follow from the recurrence for simplicial B-splines.

Unfortunately, we do not have time to properly introduce the subject of polyhedral B-splines. The Hermite-Gennochi formula follows from Peano's theorem, page 70 of (Davis, 1963) . With this result in hand, (Micchelli, 1979) is perhaps the easiest-reading introduction to simplicial B-splines. It has some examples with nice pictures. Moreover, this paper proves the theorem on which de Boor and Höllig's generalization rests. A book is being written by Dahmen and Micchelli, that should be invaluable to those studying this branch of multivariate approximation theory. In the meantime there is their survey (Dahmen, 1983) that contains quite a bit of detail, as well as an excellent guide to the literature as of that moment. de Boor's survey (de Boor, 1982) is better for putting the study of polyhedral B-splines in the context of approximation theory but, being a large scale map, it is lacking the detail of Dahmen and Micchelli's survey.

7. Epilogue

This chapter contains a few notes on references that might prove useful to someone who wants to research geometric continuity.

Note 7.0.1: Differential Geometry is difficult for the outsider to approach, so we set aside a paragraph to suggest how one might make the approach. Differential Geometry should probably be learned in several passes, with the complexity and generality of (essentially the same) objects of study being increased at each pass. (Boothby, 1975) is an easily read introduction. The main object of study is the differentiable manifold. Group structure and connection structure are added slowly. A useful second pass, both for its use of modern language and for containing important results in Differential Geometry, is (Kobayashi, 1969) . The main object of study is the principle fiber bundle with connection. A third pass would nail down some of the category theory used in differential geometry. (Vaisman, 1973) is useful for this purpose.

Note 7.0.2: In chapter 6 we mentioned that, in 2 or more dimensions, piecing together splines is really a problem that has been studied in Differential Topology, and that piecing them together in such a way to preserve geometric properties would involve mixing notions from Differential Topology and Differential Geometry together. The best reference for this that we have found is (Munkres, 1966) .

Note 7.0.3: The author spent much time trying to think of an intrinsic characterization of geometric continuity. Prof. Webb convinced him that this is not possible. In view of the fact that the atlas on $S:=\{(x,y)\colon y=|x|<1\}$ consisting of the single chart h defined in chapter 1 is a smooth manifold structure on S, manifold structure alone does not provide enough intrinsic information to give the notion of geometric continuity that we want. Apparently, no matter how much differentiability we give a curve, the only intrinsic information about it we get is the arc length between any two of its points. For a surface, there are more intrinsic, invariant measures (like curvature) but the point is that there a differences between surfaces like a sheet and a cone that are not intrinsic. Prof. Webb explained the additional intrinsic information in surfaces as follows. One can bend a piece of string all sorts of ways without breaking it (without changing the arc length between any two points), but one cannot do the equivalent thing with a piece of paper. We can roll a piece of paper into a cylinder, but we cannot bring the ends of the cylinder together to form a torus without crinkling the piece of paper. Thus the difference between a torus and a sheet can be measured intrinsically, but the difference between a sheet and a cylinder cannot. Also, the difference between any two curves with the same length cannot be measured intrinsically.

Note 7.0.4: The notion of transversality is a generalization of the notion of a regular value, which we have seen is crucial to the notion of geometric continuity. (Bruce, 1984) provides easy access to at least the introductory level of this field. (Guillemin, 1974) is a popular readable book on Differential Topology that has a description of singularity theory. Unfortunately, the author didn't find this area of study until late in his research. Singularity theory and the Monge-Taylor map are used to prove some very interesting results in curve theory. For example, the statement that "almost any compact plane curve $\gamma: I \rightarrow \mathbb{R}^2$ has only ordinary inflections and vertices" appears throughout (Bruce, 1984) and is proven on pages 177-181.

Appendix 1: VAXIMA Programs for Constructing β-splines

This appendix assmbles together the MACSYMA programs discussed in chapter 5, their output for a few examples and a small discussion of some implementation details. All programs were run with the VAXIMA version of MACSYMA on the University of Waterloo Computer Graphics Laboratory's VAX 11/780 ("WATCGL"). This machine is operated under BSD 4.2 UNIX an provides each process with 8 MB of actual memory and 24 MB of virtual memory. As VAXIMA is currently configured on WATCGL, 3 MB are taken from actual memory by the VAXIMA system before execution begins.

Program 1: Uniform β-splines with the same β's at all knots

```
/*
 *   "n" must be assigned the positive integer degree of the uniform
 * beta-splines with the same betas at all knots before executing this
 * program.
 *   The first part of this program develops the chain rule for the first
 * "n"-1 derivatives.  See "betaspl_1.v" for a description of the algorithm.
 */
n: n-1$    /* Interpret "n" as the order of the chain rule for a while. */
q(t):= sum( beta[r]*t^r/r!, r, 1, n )$
p[0](t):=  ev(  deriv[0]   +  sum( deriv[r]*q(t)^r/r!, r, 1, n )  )$
coeff[0,0]: 1$
for i: 1 thru n do
   (
   coeff[i,0]: 0,
   p[i](t):= diff( p[i-1](t), t ),
   for j: 1 thru i do
      coeff[i,j]: diff( ev(p[i](0)), deriv[j] )
   )$
/*
 *   The second part of this pgm constructs the degree "n"
 * uniform beta splines with the same betas at every knot and
 * shows that they add up to 1.
 */
n: n+1$    /* From now on, "n" is the degree of the beta-splines. */
batch("usgspline.v")$
```

