I. Classical Differential Geometry of Curves

We shall begin with a few words on background material from prerequisite courses. This course explicitly assumes prior experience with the elements of linear algebra (including matrices, dot products and determinants), the portions of multivariable calculus involving partial differentiation, and some familiarity with the a few basic ideas from set theory such as unions and intersections. For the sake of completeness, a file describing the background material (with references to standard texts used in the Department's courses) is included in the course directory and can be found in the files called `background.*`, where $*$ is one of the extensions `dvi`, `ps`, or `pdf`.

Differential geometry uses ideas from calculus and vector algebra to obtain geometrical information about curves and surfaces. At many points it is necessary to work with topics from the prerequisites in a more sophisticated manner, and it is also necessary to be more careful in our logic to make sure that our formulas and conclusions are accurate. At numerous steps it might be necessary to go back and review things from earlier courses, and in some cases it will be important to understand things in more depth than one needs to get through ordinary calculus, multivariable calculus or matrix algebra. Frequently one of the benefits of a mathematics course is that it sharpens one's understanding and mastery of earlier material, and differential geometry certainly provides many opportunities of this sort.

The origins of differential geometry

Straight lines and circles have been central objects in geometry ever since its beginnings. During the fourth century B.C.E., Greek geometers began to study more general curves, starting with the ellipse, hyperbola and parabola. In the following centuries they discovered a large number of other curves and investigated the properties of such curves in considerable detail for a variety of reasons. The development of analytic geometry and calculus, particularly during the seventeenth and eighteenth centuries, yielded powerful techniques for analyzing curves and their properties. In particular, these advances created a unified framework for understanding the work of the Greek geometers and a setting for studying new classes of curves and problems beyond the reach of classical Greek geometry. Interactions with physics played a major role in the mathematical study of curves during that time, largely because curves provided a means for analyzing the motion of physical objects. By the beginning of the nineteenth century, the differential geoemtry of curves had begun to emerge as a subject in its own right.

This unit describes the classical nineteenth century theory of curves in the plane and 3-dimensional space. Some further results from the twentieth century will be discussed in the next unit.

References for examples

Here are some web links to sites with pictures and written discussions of many curves that mathematicians have studied during the past 2500 years:

`http://www-gap.dcs.st-and.ac.uk/~history/Curves/Curves.html`

`http://www.xahlee.org/SpecialPlaneCurves_dir/specialPlaneCurves.html`

`http://facstaff.bloomu.edu/skokoska/curves.pdf`

I.1 : Cross products

(do Carmo, § 1–4)

Courses in single variable or multivariable calculus usually define the cross product of two vectors and describe some of its basic properties. Since this construction will be particularly important to us and we shall use properties that are not always emphasized in calculus courses, we shall begin with a more detailed treatement of this construction.

Note on orthogonal vectors

One way of attempting to describe the dimension of a vector space is to suggest that the dimension represents the maximum number of mutually perpendicular directions. The following elementary result provides a formal justification for this idea.

PROPOSITION. *Let* $\mathbf{S} = \{\mathbf{a}_1, \cdots, \mathbf{a}_k\}$ *be a set of nonzero vectors that are mutually perpendicular. Then* $\mathbf{S}$ *is linearly independent.*

Proof. Suppose that we have an equation of the form

$$\sum_{i=1}^{n} c_i \mathbf{a}_i \quad = \quad \mathbf{0}$$

for some scalars c_i. If $1 \leq j \leq k$ we then have

$$0 \quad = \quad \mathbf{0} \cdot \mathbf{a}_j \quad = \quad \left(\sum_{i=1}^{n} c_i \mathbf{a}_i \right) \cdot \mathbf{a}_j \quad = \quad \sum_{i=1}^{n} (c_i \mathbf{a}_i \cdot \mathbf{a}_j)$$

and since the vectors in $\mathbf{S}$ are mutually perpendicular the latter reduces to $c_j |\mathbf{a}_j|^2$. Thus the original equation implies that $c_j |\mathbf{a}_j|^2 = 0$ for all j. Since each vector $\mathbf{a}j$ is nonzero it follows that $|\mathbf{a}_j|^2 > 0$ for all j which in turn implies $c_j = 0$ for all j. Therefore $\mathbf{S}$ is linearly independent.■

Properties of cross products

Definition. If $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$ are vectors in $\mathbf{R}^3$ then their cross product or vector product is defined to be

$$\mathbf{a} \times \mathbf{b} \quad = \quad \big(a_2 b_3 - a_3 b_2,\ a_3 b_1 - a_1 b_3,\ a_1 b_2 - a_2 b_1 \big) \ .$$

If we define unit vectors in the traditional way as $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, and $\mathbf{k} = (0, 0, 1)$, then the right hand side may be written symbolically as a 3×3 deterinant:

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & a_2 & a_3 \end{vmatrix}$$

The following are immediate consequences of the definition:

(1) $\mathbf{a}\times\mathbf{b} \quad = \quad -\mathbf{b}\times\mathbf{a}$

(2) $(c\,\mathbf{a})\times\mathbf{b} \quad = \quad c\,(\mathbf{a}\times\mathbf{b})$

(3) $\mathbf{a}\times(\mathbf{b}+\mathbf{c}) \quad = \quad (\mathbf{a}\times\mathbf{b}) \;+\; (\mathbf{a}\times\mathbf{c})$

Other properties follow directly. For example, by (1) we have that $\mathbf{a}\times\mathbf{a} = -\mathbf{a}\times\mathbf{a}$, so that $2\,\mathbf{a}\times\mathbf{a} = \mathbf{0}$, which means that $\mathbf{a}\times\mathbf{a} = \mathbf{0}$. Also, if $\mathbf{c} = (c_1, c_2, c_3)$ then the triple product

$$[\mathbf{c},\mathbf{a},\mathbf{b}] \quad = \quad \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b})$$

is simply the determinant of the 3×3 matrix whose rows are $\mathbf{c}$, $\mathbf{a}$, $\mathbf{b}$ in that order, and therefore we know that

the cross product $\mathbf{a}\times\mathbf{b}$ *is perpendicular to both* $\mathbf{a}$ *and* $\mathbf{b}$.■

The basic properties of determinants yield the following additional identity involving dot and cross products:

$$[\mathbf{c},\mathbf{a},\mathbf{b}] \quad = \quad [\mathbf{a},\mathbf{b},\mathbf{c}]$$

This follows because a determinant changes sign if two rows are switched, for the latter implies

$$[\mathbf{c},\mathbf{a},\mathbf{b}] \quad = \quad -[\mathbf{a},\mathbf{c},\mathbf{b}] \quad = \quad [\mathbf{a},\mathbf{b},\mathbf{c}]\ .■$$

The following property of cross products plays an extremely important role in this course.

PROPOSITION. *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are linearly independent, then* $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{a}\times\mathbf{b}$ *form a basis for* $\mathbf{R}^3$.

Proof. First of all, we claim that if $\mathbf{a}$ and $\mathbf{b}$ are linearly independent, then $\mathbf{a}\times\mathbf{b}\neq\mathbf{0}$. To see this we begin by writing out $|\mathbf{a}\times\mathbf{b}|^2$ explicitly:

$$|\mathbf{a}\times\mathbf{b}|^2 \quad = \quad (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2$$

Direct computation shows that the latter is equal to

$$(a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) \;-\; (a_1b_1 + a_2b_2 + a_3b_3)^2 \quad = \quad |\mathbf{a}|^2\,|\mathbf{b}|^2 \;-\; (\mathbf{a}\cdot\mathbf{b})^2$$

In particular, if $\mathbf{a}$ and $\mathbf{b}$ are both nonzero then

$$|\mathbf{a}\times\mathbf{b}| \quad = \quad |\mathbf{a}|\,|\mathbf{b}|\,|\sin\theta|$$

where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. Since the sine of this angle is zero if and only if the vectors are linearly dependent, it follows that $\mathbf{a}\times\mathbf{b}\neq\mathbf{0}$ if $\mathbf{a}$ and $\mathbf{b}$ are linearly independent.

Suppose now that we have an equation of the form

$$x\,\mathbf{a} + y\,\mathbf{b} + z(\mathbf{a}\times\mathbf{b}) \quad = \quad \mathbf{0}$$

for suitable scalars x, y, z. Taking dot products with $\mathbf{a}\times\mathbf{b}$ yields the equation $z|\mathbf{a}\times\mathbf{b}|^2 = 0$, which by the previous paragraph implies that $z = 0$. One can now use the linear independence of $\mathbf{a}$ and $\mathbf{b}$

to conclude that x and y must also be zero. Therefore the three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{a}\times\mathbf{b}$ are linearly independent, and consequently they must form a basis for $\mathbf{R}^3$.■

In may situations it is useful to have formulas for more complicated expressions involving cross products. For example, we have the following identity for computing threefold cross products.

"BAC—CAB" RULE. $\mathbf{a}\times\big(\mathbf{b}\times\mathbf{c}\big) = \mathbf{b}(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}(\mathbf{a}\cdot\mathbf{b})$, *or in more standard format the left hand side is equal to* $(\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}$.

Derivation. Suppose first that $\mathbf{b}$ and $\mathbf{c}$ are linearly dependent. Then their cross product is zero, and one is a scalar multiple of the other. If $\mathbf{b} = x\,\mathbf{c}$, then it is an elementary exercise to verify that the right hand side of the desired identity is zero, and we already know the same is true of the left hand side. If on the other hand $\mathbf{c} = y\,\mathbf{b}$, then once again one finds that both sides of the desired identity are zero.

Now suppose that $\mathbf{b}$ and $\mathbf{c}$ are linearly independent, so that $\mathbf{b}\times\mathbf{c}\neq\mathbf{0}$. Note that a vector is perpendicular to $\mathbf{b}\times\mathbf{c}$ if and only if it is a linear combination of $\mathbf{b}$ and $\mathbf{c}$. The ($\Longleftarrow$) implication follows from the perpendicularity of $\mathbf{b}$ and $\mathbf{c}$ to their cross product and the distributivity of the dot product, while the reverse implication follows because every vector is a linear combination

$$x\,\mathbf{b} + y\,\mathbf{c} + z\,(\mathbf{b}\times\mathbf{c})$$

and this linear combinationn is perpendicular to the cross product if and only if $z = 0$; *i.e.*, if and only if the vector is a linear combination of $\mathbf{b}$ and $\mathbf{c}$.

Since the vector $\mathbf{b}\times\big(\mathbf{b}\times\mathbf{c}\big)$ is perpendicular to $\mathbf{b}\times\mathbf{c}$ we may write it in the form

$$\mathbf{b}\times\big(\mathbf{b}\times\mathbf{c}\big) \quad = \quad x\,\mathbf{b} + y\,\mathbf{c}$$

for suitable scalars x and y. If we take dot products with $\mathbf{b}$ and $\mathbf{c}$ we obtain the following equations:

$$\mathbf{0} \;=\; [\mathbf{b},\mathbf{b},\mathbf{b}\times\mathbf{c}] \;=\; \Big(\mathbf{b}\cdot\big(\mathbf{b}\times(\mathbf{b}\times\mathbf{c})\big)\Big) \;=\; \mathbf{b}\cdot\big(x\,\mathbf{b}+y\,\mathbf{c}\big) \;=\; x\,(\mathbf{b}\cdot\mathbf{b}) + y\,(\mathbf{b}\cdot\mathbf{c})$$

$$-|\mathbf{b}\times\mathbf{c}|^2 \;=\; -[(\mathbf{b}\times\mathbf{c}),\mathbf{b},\mathbf{c}] \;=\; [\mathbf{b},(\mathbf{b}\times\mathbf{c}),\mathbf{c}] \;=\; [\mathbf{c},\mathbf{b},(\mathbf{b}\times\mathbf{c})] \;=$$

$$\Big(\mathbf{c}\cdot\big(\mathbf{b}\times(\mathbf{b}\times\mathbf{c})\;\big)\Big) \;=\; \mathbf{c}\cdot\big(x\,\mathbf{b}+y\,\mathbf{c}\big) \;=\; x\,(\mathbf{b}\cdot\mathbf{c}) + y\,(\mathbf{c}\cdot\mathbf{c})$$

If we solve these equations for x and y we find that $x = \mathbf{b}\cdot\mathbf{c}$ and $y = -\mathbf{b}\cdot\mathbf{b}$. Therefore we have

$$\mathbf{b}\times\big(\mathbf{b}\times\mathbf{c}\big) \quad = \quad (\mathbf{b}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{b}\cdot\mathbf{b})\,\mathbf{c})\ .$$

Similarly, we also have

$$\mathbf{c}\times\big(\mathbf{b}\times\mathbf{c}\big) \quad = \quad (\mathbf{c}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{b}\cdot\mathbf{c})\,\mathbf{c}))\ .$$

Therefore, if we write $\mathbf{a} = p\,\mathbf{b} + q\,\mathbf{c} + r\big(\mathbf{b}\times\mathbf{c}\big)$ we have

$$\mathbf{a}\times\big(\mathbf{b}\times\mathbf{c}\big) \quad = \quad p\,\mathbf{b}\times\big(\mathbf{b}\times\mathbf{c}\big) \;+\; q\,\mathbf{c}\times\big(\mathbf{b}\times\mathbf{c}\big) \quad =$$

$$\Big(p\,(\mathbf{c}\cdot\mathbf{c}) + q\,(\mathbf{c}\cdot\mathbf{c})\Big)\,\mathbf{b} \;-\; \Big(p\,(\mathbf{b}\cdot\mathbf{b}) + q\,(\mathbf{b}\cdot\mathbf{c})\Big)\,\mathbf{c}\ .$$

Since $\mathbf{b}$ and $\mathbf{c}$ are perpendicular to their cross product, the right hand side of the previous equation is equal to $(\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}$.∎

The formula for $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ yields numerous other identities. Here is one that will be particularly useful in this course.

