The Calculus of Space Curves

(**MCS65C | F15: Reading Assignment 1**)

Introduction

We've been considering the notion of differentiability for vector-valued functions $\mathbf{f} : D \subseteq \mathbb{R}^d \longrightarrow \mathbb{R}^e$, where d and e are natural numbers. In this exploration, we specialize from the general multivariable setting to a case of great importance: namely, that in which $d = 1$ and $e \geq 2$.

A function $\mathbf{f} : D \subseteq \mathbb{R} \longrightarrow \mathbb{R}^e$ $(e \geq 2)$ is a *curve* — also called a *path*, *motion*, or *trajectory* — if (i) D is connected and contains at least two points, and (ii) $\mathbf{f}$ is continuous on D. The connected subsets of $\mathbb{R}$ are the empty set $\emptyset$, the singletons $\{a\}$, and the *intervals*: (a, b), $[a, b]$, $(a, b]$, $[a, b)$, (a, ∞), $[a, \infty)$, $(-\infty, b)$, $(-\infty, b]$, and $(-\infty, \infty)$, where $a < b$. Thus, the set D must be an interval. We'll typically use the letters I and J to denote intervals, and we'll write $\boldsymbol{\gamma} : I \longrightarrow \mathbb{R}^e$, $\boldsymbol{\delta} : J \longrightarrow \mathbb{R}^e$, etc., for typical curves. The single real variable of which $\boldsymbol{\gamma}$ is a function will be denoted by t or τ, or sometimes by s, and thought of physically as *time*. Thus, as time flows, the point $\boldsymbol{\gamma}(t)$ moves continuously in $\mathbb{R}^e$, tracing out an unbroken curved line that is, on an intuitive level, "one point thick" everywhere. This point-set is called the *trace* of $\boldsymbol{\gamma}$. It is simply the range of the function $\boldsymbol{\gamma}$, i.e., $\boldsymbol{\gamma}(I) \subseteq \mathbb{R}^e$.

The Velocity Vector

Fix a time $t \in I^\circ$. According to our general definition of differentiability, $\boldsymbol{\gamma}$ is differentiable at t if there exists an $e \times 1$ column matrix — i.e., a *vector* $\mathbf{v} \in \mathbb{R}^e$, which we denote by $\mathbf{v} := D\boldsymbol{\gamma}(t)$ — such that

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0 :\ \forall h \in B_\delta(0) = (-\delta, \delta), \quad \|\boldsymbol{\gamma}(t+h) - \boldsymbol{\gamma}(t) - \mathbf{v}h\| < \varepsilon|h|.$$

Dividing this inequality through by $|h|$ and using the absolute scaling property ($\|h\mathbf{a}\| = |h|\|\mathbf{a}\|$), it becomes

$$\left\| \frac{\boldsymbol{\gamma}(t+h) - \boldsymbol{\gamma}(t)}{h} - \mathbf{v} \right\| < \varepsilon.$$

This means simply that

$$\lim_{h \to 0} \frac{\boldsymbol{\gamma}(t+h) - \boldsymbol{\gamma}(t)}{h} = \mathbf{v}.$$

On the other hand, since subtraction, scaling, and limits all operate componentwise, we see that the i-th component of $\mathbf{v}$ must be

$$v_i = \lim_{h \to 0} \frac{\gamma_i(t+h) - \gamma_i(t)}{h} = \gamma_i'(t).$$

So if $\boldsymbol{\gamma}$ is differentiable at t, then each γ_i is differentiable at t in the sense of one-variable calculus, and conversely. Moreover,

$$D\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}'(t) := (\gamma_1'(t), \gamma_2'(t), \ldots, \gamma_e'(t)).$$

We call $\mathbf{v}(t) := \boldsymbol{\gamma}'(t)$ the *velocity vector* of $\boldsymbol{\gamma}$ at t, for physical reasons to be explored presently.

It's easy to see that if the velocity vector is nonzero, then the unit vector in its direction, $\mathbf{v}(t)/\|\mathbf{v}(t)\|$, is the *instantaneous direction of motion* at time t for a point-particle moving according to $\boldsymbol{\gamma}$. This can be pictured as follows: imagine that all forces acting on the particle are simultaneously "switched off" at time t; then the particle would continue moving in a straight line in the direction of $\mathbf{v}(t)$. Since we're interested only in the direction of motion for *forward time*, let's consider a small number $h > 0$, and define the *unit secant vector* from t to $t + h$ as

$$\mathbf{T}_h(t) := \frac{\boldsymbol{\gamma}(t+h) - \boldsymbol{\gamma}(t)}{\|\boldsymbol{\gamma}(t+h) - \boldsymbol{\gamma}(t)\|}, \qquad (0 < h < \delta).$$

This is the unit vector starting from $\boldsymbol{\gamma}(t)$ that passes through the nearby curve point $\boldsymbol{\gamma}(t+h)$. As $h \to 0^+$, the point $\boldsymbol{\gamma}(t+h)$ will slide back along the curve toward $\boldsymbol{\gamma}(t)$, and the vector $\mathbf{T}_h(t)$ will gradually change direction (while maintaining unit length). The instantaneous direction of motion, or *unit tangent vector*, is naturally defined as the limiting position of this gradually changing secant vector:

$$\mathbf{T}(t) := \lim_{h \to 0^+} \mathbf{T}_h(t),$$

provided the limit exists. On the other hand, since $h > 0$, we can use the absolute scaling property of norms to write

$$\mathbf{T}_h(t) = \frac{\{\boldsymbol{\gamma}(t+h) - \boldsymbol{\gamma}(t)\}/h}{\|\{\boldsymbol{\gamma}(t+h) - \boldsymbol{\gamma}(t)\}/h\|}.$$

As $h \to 0^+$, the numerator of this fraction approaches $\mathbf{v}(t)$ by definition (actually the limit in the definition of $\mathbf{v}$ is two-sided; but when a two-sided limit exists, the one-sided limits must exist and agree with it in value). Using the continuity of the norm function, the denominator of the fraction approaches $\|\mathbf{v}(t)\|$. Since we are assuming that $\mathbf{v}(t) \neq \mathbf{0}$, the quotient rule for limits implies that

$$\mathbf{T}(t) = \lim_{h \to 0^+} \mathbf{T}_h(t) = \frac{\mathbf{v}(t)}{\|\mathbf{v}(t)\|}.$$

To interpret the physical meaning of a vector, it suffices to understand both its length and its direction. We now understand the direction of the velocity vector $\mathbf{v}(t)$, and we wish to interpret its length. We will show that $\|\mathbf{v}(t)\|$ is the *instantaneous speed* at time t of a particle moving according to $\boldsymbol{\gamma}$.

Since speed is supposed to involve "distance traveled divided by time elapsed", we must first determine the total distance traveled by the particle *along the trace of the curve* from time a to time b, where $a, b \in I$ satisfy $a \leq b$. This is called the *arclength* of $\boldsymbol{\gamma}$ from a to b, and denoted by $L_a^b(\boldsymbol{\gamma})$.

Arclength

Consider a partition $\mathscr{P} = \{t_0, t_1, \ldots, t_n\}$ of $[a, b]$, so that

$$a = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = b.$$

We write $\mathscr{P} \vdash [a, b]$ to indicate this. The *mesh* of $\mathscr{P}$ is defined to be the largest subinterval length,

$$\|\mathscr{P}\| := \max_{1 \leq i \leq n} \triangle t_i,$$

where $\triangle t_i := t_i - t_{i-1}$. We can inscribe a broken line in the arc of $\boldsymbol{\gamma}$ from $t = a$ to $t = b$ by connecting each pair of consecutive curve points $\boldsymbol{\gamma}(t_{i-1})$ and $\boldsymbol{\gamma}(t_i)$ by a line segment. The total length of this broken line is given by

$$L(\boldsymbol{\gamma}, \mathscr{P}) := \sum_{i=1}^{n} \|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\|.$$

The natural definition of arclength is the limit of this total length as the broken lines become more and more tightly inscribed in the arc, i.e., as the mesh $\|\mathscr{P}\|$ tends to zero. That is, we propose to define

$$L_a^b(\boldsymbol{\gamma}) := \lim_{\substack{\|\mathscr{P}\| \to 0 \\ \mathscr{P} \vdash [a,b]}} L(\boldsymbol{\gamma}, \mathscr{P}),$$

provided we can make rigorous sense of the limiting process (and provided the limit actually exists).

The reason this is not any kind of limit we have seen before is that the set of possible partitions of $[a, b]$ does not have any simple structure, like that of a vector space, in which a suitable norm or distance concept can be defined. No partition has mesh exactly zero; and those that have a mesh within ε of 0 have no necessary relation to one another, in terms of which points t_i they contain. The limiting process is an example of the general notion

$$a = \lim_{\substack{f(\omega) \to 0 \\ \omega \in \Omega}} g(\omega),$$

where Ω is a nonempty set (not necessarily having any specific mathematical structure), and f and g are real-valued functions with domain Ω. We define the above statement to mean that for any $\varepsilon > 0$, there is a corresponding $\delta > 0$ such that for *all* elements $\omega \in \Omega$ with $|f(\omega)| < \delta$, we have $|g(\omega) - a| < \varepsilon$. This is certainly a natural extension of the ordinary limit concept. With this understood, we have at least a reasonable definition of arclength. However, we don't yet know any conditions under which the arclength is guaranteed to exist.

We'll show soon that if $\boldsymbol{\gamma}$ is C^1 on an open interval $(a-r, b+r)$ for some $r > 0$, then the arclength $L_a^b(\boldsymbol{\gamma})$ exists, and is in fact given by the integral

$$L_a^b(\boldsymbol{\gamma}) = \int_a^b \|\mathbf{v}(t)\| \; \mathrm{d}t.$$

This will also help us complete our discussion of the instantaneous speed, as follows. The distance traveled by the particle along its trace from time t to time $t+h$ (where $h > 0$) is

$$d_h(t) := L_t^{t+h}(\boldsymbol{\gamma}) = \int_t^{t+h} \|\mathbf{v}(\tau)\| \; \mathrm{d}\tau.$$

Dividing by the elapsed time h, the speed at time t should be the limit of $d_h(t)/h$ as $h \to 0^+$. Observe that

$$\begin{aligned}
\left|\frac{d_h(t)}{h} - \|\mathbf{v}(t)\|\right| &= \frac{1}{h}\Big|d_h(t) - h\|\mathbf{v}(t)\|\Big| \\
&= \frac{1}{h}\left|\int_t^{t+h} \|\mathbf{v}(\tau)\| \; \mathrm{d}\tau - \int_t^{t+h} \|\mathbf{v}(t)\| \; \mathrm{d}\tau\right| \\
&\leq \frac{1}{h}\int_t^{t+h} \Big|\|\mathbf{v}(\tau)\| - \|\mathbf{v}(t)\|\Big| \; \mathrm{d}\tau.
\end{aligned}$$

Here we have used the fact that $\int_t^{t+h} \|\mathbf{v}(t)\| \; \mathrm{d}\tau = \|\mathbf{v}(t)\| \int_t^{t+h} 1 \; \mathrm{d}\tau = h\|\mathbf{v}(t)\|$, as well as the fact that

$$\left|\int_a^b f(x) \; \mathrm{d}x\right| \leq \int_a^b |f(x)| \; \mathrm{d}x,$$

which holds whenever $|f|$ is Riemann integrable on $[a, b]$. By the Reverse Triangle Inequality, the last integral above is majorized by (i.e., is less than or equal to)

$$\frac{1}{h}\int_t^{t+h} \|\mathbf{v}(\tau) - \mathbf{v}(t)\| \; \mathrm{d}\tau.$$

Using the simple bound $\left|\int_a^b f(x) \; \mathrm{d}x\right| \leq (b-a) \cdot \sup_{[a,b]} |f|$, the above is in turn majorized by

$$\sup_{\tau \in [t, t+h]} \|\mathbf{v}(\tau) - \mathbf{v}(t)\|.$$

[Here, $1/h$ is cancelled by $(t+h) - t = h$.] Since $\boldsymbol{\gamma}$ is assumed to be C^1, $\mathbf{v} = \boldsymbol{\gamma}'$ is *continuous* at t. It follows that

$$\sup_{\tau \in [t, t+h]} \|\mathbf{v}(\tau) - \mathbf{v}(t)\| \to 0 \qquad \text{as } h \to 0^+.$$

We see therefore that $d_h(t)/h$ approaches $\|\mathbf{v}(t)\|$ as $h \to 0^+$. Thus, $\|\mathbf{v}(t)\|$ is interpretable as the instantaneous speed at time t, provided $\boldsymbol{\gamma}$ is C^1 in some neighborhood of t.