The subprogram "usgspline.v" that this program calls is useful in itself for examining the properties of splines that arise from linear systems of the same form, but that use different constraints than the geometric constraints of the first part of the program.

```
/*
 *   Construct a subclass of the degree "n" uniform polynomial g-splines
 * with the same collection of linear functionals applied at every knot.
 * These g-splines allow for the restricted class of
 * linear constraints that are used in the construction of B- and beta-
 * splines.  See the comments in the program for a description of the
 * special structure of these conditions.
 */
/* The following line imposes the condition that the functions be continuous. */
coeff[0,0]: 1$
/* The following line imposes the condition that derivatives to the right
 * of a knot do not depend on the zeroth derivative to the left of the knot.
 */
for i: 1 thru n do coeff[i,0]: 0$
np1: n+1$
nm1: n-1$
np1tn: np1*n$
for k: -np1 step 1 thru -1 do
   b[0](k,u):=  sum( a[k,i]*u^i, i, 1, n )  +  a[k,0]$
for k: -nm1 step 1 thru -1 do
   for i: 1 step 1 thru nm1 do
      b[i](k,u):= diff( b[i-1](k,u), u )$
/*    The following loop defines the equations that impose the conditions
 * that the zeroth through n-1th derivatives must be zero at the
 * endpoints of the compact support.
 */
for i: 0 step 1 thru nm1 do
   (
   eq[i+1]: b[i](-1,0) = 0,
   eq[np1tn+i+1]: sum( coeff[i,j]*b[j](-np1,1), j, 0, i ) = 0
   )$
/*    The conditions in the following loop
 * are not the most general g-spline conditions that could be
 * imposed at the knots - the ith derivative at the right of the knot
 * is a linear combination of those derivatives to the left
 * of the knot that are only of orders between 1 and i inclusive.  The
 * general case would allow for linear combinations of derivatives
 * from the left of orders from 0 up to and including order n.
 */
for k: 1 step 1 thru n do
   for i: 0 step 1 thru nm1 do
      eq[k*n+i+1]:  sum( coeff[i,j]*b[j](-k,1), j, 0, i )  -
                    b[i](-k-1,0)  =  0$
/*    This special constraint scales beta-splines to sum to 1.  Of
 * special interest to this thesis are the properties of other g-splines
 * that also have the property that forcing them to sum to 1 at the knots
 * causes them to sum to 1 everywhere.
 */
eq[np1*np1]:  sum( b[0](k,0), k, -np1, -2 )  -  1  =  0$
eqlist: makelist( eq[i], i, 1, np1*np1 )$
xlist: makelist( a[-1,i], i, 0, n )$
for k: 2 step 1 thru np1 do
   xlist: append( xlist,  makelist( a[-k,i], i, 0, n )  )$
linsolve( eqlist, xlist ), globalsolve: true$
/*typeset: true$*/
for k: 1 step 1 thru np1 do
   (
   print( k, "th piece:" ),
   disp(factor(b[0](-k,u)))
```

```
   )$
ratsimp(  sum( b[0](k,u), k, -np1, -1 )  );
```

The program has to solve an $(n+1)^2$ by $(n+1)^2$ system, although it is sparse. VAX-IMA as presently configured on WATCGL could not do anything more than the quartic case.

The results of previous program for the quartic case follow. (The last statement in the program has produced 1 for $n=1,2,3,4$.)

First piece:

$$- [6u^4]$$

/

$$[\beta_1^2\beta_3 - \beta_3 - 3\beta_1\beta_2^2 - 6\beta_1^3\beta_2 - 18\beta_1^2\beta_2 - 18\beta_1\beta_2 - 6\beta_2 - 6\beta_1^6 - 18\beta_1^5 - 30\beta_1^4 - 36\beta_1^3 - 30\beta_1^2 - 18\beta_1 - 6]$$

Second Piece:

$$[3\beta_3u^4 + 18\beta_1\beta_2u^4 + 6\beta_2u^4 + 6\beta_1^3u^4 + 6\beta_1^2u^4 + 6\beta_1u^4 + 6u^4 - 4\beta_3u^3 - 36\beta_1\beta_2u^3 - 24\beta_1^3u^3 - 12\beta_2u^2 - 36\beta_1^2u^2 - 24\beta_1u - 6]$$

/

$$[\beta_1^2\beta_3 - \beta_3 - 3\beta_1\beta_2^2 - 6\beta_1^3\beta_2 - 18\beta_1^2\beta_2 - 18\beta_1\beta_2 - 6\beta_2 - 6\beta_1^6 - 18\beta_1^5 - 30\beta_1^4 - 36\beta_1^3 - 30\beta_1^2 - 18\beta_1 - 6]$$