PROPOSITION. *If* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ *and* $\mathbf{d}$ *are arbitrary vectors in* $\mathbf{R}^3$ *then we have the following identity:*

$$(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) \quad = \quad (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) \; - \; (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})$$

Proof. By definition, the expression on the left hand side of the display is equal to the triple product $[(\mathbf{a}\times\mathbf{b}),\mathbf{c},\mathbf{d}]$. As noted above, the properties of determinants imply that the latter is equal to $[\mathbf{d},(\mathbf{a}\times\mathbf{b}),\mathbf{c}]$, which in turn is equal to

$$\mathbf{d}\cdot\big(\mathbf{a}\times(\mathbf{b}\times\mathbf{c})\big) \quad = \quad \mathbf{d}\cdot\big((\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c}\big)$$

and if we expand the final term we obtain the expression $(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})$.∎

I.2 : Parametrized curves

(do Carmo, § 1–2)

There is a great deal of overlap between the contents of this section and certain standard topics in calculus courses. One major difference in this course is the need to work more systematically with some fundamental but relatively complex theoretical points in calculus that can be overlooked when working most ordinary and multivariable calculus problems. In particular this applies to the definitions of limits and continuity, and accordingly we shall begin with some comments on this background material.

Useful facts about limits

In ordinary and multivariable calculus courses it is generally possible to get by with only a vague understanding of the concept of limit, but in this course a somewhat better understanding is necessary. In particular, the following consequences of the definition arise repeatedly.

FACT I. *Let f be a function defined at all points of the interval $(a-h, a+h)$ for some $h > 0$ except possibly at a, and suppose that*

$$\lim_{x \to a} f(x) = b > 0 \ .$$

Then there is a $\delta > 0$ such that $\delta < h$ and $f(x) > 0$ provided $x \in (a-\delta, a+\delta)$ and $x \neq a$.

FACT II. *In the situation described above, if the limit exists but is* **negative**, *then there is a $\delta > 0$ such that $\delta < h$ and $f(x) > 0$ provided $x \in (a-\delta, a+\delta)$ and $x \neq a$.*

FACT III. *Each of the preceding statements remains true if 0 is replaced by an arbitrary real number.*

Derivation(s). We shall only do the first one; the other two proceed along similar lines. By assumption b is a positive real number. Therefore the definition of limit implies there is some $\delta > 0$ such that $|f(x) - b| < b$ provided provided $x \in (a-\delta, a+\delta)$ and $x \neq a$. It then follows that

$$f(x) \;=\; b + \big(f(x) - b\big) \;\geq\; b - |f(x) - b| \;>\; b - b \;>\; 0$$

which is what we wanted to show.∎

We shall also need the following statement about infinite limits:

FACT IV. *Let f be a continuous function defined on some open interval containing 0 such that f is strictly increasing and $f(0) = 0$. Then for each positive constant C there is a positive real number h sufficiently close to zero such that $x \in (0,\, h) \implies 1/f(x) > C$ and $x \in (-h,\, 0) \implies 1/f(x) < -C$.*

Proof. Let ε be the positive number $1/C$; by continuity we know that $|f(x)| < \varepsilon$ if $x \in (-h,\, h)$ for a suitably small $h > 0$. Therefore $x \in (0,\, h) \implies 0 < f(x) < \varepsilon$ and $x \in (-h,\, 0) \implies -\varepsilon < f(x) < 0$. The desired inequalities follow by taking reciprocals in each case.∎

There are two different but related ways to think about curves in the plane or 3-dimensional space. One can view a curve simply as a set of points, or one can view a curve more dynamically as a description of the position of a moving object at a given time. In calculus courses one generally adopts the second approach to define curves in terms of parametric equations; from this viewpoint one retrieves the description of curves as sets of points by taking the set of all points traced out by the moving object. For example, the line in $\mathbf{R}^2$ defined by the equation $y = m\,x$ is the set of points traced out by the parametrized curve defined by $x(t) = t$ and $y(t) = m\,t$. Similarly, the unit circle defined by the equation $x^2 + y^2 = 1$ is the set of points traced out by the parametrized curve $x(t) = \cos t$, $y(t) = \sin t$. The set of all points expressible as $\mathbf{x}(t)$ for some $t \in J$ will be called the *image* of the parametrized curve (since it represents all point traced out by the curve this set is sometimes called the *trace* of the curve, but we shall not use this term in order to avoid confusion with the entirely different notion of the trace of a matrix). We shall follow the standard approach of calculus books here unless stated otherwise.

A parametrized curve in the plane or 3-dimensional space may be viewed as a vector-valued function γ or $\mathbf{x}$ defined on some interval of the real line and taking values in $V = \mathbf{R}^2$ or $\mathbf{R}^3$. In this course we usually want our curves to be continuous; this is equivalent to saying that each of the coordinate functions is continuous. Given that this is a course in *differential* geometry it should not be surprising that we also want our curves to have some decent differentiability properties. If $\mathbf{x}$ is the vector function defining our curve and its coordinates are given by x_i, where i runs between 1 and 2 or 1 and 3 depending upon the dimension of V, then the derivative of $\mathbf{x}$ at a point t is defined using the coordinate functions:

$$\mathbf{x}'(t) = \big(\, x_1'(t),\, x_2'(t),\, x_3'(t)\,\big)$$

Strictly speaking this is the definition in the 3-dimensional case, but the adaptation to the 2-dimensional case is immediate — one can just suppress the third coordinate or view $\mathbf{R}^2$ as the subset of $\mathbf{R}^3$ consisting of all points whose third coordinate is zero.

Definition. A curve $\mathbf{x}$ defined on an interval J and taking values in $V = \mathbf{R}^2$ or $\mathbf{R}^3$ is *differentiable* if $\mathbf{x}'(t)$ exists for all $t \in J$. The curve is said to be *smooth* if $\mathbf{x}'$ is continuous, and it is said to be a *regular smooth curve* if it is smooth and $\mathbf{x}'(t)$ is nonzero for all $t \in J$. The curve will be said to be *smooth of class* $\mathcal{C}^r$ for some integer $r \geq 1$ if $\mathbf{x}$ has an r^{th} order continuous derivative, and the curve will be said to be smooth of class $\mathcal{C}^\infty$ if it is infinitely differentiable (equivalently, $\mathcal{C}^r$ for all finite r).

The crucial property of regular smooth curves is that they have well defined tangent lines:

Definition. Let $\mathbf{x}$ be a regular smooth curve and let a be a point in the domian J of $\mathbf{x}$. The *tangent line* to $\mathbf{x}$ at the parameter value $t = a$ is the unique line passing through $\mathbf{x}(a)$ and $\mathbf{x}(a) + \mathbf{x}'(a)$. There is a natural associated parametrization of this line given by

$$T(u) = \mathbf{x}(a) \;+\; u\,\mathbf{x}'(a)\ .$$

One expects the tangent line to be the "best possible" linear approximation to a smooth curve. The following result confirms this:

PROPOSITION. *In the notation above, if $u \neq 0$ is small and $a + u \in J$ then we have*

$$\mathbf{x}(a+u) \quad = \quad \mathbf{x}(u) \;+\; u\,\mathbf{x}'(a) \;+\; u\,\Theta(u)$$

where $\lim_{u \to 0} \Theta(u) = \mathbf{0}$*. Furthermore, if* $\mathbf{p}$ *is any vector such that*

$$\mathbf{x}(a+u) \quad = \quad \mathbf{x}(u) \; + \; u\,\mathbf{p} \; + \; u\,\mathbf{W}(u)$$

where $\lim_{u \to 0} \mathbf{W}(u) = \mathbf{0}$*, then* $\mathbf{p} = \mathbf{x}'(a)$*.*

Proof. Given a vector $\mathbf{a}$ we shall denote its i^{th} coordinate by a_i.

Certainly there is no problem writing $\mathbf{x}(a+u)$ in the form $\mathbf{x}(u) + u\,\mathbf{x}'(a) + u\,\Theta(u)$ for some vector valued function Θ; the substance of the first part of the proposition is that this function goes to zero as $u \to 0$. Limit identities for vector valued functions are equivalent to scalar limit identites for every coordinate function of the vectors, so the proof of the first part of the proposition reduces to checking that the coordinates θ_i of Θ satisfy $\lim_{u \to 0} \theta_i(u) = 0$ for all i. However, by construction we have

$$\theta_i(u) \quad = \quad \frac{x_i(a+u) - x_i(a)}{u} \; - \; x_i'(a)$$

and since $\mathbf{x}$ is differentiable at a the limit of the right hand side of this equation is zero. Therefore we have where $\lim_{u \to 0} \Theta(u) = \mathbf{0}$.

Suppose now that the second equation in the statement of the proposition is valid. As in the previous paragraph we have

$$w_i(u) \quad = \quad \frac{x_i(a+u) - x_i(a)}{u} \; - \; p_i(a)$$

but this time we know that $\lim_{u \to 0} w_i(u) = 0$ for all i. The only way these equations can hold is if $p_i(a) = x_i'(a)$ for all i.■

Piecewise smooth curves

There are many important geometrical curves that that are not smooth but can be decomposed into smooth pieces. One of the simplest examples is the boundary of the square parametrized in a counterclockwise sense. Specifically, take $\mathbf{x}$ to be defined on the interval $[0,4]$ by the following rules:

$$\begin{array}{llll} (a) & \mathbf{x}(t) & = & (t,0) \quad \text{for } t \in [0,1] \\ (b) & \mathbf{x}(t) & = & (1,t-1) \text{ for } t \in [1,2] \\ (c) & \mathbf{x}(t) & = & (2-t,1) \text{ for } t \in [2,3] \\ (d) & \mathbf{x}(t) & = & (0,1-t) \text{ for } t \in [3,4] \end{array}$$

The formulas for (a) and (b) agree when $t=1$, and likewise the formulas for (b) and (c) agree when $t=2$, and finally the formulas for (c) and (d) agree when $t=3$; therefore these formulas define a continuous curve. On each of the intervals $[n,n+1]$ for $n = 0,1,2,3$ the curve is a regular smooth curve, but of course the tangent vectors coming from the left and the right at these values are perpendicular to each other. Clearly there are many other examples of this sort, and they include all broken line curves. The following definition includes both these types of curves and regular smooth curves as special cases:

Definition. A continuous curve $\mathbf{x}$ defined on an interval $[a,b]$ is said to be a *regular piecewise smooth curve* if there is a partition of the inverval given by points

$$a \; = \; p_0 \; < \; p_1 \; \cdots \; < \; p_{n-1} \; < p_n \; = b$$

such that for each i the restriction $\mathbf{x}[i]$ of $\mathbf{x}$ to the subinterval $[p_{i-1}, p_i]$ is a regular smooth curve.

For the boundary of the square parametrized in the counterclockwise sense, the partition is given by

$$0 \;<\; 1 \;<\; 2 \;<\; 3 \;<\; 4\,.$$

Calculus texts give many further examples of such curves, and the references cited at the beginning of this unit also contain a wide assortment of examples. One important thing to note is that at each of the partition points p_i one has a left hand tangent vector $\mathbf{x}'(p_i-)$ obtained from $\mathbf{x}[i]$ and a right hand tangent vector $\mathbf{x}'(p_i+)$ obtained from $\mathbf{x}[i+1]$, but these two vectors are not necessarily the same. In particular, they do not coincide at the partition points $1, 2, 3$ for the parametrized boundary curve for the square that was described above.

Taylor's Formula for vector valued functions

We shall need an vector analog of the usual Taylor's Theorem for polynomial approximations of real valued functions on an interval.

VECTOR VALUED TAYLOR'S THEOREM. *Let* $\mathbf{g}$ *be a vector valued function defined on an interval* $(a-r, a+r)$ *that has continuous derivatives of all orders less than or equal to* $n+1$ *on that interval. Then for* $|h| < r$ *we have*

$$\mathbf{g}(a+h) \;=\; \mathbf{g}(a) \;+\; \sum_{k=1}^{n} \frac{h^k}{k!}\,\mathbf{g}^{(k)}(a) \;+\; \int_a^{a+h} \frac{(a+h-t)^n}{n!}\,\mathbf{g}^{(n+1)}(t)\,dt$$

where $\mathbf{g}^{(k)}$ *as usual denotes the* $k^{\rm th}$ *derivative of* $\mathbf{g}$.

Proof. Let $R_n(h)$ be the integral in the displayed equation. Then integration by parts implies that

$$R_{n-1}(h) \;=\; \frac{h^n}{n!}\,\mathbf{g}^{(n)}(a) \;+\; R_n(h)$$

and the Fundamental Theorem of Calculus implies that

$$\mathbf{g}(a+h) \;=\; \mathbf{g}(a) \;+\; R_1(h)\;.$$

Therefore if we set $R_0 = 0$ we have

$$\mathbf{g}(a+h) \;=\; \mathbf{g}(a) \;+\; \sum_{k=1}^{n} \big(R_k(h) - R_{k-1}(h)\big) \;+\; R_n(h)$$

and if we use the formulas above to substitute for the terms $\mathrm{R}_k(h) - R_{k-1}(h)$ and $R_n(h)$ we obtain the formula displayed above.■

The following consequence of Taylor's Theorem will be particularly useful:

COROLLARY. *Given* $\mathbf{g}$ *and the other notation as above, let* $P_n(h)$ *be the sum of*

$$\mathbf{g}(a) \;+\; \sum_{k=1}^{n} \frac{h^k}{k!}\,\mathbf{g}^{(k)}(a)\;.$$

Then given $r_0 < r$ and $|h| < r_0 < r$ we have $|\mathbf{g}(a+h) - P_n(h)| \le C\,|h|^{n+1}$, for some positive constant C.

Proof. The length of the difference vector in the previous sentence is given by

$$|R_n(h)| \quad = \quad \left| \int_a^{a+h} \frac{(a+h-t)^n}{n!}\, \mathbf{g}^{(n+1)}(t)\, dt \right| \quad \le$$

$$\mathrm{sign}(h) \;\cdot\; \int_a^{a+h} \left| \frac{(a+h-t)^n}{n!}\, \mathbf{g}^{(n+1)}(t) \right| \, dt \quad \le$$

$$\left(\max_{|t-a| \le r_0} |\mathbf{g}^{(n+1)}(t)| \right) \cdot \int_0^{|h|} \frac{u^n}{n!}\, du \quad \le \quad M\, \frac{|h|^{n+1}}{(n+1)!}$$

where M is a positive constant at least as large as the maximum value of $|\mathbf{g}^{(n+1)}(t)|$ for $|t-a| < r_0$. ■

I.3 : Arc length and reparametrization

(do Carmo, § 1–3)

Given a parametrized smooth regular curve $\mathbf{x}$ defined on a closed interval $[a, b]$, as in calculus we define the *arc length* of $\mathbf{x}$ from $t = a$ to $t = b$ to be the integral

$$L = \int_a^b |\mathbf{x}'(t)| \, dt .$$

Some motivation for this definition is discused in Exercise 8 on page 10 of do Carmo. More generally, if $a \leq t \leq b$ then the length of the curve from parameter value a to parameter value t is given by

$$s(t) = \int_a^t |\mathbf{x}'(u)| \, du .$$

By the Fundamental Theorem of Calculus, the partial arc length function s is differentiable on $[a, b]$ and its derivative is equal to $|\mathbf{x}'(t)|$. If we have a regular smooth curve, this function is continuous and everywhere positive (hence $s(t)$ is a strictly increasing function of t), and the image of this function is equal to the closed interval $[0, L]$.