Let's now try to prove that the arclength exists and is equal to the relevant integral when $\boldsymbol{\gamma}$ is C^1 on $(a-r, b+r)$ for some $r > 0$. Recall that the broken line length for the partition $\mathscr{P}$ is

$$L(\boldsymbol{\gamma}, \mathscr{P}) = \sum_{i=1}^n \|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\|,$$

where

$$\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| = \sqrt{\sum_{j=1}^e \big[\gamma_j(t_i) - \gamma_j(t_{i-1})\big]^2}.$$

Since each γ_j is differentiable on (t_{i-1}, t_i) and continuous on $[t_{i-1}, t_i]$, the Mean Value Theorem allows us to write

$$\gamma_j(t_{i-1}) - \gamma_j(t_i) = \gamma_j'(t_{ij}^*) \, \triangle t_i,$$

where $t_{ij}^* \in (t_{i-1}, t_i)$. Note that $\triangle t_i$ is positive and independent of j, so we can take it out of the summation to get

$$\alpha_i := \|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| = \sqrt{\sum_{j=1}^{e} [\gamma_j'(t_{ij}^*)]^2}\, \triangle t_i.$$

By the Mean Value Theorem for Integrals, we also have points $T_i^* \in (t_{i-1}, t_i)$ such that

$$\beta_i := \int_{t_{i-1}}^{t_i} \|\mathbf{v}(t)\|\, \mathrm{d}t = (t_i - t_{i-1})\|\mathbf{v}(T_i^*)\| = \sqrt{\sum_{j=1}^{e} [\gamma_j'(T_i^*)]^2}\, \triangle t_i.$$

Note that the expressions for α_i and β_i differ only in the sample points from the i-th subinterval. We would like to upper-bound the difference $|\alpha_i - \beta_i|$. Note first that $|\alpha_i - \beta_i| = |\sqrt{A} - \sqrt{B}|\, \triangle t_i$, where A and B are certain sums of squares. Apply the elementary inequality $|\sqrt{A} - \sqrt{B}| \leq \sqrt{|A - B|}$ for $A, B \geq 0$ (which is proven by taking $A \geq B$ to eliminate absolute values, and squaring both sides). This yields, after combining the two summations:

$$\begin{aligned} |\alpha_i - \beta_i| &\leq \left| \sum_{j=1}^{e} \left\{ [\gamma_j'(t_{ij}^*)]^2 - [\gamma_j'(T_i^*)]^2 \right\} \right|^{\frac{1}{2}} \triangle t_i \\ &\leq \left(\sum_{j=1}^{e} |\gamma_j'(t_{ij}^*) + \gamma_j'(T_i^*)| |\gamma_j'(t_{ij}^*) - \gamma_j'(T_i^*)| \right)^{\frac{1}{2}} \triangle t_i \end{aligned}$$

The second line was gotten from the first by using the Triangle Inequality to push the absolute value sign inside the sum, and then factoring the difference of squares. Now $\boldsymbol{\gamma}'$ is continuous on $(a - r, b + r)$, and hence on $[a, b]$. Thus the Extreme Value Theorem gives a constant $M > 0$ such that $\|\boldsymbol{\gamma}'(t)\| \leq M$ for all $t \in [a, b]$. It follows by the Triangle Inequality that $|\gamma_j'(t_{ij}^*) + \gamma_j'(T_i^*)| \leq 2M$ for each j. At the same time, the Uniform Continuity Theorem tells us that $\boldsymbol{\gamma}'$ is uniformly continuous on $[a, b]$, because $[a, b]$ is compact. This means that for any $\varepsilon > 0$, there is a $\delta > 0$ such that as long as $|t - \tau| < \delta$ and $t, \tau \in [a, b]$, we have $\|\boldsymbol{\gamma}'(t) - \boldsymbol{\gamma}'(\tau)\| < \varepsilon$. Certainly therefore $|\gamma_j'(t_{ij}^*) - \gamma_j'(T_i^*)| < \varepsilon$ for each j, provided only that the partition $\mathscr{P}$ has mesh $\|\mathscr{P}\| < \delta$; for $|t_{ij}^* - T_i^*| \leq \triangle t_i \leq \|\mathscr{P}\| < \delta$. Putting everything together, we see that

$$|\alpha_i - \beta_i| \leq \sqrt{e \cdot 2M\varepsilon}\, \triangle t_i \quad \text{ provided } \|\mathscr{P}\| < \delta.$$

[The factor e comes from the fact that the summation has e terms.] Now we see that

$$L(\boldsymbol{\gamma}, \mathscr{P}) = \sum_{i=1}^{n} \alpha_i \quad \text{ and } \quad \int_a^b \|\mathbf{v}(t)\|\, \mathrm{d}t = \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \|\mathbf{v}(t)\|\, \mathrm{d}t = \sum_{i=1}^{n} \beta_i$$

can differ by no more than

$$\left| \sum_{i=1}^{n} (\alpha_i - \beta_i) \right| \leq \sum_{i=1}^{n} |\alpha_i - \beta_i| \leq \sqrt{e \cdot 2M\varepsilon} \sum_{i=1}^{n} \triangle t_i = (b - a)\sqrt{2Me} \cdot \sqrt{\varepsilon},$$

provided $\|\mathscr{P}\| < \delta$. Note that aside from needing to have mesh $< \delta$, the partition $\mathscr{P}$ is completely unrestricted. Since $\sqrt{\varepsilon}$ can be taken arbitrarily small, we see that the limit

$$\lim_{\substack{\|\mathscr{P}\| \to 0 \\ \mathscr{P} \vdash [a,b]}} L(\boldsymbol{\gamma}, \mathscr{P})$$

exists, and is in fact equal to $\int_a^b \|\mathbf{v}(t)\|\, \mathrm{d}t$. The existence of this integral itself is a simple consequence of the fact that every continuous function is Riemann integrable on any closed interval in its domain. So we have

$$L_a^b(\boldsymbol{\gamma}) = \int_a^b \|\mathbf{v}(t)\|\, \mathrm{d}t.$$

It should be pointed out that despite the name "arclength", $L_a^b(\boldsymbol{\gamma})$ does not always measure the true geometric length of the curve's trace. What it measures is the total distance traveled by a particle moving according to $\boldsymbol{\gamma}$. If the particle happens to come smoothly to a standstill and then smoothly accelerate in the reverse direction, then it can retrace some of its previously covered points. In this case, the total distance traveled will be strictly greater than the geometric length. To this extent, $L_a^b(\boldsymbol{\gamma})$ is a *physical* quantity, rather than a purely geometric one. However, if retracing is avoided, then we will see shortly that the number $\ell := L_a^b(\boldsymbol{\gamma})$ is a quantity inherently associated with the geometric shape of the curve trace, and is independent of the particular manner in which a particle may traverse the arc in question.

Parameter Invariance of Arclength

In addition to the "no retracing" requirement, which is obviously necessary if $L_a^b(\boldsymbol{\gamma})$ is to be meaningful on a purely geometric level, we must also restrict our attention to parametrizations (i.e., curves $\boldsymbol{\gamma}$ that trace out the given arc) which *never stop*. This means that the speed $\|\boldsymbol{\gamma}'\|$ is never zero, or equivalently that $\boldsymbol{\gamma}'(t) \neq \mathbf{0}$ for all $t \in [a,b]$.

Consider the arc

$$\alpha := \boldsymbol{\gamma}([a,b]),$$

where $\boldsymbol{\gamma} : I \longrightarrow \mathbb{R}^e$ is a C^1 curve (i.e., it is C^1 on I), and $[a,b] \subseteq I^\circ$. Assume that $\boldsymbol{\gamma}$ is *one-to-one* on $[a,b]$. That is, if $t, t^* \in [a,b]$ satisfy $t \neq t^*$, then $\boldsymbol{\gamma}(t) \neq \boldsymbol{\gamma}(t^*)$. The purpose of this assumption is to formalize the "no retracing" requirement. Assume also that $\boldsymbol{\gamma}$ never stops: $\boldsymbol{\gamma}'(t) \neq \mathbf{0}$ for all $t \in [a,b]$. Now, as we know,

$$L_a^b(\boldsymbol{\gamma}) = \int_a^b \|\boldsymbol{\gamma}'(t)\| \, \mathrm{d}t.$$

Let's suppose that the same arc can be written as

$$\alpha = \boldsymbol{\delta}([c,d]),$$

where $\boldsymbol{\delta} : J \longrightarrow \mathbb{R}^e$ is a C^1 curve with $[c,d] \subseteq J^\circ$, such that $\boldsymbol{\delta}$ is one-to-one on $[c,d]$ and $\boldsymbol{\delta}'(\tau) \neq 0$ for all $\tau \in [c,d]$. So

$$L_c^d(\boldsymbol{\delta}) = \int_c^d \|\boldsymbol{\delta}'(\tau)\| \, \mathrm{d}\tau.$$

Since $\boldsymbol{\gamma}(t)$ and $\boldsymbol{\delta}(\tau)$ cover exactly the same points as t varies through $[a,b]$ and τ varies through $[c,d]$, we see that for each $\tau \in [c,d]$ there is a value $t \in [a,b]$ such that $\boldsymbol{\gamma}(t) = \boldsymbol{\delta}(\tau)$. This value of t is uniquely determined by τ, by the one-to-oneness of both parametrizations; so there a function $\phi : [c,d] \longrightarrow [a,b]$ defined by

$$t = \phi(\tau) \quad \text{iff} \quad \boldsymbol{\gamma}(t) = \boldsymbol{\delta}(\tau).$$

Another way to put this is that

$$\boldsymbol{\delta}(\tau) = \boldsymbol{\gamma}(\phi(\tau)), \quad \text{or} \quad \boldsymbol{\delta} = \boldsymbol{\gamma} \circ \phi.$$

We claim that ϕ is continuous. This actually follows from a more general theorem which is quite important in its own right. Let's call it the *Continuous Inverse Theorem*. It says that if $\mathbf{f} : D \subseteq \mathbb{R}^d \longrightarrow \mathbb{R}^e$ is a continuous and one-to-one function (so that $\mathbf{f}^{-1}$ exists), and if D is compact, then $\mathbf{f}^{-1}$ is also continuous on its domain $E := \mathbf{f}(D)$.

We prove this as follows. First, a notation: Let $X_D := X \cap D$ for $X \subseteq \mathbb{R}^d$, and $Y_E := Y \cap E$ for $Y \subseteq \mathbb{R}^e$. Take any point $\mathbf{q} \in E$, and let $\mathbf{p} := \mathbf{f}^{-1}(\mathbf{q})$; so $\mathbf{p} \in D$ and $\mathbf{f}(\mathbf{p}) = \mathbf{q}$. Fix $\varepsilon > 0$, and let $\delta = \delta(\varepsilon) > 0$ be such that $\mathbf{f}(B_\delta(\mathbf{p})_D) \subseteq B_\varepsilon(\mathbf{q})_E$, by the continuity of $\mathbf{f}$. Now $B_\delta(\mathbf{p})^c$ is a closed set, as is D. The intersection of two closed sets is closed, so $K := B_\delta(\mathbf{p})^c{}_D$ is closed. Since $K \subseteq D$, and D is bounded, K is bounded. So K is compact. Hence, $\mathbf{f}(K) = \mathbf{f}(B_\delta(\mathbf{p})^c{}_D)$ is a compact subset of E by preservation of compactness, and is closed. It follows that $\mathbf{f}(K)^c = \mathbf{f}(B_\delta(\mathbf{p})_D) \cup E^c$ is open. The two pieces of $\mathbf{f}(K)^c$ are disjoint, as every $\mathbf{f}$-value lies in E. Note however that $\mathbf{q} = \mathbf{f}(\mathbf{p}) \in \mathbf{f}(B_\delta(\mathbf{p})_D) \subseteq \mathbf{f}(K)^c$. Since $\mathbf{f}(K)^c$ is open, there is a ball $B_\eta(\mathbf{q}) \subseteq \mathbf{f}(K)^c$. Note that $B_\eta(\mathbf{q})_E \cap E^c = \emptyset$. So $B_\eta(\mathbf{q})_E \subseteq \mathbf{f}(B_\delta(\mathbf{p})_D)$. By taking $\mathbf{f}^{-1}$ of both sides, it follows that $\mathbf{f}^{-1}(B_\eta(\mathbf{q})_E) \subseteq B_\delta(\mathbf{p})_D$. Since δ can be made arbitrarily small, but determines a corresponding $\eta = \eta(\delta) > 0$ in any case, we see that $\mathbf{f}^{-1}$ is continuous at $\mathbf{q}$.

The Continuous Inverse Theorem applies to our situation, since $\boldsymbol{\gamma}$ is continuous and one-to-one on the compact set $[a,b]$, so that $\boldsymbol{\gamma}^{-1}$ is continuous. Thus, $\phi = \boldsymbol{\gamma}^{-1} \circ \boldsymbol{\delta}$ is a composition of continuous functions, hence continuous.

It's also true that ϕ is C^1 on (c,d), and that $\phi'(\tau)$ is of constant sign on (c,d). First we'll show that ϕ is differentiable. Fix a point $\tau \in (c,d)$, and for small h, let $k := \phi(\tau + h) - \phi(\tau)$. Thus, $\phi(\tau) + k = \phi(\tau + h)$. Also, $\boldsymbol{\delta}(\tau) = \boldsymbol{\gamma}(\phi(\tau))$ and $\boldsymbol{\delta}(\tau + h) = \boldsymbol{\gamma}(\phi(\tau + h)) = \boldsymbol{\gamma}(\phi(\tau) + k)$. Consider the difference quotient

$$\frac{\boldsymbol{\delta}(\tau + h) - \boldsymbol{\delta}(\tau)}{h} = \frac{\boldsymbol{\gamma}(\phi(\tau) + k) - \boldsymbol{\gamma}(\phi(\tau))}{h} = \frac{k}{h}\left(\frac{\boldsymbol{\gamma}(\phi(\tau) + k) - \boldsymbol{\gamma}(\phi(\tau))}{k}\right).$$

Since ϕ is continuous at τ, we have $k \to 0$ as $h \to 0$. Write $\triangle\boldsymbol{\delta} := \boldsymbol{\delta}(\tau + h) - \boldsymbol{\delta}(\tau)$ and $\triangle\boldsymbol{\gamma} := \boldsymbol{\gamma}(\phi(\tau) + k) - \boldsymbol{\gamma}(\phi(\tau))$. So the above says $(\triangle\boldsymbol{\delta})/h = (k/h)[(\triangle\boldsymbol{\gamma})/k]$. Dotting with $(\triangle\boldsymbol{\gamma})/k$, we get $[(\triangle\boldsymbol{\delta})/h] \cdot [(\triangle\boldsymbol{\gamma})/k] = (k/h)\|(\triangle\boldsymbol{\gamma})/k\|^2$. Dividing by the squared norm appearing here gives

$$\frac{k}{h} = \frac{1}{\|(\triangle\boldsymbol{\gamma})/k\|^2}\left(\frac{\triangle\boldsymbol{\delta}}{h} \cdot \frac{\triangle\boldsymbol{\gamma}}{k}\right).$$

Letting $h \to 0$, we have $k \to 0$ too; so the right-hand side approaches $\{1/\|\boldsymbol{\gamma}'(\phi(\tau))\|^2\}\ \boldsymbol{\delta}'(\tau) \cdot \boldsymbol{\gamma}'(\phi(\tau))$. It follows that k/h, i.e., $(\phi(\tau+h)-\phi(\tau))/h$, has a limit as $h \to 0$. This means ϕ is differentiable at τ. Moreover, our result

$$\phi'(\tau) = \frac{1}{\|\boldsymbol{\gamma}'(\phi(\tau))\|^2}\ \boldsymbol{\delta}'(\tau) \cdot \boldsymbol{\gamma}'(\phi(\tau))$$

shows that ϕ' is continuous at τ; for $\boldsymbol{\gamma}'$, $\boldsymbol{\delta}'$, and ϕ are all continuous, and $\boldsymbol{\gamma}' \neq \mathbf{0}$. So actually ϕ is C^1. By taking limits as $h \to 0$ in the earlier equation $(\triangle\boldsymbol{\delta})/h = [(\triangle\boldsymbol{\gamma})/k](k/h)$, we also have

$$\boldsymbol{\delta}'(\tau) = \boldsymbol{\gamma}'(\phi(\tau))\,\phi'(\tau).$$

[This is a special case of the Chain Rule equation, but we couldn't derive it from the Chain Rule since we didn't begin with the knowledge that ϕ is differentiable.] If $\phi'(\tau) = 0$, it would follow that $\boldsymbol{\delta}'(\tau) = \mathbf{0}$, contrary to our "no stopping" assumption. So ϕ' is never zero. Since ϕ' is continuous, it must be of constant sign throughout (c,d). Hence ϕ is either strictly increasing on $[c,d]$, or strictly decreasing on $[c,d]$. In the first case, $\phi(c) = a$ and $\phi(d) = b$, while in the second case $\phi(c) = b$ and $\phi(d) = a$.