Third Piece:

$$[3\beta_1^2\beta_3u^4 - 3\beta_3u^4 - 9\beta_1\beta_2^2u^4 - 6\beta_1^3\beta_2u^4 - 18\beta_1^2\beta_2u^4 - 18\beta_1\beta_2u^4 - 6\beta_2u^4 - 6\beta_1^5u^4 - 6\beta_1^4u^4 - 12\beta_1^3u^4 - 6\beta_1^2u^4 - 6\beta_1u^4 - 8\beta_1^2\beta_3u^3 + 4\beta_3u^3 + 24\beta_1\beta_2^2u^3 + 24\beta_1^3\beta_2u^3 + 36\beta_1^2\beta_2u^3 + 36\beta_1\beta_2u^3 + 24\beta_1^5u^3 + 24\beta_1^4u^3 + 24\beta_1^3u^3 + 6\beta_1^2\beta_3u^2 - 18\beta_1\beta_2^2u^2 - 24\beta_1^3\beta_2u^2 + 12\beta_2u^2 - 36\beta_1^5u^2 + 36\beta_1^3u^2 + 36\beta_1^2u^2 - 36\beta_1^2\beta_2u - 48\beta_1^4u - 48\beta_1^3u + 24\beta_1u - \beta_3 - 18\beta_1\beta_2 - 6\beta_2 - 18\beta_1^3 - 30\beta_1^2 - 18\beta_1]$$

/

$$[\beta_1^2\beta_3 - \beta_3 - 3\beta_1\beta_2^2 - 6\beta_1^3\beta_2 - 18\beta_1^2\beta_2 - 18\beta_1\beta_2 - 6\beta_2 - 6\beta_1^6 - 18\beta_1^5 - 30\beta_1^4 - 36\beta_1^3 - 30\beta_1^2 - 18\beta_1 - 6]$$

Fourth Piece

$$- [\beta_1(3\beta_1\beta_3u^4 - 9\beta_2^2u^4 - 6\beta_1^2\beta_2u^4 - 18\beta_1\beta_2u^4 - 6\beta_1^5u^4 - 6\beta_1^4u^4 - 6\beta_1^3u^4 - 6\beta_1^2u^4 - 8\beta_1\beta_3u^3 + 24\beta_2^2u^3 + 24\beta_1^2\beta_2u^3 + 36\beta_1\beta_2u^3 + 24\beta_1^5u^3 + 24\beta_1^4u^3 + 24\beta_1^3u^3 + 6\beta_1\beta_3u^2 - 18\beta_2^2u^2 - 24\beta_1^2\beta_2u^2 - 36\beta_1^5u^2 - 36\beta_1^4u^2 + 36\beta_1^2u^2 - 36\beta_1\beta_2u + 24\beta_1^5u - 48\beta_1^3u - 48\beta_1^2u - \beta_1\beta_3 + 3\beta_2^2 + 6\beta_1^2\beta_2 + 18\beta_1\beta_2 + 18\beta_1^4 + 30\beta_1^3 + 18\beta_1^2)]$$

/

$$[\beta_1^2\beta_3 - \beta_3 - 3\beta_1\beta_2^2 - 6\beta_1^3\beta_2 - 18\beta_1^2\beta_2 - 18\beta_1\beta_2 - 6\beta_2 - 6\beta_1^6 - 18\beta_1^5 - 30\beta_1^4 - 36\beta_1^3 - 30\beta_1^2 - 18\beta_1 - 6]$$

Fifth Piece:

$- [6\beta_1^6(u - 1)^4]$
/
$[\beta_1^2\beta_3 - \beta_3 - 3\beta_1\beta_2^2 - 6\beta_1^3\beta_2 - 18\beta_1^2\beta_2 - 18\beta_1\beta_2 - 6\beta_2 - 6\beta_1^6 - 18\beta_1^5 - 30\beta_1^4 - 36\beta_1^3 - 30\beta_1^2 - 18\beta_1 - 6]$

Program 2: Uniform β-splines

Mainline:

```
n: 2$
n: n-1$
batch("betaspl_1.v")$
n: n+1$
batch("ubeta_2.v")$
batch("ubeta_3.v")$
quit()$
```

betaspl_1.v:

```
/*
   This program develops the chain rule for the first
"n" derivatives.  "tp[0](t)" is the first "n"+1 terms of the Taylor
expansion of p(q(t)) about 0, where p(0)=0=q(0) and p, q are "n"
times differentiable but are otherwise arbitrary.  "beta[r](k)"
is the "r"th derivative of q - for beta-splines, it
corresponds to the "r"th derivative of the change of parameter
from the "k"th piece to the ("k"+1)th piece at the knot between
them ("uu[k]").  "c[i,j](k)" is the coefficient of the
"j"th derivative in the "i"th order chain rule at knot "k".
*/
q(t):= sum( beta[r](k)*t^r/r!, r, 1, n )$
tp[0](t):=  ev(  deriv[0]   +  sum( deriv[r]*q(t)^r/r!, r, 1, n )  )$
/*
tp[0](t)=deriv[0]+deriv[1]/1! (beta[1](k) t /1!+...+beta[n](k) t^n /n!)^1
             +...+deriv[n]/n! (beta[1](k) t /1!+...+beta[n](k) t^n /n!)^n
*/
define( c[0,0](k), 1 )$
for i: 1 thru n do
   (
   define( c[i,0](k), 0 ),
   tp[i](t):= diff( tp[i-1](t), t ),
   for j: 1 thru i do
      define(  c[i,j](k),  diff( tp[i](0), deriv[j] )  )
   )$
```