Reparametrizations of curves

Given a parametrized curve $\mathbf{x}$ defined on an interval $[a, b]$, it is easy to find other parametrizations by simple changes of variables. For example, the curve $\mathbf{y}(t) = \mathbf{x}(t + a)$ resembles the original curve in many respects: For example, both have the same tangent vectors and images, and the only real difference is that $\mathbf{y}$ is defined on $[0, b - a]$ rather than $[a, b]$. Less trivial changes of variable can be extremely helpful in analyzing the image of a curve. For example, the parametrized curve $\mathbf{x}(t) = (e^t - e^{-t},\ e^t + e^{-t})$ has the same image as the the upper piece of the hyperbola $y^2 - x^2 = 4$ (*i.e.*, the graph of $y = \sqrt{4 + x^2}$); as a graph, this curve can also be parametrized using $\mathbf{y}(u) = (u,\ \sqrt{4 + u^2})$. These parametrizations are related by the change of variables $u = 2 \sinh t$; in other words, we have

$$\mathbf{x}(t) = \mathbf{y}\big(2 \sinh t\big) .$$

Note that u varies from $-\infty$ to $+\infty$ as t goes from $-\infty$ to $+\infty$, and $u'(t) = \cosh t > 0$ for all t.

More generally, it is useful to consider reparametrizations of curves corresponding to functions $u(t)$ such that $u'(t)$ is never zero. Of course the sign of u' determines whether u is strictly increasing or decreasing, and it is useful to allow both possibilities. Suppose that we are given a differentiable function u defined on $[a, b]$ such that u' is never zero on $[a, b]$. Then the image ot u is some other closed interval, say $[c, d]$; if u is increasing then $u(a) = c$ and $u(b) = d$, while if u is decreasing then $u(a) = d$ and $u(b) = c$. It follows that u has an inverse function t defined on $[c, d]$ and taking values in $[a, b]$. Furthermore, the derivatives dt/du and du/dt are reciprocals of each other by the standard formula for the derivative of an inverse function.

It is important to understand how reparametrization changes geometrical properties of a curve such as tangent lines and arc lengths. The most basic thing to consider is the effect on tangent vectors.

PROPOSITION. *Let* $\mathbf{x}$ *be a regular smooth curve defined on the closed interval* $[c,d]$*, let* $u : [a,b] \to [c,d]$ *be a function with a continuous derivative that is nowhere zero, and let* $\mathbf{y}(t) = \mathbf{x}\big(u(t)\big)$*. Then*

$$\mathbf{y}'(t) \quad = \quad u'(t) \cdot \mathbf{x}'\big(u(t)\big) \ .$$

This is an immediate consequence of the Chain Rule.■

COROLLARY. *For each* $t \in [a,b]$ *the tangent line to* $\mathbf{y}$ *at parameter value* t *is the same as the tangent line to* $\mathbf{x}$ *at* $u(t)$*. Furthermore, the standard parametrizations are related by a linear change of coordinates.*

Proof. By definition, the tangent line to $\mathbf{x}$ at $u(t)$ is the line joining $\mathbf{x}\big(u(t)\big)$ and $\mathbf{x}\big(u(t)\big) + \mathbf{x}'\big(u(t)\big)$. Similarly, the tangent line to $\mathbf{y}$ at t is the line joining $\mathbf{y}(t) = \mathbf{x}\big(u(t)\big)$ and

$$\mathbf{y}(t) + \mathbf{y}'(t) = \mathbf{x}\big(u(t)\big) + u'(t)\,\mathbf{x}'\big(u(t)\big) \ .$$

Since the line joining the distinct points (or vectors) $\mathbf{a}$ and $\mathbf{a}+\mathbf{b}$ is the same as the line joining $\mathbf{a}$ and $\mathbf{a} + c\,\mathbf{b}$ if $c \neq 0$, it follows that the two tangent lines are the same (take $\mathbf{a} = \mathbf{y}(t)$, $\mathbf{b} = \mathbf{x}'(u)$ and $c = u'(t)$).

In fact, we have obtained standard linear parametrizations of this line given by $\mathbf{f}(z) = \mathbf{a} + z\,\mathbf{b}$ and $\mathbf{g}(w) = \mathbf{a} + cw\,\mathbf{b}$. It follows that $\mathbf{g}(w) = \mathbf{f}(cw)$.■

Arc length is another property of a curve that does not change under reparmetrization.

PROPOSITION. *Let* $\mathbf{x}$ *be a regular smooth curve defined on the closed interval* $[c,d]$*, let* $u : [a,b] \to [c,d]$ *be a function with a continuous derivative that is nowhere zero, and let* $\mathbf{y}(t) = \mathbf{x}\big(u(t)\big)$*. Then*

$$\int_c^d |\mathbf{x}'(u)|\,du = \int_a^b |\mathbf{y}'(t)|\,dt$$

Proof. The standard change of variables formula for integrals implies that

$$\int_c^d |\mathbf{x}'(u)|\,du = \int_a^b |\mathbf{x}'\big(u(t)\big)\,|u'(t)|\,dt \ .$$

Some comments about this formula and the absolute value sign may be helpful. If u is increasing then the sign is positive and we have $u(a) = c$ and $u(b) = d$, so $|u'(t)| = u'(t)$; on the other hand if u is decreasing, then the Fundamental Theorem of Calculus suggests that the integral on the left hand side should be equal to

$$\int_b^a \Big|\mathbf{x}'\big(u(t)\big)\Big| \cdot u'(t)\,dt \quad = \quad -\int_a^b \ \Big|\mathbf{x}'\big(u(t)\big)\Big| \cdot u'(t)\,dt \quad = \quad \int_a^b \ \Big|\mathbf{x}'\big(u(t)\big)\Big| \cdot [-u'(t)]\ dt$$

so that the formula above holds because $u' < 0$ implies $|u'| = -u'$. In any case, the properties of vector length imply that the integrand on the right hand side of the change of variables equation is $|u'(t) \cdot \mathbf{x}'(u)|$, which by the previous proposition is equal to $|\mathbf{y}'(t)|$.■

If $\mathbf{v}$ is a regular smooth curve defined on $[a,b]$, then the arc length function

$$s(t) = \int_a^t |\mathbf{v}'(u)|\,du$$

often provides an extremely useful reparametrization because of the following result:

PROPOSITION. *Let* $\mathbf{v}$ *be as above, and let* $\mathbf{x}$ *be the reparametrization defined by* $\mathbf{x}(s) = \mathbf{v}\big(\mu(s)\big)$, *where* μ *is the inverse function to the arc length function* $\lambda : [a,b] \to [0,L]$. *Then* $|\mathbf{x}'(s)| = 1$ *for all* s.

Proof. By the Fundamental Theorem of Calculus we know that $\lambda'(t) = |\mathbf{v}'(t)|$. Therefore by the Chain Rule we know that

$$\mathbf{x}'(s) \quad = \quad \mu'(s)\,\mathbf{v}'\big(\mu(s)\big)$$

and by the differentiation formula for inverse functions we know that

$$\mu'(s) \quad = \quad \frac{1}{\lambda'(\,\mu(s)\,)} \quad = \quad T'(s) \quad = \quad \frac{1}{|\mathbf{v}'(\,T(s)\,)|}$$

and if we substitute this into the expression given by the chain rule we see that

$$|\mathbf{x}'(s)| \quad = \quad |T'(s)|\,|\mathbf{v}'\big(T(s)\big)| \quad = \quad \frac{1}{|\mathbf{v}'\big(T(s)\big)|}\cdot|\mathbf{v}'\big(T(s)\big)| \quad = \quad 1\ .\blacksquare$$

Arc length for more general curves

The geometric motivation for the definition of arc length is described in Exercises 8–0 on pages 10–11 of do Carmo; specifically, given a parametrized curve $\mathbf{x}$ defined on $[a,b]$ one picks a finite set of points t_i such that

$$a \quad = \quad t_0 \ < \ t_1 \ < \ \cdots \ < \ t_m = b$$

and views the length of the inscribed broken line joining t_0 to t_1, t_1 to t_2 *etc.* as an approximation to the length of the curve. In favorable circumstances if one refines the finite set of points by taking more and more of them and making them closer and closer together, the lengths of these broken line curves will have a limiting value which is the arc length. Exercise 9(b) on page 11 of do Carmo gives one example of a curve for which no arc length can be defined. During the time since do Carmo's book was published, a special class of such curves known as *fractal curves* has received considerable attention. The parametric equations defining such curves all have the form $\mathbf{x}(t) = \lim_{n\to\infty}\mathbf{x}_n(t)$, where each $\mathbf{x}_n$ is a piecewise smooth regular curve and for each n one obtains $\mathbf{x}_n$ from $\mathbf{x}_{n-1}$ by making some small but systematic changes. Some online references with more information on such curves are given below.

`http://mathworld.wolfram.com/Fractal.html`

`http://academy.wolfram.agnescott.edu/ lriddle/ifs/ksnow/lsnow/htm`

`http://en2.wikipedia.org/wiki/Koch_snowflake`

`http://en.wikipedia.org/wiki/Fractal_geometry`

I.4 : Curvature and torsion

(do Carmo, §§1–5, 1–6)

Many calculus courses include a brief discussion of curvature, but the approaches vary and it will be better to make a fresh start.

Definition. Let $\mathbf{x}$ be a regular smooth curve, and assume it is parametrized by arc length plus a constant (*i.e.*, $|\mathbf{x}'(s)| = 1$ for all s). The *curvature* of $\mathbf{x}$ at parameter value s is equal to $\kappa(s) = |\mathbf{x}''(s)|$.

The most immediate question about this definition is why it has anything to do with our intuitive idea of curvature. The best way to answer this is to look at some examples.

Suppose that we are given a parametrized line with an equation of the form $\mathbf{x}(t) = \mathbf{a} + t\,\mathbf{b}$ where $|\mathbf{b}| = 1$. It then follows that $\mathbf{x}$ is parametrized by arc length by means of t, and clearly we have $\mathbf{x}''(t) = \mathbf{0}$. This means that the curvature of the line is zero at all points, which is what we expect.

Consider now an example that is genuinely curved; namely, the circle of radius r about the origin. The arc length parametrization for this curve has the form

$$\mathbf{x}(s) \quad = \quad \Big(r\cos(s/r),\; r\sin(s/r) \Big)$$

and one can check directly that its first two derivatives are given as follows:

$$\mathbf{x}''(s) \quad = \quad \Big(-\sin(s/r),\; \cos(s/r) \Big)$$

$$\mathbf{x}(s) \quad = \quad \left(-\frac{\cos(s/r)}{r},\; -\frac{\sin(s/r)}{r} \right)$$

It follows that *the curvature of the circle at all points is given by the reciprocal of the radius.*∎

The following simple property of the "acceleration" function $\mathbf{x}''(s)$ turns out to be quite important for our purposes:

PROPOSITION. *The vectors* $\mathbf{x}''(s)$ *and* $\mathbf{x}'(s)$ *are perpendicular.*

Proof. We know that $|\mathbf{x}'(s)|$ is always equal to 1, and thus the same is true of its square, which is just the dot product of $\mathbf{x}'(s)$ with itself. The product rule for differentiating dot products of two functions then implies that

$$0 \quad = \quad \frac{d}{ds}\Big(\mathbf{x}'(s)\cdot\mathbf{x}'(s) \Big) \quad = \quad 2\left(\mathbf{x}'(s)\cdot\mathbf{x}''(s) \right)$$

and therefore the two vectors are indeed perpendicular.∎

Geometric interpretation of curvature

We begin with a very simple observation.

PROPOSITION. *If $\mathbf{x}(s)$ is a smooth curve (parametrized by arc length) whose curvature $\kappa(s)$ is zero for all s, then $\mathbf{x}(s)$ is a straght line curve of the form $\mathbf{x}(s) = \mathbf{x}(0) + s\,\mathbf{x}'(0)$.*

Proof. Since $\kappa(s)$ is the length of $\mathbf{x}''(s)$, if the curvature is always zero then the same is true for $\mathbf{x}''(s)$. But this means that $\mathbf{x}'(s)$ is constant and hence equal to $\mathbf{x}'(0)$ for all s, and the latter in turn implies that $\mathbf{x}(s) = \mathbf{x}(0) + s\,\mathbf{x}'(0)$.■

Given a smooth curve, the tangent line to the curve at a point t may be viewed as a first order linear approximation to the curve. The notion of curvature is related to a corresponding second order approximation to the curve at parameter value t by a line or circle. We begin by making this notion precise:

Defintion. Let n be a positive integer. Given two curves $\mathbf{a}(t)$ and $\mathbf{b}(t)$ defined on an interval J containing t_0 such that $\mathbf{a}(t_0) = \mathbf{b}(t_0)$, we say that $\mathbf{a}$ *and* $\mathbf{b}$ *are strong n^{th} order approximations to each other* if there is an $\varepsilon > 0$ such that $|h| < \varepsilon$ and $t_0 + h \in J$ imply

$$|\mathbf{b}(t_0+h) \;-\; \mathbf{a}(t_0+h)| \;\;\leq\;\; C\,|h|^{n+1}$$

for some constant $C > 0$. The analytic condition on the order of approximation is often formulated geometrically as the order of contact that two curves have with each other at a given point; as the order of contact increases, so does the speed at which the curves approach each other. The most basic visual examples here are the x-axis and the graphs of the curves x^n near the origin. Further information relating geometric ideas of high order contact and Taylor polynomial approximations is presented on pages 87–91 of the Schaum's Outline Series book on differential geometry (M. Lipschultz, *Schaum's Outlines — Differential Geometry*, Schaum's/McGraw-Hill, 1969, ISBN 0-07-037985-8).