Finally, we can prove that $L_a^b(\boldsymbol{\gamma}) = L_c^d(\boldsymbol{\delta})$. Let's first assume that $\phi' > 0$ throughout (c,d). Then we have

$$\|\boldsymbol{\delta}'(\tau)\| = \|\boldsymbol{\gamma}'(\phi(\tau))\|\,|\phi'(\tau)| = \|\boldsymbol{\gamma}'(\phi(\tau))\|\,\phi'(\tau).$$

Now

$$L_c^d(\boldsymbol{\delta}) = \int_c^d \|\boldsymbol{\delta}'(\tau)\|\ \mathrm{d}\tau = \int_c^d \|\boldsymbol{\gamma}'(\phi(\tau))\|\,\phi'(\tau)\ \mathrm{d}\tau.$$

In this integral, we'll make the change of variables $t = \phi(\tau)$. Then $\mathrm{d}t = \phi'(\tau)\,\mathrm{d}\tau$, and $c \mapsto a$, $d \mapsto b$. We get

$$L_c^d(\boldsymbol{\delta}) = \int_a^b \|\boldsymbol{\gamma}'(t)\|\ \mathrm{d}t = L_a^b(\boldsymbol{\gamma}).$$

A similar calculation unfolds when $\phi' < 0$. This time, $|\phi'(\tau)| = -\phi'(\tau)$, while $c \mapsto b$, $d \mapsto a$. Thus,

$$L_c^d(\boldsymbol{\delta}) = \int_b^a \|\boldsymbol{\gamma}'(t)\|\,(-\mathrm{d}t) = \int_a^b \|\boldsymbol{\gamma}'(t)\|\ \mathrm{d}t = L_a^b(\boldsymbol{\gamma}).$$

Because of this invariance under *regular reparametrization*, we may write

$$L(\alpha) := L_a^b(\boldsymbol{\gamma}),$$

and call it the *length* of the arc α. Clearly $L(\alpha)$ is a purely geometric quantity.

Unit Speed Parametrizations

A curve $\boldsymbol{\gamma} : I \longrightarrow \mathbb{R}^e$ is said to be *regular* if (i) $\boldsymbol{\gamma}$ is C^1 on I°, and (ii) $\boldsymbol{\gamma}'(t) \neq \mathbf{0}$ for all $t \in I^\circ$ (i.e., the "no stopping" condition). It turns out that a regular curve can be *reparametrized* as, say, $\boldsymbol{\Gamma} : J \longrightarrow \mathbb{R}^e$, such that $\boldsymbol{\Gamma}$ is C^1 and has *unit speed* at all times:

$$\|\boldsymbol{\Gamma}'(s)\| = 1 \quad \text{for all } s \in J^\circ.$$

It's traditional to use the letter 's' for the parameter of this unit speed parametrization. The quantity s is not only the "time" for the curve $\boldsymbol{\Gamma}$, but also has the property that $|s|$ is the "running arclength" between times 0 and s. For any $s \in J$ with $s \geq 0$, we have

$$L_0^s(\boldsymbol{\Gamma}) = \int_0^s \|\boldsymbol{\Gamma}'(\sigma)\|\ \mathrm{d}\sigma = \int_0^s 1\ \mathrm{d}\sigma = s = |s|,$$

while for any $s \in J$ with $s \leq 0$,

$$L_s^0(\boldsymbol{\gamma}) = \int_s^0 \|\boldsymbol{\Gamma}'(\sigma)\|\ \mathrm{d}\sigma = \int_s^0 1\ \mathrm{d}\sigma = -s = |s|.$$

For this reason, $\boldsymbol{\Gamma}$ is often called a *parametrization with respect to arclength.*

To prove the existence of $\mathbf{\Gamma}$, we argue as follows. Choose a point $a \in I^\circ$, and consider the function

$$S(t) := \int_a^t \|\boldsymbol{\gamma}'(\tau)\| \, \mathrm{d}\tau = \begin{cases} L_a^t(\boldsymbol{\gamma}), & \text{if } t \geq a \\ -L_t^a(\boldsymbol{\gamma}), & \text{if } t < a. \end{cases}$$

As $\|\boldsymbol{\gamma}'\|$ is continuous, we can use the Fundamental Theorem of Calculus, differentiating the definition of $S(t)$ to get

$$S'(t) = \|\boldsymbol{\gamma}'(t)\| > 0.$$

Here we have used condition (ii) from the definition of regularity: $\boldsymbol{\gamma}'(t) \neq \mathbf{0}$. It follows that S is a strictly increasing function of t. Hence S is one-to-one, and has an inverse function $T = S^{-1}$. This function has the following properties:

$$L_a^{T(s)}(\boldsymbol{\gamma}) = s \text{ for any } s \in [0, \ell], \qquad \text{and} \qquad L_{T(s)}^a(\boldsymbol{\gamma}) = -s \text{ for any } s \in [-\ell, 0].$$

Here, $\ell > 0$ can be any number such that $[-\ell, \ell] \subseteq S(I)$. So we've solved $s = S(t)$ for t, getting $t = T(s)$. Now set

$$\mathbf{\Gamma}(s) := \boldsymbol{\gamma}(T(s)) \qquad \text{for } s \in [-\ell, \ell].$$

Equivalently, $\boldsymbol{\gamma}(t) = \mathbf{\Gamma}(S(t))$. To differentiate $\mathbf{\Gamma}$, we need to know that $T(s)$ is C^1, and that $T'(s) = 1/S'(T(s))$. This comes from a familiar theorem of single-variable calculus, but it is typically not proven in calculus courses, we will pause below to supply a proof. Meanwhile, taking the theorem for granted momentarily, the Chain Rule gives

$$\mathbf{\Gamma}'(s) = \boldsymbol{\gamma}'(T(s))\, T'(s) = \frac{\boldsymbol{\gamma}'(T(s))}{S'(T(s))} = \frac{\boldsymbol{\gamma}'(T(s))}{\|\boldsymbol{\gamma}'(T(s))\|} = \mathbf{T}(T(s)).$$

This is clearly a unit vector for any $s \in (-\ell, \ell)$, so we have our unit speed parametrization.

To complete this discussion, we wish to prove the one-dimensional version of the *Inverse Function Theorem*. Its d-dimensional version will be proved in class. The one-dimensional case says that if f is a C^1 function on an interval I, and if $f'(a) \neq 0$ for some $a \in I^\circ$, then f has a *local inverse*, say g, defined on some interval $E = (f(a) - r, f(a) + r)$ where $r > 0$; moreover, g is C^1, and $g'(y) = 1/f'(g(y))$ for all $y \in E$.

Suppose $f'(a) > 0$. Since f' is continuous at a, $f' > 0$ on some interval $J := [a - \delta, a + \delta]$, where $\delta > 0$. If $u, v \in J$ with $u < v$, the Mean Value Theorem gives $f(v) - f(u) = f'(x^*)(v - u)$ for some $x^* \in (u, v) \subseteq J$; but note that $f'(x^*) > 0$ and $v - u > 0$, and hence $f(u) < f(v)$. So f is strictly increasing on J. Similarly, if $f'(a) < 0$ then f is strictly decreasing on a closed interval J centered at a. The restriction of f to J is therefore one-to-one, and has an inverse function g. The domain of g is $f(J)$, but since $f(a)$ is an interior point of $f(J)$, we are free to further restrict to a symmetric interval about $f(a)$, say $E := (f(a) - r, f(a) + r)$, where $r > 0$. Since J is compact, the Continuous Inverse Theorem implies that g is continuous on E. Moreover, $f(g(y)) = y$ for all $y \in E$. Choose $b \in E^\circ$, and let h be so small that $b + h \in E$. Also define $k := g(b + h) - g(b)$, so that $g(b + h) = g(b) + k$. Now we have

$$h = (b + h) - b = f(g(b + h)) - f(g(b)) = f(g(b) + k) - f(g(b)).$$

Rearranging this and multiplying through by k, we get

$$\frac{g(b + h) - g(b)}{h} = \frac{k}{h} = \frac{1}{[f(g(b) + k) - f(g(b))]/k}.$$

By the continuity of g, we have $k \to 0$ as $h \to 0$. Thus, letting $h \to 0$ in the above equation, the right-hand side tends to the limit $1/f'(g(b))$, which is well-defined since $f' \neq 0$ on J. Thus g is differentiable at b, and $g'(b) = 1/f'(g(b))$. Thus, $g'(y) = 1/f'(g(y))$ for *all* $y \in E$. Since f' and g are continuous, and $f' \neq 0$, it follows that g' is continuous.

Curvature

We would now like to quantify the amount of "bending" in the trace of a given curve $\boldsymbol{\gamma}$ near a given point $\mathbf{p}$. This is supposed to be related to the rate at which the unit tangent vector $\mathbf{T}(t)$ *rotates* (relative to its base point) at the instant a when $\boldsymbol{\gamma}(a) = \mathbf{p}$. The problem is that we may not get a true measure of the geometric shape of the curve near $\mathbf{p}$, because $\mathbf{T}(t)$ may change direction only quite slowly even around a very sharp bend, if $\boldsymbol{\gamma}(t)$ happens to be moving at a sufficiently slow crawl. To eliminate such artificial effects of variable speed, we can use the unit speed parametrization $\mathbf{\Gamma}$ whose existence was proved above (assuming $\boldsymbol{\gamma}$ is C^1 on I).

Instead of looking at the rotation rate of $\mathbf{T}(t)$ with respect to t, we use the signed running arclength s as our time parameter (so that the speed is always 1), and look at the rotation rate of $\mathbf{T}(T(s))$ *with respect to* s. Recall that $\mathbf{T}(T(s)) = \mathbf{\Gamma}'(s)$. Now $\mathbf{\Gamma}'(s)$ is always a unit vector, so it never experiences any length change. The instantaneous change in $\mathbf{\Gamma}'(s)$ at any time $s = b$ must therefore be entirely due to *directional change*. We can measure the rate of directional change at $s = b$, otherwise known as the "rotation rate", by taking the *speed* of $\mathbf{\Gamma}'$, i.e., the length of its derivative: $\|\mathbf{\Gamma}''(b)\|$. This will be called the *curvature* at $s = b$. Thus, we define the *curvature function*, κ, by

$$\kappa(s) := \|\mathbf{\Gamma}''(s)\| \quad \text{for all } s \in (-\ell, \ell).$$

Let's elaborate briefly on the logic of our interpretation of κ. Since $\mathbf{\Gamma}'(s)$ has constant length 1, we can view it as the position vector (assuming we anchor it to $\mathbf{0}$) for a curve whose trace is confined to the unit sphere $\mathbb{S} \subseteq \mathbb{R}^e$. A small arc of this curve from $s = b$ to $s = b + h$ can be approximated by an arc of the unique "great circle" of $\mathbb{S}$ through $\mathbf{\Gamma}(b)$ and $\mathbf{\Gamma}(b+h)$. Now the length of an arc in a unit circle is exactly the radian measure of the intercepting angle, by definition. So the distance traveled by the curve $\mathbf{\Gamma}'(s)$ from $s = b$ to $s = b + h$ is approximately equal to the angle measure $\triangle\theta := \measuredangle(\mathbf{\Gamma}(b), \mathbf{\Gamma}(b+h))$. Dividing by the elapsed time $\triangle s := h$, the average speed of $\mathbf{\Gamma}'$ on the interval $[b, b+h]$ is approximately $\triangle\theta/\triangle s$, which looks like an average angular speed. Since we haven't defined an "angle function" $\theta(s)$ to be differentiated, we cannot really interpret $\triangle\theta/\triangle s$ as an average angular speed in any precise sense; but the analogy is suggestive. It gives us confidence that the *limit* of the average speed $L_b^{b+h}(\mathbf{\Gamma}')/h$, i.e., the instantaneous speed $\kappa(b) = \|\mathbf{\Gamma}''(b)\|$, is the natural *definition* of "angular speed" at $s = b$, in the absence of any logically prior definition.

In order to guarantee the existence of the curvature function, we must assume that $\mathbf{\Gamma}'$ is differentiable, or that $\mathbf{\Gamma}$ is twice differentiable. In order to guarantee the *continuity* of κ, we must assume that $\mathbf{\Gamma}'$ is C^1, i.e., that $\mathbf{\Gamma}$ is C^2.

A point $\mathbf{\Gamma}(b)$ with $b \in J^\circ$ at which $\kappa(b) = 0$ is called an *inflection point* if $\kappa(b-\varepsilon)\kappa(b+\varepsilon) < 0$ for all sufficiently small $\varepsilon > 0$, and an *undulation point* if $\kappa(b-\varepsilon)\kappa(b+\varepsilon) > 0$ for all sufficiently small $\varepsilon > 0$. It is possible for κ to be zero on a subinterval $(b-\varepsilon, b+\varepsilon) \subseteq J$; as we will see, the portion of the trace swept out while $s \in (b-\varepsilon, b+\varepsilon)$ is just a line segment. Such points are neither inflection points nor undulation points.

If $\kappa(s) \equiv 0$ for all $s \in J$ (or for all s in some subinterval $J_1 \subseteq J$), then $\mathbf{\Gamma}$ (or the restriction of $\mathbf{\Gamma}$ to J_1) is a *uniform rectilinear motion*. For $\kappa = \|\mathbf{\Gamma}''\| = 0$ implies $\mathbf{\Gamma}''(s) = \mathbf{0}$ for all s, and we can integrate both sides (componentwise) to get $\mathbf{\Gamma}'(s) = \mathbf{a}$, where $\mathbf{a}$ is a constant vector. Integrating once again, $\mathbf{\Gamma}(s) = s\mathbf{a} + \mathbf{b}$ for some constant vector $\mathbf{b}$. This is precisely a uniform rectlinear motion through the point $\mathbf{b}$ with velocity vector $\mathbf{a}$.