ubeta_2.v:

```
/*
 *   This program sets up and solves the linear system for the degree ''n''
 * uniform beta splines.
 * ''b[i](k,u)'' is the ''i''th derivative of the ''k''th piece.
 * 0<=''u''<=1 is the parameter of each piece.
 */
np1: n+1$
nm1: n-1$
np1tn: np1*n$
```

```
for k: 1 thru np1 do
   b[0](k,u):=  sum( a[k,i]*u^i, i, 1, n )  +  a[k,0]$
for k: 1 thru nm1 do
   for i: 1 step 1 thru nm1 do
      b[i](k,u):= diff( b[i-1](k,u), u )$
for i: 0 thru nm1 do
   (
   eq[i+1]: b[i](1,0) = 0,
   eq[np1tn+i+1]: sum( c[i,j](np1)*b[j](np1,1), j, 0, i ) = 0
   )$
for k: 1 thru n do
   for i: 0 thru nm1 do
      eq[k*n+i+1]:  sum( c[i,j](k)*b[j](k,1), j, 0, i )  -
                    b[i](k+1,0)  =  0$
eq[np1*np1]:  sum( b[0](k,0), k, 2, np1 )  -  1  =  0$
eqlist: makelist( eq[i], i, 1, np1*np1 )$
xlist: makelist( a[1,i], i, 0, n )$
for k: 2 step 1 thru np1 do
   xlist: append(  xlist,  makelist( a[k,i], i, 0, n )  )$
linsolve( eqlist, xlist ), globalsolve: true$
```

ubeta_3.v:

```
eq[0]:   sum(  d[k]*b[0](k,0),  k,  1,  np1  )   =   1$
for i: 1 thru np1 do
   eq[i]:    ev(  sum(  d[k]*coeff( b[0](k,u), u, i ),  k,  1,  np1  )   )
          =   0$
eqlist:  makelist( eq[i], i, 0, n )$
xlist:  makelist( d[i], i, 1, np1 )$
linsolve( eqlist, xlist ),  globalsolve: true$
total: 0$
typeset: true$
for k: 1 thru np1 do
   (
   x: d[k]*b[0](k,u),
   total: ev(total+x),
   print("piece #",k,":"),
   disp(factor(x)),
   print("new piece #",k,"= old piece #",k,"times:"),
   disp(factor(d[k]))
   )$
typeset: false$
ratsimp(total);
```

The following is the output for quadratic beta splines with uniform knot spacing. As configured on WATCGL, VAXIMA could not solve the second linear system (in the "d"s) for the cubic case, although it did solve the first system.

piece #1:
[u^2]
/
[$\beta_1[2] + 1$]

new piece #1= *old piece* #1 *times*:
[$\beta_1[1]\beta_1[2] + 2\beta_1[2] + 1$]
/
[$(\beta_1[2] + 1)^2$]

piece #2:
- [$\beta_1[1]\beta_1[2]u^2 + 2\beta_1[1]u^2 + u^2 - 2\beta_1[1]\beta_1[2]u - 2\beta_1[1]u - \beta_1[2] - 1$]
/
[$(\beta_1[1] + 1)(\beta_1[2] + 1)$]

new piece #2= *old piece* #2 *times*:
[$\beta_1[1]\beta_1[2] + 2\beta_1[2] + 1$]
/
[$(\beta_1[1] + 1)(\beta_1[2] + 1)$]

piece #3:
[$\beta_1[1](u - 1)^2$]
/
[$\beta_1[1] + 1$]

new piece #3= *old piece* #3 *times*:
[$\beta_1[1](\beta_1[1]\beta_1[2] + 2\beta_1[2] + 1)$]
/
[$(\beta_1[1] + 1)^2\beta_1[2]$]

Program 3: β-splines which are designed to add up to 1

MAINLINE:

```
/*
 *    "n" must be assigned the (positive integer) degree of the beta splines
 * desired either in this file or before reading this file into VAXIMA.
 */
n: 1$
batch("betaspl_1.v")$
kill(allbut(c,n))$
batch("betaspl_2.v")$
kill(allbut(n,nm1,np1,p))$
batch("betaspl_3.v")$
kill(allbut(n,nm1,np1,p,eqlist,xlist))$
batch("betaspl_4.v")$
quit()$
```

SECOND PART (betaspl_2.v):

```
/*
Find the truncated power functions.
"p[j](i,u)" is the piece of (the truncated power fn starting at
"uu[i]") on the interval "uu[i+j]"<="u"<="uu[i+j+1]".
*/
np1: n+1$
nm1: n-1$
p[j](i,u):= (u-uu[i])^n +
            sum(sum( a[r,k](i)*(u-uu[i+r])^k,
```

```
                  k, 1, nm1 ),
                  r, 1, j)$
for j: 1 thru np1 do
   for s: nm1 step -1 thru 1 do
      a[j,s](i):=
         ev(
         (
         sum(c[s,l](i+j)  *  subst( uu[i+j], u, diff(p[j-1](i,u),u,l) ),
            l, 1, s-1)
         +(c[s,s](i+j)-1)  *  subst( uu[i+j], u, diff(p[j-1](i,u),u,s) )
         )/s!
         )$
for j: 0 thru np1 do
   define( p[j](i,u), ratsimp(p[j](i,u)) )$
```