LEMMA. *Suppose that the curves $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are defined on an interval J containing t_0 such that $\mathbf{a}(t_0) = \mathbf{b}(t_0)$, and assume also that $\mathbf{a}$ and $\mathbf{b}$ are strong n^{th} order approximations to each other at t_0. Then for each regular smooth reparametrization $t(u)$ with $t_0 = t(u_0)$ the curves $\mathbf{a}\circ t$ and $\mathbf{b}\circ t$ are strong n^{th} order approximations to each other at u_0.*

Proof. Let J_0 be the domain of the function $t(u)$, and let K_0 be a closed bounded subinterval containing u_0 such that the latter is an endpoint of K_0 if and only if it is an endpoint of J_0. Denote the maximum value of $|t'(u)|$ on this interval by M. Then by hypothesis and the Mean Value Theorem we have

$$|\mathbf{b}\big(\,t(u_0+h)\big) \;-\; \mathbf{a}\big(\,t(u_0+h)\big)| \;\;\leq\;\; C\,|t(u_0+h) - t(u_0)|^{n+1} \;\;\leq\;\; C\,M^{n+1}\cdot|h|^{n+1}$$

which proves the assertion of the lemma.■

In the terminology of n^{th} order approximations, if we are given a regular smooth curve $\mathbf{x}$ then a strong first order approximation to it is given by the tangent line with the standare linear parametrization

$$\mathbf{L}(t_0+h) = \mathbf{x}(t_0) + h\,\mathbf{x}'(t)\ .$$

Furthermore, this line is the unique strong first order linear approximation to $\mathbf{x}$.

Here is the main result on curvature and strong second order approximations.

THEOREM. *Let $\mathbf{x}$ be a regular smooth curve defined on an interval J containing 0 such that $\mathbf{x}'$ has a continuous* **second** *derivative and $|\mathbf{x}'| = 1$ (hence $\mathbf{x}$ is parametrized by arc length plus a constant).*

(i) If the curvature of $\mathbf{x}$ *at* 0 *is zero, then the tangent line is a strong second order approximation to* $\mathbf{x}$.

(ii) Suppose that the curvature of $\mathbf{x}$ *at* 0 *is nonzero, let* $\mathbf{N}$ *be the unit vector pointing in the same direction as* $\mathbf{x}''(0)$ *(the latter is nonzero by the definition and nonvanishing of the curvature at parameter value* 0*). If* Γ *is the circle through* $\mathbf{x}(0)$ *such that [1] its center is* $\mathbf{x}(0) + (\kappa(0))^{-1}\mathbf{N}$*, [2] it lies in the plane containing this center and the tangent line to the curve at parameter value zero, then* Γ *is a strong second order approximation to* $\mathbf{x}$.

For the sake of completeness, we shall describe the unique plane containing a given line and an external point explicitly as follows. If $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are noncollinear points in $\mathbf{R}^3$ then the plane containing them consists of all $\mathbf{x}$ such that $\mathbf{x} - \mathbf{a}$ is perpendicular to

$$(\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})$$

which translates to the triple product equation

$$[(\mathbf{x} - \mathbf{a}), (\mathbf{b} - \mathbf{a}), (\mathbf{c} - \mathbf{a})] \;=\; 0\,.$$

Suppose now that $\mathbf{b}_1$ and $\mathbf{c}_1$ are points on the line containing $\mathbf{b}$ and $\mathbf{c}$. Then we may write

$$\mathbf{b}_1 \;=\; u\,\mathbf{b} + (1-u)\,\mathbf{c}\,, \quad \mathbf{c}_1 \;=\; v\,\mathbf{b} + (1-v)\,\mathbf{c}$$

for suitable real numbers u and v. The equations above immediately imply the following identities:

$$(\mathbf{b}_1 - \mathbf{a}) \;=\; u\,(\mathbf{b} - \mathbf{a}) + (1-u)\,(\mathbf{c} - \mathbf{a})$$

$$(\mathbf{c}_1 - \mathbf{a}) \;=\; v\,(\mathbf{b} - \mathbf{a}) + (1-v)\,(\mathbf{c} - \mathbf{a})\,.$$

These formulas and the basic properties of determinants imply

$$[(\mathbf{x} - \mathbf{a}).\,(\mathbf{b}_1 - \mathbf{a}), (\mathbf{c}_1 - \mathbf{a})] \;=$$

$$[(\mathbf{x} - \mathbf{a}).\,u(\mathbf{b}_1 - \mathbf{a}), v(\mathbf{c}_1 - \mathbf{a})] + \big[(\mathbf{x} - \mathbf{a}).\,(1-u)\big(\mathbf{b}_1 - \mathbf{a}\big), (1-v)\big(\mathbf{c}_1 - \mathbf{a}\big)\big] \;=$$

$$uv\,[(\mathbf{x} - \mathbf{a}), (\mathbf{b} - \mathbf{a}), (\mathbf{c} - \mathbf{a})] + (1-u)\,(1-v)\,[(\mathbf{x} - \mathbf{a}), (\mathbf{c} - \mathbf{a}), (\mathbf{b} - \mathbf{a})] \;=$$

$$uv\,0 - (1-u)\,(1-v)\,0 \;=\; 0$$

and hence the equation

$$[(\mathbf{x} - \mathbf{a}), (\mathbf{b} - \mathbf{a}), (\mathbf{c} - \mathbf{a})] \;=\; 0$$

implies the corresponding equation if $\mathbf{b}$ and $\mathbf{c}$ are replaced by two arbitrary points on the line containing $\mathbf{b}$ and $\mathbf{c}$.■

Proof of Proposition. Consider first the case where $\kappa(0) = 0$. Then the tangent line to the curve has equation $\mathbf{L}(s) = s\,\mathbf{x}'(0)$ and the second order Taylor expansion for $\mathbf{x}$ has the form $\mathbf{x}(s) = s\,\mathbf{x}'(0) + \frac{1}{2}s^2\,\mathbf{x}''(0) + s^3\,\theta(s)$ where $\theta(s)$ is bounded for s sufficiently close to zero. The assumption $\kappa(0) = 0$ implies that $\mathbf{x}''(0) = 0$ and therefore we have $\mathbf{x}(s) - \mathbf{L}(s) = s^3\,\theta(s)$ where $\theta(s)$ is bounded for s sufficiently close to zero. Therefore the tangent line is a strong second order approximation to the curve if the curvature is equal to zero.

Suppose now that $\kappa(0) \neq 0$, and let $\mathbf{N}$ be the unit vector pointing in the same direction as $\mathbf{x}''(0)$. Define $\mathbf{z}$ by the formula

$$\mathbf{z} \quad = \quad \mathbf{x}(0) \; + \; \frac{1}{\kappa(0)}\,\mathbf{N}$$

and consider the circle in the plane of $\mathbf{z}$ and the tangent line to $\mathbf{x}$ at parameter value $s = 0$ such that the center is $\mathbf{z}$ and the radius is $1/\kappa(0)$. If we set r equal to $1/\kappa(0)$ and $\mathbf{T} = \mathbf{x}'(0)$, then a parametrization of this circle in terms of arc length is given by

$$\Gamma(s) \quad = \quad \mathbf{z} \; - \; r\cos(s/r)\,\mathbf{N} \; + \; r\sin(s/r)\,\mathbf{T}\ .$$

Using the standard power series expansions for the sine and cosine function and the identity $\mathbf{z} = \mathbf{x}(0) - r\,\mathbf{N}$, we may rewrite this in the form

$$\Gamma(s) \quad = \quad \mathbf{x}(0) \; + \; \frac{s^2}{2r}\mathbf{N} \; + \; s^3\alpha(s)\mathbf{N} + s\mathbf{T} + s^3\beta(s)\mathbf{T}$$

where $\alpha(s)$ and $\beta(s)$ are continuous functions and hence are bounded for s close to zero. On the other hand, using the Taylor expansion of $\mathbf{x}(s)$ near $s = 0$ we may write $\mathbf{x}(s)$ in the form

$$\mathbf{x}(0) \; + \; s\,\mathbf{x}'(0) \; + \; \frac{s^2}{2}\mathbf{x}''(0) \; + \; s^3\,\mathbf{W}(s)$$

where $\mathbf{W}(s)$ is bounded for s close to zero. But $\mathbf{x}'(0) = \mathbf{T}$ and

$$\mathbf{x}''(0) \quad = \quad \kappa(0)\,\mathbf{N} \quad = \quad \frac{1}{r}\mathbf{N}$$

so that $\Gamma(s) - \mathbf{x}(s)$ has the form $s^3\mathbf{W}_1(s)$ where $\mathbf{W}_1(s)$ is a bounded function of s. Therefore the circle defined by Γ is a strong second order approximation to the original curve at the parameter value $s = 0$.■

Notation. If the curvature of $\mathbf{x}$ is nonzero near parameter value s as in the proposition, then the center of the strong second order circle approximation

$$\mathbf{z}(s) \quad = \quad \mathbf{x}(s) \; + \; \frac{1}{(\kappa(s))^2}\,\mathbf{x}''(s)$$

is called the *center of curvature* of $\mathbf{x}$ at parameter value s. The circle itsef is called the *osculating circle* to the curve at parameter value s (in Latin, *osculare* = to kiss).

Complementary result. A more detailed analysis of the situation shows that if $\kappa(0) \neq 0$ then the circle given above is the unique circle that is a second order approximation to the original curve at the given point.■

Computational techniques

Although the description of curvature in terms of arc length parametrizations is important for theoretical purposes, it is usually not particulary helpful if one wants to compute the curvature of a given curve at a given point. One major reason for this is that the arc length function $s(t)$ can only be written down explicitly in a very restricted class of cases. In particular, if we consider the graph of the cubic polynomial $y = x^3$ with parametrization $(t,\, t^3)$ on some interval $[0, a]$ then the arc length parameter is given by the formula

$$s(t) \quad = \quad \int_0^t \sqrt{1 + 9\,u^4}\, du$$

and results of P. Chebyshev from the nineteenth century show that there is no "nice" formula for this function in terms of the usual functions one studies in first year calculus. Therefore it is important to have formulas for curvature in terms of arbitrary parametrizations of a regular smooth curve.

Remarks.

1. The statement about the antiderivative of $\sqrt{1 + 9\,x^4}$ is stronger than simply saying that no one has has been able to find a nice formula for the antiderivative. It as just as impossible to find one as it is to find two positive whole numbers a and b such that $\sqrt{2} = a/b$.

2. A detailed statement of Chebyshev's result can be found on the web link

`http://mathworld.wolfram.com/Integral.html`

and further references are also given there.

The following formula appears in many calculus texts:

FIRST CURVATURE FORMULA *Let* $\mathbf{x}$ *be a smooth regular curve, let* s *be the arc length function, let* $k(t) = \kappa(\, s(t)\,)$*, and let* $\mathbf{T}(t)$ *be the unit tangent vector function obtained by multiplying* $\mathbf{x}'(t)$ *by the reciprocal of its length. Then we have*

$$k(t) \quad = \quad \frac{|\mathbf{T}\,'(t)|}{|\mathbf{x}'(t)|} \ .$$

Derivation. We have seen that $\mathbf{T}\,(s)$ is equal to $\mathbf{x}'(s)$, and therefore by the chain rule we have

$$\mathbf{T}\,'(t) \quad = \quad s'(t)\,\mathbf{T}\,'(\,s(t)\,) \quad = \quad |\mathbf{x}'(t)|\,\mathbf{x}''(s) \ .$$

Taking lengths of the vectors on both sides of this equation we see that

$$|\mathbf{T}\,'(t)| \quad = \quad |\mathbf{x}'(t)| \cdot |\mathbf{x}''(s)| \quad = \quad |\mathbf{x}'(t)|\,k(t)$$

which is equivalent to the formula for $k(t)$ displayed above.■

Here is another formula for curvature that is often found in calculus textbooks.

SECOND CURVATURE FORMULA *Let* $\mathbf{x}$ *be a smooth regular curve, let* s *be the arc length function, let* $\mathbf{T}(t)$ *be the unit length tangent vector function, and let* $k(t) = \kappa(\, s(t)\,)$*. Then we have*

$$k(t) \quad = \quad \frac{|\mathbf{x}'(t) \ \times \ \mathbf{x}''(t)|}{|\mathbf{x}'(t)|^3} \ .$$

Derivation. As in the derivation of the First Curvature Formula we have $\mathbf{x}' = s'\mathbf{T}$. Therefore the Leibniz product rule for differentiating the product of a scalar function and a vector function yields

$$\mathbf{x}'' \quad = \quad s''\mathbf{T} \ + \ s'\mathbf{T}\,' \ .$$

Since $\mathbf{T} \times \mathbf{T} = \mathbf{0}$ the latter implies

$$\mathbf{x}' \times \mathbf{x}'' \quad = \quad (s'')^2 \left(\mathbf{T} \times \mathbf{T}' \right) .$$

Since $|\mathbf{T}| = 1$ it follows that $\mathbf{T} \cdot \mathbf{T}' = 0$; *i.e.*, the vectors $\mathbf{T}$ and $\mathbf{T}'$ are orthogonal. This in turn implies that $|\mathbf{T} \times \mathbf{T}'|$ is equal to $|\mathbf{T}| \cdot |\mathbf{T}'|$ so that

$$|\mathbf{x}' \times \mathbf{x}''| \quad = \quad |s''|^2 \, |\mathbf{T} \times \mathbf{T}'| \quad = \quad |s''|^2 \, |\mathbf{T}| \cdot |\mathbf{T}'| \quad = \quad (s'')^2 \, |\mathbf{T}| \quad = \quad |\mathbf{x}'|^2 \, |\mathbf{T}'|$$

(at the next to last step we again use the identity $|\mathbf{T}| = 1$). It follows that

$$|\mathbf{T}'| \quad = \quad \frac{|\mathbf{x}'(t) \; \times \; \mathbf{x}''(t)|}{|\mathbf{x}'(t)|^2}$$

and the Second Curvature Formula follows by substitution of this expression into the First Curvature Formula.■

Osculating planes

Thus far we have discussed lines and circles that are good approximations to a curve. Given a curve in 3-dimensional space one can also ask whether there is some plane that comes as close as possible to containing the given curve. Of course, for curves that lie entirely in a single plane, the definition should yield this plane.