Let's see how to compute the curvature κ in terms of the original time parameter t of the given curve $\boldsymbol{\gamma}$. That is, we wish to compute the function

$$K(t) := \kappa(S(t)),$$

which represents the curvature at the point $\boldsymbol{\gamma}(t) = \mathbf{\Gamma}(S(t))$, using only the given function $\boldsymbol{\gamma}$ and its derivatives. Using the Chain Rule, we have

$$\boldsymbol{\gamma}'(t) = \mathbf{\Gamma}'(S(t))S'(t) = \mathbf{\Gamma}'(S(t))\|\boldsymbol{\gamma}'(t)\| = \mathbf{\Gamma}'(S(t))v(t),$$

where we have written $v(t) := \|\mathbf{v}(t)\| = \|\boldsymbol{\gamma}'(t)\|$ for the speed function of $\boldsymbol{\gamma}$. Differentiating once more,

$$\boldsymbol{\gamma}''(t) = \{\mathbf{\Gamma}''(S(t))S'(t)\}v(t) + \mathbf{\Gamma}'(S(t))v'(t) = \mathbf{\Gamma}''(S(t))[v(t)]^2 + \mathbf{\Gamma}'(S(t))v'(t).$$

Recall that $\mathbf{T}(t) = \mathbf{\Gamma}'(S(t))$ is the unit tangent of $\boldsymbol{\gamma}$ at time t. Also, define

$$\mathbf{N}(t) := \frac{\mathbf{\Gamma}''(S(t))}{\|\mathbf{\Gamma}''(S(t))\|}$$

to be the *unit normal* of $\boldsymbol{\gamma}$ at time t, which is well-defined as long as $\mathbf{\Gamma}''(S(t)) \neq \mathbf{0}$, i.e., as long as $K(t) \neq 0$. Since $K(t) = \kappa(S(t)) = \|\mathbf{\Gamma}''(S(t))\|$, we have $\mathbf{\Gamma}''(S(t)) = K(t)\mathbf{N}(t)$. With these notations, we get

$$\boldsymbol{\gamma}''(t) = v'(t)\mathbf{T}(t) + K(t)v(t)^2\mathbf{N}(t).$$

Since every quantity present now depends on t directly, rather than indirectly through $S(t)$, we may as well suppress the variable and write

$$\boldsymbol{\gamma}'' = v'\,\mathbf{T} + Kv^2\mathbf{N}.$$

This has an interesting physical interpretation, which we will discuss shortly. We emphasize that the above is only true at times t for which $K(t) \neq 0$, since $\mathbf{N}(t)$ is undefined when $K(t) = 0$. Meanwhile, we would like to show that

$$\mathbf{T}(t) \perp \mathbf{N}(t) \qquad \text{for all } t \in I^\circ \text{ with } K(t) \neq 0.$$

Here's why. First, $\mathbf{T}(t)$ has constant length: $\|\mathbf{T}\| \equiv 1$. For any differentiable one-variable vector-valued function $\mathbf{F}$ having constant length $\|\mathbf{F}\| \equiv c$, we can write $\mathbf{F} \cdot \mathbf{F} = \|\mathbf{F}\|^2 \equiv c^2$, and then differentiate both sides to get

$$\mathbf{F}' \cdot \mathbf{F} + \mathbf{F} \cdot \mathbf{F}' \equiv 0, \qquad \text{i.e.,} \qquad \mathbf{F} \cdot \mathbf{F}' \equiv 0,$$

so that $\mathbf{F} \perp \mathbf{F}'$ for all values of the variable. Here we have used the *Product Rule for Dot Products*, namely

$$(\mathbf{F} \cdot \mathbf{G})'(t) = \mathbf{F}'(t) \cdot \mathbf{G}(t) + \mathbf{F}(t) \cdot \mathbf{G}'(t).$$

It follows that $\mathbf{T}(t) \perp \mathbf{T}'(t)$. But $\mathbf{T}'(t) = \mathbf{\Gamma}''(S(t))S'(t) = [K(t)\mathbf{N}(t)]v(t)$, which is parallel to $\mathbf{N}(t)$. So $\mathbf{T}(t) \perp \mathbf{N}(t)$.

Let's pause briefly to justify the Product Rule for Dot Products. A rule like this actually holds for any kind of "product" $A*B$, as long as $A, B, \ldots$ are mathematical objects that can be added, subtracted, scaled by real numbers, and for which there is a norm, making limits meaningful (specifically, $X \to A$ if $\|X - A\| \to 0$); for example, matrix-valued functions of one variable under matrix multiplication. This product operation need not be commutative or associative, as long as we have distributive laws $A*(B+C) = (A*B)+(A*C)$ and $(A+B)*C = (A*C)+(B*C)$, scaling laws $(cA)*B = c(A*B) = A*(cB)$, and "zero laws" $A*O = 0B = O$, where O is the zero: $A+O = A$. For any product operation of this kind, we have the identity

$$(A*B) - (A_0 * B_0) = A_0 * (B - B_0) + (A - A_0) * B_0 + (A - A_0) * (B - B_0).$$

This may be verified by using distributivity to "multiply out". Taking $A = A(t+h)$, $A_0 = A(t)$, and $B = B(t+h)$, $B_0 = B(t)$, and dividing the above identity through by h, we can take limits as $h \to 0$. Assuming $A'(t)$ and $B'(t)$ exist as limits — and hence that $B(t+h) - B(t) = h\,[(\triangle B)/h] \to 0\,B'(t) = O$ — we get

$$(A*B)'(t) = A(t) * B'(t) + A'(t) * B(t) + A'(t) * O.$$

Since $A'(t) * O = O$, the last term can be ignored, and we have

$$(A*B)'(t) = A'(t) * B(t) + A(t) * B'(t).$$

The dot product rule is one instance; but there are similar rules for cross products of vector-valued functions in $\mathbb{R}^3$, matrix products of matrix-valued functions, compositions of time-dependent linear operators on a vector space, etc.

Continuing with our calculation of the curvature K, we recall that

$$\boldsymbol{\gamma}'(t) = \mathbf{\Gamma}'(S(t))S'(t) = \mathbf{T}(t)v(t), \qquad \text{or} \qquad \boldsymbol{\gamma}' = v\mathbf{T}.$$

It follows that

$$\boldsymbol{\gamma}'' - \frac{v'}{v}\boldsymbol{\gamma}' = [v'\mathbf{T} + Kv^2\mathbf{N}] - \frac{v'}{v}(v\mathbf{T}) = Kv^2\mathbf{N}.$$

Taking norms of both sides and noting that $K \geq 0$ and $\|\mathbf{N}\| = 1$, we get

$$K = \frac{1}{v^2}\left\|\boldsymbol{\gamma}'' - \frac{v'}{v}\boldsymbol{\gamma}'\right\|.$$

We can eliminate the appearances of v in this formula as follows. First, note that $v^2 = \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$. Differentiating both sides with respect to t, we get $2vv' = \mathbf{v}' \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v}' = 2(\mathbf{v} \cdot \mathbf{v}') = 2(\boldsymbol{\gamma}' \cdot \boldsymbol{\gamma}'')$. Using $v^2 = \|\boldsymbol{\gamma}'\|^2$, we get

$$\frac{v'}{v} = \frac{vv'}{v^2} = \frac{\boldsymbol{\gamma}' \cdot \boldsymbol{\gamma}''}{\|\boldsymbol{\gamma}'\|^2}.$$

Finally, we have our formula for the curvature:

$$K = \frac{1}{\|\boldsymbol{\gamma}'\|^2}\left\|\boldsymbol{\gamma}'' - \frac{(\boldsymbol{\gamma}' \cdot \boldsymbol{\gamma}'')\boldsymbol{\gamma}'}{\|\boldsymbol{\gamma}'\|^2}\right\| = \frac{\|(\boldsymbol{\gamma}' \cdot \boldsymbol{\gamma}')\boldsymbol{\gamma}'' - (\boldsymbol{\gamma}' \cdot \boldsymbol{\gamma}'')\boldsymbol{\gamma}'\|}{\|\boldsymbol{\gamma}'\|^4}.$$

To get a feel for the curvature in a familiar setting, let's consider curves whose traces are graphs of ordinary functions: $G_f := \{(x, f(x)) \mid x \in [a,b]\}$. Assume f is C^2 on $I = (a-r, b+r)$ for some $r > 0$. We can parametrize this very easily: $\boldsymbol{\gamma}(t) = (t, f(t))$ for $t \in I$. Then $\boldsymbol{\gamma}' = (1, f')$, and $\boldsymbol{\gamma}'' = (0, f'')$. Also, $\boldsymbol{\gamma}' \cdot \boldsymbol{\gamma}' = \|\boldsymbol{\gamma}'\|^2 = 1 + f'^2$, while $\boldsymbol{\gamma}' \cdot \boldsymbol{\gamma}'' = f'f''$. The curvature is

$$K = \frac{\|(1+f'^2)(0, f'') - f'f''(1, f')\|}{(1+f'^2)^2} = \frac{\|(-f'f'', (1+f'^2)f'' - f'^2 f'')\|}{(1+f'^2)^2} = \frac{|f''|\,\|(-f', 1)\|}{(1+f'^2)^2} = \frac{|f''|}{\sqrt{(1+f'^2)^3}}.$$

Since the final denominator is always at least 1, we see that $0 \le K \le |f''|$. Also, $K = |f''|$ if and only if $f' = 0$.

Finally for now, let's prove that a curve of constant nonzero curvature in $\mathbb{R}^2$ is (part of) a circle. Say we have $\kappa(s) = \|\boldsymbol{\Gamma}''(s)\| \equiv |c| > 0$. Since $\boldsymbol{\Gamma}$ is a unit speed curve, we can write the unit vector $\boldsymbol{\Gamma}'(s)$ as $(\cos\,\theta(s), \sin\,\theta(s))$. As we'll prove later, $\theta(s)$ can be taken as a C^1 function. By the Chain Rule, $\boldsymbol{\Gamma}'' = \theta'(\cos\theta, \sin\theta)$. Thus $\kappa = \|\boldsymbol{\Gamma}''\| = |\theta'|$. It follows that $\theta' = c$, the sign of c being arbitrary (the continuity of θ' prevents it from jumping between c and $-c$). Integrating $\theta'(s) = c$ with respect to s, we get $\theta(s) = cs + k$. Thus, $\boldsymbol{\Gamma}'(s) = (\cos(cs+k), \sin(cs+k))$. Making the change of variable $u := cs + k$, so that $\mathrm{d}u = c\,\mathrm{d}s$, we can integrate componentwise with respect to s to get $\boldsymbol{\Gamma}(s) = (c^{-1}\sin(cs+k) + p, -c^{-1}\sin(cs+k) + q)$. Here, p and q are constants of integration. Letting $\mathbf{p} = (p, q)$,

$$\|\boldsymbol{\Gamma}(s) - \mathbf{p}\| = \sqrt{|c|^{-2}(\cos^2\theta(s) + \sin^2\theta(s))} = |c|^{-1}.$$

This means $\boldsymbol{\Gamma}(s)$ moves along a circle of radius $1/|c|$ centered at $\mathbf{p}$. Conversely, it is easy to show that the curvature of a circle of radius r is exactly $1/r$ at each point: use $\boldsymbol{\Gamma}(s) = (r\cos(s/r), r\sin(s/r))$ to get $\kappa(s) = \|\boldsymbol{\Gamma}''(s)\| \equiv 1/r$.

Tangential and Normal Acceleration

There is an interesting physical interpretation of the formula

$$\boldsymbol{\gamma}'' = v'\,\mathbf{T} + Kv^2\mathbf{N},$$

proved above for a C^2 curve of nonvanishing curvature. Recall that $\mathbf{v} = \boldsymbol{\gamma}'$ is the velocity of $\boldsymbol{\gamma}$; it points in the direction of motion and has length equal to the speed. For similar reasons, $\mathbf{a}(t) := \boldsymbol{\gamma}''(t)$ is called the *acceleration vector* of $\boldsymbol{\gamma}$. As we will see, $\mathbf{a}$ points in the direction toward which the particle is turning, and its length $a = \|\mathbf{a}\|$ is an overall scalar measure of the rate of change in the velocity vector. But there are two sources of change to the velocity vector, which are physically quite distinct: (i) speed change, and (ii) directional change. By analyzing the expansion $\mathbf{a} = v'\mathbf{T} + Kv^2\mathbf{N}$, we will see how to separate the two sources of velocity change, and quantify each of them in isolation from the other. Source (i) — speed change — leads to *tangential acceleration* a_{T}, while source (ii) — directional change — leads to *normal acceleration* a_{N}. The three kinds of scalar acceleration, a, a_{T}, and a_{N}, satisfy a simple relation which we will also discover.