THIRD PART (betaspl_3.v):

```
/*
Construct the beta-splines.
The beta-spline starting at knot "uu[ii]" has value
            j
           ---
b[j](ii):= >   d[i](ii) p[j-i](ii+i,u)
           ---
           i=0
for "uu[ii+j]"<="u"<="uu[ii+j+1]".
*/
betaend(ii):= ratsimp(
                     sum(d[i](ii)*p[np1-i](ii+i,u),
                        i,0,np1)
                     )$
for ii: jj thru jj+n do
   (
   t: betaend(ii),
   for i: 0 thru np1 do
      (
      t:  ev( subst(cd[i,ii],d[i](ii),t) )
      ),
   for k: 0 thru n do
      (
      eq[(ii-jj)*np1+k+1]:  ev( coeff(t,u,k) )  =  0
      )
   )$
/*
These previous equations were the "zero on the last interval" conditions.
There are "n"+1 equations, one for each power of "u" on this last
interval, and "n"+2 unknowns, "d[i](ii)", for "i"=0 to "n"+1.  We
will put "n"+1 of these linear systems together, so that there are
"n"+1 beta-splines (starting at "uu[jj]",..,"uu[jj+n]")
in the works and so that we have "n"+1 degrees of freedom,
which are enough to guarantee summation to 1 on
"uu[jj+n]"<="u"<="uu[jj+n+1]", i.e. to guarantee
     jj+n                       jj+n  jj+n-ii
    -----                      -----  -----
1 = >     b[jj+n-ii](ii)  [= >       >     d[i](ii) p[jj+n-ii-i](ii+i,u)]
    -----                      -----  -----
    ii=jj                      ii=jj   i=0
In total, there are ("n"+1)("n"+2) equations in that many unknowns,
the "d[i](ii)"s.
```

```
*/
mess: ratsimp(
            sum(sum(cd[i,ii]*p[jj+n-ii-i](ii+i,u),
                    i,0,jj+n-ii),
                ii,jj,jj+n)
            )$
eq[np1^2+1]: ev(coeff(mess,u,0))-1=0$
for i: 1 thru n do
   eq[np1^2+i+1]: ev(coeff(mess,u,i))=0$
eqlist: makelist(eq[i], i, 1, np1^2+np1 )$
grind(eqlist)$
xlist: makelist( cd[i,jj], i, 0, np1 )$
for ii: 1 thru n do
   xlist:   append(  xlist,  makelist( cd[i,ii+jj], i, 0, np1 )  )$
grind(xlist)$
```

FOURTH PART (betaspl_4.v):

```
/*
 *    Solve the linear system that was constructed in "betaspl_3.v" and
 *  print out the solution (i.e. the beta-spline pieces).
 */
linsolve( eqlist, xlist ),  globalsolve: true$
for ii: jj thru jj+n do
   for j: 0 thru n do
      (
      write("piece #",j+1," of the beta-spline starting at uu[",ii,"]:"),
      disp(factor(ratsimp(sum( cd[i,ii]*p[j-i](ii+i,u), i, 0, j ))))
      )$
```

Program 4: The Divided Difference β-splines

Mainline:

```
n: 2$
showtime: true$
nm1: n-1$
np1: n+1$
n: nm1$
batch("betaspl_1.v")$
n: n+1$
kill(allbut(c,n,nm1,np1,uu))$
batch("dd_2.v")$
kill(allbut(n,nm1,np1,ac,uu))$
batch("dd_3.v")$
quit()$
```

dd_2.v:

```
/*
SECOND PART: TRUNCATED POWER FNS
--------------------------------
"trp[j](ii,u)" is the piece of (the truncated power fn starting at
"uu[ii]") on the interval "uu[ii+j]"<="u"<="uu[ii+j+1]".
                              j  n-1
                        n   --- ---                       k
```

```
trp[j](i,u) = (u-uu[i]) + >    >  a[r,k](i)  (u-uu[i+r])
                          --- ---
                          r=1 k=1
                            n-1
                            ---                        k
trp[j](i,u)-trp[j-1](i,u) = >  a[j,k](i)  (u-uu[i+j])
                            ---
                            k=1
                                n-1
 s                              ---    k!                     k-s
D (trp[j](i,u)-trp[j-1](i,u)) = >    ------ a[j,k](i)  (u-uu[i+j])
                                --- (k-s)!
                                k=s
 s
D |        (trp[j](i,u)-trp[j-1](i,u)) = s! a[j,s](i)
   uu[i+j]
By this eqn and the defn of the c's,
 s                      s
D |        trp[j](i,u) = D |        trp[j-1](i,u) + s! a[j,s](i)
   uu[i+j]                  uu[i+j]
   s
  ---             p
= >  c[s,p](i+j) D |        trp[j-1](i,u)       i.e.
  ---               uu[i+j]
  p=1
               s-1
            1  ---             p
a[j,s](i) = --- >  c[s,p](i+j) D |        trp[j-1](i,u)
            s! ---                uu[i+j]
               p=1
                                  s
          + (c[s,s](i+j) - 1) D |        trp[j-1](i,u)
                                 uu[i+j]
This equation is useful as a recurrence if one goes from "j"=1 to "n"+1
and, for each "j", from "s"="n"-1 to 1.
*/
trp[j](ii,u):= (u-uu[ii])^n +
               sum(sum( a[r,k](ii)*(u-uu[ii+r])^k,
                   k, 1, nm1 ),
                   r, 1, j)$
for j: 1 thru np1 do
   for s: nm1 step -1 thru 1 do
      define(a[j,s](ii),
            ev(
              (
              sum(c[s,l](ii+j)*subst( uu[ii+j], u, diff(trp[j-1](ii,u),u,l) ),
                 l, 1, s-1)
              +(c[s,s](ii+j)-1)*subst( uu[ii+j], u, diff(trp[j-1](ii,u),u,s) )
              )/s!
              )
            )$
/*
"ac[j,k](ii)" is defined by
               n
              ---             k
trp[j](ii,u) =: >   ac[j,k](ii) u
              ---
              k=0
(Note that "ac[j,n](ii)"=1.)
```

```
*/
for j: 0 thru np1 do
   (
   define( ac[j,n](ii), 1 ),
   define( trp[j](ii,u), ratsimp(trp[j](ii,u)) ),
   for k: 0 thru nm1 do
      define(ac[j,k](ii),
            ev(ratcoef( trp[j](ii,u), u, k ))
            )
   )$
```