Given a continuous curve $\mathbf{x}(t)$, and a plane Π, one way of making this notion precise is to consider the function $\Delta(t)$ giving the distance from $\mathbf{x}(t)$ to Π. If the point $\mathbf{x}(t_0)$ lies on Π, then $\Delta(t_0) = 0$ and one test of how close the curve comes to lying in the plane is to determine the extent to which the zero function is an n^{th} order approximation to $\Delta(t)$ for various choices of n. In fact, if $\kappa(t_0) \neq 0$ then there is a unique plane such that the zero function is a second order approximation to $\Delta(t)$, and this plane is called the *osculating plane* to $\mathbf{x}$ at parameter value $t = t_0$. Formally, we proceed as follows:

Definition. Let $\mathbf{x}(s)$ be a regular smooth curve parametrized by arc length (so that $|\mathbf{x}'| = 1$), and assume that $\kappa(s_0) \neq 0$. Let $\mathbf{a} = \mathbf{x}(0)$, let $\mathbf{T} = \mathbf{x}'(s_0)$, and let $\mathbf{N}$ be the unit vector pointing in the same direction as $\mathbf{x}''(s_0)$. The *osculating plane* to the curve at parameter value s_0 is the unique plane containing the three noncollinear vectors $\mathbf{a}$, $\mathbf{a} + \mathbf{T}$, and $\mathbf{a} + \mathbf{N}$.

It follows that the equation defining the osculating plane may be written in the form

$$[(\mathbf{y} - \mathbf{a}), \, \mathbf{T}, \, \mathbf{N}] \quad = \quad 0 \, .$$

We can now state the result on the order of contact between curves and their osculating planes.

PROPOSITION. *Let $\mathbf{x}$ be a regular smooth curve parametrized by arc length (hence $|\mathbf{x}'| = 1$), assume that $\mathbf{x}$ has a continuous third derivative, and assume also that $\kappa(s_0) \neq 0$. Let Π be the osculating plane of $\mathbf{x}$ at parameter value s_0, and let $\Delta(s)$ denote the distance between $\mathbf{x}(s)$ and Π. Then the osculating plane is the unique plane through $\mathbf{x}(s_0)$ such that the zero function is a second order approximation to the distance function $\Delta(s)$ at s_0.*

Proof. Let $\mathbf{a} = \mathbf{x}(s_0)$, let $\mathbf{T} = \mathbf{x}'(s_0)$, let $\mathbf{N}$ be the unit vector pointing in the same direction as $\mathbf{x}''(s_0)$, and let $\mathbf{B}$ be the cross product $\mathbf{T} \times \mathbf{N}$. Then the oscularing plane is the unique plane containing $\mathbf{a}$, $\mathbf{a}+\mathbf{T}$, and $\mathbf{a}+\mathbf{N}$, and the distance between a point $\mathbf{y}$ and the osculating plane is the absolute value of the function $\widetilde{D}\,(\mathbf{y}) = (\mathbf{y}-\mathbf{a})\cdot\mathbf{B}$. The second order Taylor approximation to $\mathbf{x}(s)$ with respect to s_0 is then given by the formula

$$\mathbf{x}(s) = \mathbf{a} \;+\; (s-s_0)\cdot\mathbf{T} \;+\; \frac{(s-s_0)^2\,\kappa(s_0)}{2}\cdot\mathbf{N} \;+\; (s-s_0)^3\,\mathbf{W}(s)$$

where $\mathbf{W}(s)$ is bounded for s sufficiently close to s_0. Therefore since $\mathbf{B}$ is perpendicular to $\mathbf{T}$ and $\mathbf{N}$ we have

$$\widetilde{D}\,\big(\mathbf{x}(s))\big) \;=\; (s-s_0)^3\,\mathbf{W}(s)\cdot\mathbf{B}$$

where $\mathbf{W}(s)\cdot\mathbf{B}$ is bounded for s sufficiently close to s_0. Therefore the given curve has order of contact at least two with respect to its osculating plane.

Suppose now that we are given some other plane through $\mathbf{a}$; then one has a normal vector $\mathbf{V}$ to the plane of the form $\mathbf{B}+p\,\mathbf{T}+q\,\mathbf{N}$ where p and q are not both zero. The distance between $\mathbf{x}(s)$ and plane through $\mathbf{a}$ with normal vector $\mathbf{V}$ will then be the absolute value of a nonzero multiple of the function

$$\Big(\,(\mathbf{x}(s) \;-\; \mathbf{a})\cdot\mathbf{V}\Big)$$

which is equal to

$$g(s-s_0) \;=\; (s-s_0)\,\big(\mathbf{T}\cdot\mathbf{V}\big) \;+\; \frac{(s-s_0)^2\,\kappa(s_0)}{2}\,\big(\mathbf{N}\cdot\mathbf{V}\big) \;+\; (s-s_0)^3\,\big(\mathbf{W}(s)\cdot\mathbf{V}\big)\;.$$

We then have

$$\frac{g(s-s_0)}{(s-s_0)^3} \;=\; \frac{p}{(s-s_0)^2} \;+\; \frac{q}{(s-s_0)} \;+\; \big(\mathbf{W}(s)\cdot\mathbf{V}\big)$$

where the third term on the right is bounded. But since at least one of p and q is nonzero, it follows that the entire sum is not a bounded function of s if s is close to s_0. Therefore the curve cannot have order of contact at least two with any other plane through $\mathbf{a}$.■

Torsion

Curvature may be viewed as reflecting the rate at which a curve moves off its tangent line. The notion of torsion will reflect the rate at which a curve moves off its osculating plane. In order to define this quantity we first need to give some definitions that play an important role in the theory of curves.

Definitions. Let $\mathbf{x}$ be a regular smooth curve parametrized by arc length plus a constant (hence $|\mathbf{x}'| = 1$), assume that $\mathbf{x}$ has a continuous third derivative, and assume also that $\kappa \neq 0$ near the parameter value s_0. The *principal unit normal vector* at parameter value s is $\mathbf{N}(s) = |\mathbf{x}''(s)|^{-1}\,\mathbf{x}''(s)$. We have already encountered a special case of this vector in the study of curvatures and osculating planes, and if $\mathbf{T}(s) = \mathbf{x}'(s)$ denotes the unit tangent vector then we know that $\{\,\mathbf{T}(s),\,\mathbf{N}(s)\,\}$ is a set of perpedicular vectors with unit length (an *orthonormal* set).

If $\mathbf{x}$ is a space curve (*i.e.*, its image lies in 3-space), the *binormal* vector at parameter value s is defined to be $\mathbf{B}(s) = \mathbf{T}(s)\times\mathbf{N}(s)$. It then follows that $\{\,\mathbf{T}(s),\,\mathbf{N}(s),\,\mathbf{B}(s)\}$ is an orthonormal basis for $\mathbf{R}^3$, and it is called the *Frenet trihedron* (or frame) at parameter value s.

One can frequently define a Frenet trihedron at a parameter value s_0 even if the curvature vanishes at s_0, but there are examples where it is not possible to do so. In particular, consider the curve given by $\mathbf{x}(t) = \big(t,\, 0,\, \exp(-1/t^2)\big)$ if $t > 0$ and $\mathbf{x}(t) = \big(t,\, \exp(-1/t^2)\, 0\big)$ if $t > 0$. If we set $\mathbf{x}(t) = \mathbf{0}$, then $\mathbf{x}$ will be infinitely differentiable because for each $k \geq 0$ we have

$$\lim_{t\to 0} \frac{d^k}{dt^k} \exp(-1/t^2) \quad = \quad 0$$

(this is true by repeated application of L'Hospital's Rule) and in fact the curvature is also nonzero if $t \neq 0$ and $t^2 \neq 2/3$. Therefore one can define a principal unit normal vector $\mathbf{N}(t)$ when $t \neq 0$ but, say, $|t| < \frac{1}{2}$. However, if $t > 0$ this vector lies in the xz-plane while if $t < 0$ it lies in the xy-plane, and if one could define a continuous unit normal at $t = 0$ it would have to lie in both of these planes. Now the unit tangent at $t = 0$ is the unit vector $\mathbf{e}_1$, and there are no unit vectors that are perpendicular to $\mathbf{e}_1$ that lie in both the xy- and xz-planes. Therefore there is no way to define a continuous extension of $\mathbf{N}$ to all values of t. On the other hand, Problem 4.15 on pages 75–76 of Schaum's Outline Series on Differential Geometry provides a way to define principal normals in some situations when the curvature vanishes at a given parameter value.■

One can retrieve the Frenet trihedron from an arbitrary regular smooth reparametrization with a continuous second derivative.

LEMMA. *In the setting above, suppose that we are given an arbitrary reparametrization with continuous second derivative, and let $s(t)$ denote the arc length function. Then the Frenet trihedron at parameter value t_0 is given by the unit vectors pointing in the same directions as $\mathbf{T}(t)$, $\mathbf{T}\,'(t)$, and their cross product. Furthermore, if one considers the reoriented curve $\mathbf{y}$ with parametrization $\mathbf{y}(t) = \mathbf{x}(-t)$, then the effect on the Frenet trihedron is that the first two unit vectors are sent to their negatives and the third remains unchanged.*

Proof. It follows immediately from the Chain Rule that the unit tangent $\mathbf{T}$ remains unchanged under a standard reparametrization with $s' > 0$. Furthermore, the derivation of the formulas for curvature under reparametrization show that $\mathbf{T}\,'(t)$ is a positive multiple of $\mathbf{x}''(s)$. this proves the assertion regarding the principal normals. Finally, if we are given two ordered sets of vectors $\{\,\mathbf{a},\, \mathbf{b}\,\}$ and $\{\,\mathbf{c},\, \mathbf{d}\,\}$ such that $\mathbf{c}$ and $\mathbf{d}$ are positive multiples of $\mathbf{a}$ and $\mathbf{b}$ respectively, then $\mathbf{c} \times \mathbf{d}$ is a positive multiple of $\mathbf{a} \times \mathbf{b}$, and this implies the statement regarding the binormals.

If one reverses orientations by the reparametrization $t \longmapsto -t$, then the Chain Rule implies that $\mathbf{T}$ and its derivative are sent to their negatives, and this proves the statement about the first two vectors in the trihedron. The statement about the third vector follows from these and the cross product identity $\mathbf{a} \times \mathbf{b} = (-\mathbf{a}) \times (-\mathbf{b})$.■

We are finally ready to define torsion.

Definition. In the setting above the *torsion* of the curve is given by $\tau(s) = \mathbf{B}\,'(s) \cdot \mathbf{N}(s)$.

This is not quite the same as the definition in do Carmo, so we shall show that our formulation is equivalent.

LEMMA. *The torsion of the curve is given by the formula $\mathbf{B}\,'(s) = \tau(s)\,\mathbf{N}(s)$.*

Proof. If we can show that the left hand side is a multiple of $\mathbf{N}(s)$, then the formula will follow by taking dot products of both sides of the equation with $\mathbf{N}(s)$ (note that the dot product of the latter with itself is equal to 1). To show that the left hand side side is a multiple of $\mathbf{N}(s)$, it suffices to show that it is perpendicular to $\mathbf{T}(s)$ and $\mathbf{B}(s)$. The second of these follows because

$$0 \quad = \quad \frac{d}{ds}(1) \quad = \quad \frac{d}{ds}(\mathbf{B}\cdot\mathbf{B}) \quad = \quad 2\,\mathbf{B}\cdot\left(\frac{d\,\mathbf{B}}{ds}\right)$$

and the first follows because

$$\frac{d\mathbf{B}}{ds} \;=\; \frac{d}{ds}\left(\mathbf{T}\times\mathbf{N}\right) \;=\; (\kappa\,\mathbf{N}\times\mathbf{N}) \;+\; \left(\mathbf{T}\times\frac{d\,\mathbf{N}}{ds}\right) \;=\; \mathbf{T}\times\left(\frac{d\,\mathbf{N}}{ds}\right)$$

which implies that the left hand side is perpendicular to $\mathbf{T}$.∎

We had mentioned that the torsion of a curve is related to the rate at which a curve moves away from its osculating plane. Here is a more precise statement about the relationship:

PROPOSITION. *Let* $\mathbf{x}$ *be a regular smooth curve parametrized by arc length plus a constant (hence* $|\mathbf{x}'| = 1$*), assume that* $\mathbf{x}$ *has a continuous third derivative, and assume also that* $\kappa(s_0) \neq 0$*. Let* Π *be the osculating plane of* $\mathbf{x}$ *at parameter value* s_0*. Then the image of* $\mathbf{x}$ *is contained in* Π *for all* s *sufficiently close to* s_0 *if and only if the torsion vanishes for these parameter values.*

Proof. Suppose first that the curve is entirely contained in the osculating plane for s close to s_0. The osculating plane at s_0 is defined by the equation

$$[(\mathbf{y}-\mathbf{a}),\,\mathbf{T}_0,\,\mathbf{N}_0] \;=\; 0$$

where $\mathbf{a} = \mathbf{x}(s_0)$ and $\mathbf{T}_0$ and $\mathbf{N}_0$ represent the unit tangent and principal normal vectors at parameter value s_0. If we set $\mathbf{y} = \mathbf{x}(s)$ and simplify this expression, we see that the curve $\mathbf{x}$ satisfies the equation

$$\mathbf{x}(s)\cdot\mathbf{B}_0 \;=\; \mathbf{a}\cdot\mathbf{B}_0$$

where $\mathbf{B}_0 = \mathbf{T}_0\times\mathbf{N}_0$. If we differentiate both sides with respect to s we obtain the equation $\mathbf{x}'(s)\cdot\mathbf{B}_0 = 0$. Differentiating once again we see that $\mathbf{x}''(s)\cdot\mathbf{B}_0 = 0$. Since $\mathbf{x}'(s) = \mathbf{T}(s)$ and $\mathbf{N}(s)$ is a positive multiple of $\mathbf{x}''(s)$ for s close to s_0 (specifically at least close enough so that $\kappa(s)$ is never zero), then $\mathbf{B}_0$ is perpendicular to both $\mathbf{T}(s)$ and $\mathbf{N}(s)$. Therefore $\mathbf{B}(s)$ must be equal to $\pm\,\mathbf{B}_0$. By continuity we must have that $\mathbf{B}(s) = \mathbf{B}_0$ for all s close to s_0 (Here are the details: Look at the function $\mathbf{B}(s)\cdot\mathbf{B}_0$ on some small interval containing s_0; its value is ± 1, and its value at s_0 is $+1$ — if its value somewhere else on the interval were -1, then by the Intermediate Value Theorem there would be someplace on the interval where its value would be zero, and we know this is impossible). Thus $\mathbf{B}(s)$ is constant, and by the preceding formulas this means that its torsion must be equal to zero.

