## Chapter 2

## Curves

## Topics :

1. Tangent Vectors and Frames
2. Directional Derivatives
3. Curves in Euclidean 3-Space $\mathbb{R}^{3}$
4. Serret-Frenet Formulas
5. The Fundamental Theorem for Curves
6. Some Remarks

Copyright © Claudiu C. Remsing, 2006.
All rights reserved.

### 2.1 Tangent Vectors and Frames

## Tangent vectors

The basic method used to investigate curves (in Euclidean 3 -space $\mathbb{R}^{3}$ ) consists in assigning at each point (along the curve) a certain frame (i.e., a set of three mutually orthogonal unit vectors) and then express the rate of change of the frame in terms of the frame itself. In a real sense, the geometry of curves is merely a corollary of these basic results.

Note : A frame consists of vectors located at some specific point. These vectors are not free vectors (viewed as translations) but fixed vectors. We need to make this distinction precise by "re-thinking" the representation of a geometric vector. To obtain a concept that is both practical and precise, we shall describe an "arrow" by giving the starting (fixed) point $p$ and the change (vector) $v$, necessary to reach its terminal point $p+v$.

We make the following definition.
2.1.1 Definition. A tangent vector to $\mathbb{R}^{3}$ at a point p , denoted by $v_{p}$, is an ordered pair $(p, v) . p$ is the point of application of $v_{p}$, and $v$ is the vector part.

NOTE : $p+v$ is considered as the position vector of a point.
We shall always picture $v_{p}$ as the arrow (directed line segment) from the point $p$ to the point $p+v$.
2.1.2 Example. If $p=(1,1,3)$ and $v=(2,3,2)$ (in fact, $v=\left[\begin{array}{l}2 \\ 3 \\ 2\end{array}\right]$ ), then $v_{p}$ "runs" from $(1,1,3)$ to $(3,4,5)$.

We emphasize that tangent vectors $v_{p}$ and $w_{q}$ are equal if and only if they have the same vector part and the same point of application :

$$
v_{p}=w_{q} \Longleftrightarrow(v=w \quad \text { and } \quad p=q) .
$$

Tangent vectors $v_{p}$ and $v_{q}$ with the same vector part, but different points of application, are said to be parallel.

Note : It is essential to recognize that $v_{p}$ and $v_{q}$ are different tangent vectors if $p \neq q$. In physics, the concept of moment of a force shows this clearly : the same force $v$ applied at different points $p$ and $q$ of a rigid body can produce quite different rotational effects.

Let $p$ be a point of $\mathbb{R}^{3}$. The set $T_{p} \mathbb{R}^{3}$ of all tangent vectors to $\mathbb{R}^{3}$ at $p$ is called the tangent space of $\mathbb{R}^{3}$ at $p$.

Note: $\mathbb{R}^{3}$ has a different tangent space at each and every one of its points.
Since all the tangent vectors in a given tangent space have the same point of application, we can borrow the vector addition and scalar multiplication of $\mathbb{R}^{3}$ to turn $T_{p} \mathbb{R}^{3}$ into a vector space. Explicitly, we define (for $v_{p}, w_{p} \in T_{p} \mathbb{R}^{3}$ and $\lambda \in \mathbb{R}$ )

$$
v_{p}+w_{p}:=(v+w)_{p} \quad \text { and } \quad \lambda v_{p}:=(\lambda v)_{p} .
$$

This is just the usual "parallelogram law" for addition of vectors, and scalar multiplication by $\lambda$ merely stretches a tangent vector by a factor $|\lambda|$, reversing its direction if $\lambda<0$.
$\diamond$ Exercise 52 Show that, for a fixed point $p$, the vector spaces $\mathbb{R}^{3}$ and $T_{p} \mathbb{R}^{3}$ are isomorphic.

## Vector fields

2.1.3 Definition. A vector field $X$ on $\mathbb{R}^{3}$ is a mapping

$$
p \in \mathbb{R}^{3} \mapsto X(p) \in T_{p} \mathbb{R}^{3} .
$$

Let $X$ and $Y$ be vector fields on $\mathbb{R}^{3}$. Then we can define $X+Y$ to be the vector field on $\mathbb{R}^{3}$ such that

$$
(X+Y)(p):=X(p)+Y(p)
$$

for all $p \in \mathbb{R}^{3}$. Similarly, if $f$ is a (real-valued) function on $\mathbb{R}^{3}$ and $X$ is a vector field on $\mathbb{R}^{3}$, then we can define $f X$ to be the vector field on $\mathbb{R}^{3}$ such that

$$
(f X)(p):=f(p) X(p)
$$

for all $p \in \mathbb{R}^{3}$.
NOTE : Both operations were defined "pointwise". This scheme is general. For convenience, we shall call it the pointwise principle : if a certain operation can be performed on the values of two functions at each point, then that operation can be extended to the functions themselves; simply apply it to their values at each point. By means of the pointwise principle we can automatically extend other operations on individual tangent vectors (like dot product and cross product) to operations on vector fields.