Consider the velocity vectors at three times: t, $t+h$, and $t+2h$, where $h > 0$ is small. The directions of motion at times t and $t+h$ are approximately codirectional with $\triangle\boldsymbol{\gamma} := \boldsymbol{\gamma}(t+h) - \boldsymbol{\gamma}(t)$ and $\triangle^+\boldsymbol{\gamma} := \boldsymbol{\gamma}(t+2h) - \boldsymbol{\gamma}(t+h)$. The direction in which the direction of motion is *changing* — what might be termed the *turning direction* — is therefore closely approximated by the unit vector codirectional with the difference

$$\triangle^+\boldsymbol{\gamma} - \triangle\boldsymbol{\gamma} = \boldsymbol{\gamma}(t+2h) + \boldsymbol{\gamma}(t) - 2\boldsymbol{\gamma}(t+h).$$

Now consider the first component function γ_1 of $\boldsymbol{\gamma}$. By Taylor's Theorem from single-variable calculus, since γ_1 is C^2 near t, we can write $\gamma_1(t+h) = \gamma_1(t) + h\gamma_1'(t) + \frac{1}{2}h^2\gamma_2''(t) + h^2\varepsilon_1(t,h)$, where $\varepsilon_1(t,h) \to 0$ as $h \to 0$. Applying this to all component functions, and putting the resulting equations together into a single vector equation, we get

$$\boldsymbol{\gamma}(t+h) = \boldsymbol{\gamma}(t) + h\boldsymbol{\gamma}'(t) + \tfrac{1}{2}h^2\boldsymbol{\gamma}''(t) + h^2\boldsymbol{\varepsilon}(t,h),$$

where $\boldsymbol{\varepsilon}(t,h) \to \mathbf{0}$ as $h \to 0$. Similarly,

$$\boldsymbol{\gamma}(t+2h) = \boldsymbol{\gamma}(t) + 2h\boldsymbol{\gamma}'(t) + \tfrac{1}{2}(2h)^2\boldsymbol{\gamma}''(t) + (2h)^2\boldsymbol{\varepsilon}(t,2h).$$

Putting these results into the formula for the approximate turning direction, we find that

$$\triangle^+\boldsymbol{\gamma} - \triangle\boldsymbol{\gamma} = h^2\boldsymbol{\gamma}''(t) + 2h^2[\boldsymbol{\varepsilon}(t,h) + 2\boldsymbol{\varepsilon}(t,2h)].$$

Dividing through by h^2, we can rearrange and take norms to get

$$\left\| \frac{\triangle^+\boldsymbol{\gamma} - \triangle\boldsymbol{\gamma}}{h^2} - \boldsymbol{\gamma}''(t) \right\| = 2\|\boldsymbol{\varepsilon}(t,h) + 2\boldsymbol{\varepsilon}(t,2h)\| \to 0$$

as $h \to 0$. Thus, $(\triangle^+\boldsymbol{\gamma} - \triangle\boldsymbol{\gamma})/h^2 \to \boldsymbol{\gamma}''(t)$ as $h \to 0$. It follows that

$$\frac{\triangle^+\boldsymbol{\gamma} - \triangle\boldsymbol{\gamma}}{\|\triangle^+\boldsymbol{\gamma} - \triangle\boldsymbol{\gamma}\|} = \frac{(\triangle^+\boldsymbol{\gamma} - \triangle\boldsymbol{\gamma})/h^2}{\|(\triangle^+\boldsymbol{\gamma} - \triangle\boldsymbol{\gamma})/h^2\|} \to \frac{\boldsymbol{\gamma}''(t)}{\|\boldsymbol{\gamma}''(t)\|} = \frac{\mathbf{a}(t)}{\|\mathbf{a}(t)\|},$$

This makes precise our earlier claim that $\mathbf{a}$ points in the *instantaneous turning direction*. That $a = \|\mathbf{a}\|$ represents the instantaneous scalar acceleration is clear, because $a = \|\mathbf{a}\| = \|\mathbf{v}'\|$, and this is known from an earlier analysis to be the *speed* at which $\mathbf{v}$ is changing. The speed of velocity change is precisely what we mean by "scalar acceleration".

Next, recall that for a C^2 curve of nonvanishing curvature, $\|\mathbf{T}\| = 1$, $\|\mathbf{N}\| = 1$, and $\mathbf{T} \perp \mathbf{N}$. Also, the acceleration vector $\mathbf{a}$ is a linear combination of $\mathbf{T}$ and $\mathbf{N}$, and so lies in the plane $\Pi(\mathbf{T},\mathbf{N}) := \{\alpha\mathbf{T} + \beta\mathbf{N} \mid \alpha, \beta \in \mathbb{R}\}$. We will now show that for any vector $\mathbf{x} \in \Pi(\mathbf{T},\mathbf{N})$, the expansion of $\mathbf{x}$ as a linear combination of $\mathbf{T}$ and $\mathbf{N}$ is given by

$$\mathbf{x} = (\mathbf{x}\cdot\mathbf{T})\mathbf{T} + (\mathbf{x}\cdot\mathbf{N})\mathbf{N}.$$

Writing $\mathbf{x} = \alpha\mathbf{T} + \beta\mathbf{N}$, we can dot both sides with $\mathbf{T}$ to get $\mathbf{x}\cdot\mathbf{T} = \alpha(\mathbf{T}\cdot\mathbf{T}) + \beta(\mathbf{N}\cdot\mathbf{T}) = \alpha$. Similarly, $\beta = \mathbf{x}\cdot\mathbf{N}$.

Moreover, we may interpret $\mathbf{x}\cdot\mathbf{T}$ as the *component of* $\mathbf{x}$ *along* $\mathbf{T}$. This means the signed length of the orthogonal projection of $\mathbf{x}$ onto the line $\{c\mathbf{T} \mid c \in \mathbb{R}\}$, signed positively if the projection is codirectional with $\mathbf{T}$, and negatively otherwise. The projection vector $\mathbf{p}$ is characterized by $\mathbf{p} = c\mathbf{T}$ and $(\mathbf{x} - \mathbf{p}) \perp \mathbf{T}$. This means $(\mathbf{x} - c\mathbf{T})\cdot\mathbf{T} = 0$, or $\mathbf{x}\cdot\mathbf{T} = c(\mathbf{T}\cdot\mathbf{T}) = c$. So $\mathbf{p} = (\mathbf{x}\cdot\mathbf{T})\mathbf{T}$. The length of the projection is $\|\mathbf{p}\| = |\mathbf{x}\cdot\mathbf{T}|\|\mathbf{T}\| = |\mathbf{x}\cdot\mathbf{T}|$. Finally note that $\operatorname{sgn}(\mathbf{x}\cdot\mathbf{T}) = \operatorname{sgn}(c)$, which is positive if and only if $\mathbf{T}$ and $\mathbf{p} = c\mathbf{T}$ are codirectional.

These observations tell us that if a vector $\mathbf{x}$ has the expansion $\alpha\mathbf{T} + \beta\mathbf{N}$, then $\alpha = \mathbf{x}\cdot\mathbf{T}$ represents the component of $\mathbf{x}$ in the $\mathbf{T}$-direction, and $\beta = \mathbf{x}\cdot\mathbf{N}$ the component of $\mathbf{x}$ in the $\mathbf{N}$-direction. Applying this to the acceleration vector $\mathbf{a} = v'\mathbf{T} + Kv^2\mathbf{N}$, we can interpret v' as the *tangential component of acceleration*, and Kv^2 as the *normal component of acceleration*. That is,

$$a_{\mathrm{T}} := \mathbf{a}\cdot\mathbf{T} = v' \quad \text{and} \quad a_{\mathrm{N}} := \mathbf{a}\cdot\mathbf{N} = Kv^2.$$

The three kinds of scalar acceleration satisfy the relation

$$a = \sqrt{a_{\mathrm{T}}^2 + a_{\mathrm{N}}^2}.$$

We prove it as follows:

$$a^2 = (a_{\mathrm{T}}\mathbf{T} + a_{\mathrm{N}}\mathbf{N})\cdot(a_{\mathrm{T}}\mathbf{T} + a_{\mathrm{N}}\mathbf{N}) = a_{\mathrm{T}}^2\|\mathbf{T}\|^2 + 2a_{\mathrm{T}}a_{\mathrm{N}}(\mathbf{T}\cdot\mathbf{N}) + a_{\mathrm{N}}^2\|\mathbf{N}\|^2 = a_{\mathrm{T}}^2 + a_{\mathrm{N}}^2.$$

This relation shows exactly how the total scalar acceleration is composed of contributions from the two sources of velocity change. The tangential acceleration $a_{\mathrm{T}} = v'$ is exactly the rate of change in v, accounting perfectly for speed change. Directional change makes the contribution $a_{\mathrm{N}} = Kv^2$. It is interesting that a_{N} is proportional to the curvature K at constant speed (so the sharper the turn, the more one accelerates sideways, perpendicular to the direction of motion), but turns out to be proportional to the *square* of the speed at constant curvature. To visualize this, imagine driving on a circular racetrack, so that the curvature is constant, say $K(t) \equiv k_0$. If you also drive at constant speed v_0, and your mass is m, then you will feel a constant sideways force of $F = ma_{\mathrm{N}} = mk_0v_0^2$.

Polar Representation of a Curve

In determining the nature of constant curvature motion in $\mathbb{R}^2$ (which turned out to be uniform circular or uniform rectilinear motion), we appealed to an important fact that we would now like to prove. In a slightly generalized form, it says the following. Let $\boldsymbol{\gamma} : I \longrightarrow \mathbb{R}^2$ be a C^1 curve which avoids the origin: $\boldsymbol{\gamma}(t) \neq \mathbf{0}$ for all $t \in I$. Then $\boldsymbol{\gamma}$ can be expressed in terms of "polar coordinate functions":

$$\boldsymbol{\gamma}(t) = r(t)(\cos\,\theta(t),\, \sin\,\theta(t)),$$

where r and θ are C^1 on I°, and $r(t) > 0$ for all $t \in I$. What is surprising here is not that the functions r and θ exist, since each point $\boldsymbol{\gamma}(t)$ can be expressed individually in polar coordinates. However, polar coordinates are not unique: the angular coordinate may always be changed by an integer multiple of 2π. So if we express each point $\boldsymbol{\gamma}(t)$ individually in polar form, say as $r_t(\cos\theta_t, \sin\theta_t)$, then there is no guarantee that the different values θ_t will "sync up" properly as t varies so as to create a smooth (or even continuous) angle function $\theta(t)$. The theorem asserts, however, that such choices of angular coordinates can always be made so that there is a globally smooth (i.e., C^1) angle function. This *polar representation* has important applications, as we have already seen.

There is no difficulty in defining the function r, or in proving that it is C^1. We simply take

$$r(t) := \|\boldsymbol{\gamma}(t)\|.$$

Then $r(t) > 0$, from our assumption that the origin is avoided. That r is C^1 follows from (i) the assumed C^1-ness of $\boldsymbol{\gamma}$, and (ii) the C^1-ness of the norm function away from $\mathbf{0}$.

To elaborate on point (ii), let $g(\mathbf{x}) := \|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_d^2}$ on $\mathbb{R}^d$. Note that

$$\partial_{x_j} g(\mathbf{x}) = \frac{1}{2\sqrt{x_1^2 + \cdots + x_d^2}} \cdot 2x_j = \frac{x_j}{\|\mathbf{x}\|},$$

so by putting all the components together into a vector,

$$\boldsymbol{\nabla} g(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}.$$

Since $\|\mathbf{x}\|$ is continuous and nonzero away from $\mathbf{x} = \mathbf{0}$, this expression shows that g is C^1 on $\mathbb{R}^d \setminus \{\mathbf{0}\}$.

The hard part is defining θ. As a first step, let $\boldsymbol{\delta}(t) := \boldsymbol{\gamma}(t)/\|\boldsymbol{\gamma}(t)\|$, which is well-defined on I because $\boldsymbol{\gamma}$ avoids the origin. Note that $\|\boldsymbol{\delta}(t)\| \equiv 1$, and that $\boldsymbol{\gamma}(t) = r(t)\boldsymbol{\delta}(t)$. Henceforth we can focus exclusively on $\boldsymbol{\delta}$, which we note is C^1, using the C^1 nature of the norm function away from $\mathbf{0}$. Write $\boldsymbol{\delta}(t) = (x(t), y(t))$, so x and y are C^1 functions of t on I°. For any $c \in I^\circ$, either $x(c) \neq 0$, or else $x(c) = 0$ and $y(c) = \pm 1$; this is from $\|\boldsymbol{\delta}(c)\|^2 = x(c)^2 + y(c)^2 = 1$.

If $x(c) \neq 0$, then on some neighborhood of c to be determined, say $I_c = (c - \delta, c + \delta)$, define

$$\theta_c(t) := \begin{cases} \arcsin y(t), & \text{if } x(c) > 0 \\ \pi - \arcsin y(t), & \text{if } x(c) < 0 \end{cases}$$

We choose $\delta > 0$ so that (i) $-1 < y(t) < 1$, and (ii) $\operatorname{sgn} x(t) = \operatorname{sgn} x(c)$, for all $t \in I_c$. Condition (ii) can be satisfied by the continuity of $x(t)$, and condition (ii) can be satisfied by the continuity of $y(t)$ together with the fact that $|y(c)| = \sqrt{1 - x(c)^2} < 1$, as $x(c) \neq 0$. The inequalities $-1 < y(t) < 1$ place $y(t)$ in the *interior* of the domain of the arcsine function, namely $[-1, 1]^\circ$, where the arcsine is C^1. It follows that θ_c is C^1 on its domain I_c. Now clearly $\sin\theta_c(t) = \sin(\arcsin y(t)) = y(t)$. Also note that

$$\cos\theta_c(t) = \cos(\arcsin y(t)) = \sqrt{1 - y(t)^2} = \sqrt{x(t)^2} = |x(t)|.$$

Recall that $\operatorname{sgn} x(t) = \operatorname{sgn} x(c)$ for all $t \in I_c$. If $x(c) > 0$, then $|x(t)| = x(t)$, and we have $\cos\theta_c(t) = x(t)$. Otherwise, we use $\cos(\pi - \theta) = -\cos\theta$. This shows that in the case when $x(c) < 0$, we have $\cos\theta_c(t) = -|x(t)| = x(t)$. In either case, $\boldsymbol{\delta}(t) = (\cos\theta_c(t), \sin\theta_c(t))$ for all $t \in I_c$.

We need a different construction when $x(c) = 0$ and $y(c) = \pm 1$. This time we can define

$$\theta_c(t) := \begin{cases} \arccos x(t), & \text{if } y(c) = 1 \\ -\arccos x(t), & \text{if } y(c) = -1 \end{cases}$$

on an interval $I_c = (c - \delta, c + \delta)$, where $\delta > 0$ is chosen so that (i) $-1 < x(t) < 1$, and (ii) $\operatorname{sgn} y(t) = \operatorname{sgn} y(c)$, for all $t \in I_c$. Again, θ_c is C^1 on I_c, and again we have $\boldsymbol{\delta}(t) = (\cos\theta_c(t), \sin\theta_c(t))$ for all $t \in I_c$.

Take some closed interval $[a, b] \subseteq I^\circ$. Since the family $\{I_c \mid c \in [a, b]\}$ is a covering of the compact set $[a, b]$ by open sets, the Heine-Borel Theorem tells us that there is a finite set $\{c_1, c_2, \ldots, c_N\} \subseteq [a, b]$ for which $I_{c_1}, I_{c_2}, \ldots, I_{c_N}$ cover $[a, b]$. With a finite collection of open intervals, it is easy to see that if some triple of them have a nonempty intersection, then it is possible to shrink one of the three so that there is no longer a threefold intersection. For instance, say we have the intervals (a_1, b_1), (a_2, b_2), and (a_3, b_3), with $a_1 \le a_3 \le a_2 \le b_1 \le b_3 \le b_2$. Then replace (a_1, b_1) with (a_1, a_2), so $(a_1, a_2) \cap (a_2, b_2) = \emptyset$. This shrinking process will have to be repeated at most $\binom{N}{3}$ times on our collection of N intervals, but notice that it does not destroy the covering property of the entire collection with respect to $[a, b]$. Assuming this process has already been done, we have a finite collection of intervals I_{c_j} which cover $[a, b]$, such that no three overlap. We can also reassign the subscripts $1, 2, \dots, N$ so that the nonempty intersections $I_{c_i} \cap I_{c_j}$ are precisely those with $j = i + 1$.