Conversely, suppose that the torsion is identially zero. Then by alternate description of torsion in the lemma we know that $\mathbf{B}\,'(s) \equiv \mathbf{0}$, So that $\mathbf{B}(s) \equiv \mathbf{B}_0$. We then have the string of equations

$$0 \;=\; \mathbf{T}\cdot\mathbf{B}_0 \;=\; \mathbf{x}'\cdot\mathbf{B}_0 \;=\; \frac{d}{ds}\left(\mathbf{x}\cdot\mathbf{B}_0\right)$$

which in turn implies that $\mathbf{x}\cdot\mathbf{B_0}$ is a constant. Therefore the curve $\mathbf{x}$ lies entirely in the unique plane containing $\mathbf{x}(s_0)$ with normal direction $\mathbf{B}_0$.∎

Other planes associated to a curve

In addition to the osculating plane, there are two other associated planes through a point on the curve $\mathbf{x}$ at parameter value s_0 that are mentioned frequently in the literature. As above we assume that the curve is a regular smooth curve with a continuous third derivative i arc length parametrization, and nonzero curvature at parameter value s_0.

Definitions. In the above setting the *normal plane* is the unique plane containing $\mathbf{x}(s_0)$, $\mathbf{x}(s_0) + \mathbf{N}(s_0)$, and $\mathbf{x}(s_0) + \mathbf{B}(s_0)$, and the *rectifying plane* is the unique plane containing $\mathbf{x}(s_0)$, $\mathbf{x}(s_0) + \mathbf{T}(s_0)$, and $\mathbf{x}(s_0) + \mathbf{B}(s_0)$. These three mutually perpendicular planes meet at the point $\mathbf{x}(s_0)$ in the same way that the usual xy-, yz-, and xz-planes meet at the origin.

Oriented curvature for plane curves

For an arbitrary regular curve in 3-space one does not necessarily have normal directions when the curvature is zero, but for plane curves there is a unique normal direction up to sign. Specifically, if $\mathbf{x}$ is a regular smooth plane curve parametrized by arc length and $\mathbf{B}$ is a unit normal vector to a plane Π containing the image of $\mathbf{x}$, then one has an *associated oriented principal normal direction* at parameter value given by the cross product formula

$$\widehat{\mathbf{N}}\,(s) \quad = \quad \mathbf{B} \times \mathbf{x}'(s)$$

and by construction Π is the unique plane passing through $\mathbf{x}(s)$, $\mathbf{x}(s) + \mathbf{x}'(s)$, and $\mathbf{x}(s) = \widehat{\mathbf{N}}\,(s)$. There are two choices of $\mathbf{B}$ (the two unit normals for π are negatives of each other) and thus there are two choices for $\widehat{\mathbf{N}}\,(s)$ such that each is the negative of the other. One can then define a *signed curvature* associated to the oriented principal normal $\widehat{\mathbf{N}}$ given by the formula

$$k(s) \quad = \quad \left(\mathbf{x}''(s) \cdot \widehat{\mathbf{N}}\,(s)\right)$$

and Since $\mathbf{x}''(s)$ is perpendicular to $\mathbf{x}'(s)$ and $\mathbf{B}$ this may be rewritten in the form

$$\mathbf{x}''(s) \quad = \quad k(s)\,\widehat{\mathbf{N}}\,(s)\ .$$

An obvious question is to ask what happens if $\kappa(s_0) = 0$ (which also equals $k(s)$ in this case) and the sign of $k(s)$ is negative for $s < s_0$ and positive for $s > s_0$. A basic example of this sort is given by the graph of $f(x) = x^3$ near $x = 0$, whose standard parametrization is given by $\mathbf{x}(t) = (t,\, t^3)$. In this situation the graph lies in the lower half plane $y < 0$ for $t < 0$ and in the in the upper half plane $y > 0$ for $t > 0$, and the curve switches from being concave upward for $t < 0$ to concave downward (generally called *convex* beyond first year calculus courses). More generally, one usually says that f has a *point of inflection* in such cases. The following result shows that more general plane curves behave similarly provided the curvature has a nonvanishing derivative:

PROPOSITION. *Let $\mathbf{x}$ be a regular plane smooth curve parametrized by arc length plus a constant (hence $|\mathbf{x}'| = 1$), assume that $\mathbf{x}$ has a continuous fourth derivative, let $\widehat{\mathbf{N}}$ define a family of oriented principal normals for $\mathbf{x}$, and assume that that $k(s_0) = 0$ but $k'(s_0) > 0$. Then $\mathbf{x}(s)$ is contained in the half plane*

$$\widehat{\mathbf{N}}\,(s_0) \cdot \big(\mathbf{y} - \mathbf{x}(s_0)\big) \quad < \quad 0$$

for s sufficiently close to s_0 satisfying $s < s_0$, and $\mathbf{x}(s)$ is contained in the half plane

$$\widehat{\mathbf{N}}\,(s_0) \cdot \big(\mathbf{y} - \mathbf{x}(s_0)\big) \quad > \quad 0$$

for s sufficiently close to s_0 satisfying $s > s_0$.

A similar result holds if $k'(s_0) < 0$, and the necessary modifications of the statement and proof for that case are left to the reader as an exercise.

Proof. To simplify the computations we shall choose coordinate systems such that $\mathbf{x}(s_0) = \mathbf{0}$ and the plane is the standard coordinate plane through the origin with chosed unit normal vector $\mathbf{e}_3$. It will also be convenient to denote the unit vector $\mathbf{x}'(s)$ by $\mathbf{T}(s)$. We shall need to work with a third order approximation to the curve, which means that we are goint to need some information about $\mathbf{x}'''(s_0)$. Therefore the first step will be to establish the following formula:

$$k'(s_0) \quad = \quad \mathbf{x}'''(s_0) \ \cdot \ \widehat{\mathbf{N}}(s_0)$$

To see this, note that

$$k'(s) \quad = \quad \frac{d}{ds}\Big(\mathbf{x}''' \ \cdot \ \widehat{\mathbf{N}}\Big) \quad =$$

$$\Big(\mathbf{x}'''(s) \ \cdot \ \widehat{\mathbf{N}}(s)\Big) \ + \ \Big(\mathbf{x}''(s) \ \cdot \ \widehat{\mathbf{N}}'(s)\Big) \quad = \quad \Big(\mathbf{x}'''(s) \ \cdot \ \widehat{\mathbf{N}}(s)\Big) \ + \ \Big(\widehat{\mathbf{N}}(s) \ \cdot \ \widehat{\mathbf{N}}'(s)\Big)$$

and the second summand in the right hand expression vanishes because $|\widehat{\mathbf{N}}|^2$ is always equal to 1 (this is the same argument which implies that the unit tangent vector function is perpendicular to its derivative).

Turning to the proof of the main result, the preceding paragraph and earlier consideration show that the curve $\mathbf{x}$ is given near s_0 by the formula

$$\mathbf{x}(s) \quad = \quad (s - s_0)\,\mathbf{T}(s_0) \ + \ \frac{k(s)\,(s-s_0)^2}{2}\,\widehat{\mathbf{N}}(s_0) \ + \ \frac{(s-s_0)^3}{3!}\,\mathbf{x}'''(s_0) \ + \ (s-s_0)^4\,\theta(s)$$

where $\theta(s)$ is bounded for s suffieicntly close to zero. To simplify notation further we shall write $\Delta s = s - s_0$.

If we take the dot product of the preceding equation with $\widehat{\mathbf{N}}(s_0)$ we obtain the formula, in which $y(s)$ is the dot product of $\theta(s)$ and $\widehat{\mathbf{N}}(s_0)$, so that $y(s)$ is also bounded for s sufficiently close to s_0:

$$\Big(\mathbf{x}(s) \cdot \widehat{\mathbf{N}}(s_0)\Big) \quad = \quad \frac{k'(s_0)}{3!}\,(\Delta s)^3 + y(s)\,(\Delta s)^4$$

If s is nonzero but sufficiently close to zero then the sign of the right hand side is equal to the sign of Δs because

(i) the sign of the first term is equal to the sign of Δs,

(ii) if we let M be a positive upper bound for $|y(s)|$ and further restrict Δs so that

$$|\Delta s| \quad < \quad \frac{k'(s_0)}{6B}$$

then the absolute value of the second term in the dot product formula will be less than the absolute value of the first term.

It follows that the sign of the dot product

$$\Big(\mathbf{x}(s) \cdot \widehat{\mathbf{N}}(s_0)\Big)$$

is the same as the sign of the inital term

$$\frac{k'(s_0)}{3!}\,(\Delta s)^3$$

which in turn is equal to the sign of Δs. Since the dot product has the same sign as Δs for $s \neq 0$ and s sufficiently small, it follows that $\mathbf{x}(s)$ lies on the half plane defined by the inequality $\mathbf{y}\cdot\widehat{\mathbf{N}}(s_0) < 0$ if $s < s_0$ and $\mathbf{x}(s)$ lies on the half plane defined by the inequality $\mathbf{y}\cdot\widehat{\mathbf{N}}(s_0) > 0$ if $s > s_0$.■

In fact, the center of the osculating circle also switches sides when one goes from values of s that are less than s_0 to values of s that are greater than s_0. However, the proof takes considerably more work.

COMPLEMENT. *In the setting above, let $\mathbf{z}(s)$ denote the center of the osculating circle to $\mathbf{x}$ at parameter value at parameter value $s \neq s_0$ close to s_0 (this exists because the curvature is nonzero at such points). Then $\mathbf{z}(s)$ is contained in the half plane*

$$\widehat{\mathbf{N}}(s_0)\cdot\big(\mathbf{y}-\mathbf{x}(s_0)\big) < 0$$

for s sufficiently close to s_0 satisfying $s < s_0$, and $\mathbf{z}(s)$ is contained in the half plane

$$\widehat{\mathbf{N}}(s_0)\cdot\big(\mathbf{y}-\mathbf{x}(s_0)\big) > 0$$

for s sufficiently close to s_0 satisfying $s > s_0$.

Proof. We need to establish similar inequalities to those derived above if $\mathbf{x}(s)$ is replaced by $\mathbf{z}(s)$; note that the latter is not defined for parameter value s_0 because the formula involves the reciprocal of the curvature and the latter is zero at s_0.

The center of the osculating circle at parameter value $s \neq s_0$ was defined to be $\mathbf{x} + \kappa^{-1}\mathbf{N}$, where $\mathbf{N}$ is the ordinary principal normal; we claim that the latter is equal to $\mathbf{x} + k^{-1}\widehat{\mathbf{N}}$. By definition we have

$$\mathbf{x}'' \;=\; \kappa\,\mathbf{N} \;=\; k\,\widehat{\mathbf{N}}$$

and since $\kappa = \pm k$ is nonzero we know that $\kappa^2 = k^2$. Dividing the displayed equation by this common quantity yields the desired formula

$$\kappa^{-1}\mathbf{N} \;=\; k^{-1}\widehat{\mathbf{N}}\;.$$

Therefore the proof reduces to showing that the sign of

$$\left(\mathbf{x}(s) \;+\; \frac{1}{k(s)}\,\widehat{\mathbf{N}}(s)\right)\cdot\widehat{\mathbf{N}}(s_0)$$

is equal to the sign of Δs.

Using the formula for $\mathbf{x}(s)$ near s_0 that was derived before, we may rewrite the preceding expression as

$$h(s) \;=\; \frac{k'(s_0)}{3!}\,(\Delta s)^3 + y(s)\,(\Delta s)^4 \;+\; \frac{1}{k(s)}\,\widehat{\mathbf{N}}(s)\cdot\widehat{\mathbf{N}}(s_0)\;.$$

We need to show that $h(s)$ has the same sign as $k(s)$ and its reciprocal, and this will happen if

$$\ell(s) \;=\; h(s) - \frac{1}{k(s)} \;=\; \frac{k'(s_0)}{3!}\,(\Delta s)^3 + y(s)\,(\Delta s)^4 \;+\; \frac{1}{k(s)}\,\widehat{\mathbf{N}}(s)\cdot\Big(\widehat{\mathbf{N}}(s_0) \;-\; \widehat{\mathbf{N}}(s)\Big)$$

is bounded for $s \neq s_0$ sufficiently close to zero. To see, this, suppose that $|\ell(s)| \leq A$ for some $A > 0$. If we then choose $\delta > 0$ so that $|k(s)| < 1/A$ for for $|\Delta s| < \delta$ but $\Delta s \neq 0$, if will follow that

$$\Delta s \ > \ 0 \ \Longrightarrow \ h(s) \ = \ \frac{1}{k(s)} + \left(h(s) - \frac{1}{k(s)} \right) \ > \ A \ + \ (-A) \ > \ 0$$

and similarly with all inequalities reversed and A switched with $-A$ if $\Delta s < 0$.

In order to prove that $\ell(s)$ is bounded, it suffices to prove that each of the three summands is bounded for, say, $|\Delta s| \leq r$. The absolute value of the first is bounded by $k'(s_0)\, r^3/6$ and the absolute value of the second is bounded by $B\, r^4$ where B is a positive upper bound for $|y(s)|$. By the Cauchy-Schwarz inequality the absolute value of the third is bounded from above by

$$\frac{\left| \widehat{\mathbf{N}}\,(s) \ - \ \widehat{\mathbf{N}}\,(s_0) \right|}{|k(s)|}$$

and using the Mean Value Theorem we may estimate the numerator and denominator of this expression separately as follows:

(*i*) $\left| \widehat{\mathbf{N}}\,(s) - \widehat{\mathbf{N}}\,(s_0) \right| \leq P \cdot |\Delta s|$, where P is the maximum value of $|\widehat{\mathbf{N}}'|$ on $[s_0 - r,\, s_0 + r]$.

(*ii*) $k(s) = k'(S_1)\,\Delta s$ for some S_1 between s_0 and s, so if we choose r so small that $k' > 0$ on $[s_0 - r,\, s_0 + r]$, then $|k(s)| \geq Q\,\Delta s$, where $Q > 0$ is the minimum of k' on that interval.