dd_3.v:

```
/*
THIRD PART: CONSTRUCT THE 'DIVIDED DIFFERENCE' BETA SPLINES
-----------------------------------------------------------
"aa[0,i,k](ii)" is the coefficient, on "uu[ii+n+1]" <= "u"
<="uu[ii+n+2]", of "u^k" in "p[n+1-i](ii+i,u)" i.e. "aa[0,i,k](ii)"
= "ac[n+1-i,k](ii+i)".

Construct the beta spline which starts at "uu[i]".  Find
"c[it](i)" so that
n+1
---
>   c[it](i) aa[0,it,k](i) = 0
---
it=0
by 'divided differencing'.
Define the "ndd"th 'divided difference' "lp[ndd,nfn](i,u)", on
"uu[i+n+1]"<="u"<="uu[i+n+2]", of the truncated power functions
starting at "uu[nfn]",..,"uu[nfn+ndd]" by:
                                         n
                                        ---                  k
lp[0,nfn](i,u):= p[np1-nfn](i+nfn,u) =  >    aa[0,nfn,k](i) u
                                        ---
  n                                     k=0
 ---                         k
= >    ac[0,n+1-nfn,k](i+nfn) u
 ---
 k=0
for 0<="nfn"<="n+1", and:

                        lp[ndd-1,nfn+1](i,u) - lp[ndd-1,nfn](i,u)
lp[ndd,nfn](i,u):=  -------------------------------------------------     (1)
                    aa[ndd-1,nfn+1,n-ndd](i) - aa[ndd-1,nfn,n-ndd](i)

for 1<="ndd"<="n+1" and 0<="nfn"<="n+1-ndd", where "aa[ndd,nfn,k](i)"
for 1<="ndd"<="n+1" is defined by:
                    n-ndd
                    -----                k
lp[ndd,nfn](i,u) =: >      aa[ndd,nfn,k](i) u                             (2)
                    -----
                     k=0
"aa[ndd,nfn,n-ndd]"=1 as in the "ndd"=0 case so
for "ndd"="n+1" the recurrence for "lp" reduces to:

lp[n+1,0](i,u):= lp[n,1] - lp[n,0]  (=1-1=0)
```

```
Putting (2) into (1):

lp[ndd,nfn](i,u) =
n-ndd
-----       aa[ndd-1,nfn+1,k](i) - aa[ndd-1,nfn,k](i)        k
>       ------------------------------------------------   u
-----   aa[ndd-1,nfn+1,n-ndd](i) - aa[ndd-1,nfn,n-ndd](i)
 k=0
*/
for ndd: 1 thru n do
   for nfn: 0 thru np1-ndd do
      (
      define(den[ndd,nfn](ii),
            aa[ndd-1,nfn+1,n-ndd](ii) - aa[ndd-1,nfn,n-ndd](ii)
            ),
      define(place[ndd,nfn](ii),
            ( place[ndd-1,nfn+1](ii) - place[ndd-1,nfn](ii) )
            / den[ndd,nfn](ii)
            ),
      for power: 0 thru nm1-ndd do
         (
         define(aa[ndd,nfn,power](ii),
               (
               aa[ndd-1,nfn+1,power](ii) - aa[ndd-1,nfn,power](ii)
               )/den[ndd,nfn](ii)
               )
         )
      )$
define( place[np1,0](ii), place[n,0](ii)-place[n,1](ii) )$
ptop: ratnumer(place[np1,0](ii))$
define( pbot(ii), ratdenom(place[np1,0](ii)) )$
for i: 0 thru np1 do
   define(ct[i](ii),
         ratdiff( ptop, place[0,i](ii) )
         )$
/*
Tell VAXIMA what the "aa"s are so it can evaluate the "ct"s and
"pbot" in gory detail.
*/
for i: 0 thru np1 do
   (
   define( aa[0,i,n](ii), 1 ),
   for k: 0 thru nm1 do
      define( aa[0,i,k](ii), ev(ac[np1-i,k](ii+i)) )
   )$
for i: 0 thru n do
   define( ct[i](ii), ratsimp(ct[i](ii)) )$
define( pbot(ii), ratsimp(pbot(ii)) );
kill(allbut(n,nm1,np1,ct,pbot,ac,uu))$
/*
FOURTH PART:  CHECK WHETHER THE 'DD' BETA SPLINES ADD UP TO 1
--------------------------------------------------------------
Let "d[j](i,u)" denote the "j+1"th piece of (the "divided difference"
beta spline starting at "uu[i]").  Then,
            j                             j            n
           ---                           ---          ---                     k
d[j](i,u) = >   c[ti](i) p[j-ti](i+ti,u) = >   c[ti](i) >   ac[j-ti,k](i+ti) u
           ---                           ---          ---
           ti=0                          ti=0         k=0
```

```
    n      j
   --- ( ---                         )  k
 = >    ( >    c[ti](i) ac[j-ti,k](i+ti) ) u
   --- ( ---                         )
   k=0    ti=0
                               n
                              ---
We want to check whether >    d[j](jj+n-j,u) = 1, i.e whether
                              ---
                              j=0
 n    j
---  ---                                     (1 if k=1
>     >   c[ti](jj+n-j) ac[j-ti,k](jj+n-j+ti) = (
---  ---                                     (0 if 1<k<=n
j=0  ti=0
c[ti](i) = ct[ti](i)/pbot(i)
*/
/*
for k: n step -1 thru 0 do
   (
   t: 0,
   for j: 0 thru n do
      for ti: 0 thru j do
         t: t+ev(ct[ti](jj+n-j))*ev(ac[j-ti,k](jj+n-j+ti))/ev(pbot(jj+n-j)),
   print("coefficient of ",k,"th degree term"),
   disp(ratsimp(t))
   )$
*/
/*
We could have avoided the computational hardships of polynomial
division by multiplying through by
 n
---
| | pbot(jj+j)
j=0
but VAXIMA runs quickly out of memory.
*/
for j: 0 thru n do
   (
   print( "piece #", j+1 ),
   grind(factor(ratsimp(
        sum(
        sum( ct[ti](ii)*ac[j-ti,k](ii+ti)/pbot(ii), ti, 0, j )*u^k,
        k, 0, n)
        )      )       )
   )$
```