Now consider θ_{c_i} and $\theta_{c_{i+1}}$ on their common domain, $J_i := I_{c_i} \cap I_{c_{i+1}}$. Since they are both radian measures for the same oriented angle $\angle(\mathbf{i}, \mathbf{0}, \boldsymbol{\delta}(t))$, where $\mathbf{i} := (1, 0)$, we must have

$$\theta_{c_{i+1}}(t) = \theta_{c_i}(t) - 2\pi\, n_i(t), \qquad \text{for each } t \in J_i.$$

Here $n_i(t)$ is an *integer-valued function*. However, $n_i(t) = [\theta_{c_i}(t) - \theta_{c_{i+1}}(t)]/2\pi$ is clearly a continuous function of t. A continuous integer-valued function must be constant. So we can write $\theta_{c_{i+1}}(t) = \theta_{c_i}(t) - 2\pi n_i$ for all $t \in J_i$, where $n_i \in \mathbb{Z}$ is a fixed constant. Finally, we glue all the pieces together to get

$$\theta_{a,b}(t) := \begin{cases} \theta_{c_1}(t), & \text{if } t \in I_{c_1} \\ \theta_{c_2}(t) + 2\pi n_1, & \text{if } t \in I_{c_2} \\ \theta_{c_3}(t) + 2\pi n_1 + 2\pi n_2, & \text{if } t \in I_{c_3} \\ \vdots & \vdots \\ \theta_{c_N}(t) + 2\pi n_1 + \cdots + 2\pi n_{N-1}, & \text{if } t \in I_{c_N} \end{cases}$$

Because each two consecutive pieces coincide on the overlap of their domains, $\theta_{a,b}$ is continuous on $[a, b]$. Moreover, it is C^1 at every $t \in (a, b)$, since in some small neighborhood of any such t, $\theta_{a,b}$ simply *coincides* with at least one of the functions θ_{c_i}; and all of these are known to be C^1. For the same reason (i.e., coinciding locally with some θ_{c_i}),

$$\boldsymbol{\delta}(t) = (\cos\theta_{a,b}(t), \sin\theta_{a,b}(t)) \qquad \text{for all } t \in [a, b].$$

If $(a, b) \cap (c, d) \ne \emptyset$, then $\theta_{a,b}(t) \equiv \theta_{c,d}(t)$ for all $t \in (a, b) \cap (c, d)$, because the pieces from which either $\theta_{a,b}$ or $\theta_{c,d}$ are glued together have precisely the same formulas defining them. We can therefore define a single continuous function θ on I° as follows. For any $t \in I^\circ$, choose any a and b such that $t \in (a, b) \subseteq I^\circ$, and define $\theta(t) := \theta_{a,b}(t)$. This value is independent of the particular a and b chosen, so long as $a < t < b$. Moreover, the function θ is C^1, since it coincides locally with C^1 functions near each point of I°.

The Osculating Plane

For a C^2 curve $\boldsymbol{\gamma}$ of nonvanishing curvature, the moving plane $\Pi(\mathbf{T}, \mathbf{N})$ spanned by its unit tangent and unit normal, or equivalently by its velocity $\mathbf{v}$ and acceleration $\mathbf{a}$, is called its *osculating plane* (i.e., “kissing plane”). In this section, we will show that for any fixed point on the trace, the osculating plane at that point is unique among all possible planes in fitting the local portion of the curve *superquadratically*, in the same way that the tangent line is unique among all possible lines in fitting the local portion of the curve *superlinearly.*

We use the unit speed parametrization $\boldsymbol{\Gamma} : J \to \mathbb{R}^e$. Consider the situation at a specific point $\mathbf{p} = \boldsymbol{\Gamma}(c)$, where $c \in J^\circ$. The first condition for fitting a plane to the portion of the curve near $\mathbf{p}$ is that the plane should pass through $\mathbf{p}$; i.e., it should be of the form $\mathbf{p} + \Pi_0$, where Π_0 is some plane through $\mathbf{0}$.

The second condition is that the unit tangent $\mathbf{T}_0 := \mathbf{T}(c)$, viewed as a position vector anchored to $\mathbf{0}$, should lie in Π_0. The reason is that the tangent line at $\mathbf{p}$ is already known to be the “line of best fit” to the curve at $\mathbf{p}$. It stands to reason that the line of best fit should be contained within the plane of best fit. Thus, we consider planes of the form $\mathbf{p} + \Pi(\mathbf{T}_0, \mathbf{M}_0) = \{\mathbf{p} + \alpha\mathbf{T}_0 + \beta\mathbf{M}_0 \mid \alpha, \beta \in \mathbb{R}\}$, where $\mathbf{M}_0$ is any vector linearly independent from $\mathbf{T}_0$. There is no loss of generality, in fact, if we assume that $\|\mathbf{M}_0\| = 1$ and $\mathbf{M}_0 \perp \mathbf{T}_0$.

By making a translation of the time axis (i.e., $s \mapsto s - c$) we can always arrange that $\mathbf{p} = \boldsymbol{\Gamma}(0)$. Assume that this has been done. Using the assumption that $\boldsymbol{\Gamma}$ is C^2, we have a Taylor expansion of order 2 at $s = 0$, namely

$$\boldsymbol{\Gamma}(s) = \boldsymbol{\Gamma}(0) + s\boldsymbol{\Gamma}'(0) + \tfrac{1}{2}s^2\boldsymbol{\Gamma}''(0) + s^2\boldsymbol{\varepsilon}(s),$$

where $\boldsymbol{\varepsilon}(s) \to \mathbf{0}$ as $s \to 0$. This can be rewritten as

$$\boldsymbol{\Gamma}(s) = \mathbf{p} + s\mathbf{T}_0 + \tfrac{1}{2}\kappa_0 s^2 \mathbf{N}_0 + s^2\boldsymbol{\varepsilon}(s),$$

where $\kappa_0 = \kappa(0)$ is the curvature of $\boldsymbol{\Gamma}$ at $\mathbf{p}$ (for recall that $\boldsymbol{\Gamma}'' = \kappa\mathbf{N}$; in particular, $\boldsymbol{\Gamma}''(0) = \kappa_0\mathbf{N}_0$).

Now for a fixed value of s close to 0, let $\alpha(s)$ and $\beta(s)$ be real numbers, and define

$$d(s) := \big\|\boldsymbol{\Gamma}(s) - \{\mathbf{p} + \alpha(s)\mathbf{T}_0 + \beta(s)\mathbf{M}_0\}\big\|$$

Writing $d_{\min}(s)$ for the minimum of $d(s)$ over all choices of $\alpha(s)$ and $\beta(s)$, and writing $\alpha_{\min}(s)$ and $\beta_{\min}(s)$ for a minimizing pair of such choices, we ask whether it is ever true that

$$\frac{d_{\min}(s)}{s^2} \to 0 \qquad \text{as } s \to 0.$$

This is the *superquadratic decay condition.* It says that the distance from the curve to the "fitting plane" vanishes at a faster-than-quadratic rate as a function of s, when s approaches its specific value at the contact point.

Plugging in the Taylor expansion of $\boldsymbol{\Gamma}(s)$, the superquadratic decay condition becomes

$$\left\| \frac{s - \alpha_{\min}(s)}{s^2}\mathbf{T}_0 + \frac{\kappa_0}{2}\mathbf{N}_0 - \frac{\beta_{\min}(s)}{s^2}\mathbf{M}_0 + \boldsymbol{\varepsilon}(s) \right\| \to 0$$

as $s \to 0$. Here, the term $\boldsymbol{\varepsilon}(s)$ may be ignored, since it is known to vanish; thus, if its sum with some other terms has any chance of vanishing, the sum of those other terms must vanish too. Moreover, since both $\mathbf{N}_0$ and $\mathbf{M}_0$ are orthogonal to $\mathbf{T}_0$, so is their linear combination $(\kappa_0/2)\mathbf{N}_0 - (\beta_{\min}(s)/s^2)\mathbf{M}_0$. It follows by the Pythagorean Theorem that the condition above is equivalent to

$$\left\| \frac{s - \alpha_{\min}(s)}{s^2}\mathbf{T}_0 \right\|^2 + \left\| \frac{\kappa_0}{2}\mathbf{N}_0 - \frac{\beta_{\min}(s)}{s^2}\mathbf{M}_0 \right\|^2 \to 0.$$

This can happen only if both norms above approach zero. In particular, we must have

$$\frac{\beta_{\min}(s)}{s^2}\mathbf{M}_0 \to \frac{\kappa_0}{2}\mathbf{N}_0 \qquad \text{as } s \to 0.$$

By hypothesis, $\kappa_0 > 0$; so $(\kappa_0/2)\mathbf{N}_0 \neq \mathbf{0}$. The only way a scalar function times the nonzero constant vector $\mathbf{M}_0$ can approach the nonzero constant vector $(\kappa_0/2)\mathbf{N}_0$ is to have $\mathbf{M}_0 \parallel \mathbf{N}_0$. Hence, $\Pi(\mathbf{T}_0, \mathbf{M}_0) = \Pi(\mathbf{T}_0, \mathbf{N}_0)$. We see that there is only one plane satisfying the superquadratic approximation property, and it is the osculating plane.

Curves in $\mathbb{R}^3$: Torsion

Consider a C^3 curve of nonvanishing curvature in $\mathbb{R}^3$, say with unit speed parametrization $\boldsymbol{\Gamma} : J \longrightarrow \mathbb{R}^3$. At each $s \in J$, the osculating plane $\Pi(\mathbf{T}(s), \mathbf{N}(s))$ has a unique unit normal vector $\mathbf{B}(s)$ for which the frame $(\mathbf{T}(s), \mathbf{N}(s), \mathbf{B}(s))$ is *positively oriented.* Roughly speaking, this means that the sense of rotational direction along the shortest circular arc from $\mathbf{T}(s)$ toward $\mathbf{N}(s)$ appears *counterclockwise* when viewed from the tip of $\mathbf{B}(s)$. We will make this more precise later, but what it will turn out to mean is that

$$\mathbf{B}(s) := \mathbf{T}(s) \times \mathbf{N}(s).$$

This is the *cross product* of vectors, which yields a vector orthogonal to its two factors (whether or not the two factors are orthogonal to one another, as just so happens to be the case here).

The vector $\mathbf{B}(s)$ is called the *binormal* at s. The rate at which $\mathbf{B}(s)$ changes with respect to s is called the *torsion* of the curve at s (up to a sign, whose significance will be discussed later). The torsion at $s = c$ tells us how sharply the curve twists away from its osculating plane at $s = c$ at the instant when s increases past c; in other words, it is a measure of how far the curve is from being *planar* near s, just as the curvature is a measure of how far the curve is from being rectilinear near s.

We need first to say a little about the cross product. Its definition is based on the 3×3 *determinant function*,

$$\det A = \det \begin{bmatrix} \mathbf{a} & \mathbf{b} & \mathbf{c} \end{bmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

We will sometimes denote this by $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle$, when we wish to view it as a function of three column vectors of A. Its definition is as follows:

$$\det A := \langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle := \sum_{i,j,k} \epsilon_{ijk}\, a_i b_j c_k,$$

where the function $(i, j, k) \mapsto \epsilon_{ijk}$ is the 3-*index alternator*, defined by

$$\epsilon_{ijk} := \begin{cases} 1, & \text{if } (i,j,k) \text{ is an even permutation} \\ -1, & \text{if } (i,j,k) \text{ is an odd permutation} \\ 0, & \text{otherwise.} \end{cases}$$

If i, j, k are distinct, so that (i, j, k) is a permutation of $\{1, 2, 3\}$, we say that this permutation is *even* if

$$\frac{(x_i - x_j)(x_i - x_k)(x_j - x_k)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = 1,$$

and *odd* if

$$\frac{(x_i - x_j)(x_i - x_k)(x_j - x_k)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = -1.$$

What this means is that an even permutation can be restored to increasing order by an even number of two-element swaps, while an odd permutation requires an odd number of two-element swaps. The even permutations of $\{1, 2, 3\}$ are $(1, 2, 3)$, $(2, 3, 1)$, and $(3, 1, 2)$; the odd ones are $(1, 3, 2)$, $(3, 2, 1)$, and $(2, 1, 3)$. Thus, in long form,

$$\det A = a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_1 b_3 c_2 - a_3 b_2 c_1 - a_2 b_1 c_3.$$

Why is this a "natural" function of a 3×3 matrix? It is straightforward to prove that the determinant function has the following properties. First, it is *alternating* in its three vector arguments:

$$\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \langle \mathbf{b}, \mathbf{c}, \mathbf{a} \rangle = \langle \mathbf{c}, \mathbf{a}, \mathbf{b} \rangle = -\langle \mathbf{a}, \mathbf{c}, \mathbf{b} \rangle.$$

Secondly, it is *additive in each vector input slot*:

$$\langle \mathbf{a}_1 + \mathbf{a}_2, \mathbf{b}, \mathbf{c} \rangle = \langle \mathbf{a}_1, \mathbf{b}, \mathbf{c} \rangle + \langle \mathbf{a}_2, \mathbf{b}, \mathbf{c} \rangle,$$

with similar laws for the second and third input slots. Thirdly, it is *homogenous in each input slot*:

$$\langle \alpha \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \alpha \langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle,$$

with similar laws for the second and third input slots. Finally, it has the *unit property*

$$\langle \mathbf{i}, \mathbf{j}, \mathbf{k} \rangle = \det I = 1,$$

where $\mathbf{i} := (1, 0, 0)$, $\mathbf{j} := (0, 1, 0)$, and $\mathbf{k} := (0, 0, 1)$ are the standard basis vectors in $\mathbb{R}^3$.