It then follows that the quotient P/Q is an upper bound for the absolute value of the third term in the formula for $\ell(s)$, and therefore the latter itself is bounded. This completes the proof that $z(s)$ lies on the half plane described in the statement of the result.■

I.5 : Frenet-Serret Formulas

(do Carmo, §§1–5, 1–6, 4–Appendix)

The Frenet-Serret Formulas describe the derivatives of the three fundamental unit vectors in the Frenet trihedron associate to a curve.

FRENET–SERRET FORMULAS. *Let* $\mathbf{x}$ *be a regular smooth curve parametrized by arc length (hence* $|\mathbf{x}'| = 1$*), assume that* $\mathbf{x}$ *has a continuous third derivative, and assume also that* $\kappa(s_0) \neq 0$*. Let* $\mathbf{T}(s)$*,* $\mathbf{N}(s)$ *and* $\mathbf{B}(s)$ *be the tangent, principal normal and binormal vectors in the Frenet trihedron for* $\mathbf{x}$ *at parameter value* s_0*. Then the following equations describe the derivatives of the vectors in the Frenet trihedron:*

$$\begin{aligned} \mathbf{T}' &= \qquad\quad \kappa\,\mathbf{N} \\ \mathbf{N}' &= -\kappa\,\mathbf{T} \qquad\quad - \tau\,\mathbf{B} \\ \mathbf{B}' &= \qquad\quad \tau\,\mathbf{N} \end{aligned}$$

Proof. We have already noted that the first and third equations are direct consequences of the definition of curvature and torsion. To derive the second equation, we take the identity $\mathbf{N} = \mathbf{B} \times \mathbf{T}$ and differentiate it with respect to s:

$$\mathbf{N}'(s) \;=\; \mathbf{B}'(s) \times \mathbf{T}(s) \;+\; \mathbf{B}(s) \times \mathbf{T}'(s) \;=$$

$$\tau(s)\,\big(\mathbf{N}(s) \times \mathbf{T}(s)\big) \;+\; \kappa\,\big(\mathbf{B}(s) \times \mathbf{N}(s)\big)$$

Since $\mathbf{T}$, $\mathbf{N}$ and $\mathbf{B}$ are mutually perpendicular unit vectors such that $\mathbf{B} = \mathbf{T} \times \mathbf{N}$, as usual the "BAC–CAB" rule for threefold cross products implies that $\mathbf{N} \times \mathbf{T} = -\mathbf{B}$ and $\mathbf{B} \times \mathbf{N} = -\mathbf{T}$. If we make these substitions into the displayed equations we obtain the second of the Frenet-Serret Formulas.∎

The signifiance of the Frenet-Serret formulas is that they allow one to describe a curve in terms of its curvature and torsion in an essentially complete manner.

LOCAL UNIQUENESS FOR CURVES. *Suppose that we are given two regular smooth curves* $\mathbf{x}$ *and* $\mathbf{y}$ *defined on the same interval containing* s_0*, where both curves are parametrized by arc length, both have continuous third derivatives and everywhere nonzero curvatures, and their curvature and torsion functions of both curves are equal. Then there is a rigid motion* Φ *of* 3*-dimensional space such that* $\mathbf{y} = \Phi \circ \mathbf{x}$*.*

A rigid motion of $\mathbf{R}^2$ or $\mathbf{R}^3$ is a 1–1 and onto mapping φ such that

$$|\Phi(\mathbf{b}) \;-\; \Phi(\mathbf{a})| \;=\; |\mathbf{b} \;-\; \mathbf{a}|$$

for all vectors $\mathbf{b}$ and $\mathbf{a}$. In linear algebra it is shown that every such rigid motion has the form

$$\Phi(\mathbf{x}) \;=\; A\,\mathbf{x} + \mathbf{c}$$

where $\mathbf{c}$ is some fixed vector and A is an orthogonal matrix (*i.e.*, its rows and columns are orthonormal sets — actually, the rows are orthonormal if and only if the columns are, but we do not need this right now).

Proof. Let $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ be the standard unit vectors. We shall only consider the simplified situation where $\mathbf{x}(s_0) = \mathbf{y}(0) = \mathbf{0}$ and the Frenet trihedra for $\mathbf{x}$ and $\mathbf{y}$ at parameter value s_0 are given by $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ (one can always use a rigid motion to move the original curves into such positions, and the motion will not change the curvature or torsion of either curve — this is not really difficult to prove but it is a bit tedious and distracting).

Let $\{\mathbf{T}_\mathbf{x}(s), \mathbf{N}_\mathbf{x}(s), \mathbf{B}_\mathbf{x}(s)\}$ and $\{\mathbf{T}_\mathbf{y}(s), \mathbf{N}_\mathbf{y}(s), \mathbf{B}_\mathbf{y}(s)\}$ be the Frenet trihedra for $\mathbf{x}$ and $\mathbf{y}$ respectively, and let

$$g(s) \;=\; |\mathbf{T}_\mathbf{x}(s) - \mathbf{T}_\mathbf{y}(s)|^2 \;+\; |\mathbf{N}_\mathbf{x}(s) - \mathbf{N}_\mathbf{y}(s)|^2 \;+\; |\mathbf{B}_\mathbf{x}(s) - \mathbf{B}_\mathbf{y}(s)|^2 \;.$$

By the Frenet-Serret Formulas we then have that g' is equal to

$$2\left(\Big((\mathbf{T}_\mathbf{x} - \mathbf{T}_\mathbf{y})\cdot(\mathbf{T}'_\mathbf{x} - \mathbf{T}'_\mathbf{y})\Big) + \Big((\mathbf{N}_\mathbf{x} - \mathbf{N}_\mathbf{y})\cdot(\mathbf{N}'_\mathbf{x} - \mathbf{N}'_\mathbf{y})\Big) + \Big((\mathbf{B}_\mathbf{x} - \mathbf{B}_\mathbf{y})\cdot(\mathbf{B}'_\mathbf{x} - \mathbf{B}'_\mathbf{y})\Big) \right) =$$

$$2\Bigg(\Big(\kappa\,(\mathbf{T}_\mathbf{x} - \mathbf{T}_\mathbf{y})\cdot(\mathbf{N}_\mathbf{x} - \mathbf{N}_\mathbf{y})\Big) + \Big(\tau\,(\mathbf{B}_\mathbf{x} - \mathbf{B}_\mathbf{y})\cdot(\mathbf{N}_\mathbf{x} - \mathbf{N}_\mathbf{y})\Big) -$$

$$\Big(\kappa\,(\mathbf{N}_\mathbf{x} - \mathbf{N}_\mathbf{y})\cdot(\mathbf{T}_\mathbf{x} - \mathbf{T}_\mathbf{y})\Big) - \Big(\tau\,(\mathbf{N}_\mathbf{x} - \mathbf{N}_\mathbf{y})\cdot(\mathbf{B}_\mathbf{x} - \mathbf{B}_\mathbf{y})\Big) \Bigg) .$$

It is an elementary but clearly messy exercise in algebra to simplify the right hand side of the preceding equation, and the expression in question turns out to be zero. Therefore the function g must be a constant, and since our assumptions imply $g(s_0) = 0$, it follows that $g(s) = 0$ for all s. The latter in turn implies that each summand

$$|\mathbf{T}_\mathbf{x} - \mathbf{T}_\mathbf{y}|^2 \;,\; |\mathbf{N}_\mathbf{x} - \mathbf{N}_\mathbf{y}|^2 \;,\; |\mathbf{B}_\mathbf{x} - \mathbf{B}_\mathbf{y}|^2$$

must be zero and hence that the Frenet trihedra for $\mathbf{x}$ and $\mathbf{y}$ must be the same. The first Frenet-Serret Formula then implies $\mathbf{x}' = \mathbf{y}'$, and since the two curves both go through the origin at parameter value s_0 it follows that $\mathbf{x}$ and $\mathbf{y}$ must be identical.■

There is in fact a converse to the preceding result.

FUNDAMENTAL EXISTENCE THEOREM OF LOCAL CURVE THEORY. *Given sufficiently differentiable functions κ and τ on some interval $(-c, c)$ then there is an $h \in (0, c)$ and a sufficiently differentiable curve $\mathbf{x}$ defined on $(0, h)$ such that $\mathbf{x}(0) = \mathbf{0}$, the tangent vectors to $\mathbf{x}$ at all point have unit length, the Frenet trihedron of $\mathbf{x}$ at 0 is given by the standard unit vectors*

$$\Big(\mathbf{T}(0),\, \mathbf{N}(0),\, \mathbf{B}(0)\Big) \;=\; \Big(\mathbf{e}_1,\, \mathbf{e}_2,\, \mathbf{e}_3\Big)$$

and the curvature and torsion functions are given by the restrictions of κ and τ respectively.■

This is a consequence of the fundamental existence theorem for systems of linear differential equations. If the curve exists, then the Frenet-Serret formulas yield a system of nine first order linear differential equations for the vector valued functions $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$ in the Frenet trihedron

$$\begin{aligned} \mathbf{T}' &= \kappa\,\mathbf{N} \\ \mathbf{N}' &= -\kappa\,\mathbf{T} - \tau\,\mathbf{B} \\ \mathbf{B}' &= \tau\,\mathbf{N} \end{aligned}$$

and if one is given κ and τ the goal is to see whether this system of first order linear differential equations can be solved for $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$, at least on some smaller interval $(-h,\, h)$. If one has such a solution then the curve $\mathbf{x}$ can be retrieved using the elementary formula

$$\mathbf{x}(s) \quad = \quad \int_0^s \mathbf{T}(u)\, du$$

where $|s| < h$ (with the usual convention that $\int_0^s \;=\; -\int_s^0$ if $s < 0$). A proof of the existence of a solution to the system of differential equations is given on pages 309–311 in the Appendix to Chapter 4 of do Carmo.■

The preceding two results combine to yield the **Fundamental Theorem of Local Curve Theory:**

Given κ and τ as in the statement of the Existence Theorem, an initial vector $\mathbf{x}_0$ and an orthonormal set of vectors $(\mathbf{a},\, \mathbf{b},\, \mathbf{c})$ such that $\mathbf{a} \times \mathbf{b} = \mathbf{c}$, then there is a positive real number h_1 and a unique (sufficiently differentiable) curve $\mathbf{x}$ such that the tangent vectors to $\mathbf{x}$ at all point have unit length, the Frenet trihedron of $\mathbf{x}$ at 0 is given by the standard unit vectors

$$\Big(\mathbf{T}(0),\, \mathbf{N}(0),\, \mathbf{B}(0)\Big) \quad = \quad (\mathbf{a},\, \mathbf{b},\, \mathbf{c})$$

and the curvature and torsion functions are respectively given by the restrictions of κ and τ to $(-h_1,\, h_1)$.■

Since plane curves may be viewed as space curves whose third coordinates are zero (and whose torsion functions are zero), the Fundamental Theorem of Local Curve Theory also applies to plane curves, and in fact the Fundamental Theorem amounts to saying that there is a unique curve with a given (nonzero) curvature function κ, initial value $\mathbf{x}_0$ and initial unit tangent vector $\mathbf{T}_0$; in this case the principal normal $\mathbf{N}_0$ is completely determined by the. perpendicularity condition and the Frenet-Serret Formulas.

Local canonical forms

One application of the Frenet-Serret formulas is a description of a strong third order approximation to a curve in terms of curvature and torsion.

PROPOSITION. *Let $\mathbf{x}$ be a regular smooth curve parametrized by arc length plus a constant (hence $|\mathbf{x}'| = 1$) such that $\mathbf{x}$ has a continuous fourth derivative and $\kappa(0) \neq 0$, and let $\{\, \mathbf{T},\, \mathbf{N},\, \mathbf{B}\,\}$ be the Frenet trihedron at parameter value $s = 0$. Then a strong third order approximation to $\mathbf{x}$ is given by*

$$\mathbf{x}(0) \;+\; \left(s - \frac{s^2\kappa^2}{3!}\right)\mathbf{T} \;+\; \left(\frac{s^2\kappa}{2} + \frac{s\kappa'}{3!}\right)\mathbf{N} \;+\; \frac{s^3\kappa\tau}{3!}\mathbf{B}\,.$$

Proof. We already know that $\mathbf{x}'(0) = \mathbf{T}$ and $\mathbf{x}''(0) = \kappa\,\mathbf{N}$. It suffices to compute $\mathbf{x}'''(0)$, and the latter is given by

$$\big(\kappa\,\mathbf{N}\big)' \quad = \quad \kappa'\,\mathbf{N} \;+\; \kappa\,\mathbf{N}' \quad = \quad \kappa'\,\mathbf{N} \;-\; \kappa^2\,\mathbf{T} \;-\; \kappa\,\tau\mathbf{B}$$

where the last is derived using the Frenet-Serret Formulas.■

Two significant applications of the canonical form for the strong third order approximation appear on pages 28–29 of do Carmo. The proofs are elementary and contained on these pages of the text.

APPLICATION 1. *In the setting above, if* $\tau(0) > 0$ *then the point* $\mathbf{x}(s)$ *lies on the side of the osculating plane defined by the inequality* $(\mathbf{y} - \mathbf{x}(0)) \cdot \mathbf{B} < 0$*, when* $s < 0$ *and* s *is sufficiently close to* 0*, and* $\mathbf{x}(s)$ *lies on the side of the osculating plane defined by the inequality* $(\mathbf{y} - \mathbf{x}(0)) \cdot \mathbf{B} > 0$ *when* $s > 0$ *and* s *is sufficiently close to* 0*. Similarly, if* $\tau(0) > 0$ *then the point* $\mathbf{x}(s)$ *lies on the side of the osculating plane defined by the inequality* $(\mathbf{y} - \mathbf{x}(0)) > 0$*, when* $s < 0$*, and* $\mathbf{x}(s)$ *lies on the side of the osculating plane defined by the inequality* $(\mathbf{y} - \mathbf{x}(0)) < point\mathbf{x}(s) lies on the other side of the osculating plane 0$.■

APPLICATION 2. *In the setting above, if* $s \neq 0$ *is sufficiently close to zero then* $\mathbf{x}(s)$ *lies on the side of the rectifying plane defined by the inequality*

$$\big(\mathbf{y} - \mathbf{x}(0)\big) \cdot \mathbf{N} > 0 \;.■$$

Regular smooth curves in hyperspace

During the nineteenth century mathematicians and physicists encountered numerous questions that had natural interpetations in terms of spaces of dimension greater than three (incidentally, in physics this began long before the viewing of the universe as a 4-dimensional space-time in relativity theory). In particular, coordinate geometry gave a powerful means of dealing with such objects by analogy. For example, Euclidean n-space for and arbitrary finite n is given by the vector space $\mathbf{R}^n$, lines, planes, and various sorts of hyperplanes can be defined and studied by algebraic methods (although geometric intuition often plays a key role in formulating, proving, and interpreting results!), and distances and angles can be defined using a simple generalization of the standard dot product. Furthermore, objects like a 4-dimensional hypercube or a 3-dimensional hypersphere can be described using familiar sorts of equations. For example, a typical hypercube is given by all points $\mathbf{x} = (x_1, x_2, x_3, x_4)$ such that $0 \leq x_i \leq 1$ for all i, and a typical hypersphere is given by all points $\mathbf{x}$ such that

$$|\mathbf{x}|^2 \;=\; x_1^2 + x_2^2 + x_3^2 + x_4^2 \;=\; 1\,.$$

A full investigation of differential geometry in Euclidean spaces of dimension ≥ 4 is beyond the scope of this course, but some comments about the differential geometry of curves in 4-space seem worth mentioning.