As configured on WATCGL, this program verifies that these divided difference β-splines add up to 1 for the quadratic case but runs out of memory for the cubic case. The pieces for the quadratic case follow.

piece #1

$$\frac{(u_i - u)^2}{(u_i - u_{i+1})(\beta_1[i+1](u_{i+1} - u_{i+2}) - u_{i+1} + u_i)}$$

piece #2

$[\ \beta_1[i+2]\ (\beta_1[i+1]\ (u_{i+3}\ u^2\ +\ u_{i+2}\ (-u^2\ -\ 2u_{i+3}\ u)\ +\ 2u_{i+2}^2\ u\ +\ u_{i+1}\ (2u_{i+2}\ u_{i+3}\ -\ 2u_{i+2}^2)\ +\ u_{i+1}^2\ (u_{i+2}\ -\ u_{i+3}))\ +\ u_i\ (u_{i+2}\ u_{i+3}\ +\ u_{i+1}\ (u_{i+2}\ -\ u_{i+3})\ -\ u_{i+2}^2)\ +\ u_{i+1}\ (u_{i+2}^2\ -\ u_{i+2}\ u_{i+3})\ +\ u_{i+1}^2\ (u_{i+3}\ -\ u_{i+2}))\ +\ \beta_1[i+1]\ (2u_{i+2}\ u^2\ +\ u_{i+1}\ (2u_{i+2}^2\ -\ 2u^2)\ +\ u_{i+1}^2\ (2u\ -\ 2u_{i+2})\ -\ 2u_{i+2}^2\ u)\ +\ u_{i+1}\ (u^2\ -\ u_{i+2}^2)\ +\ u_i\ (-u^2\ +\ u_{i+1}\ (2u\ -\ 2u_{i+2})\ +\ u_{i+2}^2)\ +\ u_{i+1}^2\ (2u_{i+2}\ -\ 2u)\ ]$

/

$[\ (\beta_1[i+1]\ (u_{i+1}\ -\ u_{i+2})\ -\ u_{i+1}\ +\ u_i)\ (u_{i+1}\ -\ u_{i+2})\ (\beta_1[i+2]\ (u_{i+2}\ -\ u_{i+3})\ -\ u_{i+2}\ +\ u_{i+1})\ ]$

piece #3

$$\frac{\beta_1[i+2](u_{i+3} - u)^2}{(\beta_1[i+2](u_{i+2} - u_{i+3}) - u_{i+2} + u_{i+1})(u_{i+2} - u_{i+3})}$$

References

Barsky, 1981.

B. Barsky, *The Beta-spline: A Local Representation Based on Shape Parameters and Fundamental Geometric Measures,* University of Utah (December, 1981). PhD dissertation, Department of Computer Science

Barsky, 1984.

B. Barsky and T. DeRose, Geometric Continuity of Parametric Curves, UCB/CSD 84/205, University of California at Berkeley, Berkeley (1984).

Bartels, 1984.

R. Bartels, Splines in Interactive Computer Graphics, pp. 1-29 in *Numerical Analysis*, ed. D. Griffiths, Springer-Verlag, Berlin (1984). Proceedings, June 28-July 1, 1983

Bartels, 1983.