The alternating property tells us that $\langle \mathbf{a}, \mathbf{b}, \mathbf{b} \rangle = -\langle \mathbf{a}, \mathbf{b}, \mathbf{b} \rangle$, simply by exchanging the vectors in the second and third slots. Thus we find that a determinant with a repeated input vector is necessarily 0. The same is true for a determinant in which one input vector is linearly dependent on the other two; for

$$\langle \mathbf{a}, \mathbf{b}, \alpha \mathbf{a} + \beta \mathbf{b} \rangle = \alpha \langle \mathbf{a}, \mathbf{b}, \mathbf{a} \rangle + \beta \langle \mathbf{a}, \mathbf{b}, \mathbf{b} \rangle = 0 + 0 = 0.$$

Conversely, if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent, then $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle \neq 0$. Assume for contradiction that $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = 0$. Now each of $\mathbf{i}, \mathbf{j}, \mathbf{k}$ can be expressed in the form $x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$, by the assumed linear independence of $\mathbf{a}, \mathbf{b}, \mathbf{c}$. By expanding $\langle \mathbf{i}, \mathbf{j}, \mathbf{k} \rangle$ in terms of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ using additivity and homogeneity, each term will either contain a repeated vector, or will be a multiple of $\pm \langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle$. So all terms zero out, and we get the contradiction $1 = \langle \mathbf{i}, \mathbf{j}, \mathbf{k} \rangle = 0$. Thus, $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle \neq 0$.

Notice that

$$\det A = c_1 \sum_{i,j} \epsilon_{ij1}\, a_i b_j + c_2 \sum_{i,j} \epsilon_{ij2}\, a_i b_j + c_3 \sum_{i,j} \epsilon_{ij3}\, a_i b_j.$$

Here, $\epsilon_{ij1} = \epsilon_{1ij}$ is $+1$ for $(i,j) = (2,3)$, -1 for $(i,j) = (3,2)$, and 0 otherwise. So it coincides with $\epsilon_{i-1,j-1}$, where the 2-index alternator ϵ_{pq} is defined to be $+1$ if $(p,q) = (1,2)$, -1 if $(p,q) = (2,1)$, and 0 otherwise. It follows that the first term in the expansion above is just $c_1 \det A_{13} = c_1(a_2 b_3 - a_3 b_2)$, where

$$A_{13} := \begin{bmatrix} a_2 & b_2 \\ a_3 & b_3 \end{bmatrix}.$$

[The notation 'A_{13}' indicates that we delete row 1 and column 3 from A.] Similarly, the second term above is $-c_2 \det A_{23}$, and the third term is $c_3 \det A_{33}$. The minus sign in $-c_2 \det A_{23}$ comes from the fact that $\epsilon_{ij2} = -\epsilon_{i2j}$. The equation

$$\det A = c_1 \det A_{13} - c_2 \det A_{23} + c_3 \det A_{33}$$

is called a *cofactor expansion.* Notice that we can write it in the form of a dot product:

$$\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \mathbf{c} \cdot (\det A_{13}, -\det A_{23}, \det A_{33}).$$

Coming back to the cross product, the goal is to find $\mathbf{n}$ such that $\mathbf{n} \perp \mathbf{a}$ and $\mathbf{n} \perp \mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are given vectors. By taking $\mathbf{c} = \mathbf{a}$ or $\mathbf{c} = \mathbf{b}$ in the above, we find that

$$\mathbf{a} \cdot (\det A_{13}, -\det A_{23}, \det A_{33}) = \langle \mathbf{a}, \mathbf{b}, \mathbf{a} \rangle = 0$$

and

$$\mathbf{b} \cdot (\det A_{13}, -\det A_{23}, \det A_{33}) = \langle \mathbf{a}, \mathbf{b}, \mathbf{b} \rangle = 0.$$

It is clear, therefore, that the vector

$$\mathbf{n} = \left(\begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}, -\begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix}, \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \right) =: \mathbf{a} \times \mathbf{b}$$

does the job: it is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$. We can write the definition purely schematically as

$$\text{“}\mathbf{a} \times \mathbf{b} = \begin{vmatrix} a_1 & b_1 & \mathbf{i} \\ a_2 & b_2 & \mathbf{j} \\ a_3 & b_3 & \mathbf{k} \end{vmatrix}\text{”}$$

and expand along the last column as we did above. This is only a useful mnemonic; it should not be taken literally. A convenient formula for the k-th component of $\mathbf{a} \times \mathbf{b}$ is

$$[\mathbf{a} \times \mathbf{b}]_k = \sum_{i,j} \epsilon_{ijk}\, a_i b_j.$$

The basic properties of determinants cited above manifest themselves immediately as properties of the cross product. In particular, we have additivity and homogeneity laws in both factors:

$$\mathbf{a} \times (\alpha\mathbf{b} + \beta\mathbf{c}) = \alpha(\mathbf{a} \times \mathbf{b}) + \beta(\mathbf{a} \times \mathbf{c}) \quad \text{and} \quad (\alpha\mathbf{a} + \beta\mathbf{b}) \times \mathbf{c} = \alpha(\mathbf{a} \times \mathbf{c}) + \beta(\mathbf{b} \times \mathbf{c}).$$

The alternating property of the determinant yields an *anticommutative law* for the cross product:

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

The property about repeated inputs tells us that $\mathbf{a} \times \mathbf{a} = \mathbf{0}$, and more generally that

$$\mathbf{a} \times \mathbf{b} = \mathbf{0} \quad \text{if and only if} \quad \mathbf{a} \parallel \mathbf{b}.$$

A special case of this is $\mathbf{a} \times \mathbf{0} = \mathbf{0}$, since $\mathbf{a} \parallel \mathbf{0}$ for every $\mathbf{a}$. Notice finally that by construction,

$$\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}.$$

The alternating property shows that $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \langle \mathbf{b}, \mathbf{c}, \mathbf{a} \rangle = (\mathbf{b} \times \mathbf{c}) \cdot \mathbf{a} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$.

It turns out that if $\mathbf{a} \nparallel \mathbf{b}$, then the frame $(\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b})$ is *positively oriented.* [A *frame* is simply an ordered list of d linearly independent vectors in d-space.]

Two frames in $\mathbb{R}^3$, $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ and $(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*)$, are *co-oriented* if there are continuous curves $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} : [0, 1] \longrightarrow \mathbb{R}^3$ such that (i) $\boldsymbol{\alpha}(0) = \mathbf{a}$, $\boldsymbol{\alpha}(1) = \mathbf{a}^*$; (ii) $\boldsymbol{\beta}(0) = \mathbf{b}$, $\boldsymbol{\beta}(1) = \mathbf{b}^*$; (iii) $\boldsymbol{\gamma}(0) = \mathbf{c}$, $\boldsymbol{\gamma}(1) = \mathbf{c}^*$; and (iv) $\langle \boldsymbol{\alpha}(t), \boldsymbol{\beta}(t), \boldsymbol{\gamma}(t) \rangle \neq 0$ for all $t \in [0, 1]$. In essence, $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ can be continuously deformed into $(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*)$ without ever collapsing the transitional frames into a non-frame (i.e., a coplanar triple). Suppose this is true for two given frames. Since $\langle \boldsymbol{\alpha}(t), \boldsymbol{\beta}(t), \boldsymbol{\gamma}(t) \rangle$ is a continuous function of t that never takes the value 0, it must have the same sign at $t = 0$ and at $t = 1$. So, if $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ and $(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*)$ are co-oriented, then the determinants $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle$ and $\langle \mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^* \rangle$ must have the same sign. Conversely, two frames whose determinants have the same sign can always be connected by a continuous moving frame $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ over the time interval $[0, 1]$, where $\langle \boldsymbol{\alpha}(t), \boldsymbol{\beta}(t), \boldsymbol{\gamma}(t) \rangle \neq 0$ for all $t \in [0, 1]$. There are many ways to produce such a moving frame. To outline one messy but elementary way, the main idea will be to first deform $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ into a regular tetrahedron, so that each vector in the frame becomes a unit vector, and any two of them have unit distance between their tips. We can then rotate this tetrahedral frame through some angle about each coordinate axis, getting the image of $\mathbf{a}$ to be codirectional with $\mathbf{a}^*$, the image of $\mathbf{b}$ to lie in the plane $\Pi(\mathbf{a}^*, \mathbf{b}^*)$, and the image of $\mathbf{c}$ to lie in the halfspace of $\Pi(\mathbf{a}^*, \mathbf{b}^*)$ containing $\mathbf{c}^*$. Then we can deform this final tetrahedral frame into $(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*)$. The moves that make up the transition from $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ to its initial tetrahedral image, as well as the moves that "unwind" the final repositioned tetrahedral frame back into $(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*)$, will be made up of uniform rectilinear motions of one vector at a time along lines parallel or perpendicular to the plane of the other two vectors. So there is no danger of passing through a zero determinant during these stages. As for the rotations that move the thetrahedral frame into its final position, it is an easy algebraic problem to show that they actually preserve the exact value of the determinant all along their motions; so these rotation stages also never allow the determinant to zero out. The precise details are simple enough to work out, but we will leave this as an exercise for you to consider later (with appropriate hints).

A frame $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is said to be *positively oriented* if it is co-oriented with the *standard frame* $(\mathbf{i}, \mathbf{j}, \mathbf{k})$, i.e., if $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle > 0$. Otherwise it is *negatively oriented.*

To see that the specific frame $(\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b})$ is positively oriented, we need only check the sign of its determinant:

$$\langle \mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b} \rangle = (\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{a} \times \mathbf{b}) = \|\mathbf{a} \times \mathbf{b}\|^2 > 0,$$

where we have used the assumption $\mathbf{a} \nparallel \mathbf{b}$ to say that $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$.

The last property of the cross product we need to mention here is that

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

where $\theta = \measuredangle(\mathbf{a}, \mathbf{b})$ is determined uniquely by the conditions $\theta \in [0, \pi]$ and

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta.$$

An equivalent way to say this is that

$$\|\mathbf{a} \times \mathbf{b}\|^2 + (\mathbf{a} \cdot \mathbf{b})^2 = \|\mathbf{a}\|^2 \|\mathbf{b}\|^2,$$

since $\sin \theta \geq 0$ for $\theta \in [0, \pi]$. This identity can be motivated nicely by appealing to the theory of oriented volume for prisms in $\mathbb{R}^3$, but to save time here we give a straightforward if unenlightening verification. The left-hand side is

$$(a_2 b_3 - a_3 b_2)^2 + (a_1 b_3 - a_3 b_1)^2 + (a_1 b_2 - a_2 b_1)^2 + (a_1 b_1 + a_2 b_2 + a_3 b_3)^2.$$

In multiplying this out, all the "cross terms" will cancel. The middle term in the expansion of the first squared binomial above is $-2(a_2 b_3)(a_3 b_2)$; from the second squared binomial we get a middle term of $-2(a_1 b_3)(a_3 b_1)$, and from the third, $-2(a_1 b_2)(a_2 b_1)$. On the other hand, the squared trinomial appearing as the last term above has cross terms $2(a_1 b_1)(a_2 b_2) + 2(a_1 b_1)(a_3 b_3) + 2(a_2 b_2)(a_3 b_3)$. So all these terms cancel out perfectly. What remains are nine distinct "pure square terms" of the form $(a_i b_j)^2$; and since i and j each have exactly three possible values, these nine terms must include all possible combinations of i and j. So we are left with

$$\sum_{i=1}^{3} \sum_{j=1}^{3} a_i^2 b_j^2 = \left(\sum_{i=1}^{3} a_i^2 \right) \left(\sum_{j=1}^{3} b_j^2 \right) = \|\mathbf{a}\|^2 \|\mathbf{b}\|^2.$$

As a final note on determinants and cross products, we mention that the preceding theory can be generalized straightforwardly to d-space, with the understanding that for $d-1$ linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{d-1} \in \mathbb{R}^d$, their "cross product" $\mathbf{n} = \mathbf{a}_1 \times \mathbf{a}_2 \times \cdots \times \mathbf{a}_{d-1}$ gives the unique normal direction to the $(d-1)$-flat that they span. This $\mathbf{n}$ is defined by the requirement that for all $\mathbf{c} \in \mathbb{R}^d$,

$$\langle \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{d-1}, \mathbf{c} \rangle = \mathbf{c} \cdot \mathbf{n},$$

where the left-hand side is a $d \times d$ determinant. The components of $\mathbf{n}$ are found by cofactor-expanding this $d \times d$ determinant along its last column. Everything works just as expected, except for the slightly odd fact that the cross product is now a $(d-1)$-ary operation, rather than a binary operation. There are further generalizations of the concept, too. For instance, the cross product of $d-2$ vectors in d-space would be a $d \times 2$ matrix whose two columns are linearly independent spanning vectors for the 2-plane orthogonal to the original $(d-2)$-flat. In general, the cross product of $d-k$ vectors in d-space is a $d \times k$ matrix whose columns form a basis for the k-flat orthogonal to a given $(d-k)$-flat. We will not go into this here.

Coming back to the binormal $\mathbf{B}$ of a curve in $\mathbb{R}^3$, we note that because it is defined as the cross product of $\mathbf{T}$ and $\mathbf{N}$, it is orthogonal to the osculating plane $\Pi(\mathbf{T}, \mathbf{N})$. Also, it is a unit vector:

$$\|\mathbf{B}\| = \|\mathbf{T} \times \mathbf{N}\| = \|\mathbf{T}\|\|\mathbf{N}\| \sin \measuredangle(\mathbf{T}, \mathbf{N}) = 1 \cdot 1 \cdot \sin(\pi/2) = 1.$$

Finally, the moving frame $(\mathbf{T}, \mathbf{N}, \mathbf{B}) = (\mathbf{T}, \mathbf{N}, \mathbf{T} \times \mathbf{N})$ — so called because it changes with time — is always positively oriented. We call it the *Frenet frame* of the curve.