One can define regular smooth curves, arc length and curvature for parametrized 4-dimensional curves exactly as for curves in 3-dimensional space. In fact, there are generalizations of the Frenet-Serret formula and the Fundamental Theorem of Local Curve Theory. One complicating factor is that the 3-dimensional cross product does not generalize to higher dimensions in a particularly neat fashion, but one can devolop algebraic techniques to overcome this obstacle. In any case, in four dimensions if a sufficiently differentiable regular smooth curve $\mathbf{x}$ is parametrized by arc length plus a constant and has nonzero curvature and a nonzero secondary curvature (which is similar to the torsion of a curve in 3-space), then for each parameter value s there is an ordered orthonormal set of vectors $\mathbf{F}_i(s)$, where $1 \leq i \leq 4$, such that $\mathbf{F}_1$ is the unit tangent vector and the sequence of vector valued functions (the *Frenet frame* for the curve) satisfies the following system of differential equations, where κ_1 is curvature, κ_2 is positive valued, and the functions κ_1, κ_2, κ_3, all have sufficiently many derivatives:

$$\begin{array}{lcrcccc}
\mathbf{F}_1' & = & & \kappa_1\,\mathbf{F}_2 & & & \\
\mathbf{F}_2' & = & -\;\kappa_1\,\mathbf{F}_1 & & +\;\;\kappa_2\,\mathbf{F}_3 & & \\
\mathbf{F}_3' & = & & -\kappa_2\,\mathbf{F}_2 & & + & \kappa_3\,\mathbf{F}_4 \\
\mathbf{F}_4' & = & & & -\kappa_3\,\mathbf{F}_3 & &
\end{array}$$

The Fundamental Theorem of Local Curve Theory in 4-dimensional space states that locally there is a unique curve with prescribed higher curvature functions $\kappa_1 > 0$, $\kappa_2 > 0$ and κ_3, prescribed initial value $\mathbf{x}(s_0)$, and whose Frenet orthonormal frame satisfies $\mathbf{F}_i(s_0) = \mathbf{v}_i$ for some orthonormal basis $\{\,\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\mathbf{v}_4\,\}$. An online description and derivation of such formulas in arbitrary dimensions is available at the site

`http://www.math.technion.ac.il/ rbrooks/dgeo1.7.ps`

and a discussion of such formulas in complete generality (*i.e.*, appropriate for a graduate level course) appears on page 74 of Hicks, *Notes on Differential Geometry.*

II. Closed Cuves as Boundaries

Nineteenth century techniques provided powerful means for analyzing curves quantitatively and locally. However, by the end of the nineteenth century mathematicians and users of mathematics encountered several questions of a more qualitative nature for closed curves. The difference between local and global properties is characterized very well on the first paragraph of page 34 of Stoker, *Differential Geometry*: A global property is one that cannot be studied by only examining the curve near each point or parameter value.

Let us agree that a regular smooth closed curve of class $\mathcal{C}^r$ (where $1 \leq r \leq \infty$) is a regular smooth curve $\mathbf{x} : [a, b] \to \mathbf{R}^n$, where $n = 2$ or 3, such that for each $k \leq r$ the k^{th} derivatives at a and b are equal (if $r = \infty$ the inequality should be $k < r$). Such a curve is said to be a *simple closed curve* if $\mathbf{x}(t_1) \neq \mathbf{x}(t_2)$ for $t_1 \neq t_2$ unless $t_1 = a$ and $t_2 = b$ or vice versa; in other words, it never goes through the same point twice except at the endpoints.

Here is an example of a global property that is "intuitively obvious" but definitely not easy to prove from scratch. For plane curves, examples strongly suggest that the following statement is true:

JORDAN CURVE THEOREM. *Let C be the image of a regular smooth closed curve in $\mathbf{R}^2$. Then the points of the plane not in C are contained in one of two regions A and B such that*

(i) if two points are either in A or in B, then there is a regular smooth curve joining them which lies entirely in A or B respectively,

(ii) there is a suitably large disk about the origin that contains one of the regions (the **inside** *region) and the curve itself, but there is no such disk that contains the other one (the* **inside** *region),*

(iii) if one point lies in A and the other lies in B, then every regular smooth curve joining the two points must also pass through a point of C.

The best way to understand the meaning of this result is to sit down and draw all sorts of closed curves in the plane. Each of them looks as if it has an inside region and an outside region, just like a circle. For a circle it is easy to describe the inside and outside explicitly; one is the set of all point whose distance from the center less than the radius, and the other is the set of point whose distance from the center is greater than the radius. In other relatively simple cases one can similarly describe the inside and outside regions using explicit inequalities (the reader is invited to try this with some basic examples and also to do so for *piecewise smooth* closed curves such as triangles), but in general it is hopeless to search for such descriptions. An excellent example in this connection is given online at the site

`http://ccins.camosun.bc.ca/~jbritton/fishmaze.pdf`

which also indicates the relative difficulty of determing whether two points not on the curve lie in the same or different regions. Clearly one needs additional means to attack such questions successfully.

Although the validity of the Jordan Curve Theorem was surely believed by many mathematicians and users of the subject for a long time and proofs in wide ranges of special cases were well understood, attempts to give a mathematically rigorous proof of a general result along these lines did not begin until quite late in the nineteenth century, and the first complete proof was given by O. Veblen during the first decade of the twentieth century. In fact, he proved the result under the weaker hypothesis that the simple closed curve is merely continuous. A subsequent result of A. Schönflies yielded much stronger conclusions, including a fairly explicit description of the inside region associated to the simple closed curve. One objective of this section is to review and augment the concepts and results from multivariable calculus that are needed to formulate one version of the Schönflies Theorem. Another is to consider some applications of the Jordan Curve and Schönflies Theorems to studying the global properties of smooth regular closed curves.

The Jordan Curve Theorem is proved in many topology textbooks. A proof for smooth curves using ideas from differential geometry is given in Section 16 of the clasic text by J. J. Stoker (*Differential Geometry, etc.*). The Schönflies Theorem is somewhat more difficult to locate in textbooks; we shall discuss one proof of this result in the files `schoenflies.*` that are stored in the course directory, but this proof requires material from the texts for the Department's graduate level courses in topology and complex analysis. A concise but very informative summary of the history of the Jordan Curve Theorem, the Schönflies Theorem and their analogs in higher dimensions is available online at the site

`http://math.ohio-state.edu/ fiedorow/math655/Jordan.html`

II.1 : Regions, limits and continuity

(do Carmo, 2–Appendix A, 5–Appendix)

In the discussion of the Jordan Curve Theorem we mentioned the concept of a region without saying exactly what we meant. The first order of business is to make this precise.

When working with continuous functions of a single real variable, there are only a few reasonable choices for the sets on which most functions in elementary calculus are defined. Namely, for most purposes the right sets to consider are intervals of some sort, the main choices being whether the intervals have left or right hand endpoints or are bounded or unbounded either to the left or to the right. For continuous functions of two or more variables, the situation is far more complicated even if we restrict ourselves to sets defined by reasonable inequalities (note that intervals are defined by one or more inequalities, each of which may or may not be strict). Further discussion and an example appear on pages 838–839 of *Calculus* (Seventh Edition), by Larson, Hostetler and Edwards, and Exercises 17–28 on page 846 of that book provide further examples for consideration.

In two or more dimensions, boundary sets can be extremely difficult to analyze. This contrasts sharply with the 1-dimensional situation, where the boundary generally consists of finitly many points. Most of the special plane curves described in the links at the beginning of Unit I are boundaries (or pieces from the boundaries) of fairly decent regions, but one can also go further and view most of the fractal curves from the references in Section I.3 as boundaries of regions that are not all that terrible. Mathematicians want and need to understand just how bad boundaries can be, and during the past 80 years they have developed a large array of methods for constructing regions whose boundaries are extremely wild. The subject of fractal geometry deals with some special types of irregular boundaries that are "not too wild" in an appropriate sense. If one goes from two dimensions to three, the variety of bizarre possibilities increases dramatically (the previously mentioned site on the Jordan Curve Theorem has some particularly striking pictures). For these reasons it is frequently convenient to focus on the interior (non-boundary) points of regions, and formally one does this as follows:

Definition. Let $n = 2$ or 3 (actually,everything works for all $n \geq 2$, but in this course we are mainly interested in objects that exist in 2- or 3-dimensional space. A subset U of $\mathbf{R}^n$ will be called a *connected open domain* provided

(i) for each $\mathbf{p} \in U$ there is an $r > 0$ such that the **open disk** or **ball** or **neighborhood** centered at $\mathbf{p}$ with **radius** equal to r

$$N_r(\mathbf{p}) \quad = \quad \left\{ \mathbf{q} \in \mathbf{R}^n \;\middle|\; |\mathbf{q} \ - \ \mathbf{p}| \ < \ r \right\}$$

is entirely contained in U,

(ii) for each pair of points $\mathbf{p}$ and $\mathbf{q}$ in U there is a piecwise smooth curve Γ defined on $[0,1]$ and taking values entirely in U such that $\Gamma(0) = \mathbf{p}$ and $\Gamma(1) = \mathbf{q}$.

Most of the subsets of $\mathbf{R}^n$ that are defined by finitely many strict polynomial inequalities of the form $p_i(\mathbf{y}) \leq a_i$ satisfy (i), and either they satisfy (ii) or else they can be split into pairwise disjoint pieces such that (ii) holds on each of the pieces. Of course, these ssets include a vast number of central examples in multivariable calculus.

It is only necessary to make a few minor adjustments in order to work with limits and continuity on connected open domains. Once again the basic idea is that a function f is defined for all points sufficiently close to a pont **a** in some regions except perhaps at **a** itself. Since we are dealing with an open domain U this amounts to saying that there is some $r > 0$ such that the function is defined on the deleted neighborhood

$$N_r(\mathbf{a}) \;-\; \{\mathbf{a}\} \;=\; \left\{\, \mathbf{x} \in \mathbf{R}^n \;\middle|\; 0 \;<\; |\mathbf{x} \;-\; \mathbf{a}| \;<\; r \right\} .$$

One may then say that

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) \;=\; b$$

if and only if

for every $\varepsilon > 0$ there is a $\delta > 0$ such that $\mathbf{x} \neq \mathbf{a}$ and $|\mathbf{x} - \mathbf{a}| < \delta$ imply $|f(\mathbf{x}) - b| < \varepsilon$.

Continuity at **a** then means that

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) \;=\; f(\mathbf{a})$$

exactly as in the (opaque to the beginner) definitions from first year calculus.

Further discussion, including examples, pictures and exercises for review, may be found in Section 12.2 of *Calculus* (Seventh Edition), by Larson, Hostetler and Edwards, and pages 115–124 of *Basic Multivariable Calculus*, by Marsden, Tromba and Weinstein (full bibliographical information on these books appears in the online files `background.*` mentioned at the beginnning of these notes.

Limits and continuity for vector valued functions will also play an important role in this course. The quickest way to address this point is to say that a vector valued function has a limit if and only if each of its coordinate functions does, and in this case the limit of the vector valued function is the vector whose coordinates are the limits of the coordinate functions.

VECTOR LIMIT FORMULA. *Let* **F** *be a vector valued function defined on a deleted neighborhood of* **a**, *let f_i denote the i^{th} coordinate function of* **F**, *and suppose that*

$$\lim_{\mathbf{x}\to\mathbf{a}} f_i(x) \;=\; b_i$$

holds for all i. As usual let $\mathbf{e}_i$ denote the i^{th} unit vector in $\mathbf{R}^n$. Then we have

$$\lim_{\mathbf{x}\to\mathbf{a}} F_i(x) \;=\; \sum_{i=1}^{n} b_i\, \mathbf{e}_i \;.\blacksquare$$

This has an important consequence:

CONTINUITY AND COORDINATE FUNCTIONS. *In the notation as above, assume that all fucntions are also defined at* **a**. *Then* **F** *is continuous at* **a** *if and only if for each i, the coordinate function f_i is continuous at* **a**.■

Alternative approach

Here is another way of looking at limits that is an even more direct generalization of ideas from single variable calculus. Given a sequence $\{\,\mathbf{x}_k$ of vectors in $\mathbf{R}^n$ we can define the limit of the sequence to be the vector whose i^{th} coordinate is the limit of the numerical sequence $\{ P_i(\mathbf{x}_k) \}$ where P_i denotes the operation of taking the i^{th} coordinate of a vector. We then have the following characterization:

LIMITS AND SEQUENCES. *Let* $\mathbf{F}$ *be a vector valued function defined on a deleted neighborhood of* $\mathbf{a}$. *Then*

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{F}(\mathbf{x}) \quad = \quad \mathbf{b}$$

if and only if for each sequence of vectors $\{ \mathbf{x}_k$ *in the deleted neighborhood of* $\mathbf{a}$ *whose limit is* $\mathbf{a}$ *we have*

$$\lim_{k\to\infty} \mathbf{F}(\mathbf{x}_k) \quad = \quad \mathbf{b}\ .\blacksquare$$

There is a similar statement for continuity; writing down the precise statement of the result is left to the reader.■

II.2 : Smooth mappings

(do Carmo, 2–Appendix B)

II.3 : Inverse and implicit function theorems

(do Carmo, 2–Appendix B)

II.4 : Global properities of plane curves

(do Carmo, §1–7)