R. Bartels and J. Beatty, *Beta-splines With A Difference,* Department of Computer Science, University of Waterloo, Waterloo (1983). TR CS-83-40

Bartels, 1983.

R. Bartels, J. Beatty, and B. Barsky, *An Introduction to the Use of Splines in Computer Graphics,* Department of Computer Science, University of Waterloo, Waterloo (1983). TR CS-83-09

Bishop, 1964.

R. Bishop and R. Crittenden, *Geometry of Manifolds,* Academic Press (1964).

Boothby, 1975.

W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry,* Academic Press (1975).

Bruce, 1984.

J. Bruce and P. Giblin, *Curves and Singularities,* Cambridge University Press, Cambridge (1984).

Conte, 1980.

S. Conte and C. de Boor, *Elementary Numerical Analysis,* McGraw-Hill Book Company, New York (1980).

Dahmen, 1983.

W. Dahmen and C. Micchelli, Recent Progress in Multivariate Splines, pp. 27-121 in *Approximation Theory IV*, ed. J. Ward, Academic Press, Inc., New York (1983). Proceedings, January 10-14, 1983

Davis, 1963.

P. Davis, *Interpolation and Approximation,* Blaisdell Publishing Company, New York (1963).

Dugundji, 1970.

J. Dugundji, *Topology,* Allyn and Bacon, Inc., Boston (1970).

Farin, 1983.

G. Farin, Smooth Interpolation to Scattered 3D Data, pp. 43-64 in *Surfaces in Computer Aided Geometric Design*, ed. W. Boehm, North-Holland Publishing Company, Amsterdam (1983). Proceedings; April 25-30, 1982

Faux, 1979.

I. Faux and M. Pratt, *Computational Geometry for Design and Manufacture,* John Wiley and Sons (1979).

Gardner, .

R. Gardner, *Notes*.

Goetz, 1970.

A. Goetz, *Introduction to Differential Geometry,* Addison-Wesley, Reading, Massachusetts (1970).

Goodman, 1985.

T. Goodman, Properties of β-Splines, *Journal of Approximation Theory (to appear),* (1985).

Groetsch, 1980.

C. Groetsch, *Elements of Applicable Functional Analysis,* Marcel Dekker, Inc., New York (1980).

Guillemin, 1974.

V. Guillemin and A. Pollack, *Differential Topology,* Prentice-Hall, Inc., Englewood Cliffs (1974).

Hobson, 1957.

E. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series,* Dover Publications, Inc., New York (1957).

Kahmann, 1983.

J. Kahmann, Continuity of Curvature between Adjacent Bezier Patches, pp. 65-76 in *Surfaces in Computer Aided Geometric Design*, ed. W. Boehm, North-Holland Publishing Company, Amsterdam (1983). Proceedings; April 25-30, 1982

Keller, 1974.

H. Keller, *Differental Calculus in Locally Convex Spaces,* Springer-Verlag, Berlin (1974).

Kergin, 1980.

P. Kergin, A Natural Interpolation of C^k Functions, *Journal of Approximation Theory* **29** pp. 278-293 Academic Press, (1980).

Klingenberg, 1978.

W. Klingenberg, *A Course in Differential Geometry,* Springer-Verlag, New York (1978).

Kobayashi, 1969.

S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry - Volume I,* Interscience Publishers, New York (1969).

Maurin, 1976.

K. Maurin, *Analysis - part I: Elements,* PWN - Polish Scientific Publishers, New York (1976).

Munkres, 1966.

J. R. Munkres, *Elementary Differential Topology,* Princeton University Press - Annals of Mathematics Studies, Princeton, New Jersey (1966). Lectures given at M.I.T. in Fall, 1961

Pogorelov, .

A. Pogorelov, *Differential Geometry,* P. Noordhoff N.V., Groningen, The Netherlands ().

Ramshaw, 1984.

L. Ramshaw, *A Euclidean View of Joints between Bézier Curves,* Xerox Palo Alto Research Center (1984).

Rice, 1964.

J. Rice, *The Approximation of Functions. Volume 1 - Linear Theory.,* Addison-Wesley Publishing Company, Inc., Reading, Massachusetts (1964).

Schumaker, 1981.

L. Schumaker, *Spline Functions Basic Theory,* John Wiley & Sons, New York (1981).

Vaisman, 1973.

I. Vaisman, *Cohomology and Differential Forms,* Marcel Dekker, Inc., New York (1973).

Vaisman, 1984.

I. Vaisman, *A First Course in Differential Geometry,* Marcel Dekker, Inc., New York (1984).

de Boor, 1978.

C. de Boor, *A Practical Guide to Splines,* Springer-Verlag, New York (1978).

de Boor, 1982.

C. de Boor, Topics in Multivariate Approximation Theory, pp. 39-78 in *Topics in Numerical Analysis*, ed. B. Eckmann, Springer-Verlag, Berlin (1982). Proceedings, July 19-August 21, 1981

de Boor, 1982.

C. de Boor and K. Höllig, Recurrence Relations for Multivariate B-splines, *Proceedings of the American Mathematical Society* **85** pp. 397-400 (1982).

do Carmo, 1976.

M. do Carmo, *Differential Geometry of Curves and Surfaces,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1976).

Glossary