As mentioned earlier, the "product rule" of differentiation is true for a wide variety of binary operations, and one of them is the cross product (which we have seen satisfies left- and right-distributivity as well as the scaling laws). Applying this rule to $\mathbf{B}$, we get

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

We have already encountered the fact that $\mathbf{T}' = \mathbf{\Gamma}'' = \kappa \mathbf{N}$. It follows that the first term on the right-hand side is

$$\mathbf{T}' \times \mathbf{N} = (\kappa \mathbf{N}) \times \mathbf{N} = \kappa(\mathbf{N} \times \mathbf{N}) = \mathbf{0}.$$

Thus, $\mathbf{B}' = \mathbf{T} \times \mathbf{N}'$. Now $\mathbf{N}'$ is always orthogonal to $\mathbf{N}$, because $\mathbf{N}$ is a vector function of constant length (as we saw earlier). This forces $\mathbf{N}'$ to lie in the plane of $\mathbf{T}$ and $\mathbf{B}$, since that plane has $\mathbf{N}$ as its normal vector. Thus, for any s there are numbers $k = k(s)$ and $\tau = \tau(s)$ such that

$$\mathbf{N}' = -k\mathbf{T} + \tau \mathbf{B}.$$

These numbers are uniquely determined, since $\{\mathbf{T}, \mathbf{B}\}$ is a basis for the plane in question (the expansion coefficients of any vector with respect to a given basis are unique). Notice now that

$$\mathbf{B}' = \mathbf{T} \times \mathbf{N}' = \mathbf{T} \times (-k\mathbf{T} + \tau \mathbf{B}) = -k(\mathbf{T} \times \mathbf{T}) + \tau(\mathbf{T} \times \mathbf{B}) = -\tau(\mathbf{B} \times \mathbf{T}).$$

Consider the frame $(\mathbf{B}, \mathbf{T}, \mathbf{N})$. This is certainly orthonormal, and it is also positively oriented, since

$$\langle \mathbf{B}, \mathbf{T}, \mathbf{N} \rangle = \langle \mathbf{T}, \mathbf{N}, \mathbf{B} \rangle = 1.$$

The frame $(\mathbf{B}, \mathbf{T}, \mathbf{B} \times \mathbf{T})$ is also positively oriented, and also orthonormal, since $\|\mathbf{B} \times \mathbf{T}\| = 1 \cdot 1 \cdot \sin(\pi/2) = 1$. However, given two independent vectors in $\mathbb{R}^3$ in a particular order, there is one and only one unit vector orthogonal to them both and forming a positively oriented frame with them, in the given order. So we conclude that $\mathbf{B} \times \mathbf{T} = \mathbf{N}$. Returning to the calculation above, this means that

$$\mathbf{B}' = -\tau \mathbf{N}.$$

The quantity $\tau(s)$ is called the *torsion* of the curve $\mathbf{\Gamma}$ at s. Its absolute value is simply the rate at which the binormal $\mathbf{B}$ rotates (i.e., the rate at which the osculating plane changes direction):

$$|\tau| = |\tau| \|\mathbf{N}\| = \|-\tau \mathbf{N}\| = \|\mathbf{B}'\|.$$

The *sign* of $\tau(c)$ tells us whether the curve $\mathbf{\Gamma}$ is twisting "up", into the halfspace of the osculating plane $\Pi(\mathbf{T}(c), \mathbf{N}(c))$ containing $\mathbf{B}(c)$, or "down", into the halfspace containing $-\mathbf{B}(c)$, as s increases past c. This will be justified shortly.

We must still identify the function $k(s)$ appearing in the relation

$$\mathbf{N}' = -k\mathbf{T} + \tau\mathbf{B}.$$

Observe that

$$\mathbf{N}' \cdot \mathbf{T} = (-k\mathbf{T} + \tau\mathbf{B}) \cdot \mathbf{T} = -k(\mathbf{T} \cdot \mathbf{T}) + \tau(\mathbf{B} \cdot \mathbf{T}) = -k.$$

Here we have use the facts that $\|\mathbf{T}\| = 1$ and $\mathbf{T} \perp \mathbf{B}$. On the other hand, since $\mathbf{N} \cdot \mathbf{T} \equiv 0$, we have

$$0 = (\mathbf{N} \cdot \mathbf{T})' = (\mathbf{N}' \cdot \mathbf{T}) + (\mathbf{N} \cdot \mathbf{T}') = -k + (\mathbf{N} \cdot \kappa\mathbf{N}) = -k + \kappa.$$

So $k(s)$ is not a new function; it simply coincides with the curvature function: $k(s) = \kappa(s)$.

We now have a system of first-order vector-valued differential equations for the time evolution of the Frenet frame:

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

These are called the *Frenet-Serret Equations*. We can write them in the matrix form

$$\begin{bmatrix} \mathbf{T}' & \mathbf{N}' & \mathbf{B}' \end{bmatrix} = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{bmatrix} \begin{bmatrix} \mathbf{T} & \mathbf{N} & \mathbf{B} \end{bmatrix}.$$

Observe that the first matrix on the right-hand side is *skew-symmetric*, meaning that $a_{ji} = -a_{ij}$ for all $i, j \in \{1, 2, 3\}$.

The Frenet-Serret equations imply the *Fundamental Theorem of Curve Geometry*: Given any continuous functions $K : [0, \ell] \longrightarrow (0, \infty)$ and $T : [0, \ell] \longrightarrow \mathbb{R}$, and given an initial position $\mathbf{p}$ and an initial orthonormal and positively oriented frame $(\mathbf{t}_0, \mathbf{n}_0, \mathbf{b}_0)$, there is exactly one unit speed curve $\mathbf{\Gamma} : [0, \ell] \longrightarrow \mathbb{R}^3$ such that

(i) $\mathbf{\Gamma}(0) = \mathbf{p}$,

(ii) the curvature κ and torsion τ of $\mathbf{\Gamma}$ satisfy $\kappa(s) \equiv K(s)$ and $\tau(s) \equiv T(s)$ for all $s \in [0, \ell]$, and

(iii) the Frenet frame $(\mathbf{T}, \mathbf{N}, \mathbf{B})$ of $\mathbf{\Gamma}$ satisfies $\mathbf{T}(0) = \mathbf{t}_0$, $\mathbf{N}(0) = \mathbf{n}_0$, and $\mathbf{B}(0) = \mathbf{b}_0$.

Ignoring the given data on the initial position and initial frame directionality (which can always be achieved through Euclidean motions of $\mathbb{R}^3$ — translations, rotations, and reflections), what this means is that the *shape* of the curve is completely determined by the two prescribed functions K and T. This is a remarkable fact.

The proof of the Fundamental Theorem of Curve Geometry is an application of a basic existence theorem for solutions of *initial value problems* in differential equations. The Frenet-Serret equations — with κ replaced by the given function K, and τ replaced by the given function T — give a first-order system of linear differential equations for the nine scalar component functions of $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$. We also have initial values for these nine functions at time $s = 0$, based on the given $\mathbf{t}_0$, $\mathbf{n}_0$, and $\mathbf{b}_0$. The existence theorem says that there is a unique solution to such a linear initial value problem, which extends across the entire domain $[0, \ell]$ of the coefficient functions K and T, provided those coefficient functions are continuous (which we have assumed as a hypothesis). Once we have the unique solution $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ of the initial value problem, we check that the quantities $\|\mathbf{t}\|^2$, $\|\mathbf{n}\|^2$, $\|\mathbf{b}\|^2$, $\mathbf{t} \cdot \mathbf{n}$, $\mathbf{t} \cdot \mathbf{b}$, and $\mathbf{n} \cdot \mathbf{b}$ are "constants of motion", i.e., their derivatives with respect to s are identically zero, so that they remain constant as s increases. Since their initial values are given to be $1, 1, 1, 0, 0, 0$, by the assumed orthonormality of the initial frame $(\mathbf{t}_0, \mathbf{n}_0, \mathbf{b}_0)$, these values are maintained over time, and we find that the moving frame $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ always remains orthonormal. The verifications that the relevant s-derivatives are identically zero is straightforward; it comes directly from the given differential equations. Finally, since the function $\mathbf{t}(s)$ is now uniquely determined on $[0, \ell]$, we can consider another initial value problem: $\mathbf{\Gamma}' = \mathbf{t}$ and $\mathbf{\Gamma}(0) = \mathbf{p}$, where the initial point $\mathbf{p}$ has been given. Again, this is a first-order linear system in the three component functions of $\mathbf{\Gamma}$, and it has a unique solution $\mathbf{\Gamma}(s)$ defined on $[0, \ell]$. It is then easy to see that $\mathbf{\Gamma}$ is a unit speed curve (since $\mathbf{T} := \mathbf{\Gamma}' = \mathbf{t}$, which is known to be a unit vector). We can use the definitions of $\mathbf{T}$, $\mathbf{N}$, $\mathbf{B}$, κ, and τ in terms of $\mathbf{\Gamma}$ and its derivatives to verify that, indeed, $\kappa \equiv K$, $\tau \equiv T$, $\mathbf{T}(0) = \mathbf{t}_0$, $\mathbf{N}(0) = \mathbf{n}_0$, and $\mathbf{B}(0) = \mathbf{b}_0$. This is only an outline of the proof, but it conveys the essential ideas.

To get a practical formula for the torsion τ, we note first that dotting the equation $\mathbf{B}' = -\tau\mathbf{N}$ with $\mathbf{N}$ gives

$$\tau = -\mathbf{B}' \cdot \mathbf{N} = -(\mathbf{T} \times \mathbf{N}') \cdot \mathbf{N} = (\mathbf{N}' \times \mathbf{T}) \cdot \mathbf{N} = \langle \mathbf{N}', \mathbf{T}, \mathbf{N} \rangle = \langle \mathbf{T}, \mathbf{N}, \mathbf{N}' \rangle.$$

Now $\mathbf{T} = \mathbf{\Gamma}'$, and $\mathbf{N} = \mathbf{T}'/\kappa = \mathbf{\Gamma}''/\kappa$. By the quotient rule,

$$\mathbf{N}' = \left(\frac{\mathbf{\Gamma}''}{\kappa}\right)' = \frac{\kappa\mathbf{\Gamma}''' - \kappa'\mathbf{\Gamma}''}{\kappa^2}.$$

Hence,

$$\tau = \left\langle \mathbf{\Gamma}',\, \frac{\mathbf{\Gamma}''}{\kappa},\, \frac{\kappa\mathbf{\Gamma}''' - \kappa'\mathbf{\Gamma}''}{\kappa^2} \right\rangle = \frac{\left\langle \mathbf{\Gamma}', \mathbf{\Gamma}'', \kappa\mathbf{\Gamma}''' - \kappa'\mathbf{\Gamma}'' \right\rangle}{\kappa^3}.$$

Here we have used the homogeneity property in each input slot of the determinant function to pull the factors of $1/\kappa$ out. We can use additivity in each slot to rewrite the above result as

$$\tau = \frac{\kappa\langle \mathbf{\Gamma}', \mathbf{\Gamma}'', \mathbf{\Gamma}'''\rangle}{\kappa^3} - \frac{\kappa'\langle \mathbf{\Gamma}', \mathbf{\Gamma}'', \mathbf{\Gamma}''\rangle}{\kappa^3} = \frac{\langle \mathbf{\Gamma}', \mathbf{\Gamma}'', \mathbf{\Gamma}'''\rangle}{\kappa^2}.$$

Here we used the fact that a determinant with a repeated input must be zero. So a simple formula for τ is

$$\tau = \frac{\langle \mathbf{\Gamma}', \mathbf{\Gamma}'', \mathbf{\Gamma}'''\rangle}{\|\mathbf{\Gamma}''\|^2}.$$

To work with this formula, one must remember to first reparametrize a given curve with respect to arclength.

All that remains is to establish the geometric interpretation of the sign of τ we mentioned earlier. By a time translation, there is no loss of generality in studying the situation at $s = 0$. Suppose $\tau(0) > 0$. What we want to prove is that for sufficiently small values $s > 0$, the unit tangent $\mathbf{T}(s)$ has a positive component along $\mathbf{B}(0)$, so that the instantaneous direction of motion for the curve points at least somewhat "upward" from the osculating plane $\Pi(\mathbf{T}(0), \mathbf{N}(0))$ — upward being defined by the direction of $\mathbf{B}(0)$ — at least for some amount of time immediately following time $s = 0$. We can use the second order Taylor expansion

$$\mathbf{T}(s) = \mathbf{T}(0) + s\mathbf{T}'(0) + \tfrac{1}{2}s^2\mathbf{T}''(0) + s^2\boldsymbol{\varepsilon}(s), \qquad \boldsymbol{\varepsilon}(s) \to \mathbf{0} \text{ as } s \to 0.$$

To find the component of this along the unit vector $\mathbf{B}(0)$, we simply dot it with $\mathbf{B}(0)$. Since $\mathbf{B}(0)$ is orthogonal to both $\mathbf{T}(0)$ and $\mathbf{T}'(0) = \kappa(0)\mathbf{N}(0)$, we get

$$\mathbf{T}(s) \cdot \mathbf{B}(0) = \tfrac{1}{2}s^2(\mathbf{T}''(0) \cdot \mathbf{B}(0)) + s^2(\boldsymbol{\varepsilon}(s) \cdot \mathbf{B}(0)).$$

Now

$$\mathbf{T}'' = (\kappa\mathbf{N})' = \kappa\mathbf{N}' + \kappa'\mathbf{N} = \kappa(-\kappa\mathbf{T} + \tau\mathbf{B}) + \kappa'\mathbf{N} = -\kappa^2\mathbf{T} + \kappa'\mathbf{N} + \kappa\tau\mathbf{B}.$$

Evaluating at $s = 0$ and dotting with $\mathbf{B}(0)$, the only surviving term is $\kappa(0)\tau(0)\|\mathbf{B}(0)\|^2 = \kappa(0)\tau(0)$. Thus,

$$\mathbf{T}(s) \cdot \mathbf{B}(0) = \left\{\tfrac{1}{2}\kappa(0)\tau(0) + \boldsymbol{\varepsilon}(s) \cdot \mathbf{B}(0)\right\}s^2.$$

As $s \to 0$, we have $|\boldsymbol{\varepsilon}(s) \cdot \mathbf{B}(0)| \leq \|\boldsymbol{\varepsilon}(s)\| \to 0$ by Cauchy-Schwarz. So that term becomes unimportant compared with the term $\frac{1}{2}\kappa(0)\tau(0)$, which is positive ($\kappa > 0$ always, and $\tau(0) > 0$ by hypothesis). Accounting for the positive factor s^2, we see that $\mathbf{T}(s)$ has a positive component along $\mathbf{B}(0)$ for sufficiently small $s > 0$. [Since only s^2 is involved here, you may wonder why the argument doesn't work in backward time, too. The answer is that the unit tangent already reflects the forward direction of time, since that's the direction in which the particle moves in forward time.]

It will be left as an exercise (with hints to come soon) for you to determine all curves in $\mathbb{R}^3$ having both constant nonzero curvature and constant torsion.

There is a great deal more that one could say about the calculus and geometry of curves, but we have managed to give a pretty good sample of the theory here.