Let $E_{1}, E_{2}, E_{3}$ be the vector fields on $\mathbb{R}^{3}$ such that

$$
E_{1}(p):=(1,0,0)_{p}, \quad E_{2}(p):=(0,1,0)_{p}, \quad \text { and } \quad E_{3}(p):=(0,0,1)_{p}
$$

at each point $p$ of $\mathbb{R}^{3}$. Thus $E_{i}$ is the unit vector field in the positive $x_{i^{-}}$ direction. We shall refer to the ordered set $\underline{E}=\left(E_{1}, E_{2}, E_{3}\right)$ as the natural frame field on $\mathbb{R}^{3}$.
2.1.4 Proposition. If $X$ is a vector field on $\mathbb{R}^{3}$, then there exist uniquely determined real-valued functions $X_{1}, X_{2}, X_{3}$ on $\mathbb{R}^{3}$ such that

$$
X=X_{1} E_{1}+X_{2} E_{2}+X_{3} E_{3}
$$

Proof : By definition, vector field $X$ assigns to each point $p$ a tangent vector $X(p)$ at $p$. Thus the vector part of $X(p)$ depends on $p$, so we write it

$$
\left(X_{1}(p), X_{2}(p), X_{3}(p)\right)
$$

This defines $X_{1}, X_{2}$ and $X_{3}$ as (real-valued) functions on $\mathbb{R}^{3}$. Hence

$$
\begin{aligned}
X(p) & =\left(X_{1}(p), X_{2}(p), X_{3}(p)\right)_{p} \\
& =X_{1}(p)(1,0,0)_{p}+X_{2}(p)(0,0,1)_{p}+X_{3}(p)(0,0,1)_{p} \\
& =X_{1}(p) E_{1}(p)+X_{2}(p) E_{2}(p)+X_{3}(p) E_{3}(p)
\end{aligned}
$$

for each point $p$. This means that the vector fields $X$ and $X_{1} E_{1}+X_{2} E_{2}+$ $X_{3} E_{3}$ have the same (tangent vector) value at each point, and hence they are equal.

The functions $X_{1}, X_{2}$, and $X_{3}$ are called the Euclidean coordinate functions of $X$. We write

$$
X=\left(X_{1}, X_{2}, X_{3}\right) \quad \text { or sometimes } \quad X=\left[\begin{array}{c}
X_{1} \\
X_{2} \\
X_{3}
\end{array}\right] .
$$

Note: A vector field $X$ on $\mathbb{R}^{3}$ is a mapping not from $\mathbb{R}^{3}$ to $\mathbb{R}^{3}$ but from $\mathbb{R}^{3}$ to (the union) $\bigcup_{p \in \mathbb{R}^{3}} T_{p} \mathbb{R}^{3}$. So $X(p)=\left(p,\left(X_{1}(p), X_{2}(p), X_{3}(p)\right)\right)=\left(X_{1}(p), X_{2}(p), X_{3}(p)\right)_{p}$.

Computations involving vector fields may always be expressed in terms of their Euclidean coordinate functions. A vector field $X$ is differentiable provided its Euclidean coordinate functions are differentiable. Henceforth, we shall understand "vector field" to mean "differentiable vector field".

NOTE : Since the subscript notation $v_{p}$ for a tangent vector is somewhat cumbersome, from now on we shall omit the point of application $p$ from the notation if no confusion is caused. However, in many situations the point of application is crucial, and will be indicated by using the old notation $v_{p}$ or the phrase "a tangent vector $v$ to $\mathbb{R}^{3}$ at $p^{\prime \prime}$.
$\diamond$ Exercise 53 Sketch the following vector fields on (the Euclidean plane) $\mathbb{R}^{2}$ :
(a) $X(p)=(1,0)$;
(b) $X(p)=p$;
(c) $X(p)=-p$;
(d) $X\left(x_{1}, x_{2}\right)=\left(x_{2}, x_{1}\right)$;
(e) $X\left(x_{1}, x_{2}\right)=\left(-x_{2}, x_{1}\right)$.

## Frames

Using the isomorphism $v \mapsto v_{p}$ between $\mathbb{R}^{3}$ and $T_{p} \mathbb{R}^{3}$, the dot product on $\mathbb{R}^{3}$ may be transferred to each of its tangent spaces.
2.1.5 Definition. The dot product of tangent vectors $v_{p}$ and $w_{p}$ at the same point of $\mathbb{E}^{3}$ is the number

$$
v_{p} \bullet w_{p}:=v \bullet w
$$

Note : This definition provides a dot product on each tangent space $T_{p}\left(\mathbb{R}^{3}\right)$ with the same properties as the original dot product on $\mathbb{R}^{3}$. In particular, each tangent vector $v_{p}$ has a norm (or length) $\left\|v_{p}\right\|:=\|v\|$. A vector of length 1 is called a unit tangent vector. Two tangent vectors $v_{p}$ and $w_{p}$ are orthogonal if and only if $v_{p} \bullet w_{p}=0$.
2.1.6 Definition. An ordered set $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ of three mutually orthogonal unit tangent vectors to $\mathbb{R}^{3}$ at the point $p$ is called a frame (at $p)$.

Thus $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ is a frame if and only if

$$
u_{i} \bullet u_{j}=\delta_{i j}, \quad i, j=1,2,3
$$

$\diamond$ Exercise 54 Check that the tangent vectors

$$
u_{1}=\frac{1}{\sqrt{6}}(1,2,1)_{p}, \quad u_{2}=\frac{1}{2 \sqrt{2}}(-2,0,2)_{p}, \quad \text { and } \quad u_{3}=\frac{1}{\sqrt{3}}(1,-1,1)_{p}
$$

constitute a frame at $p$. Express $v=(6,1,-1)_{p}$ as a linear combination of these vectors. (Check the result by direct computation.)
2.1.7 Example. At each point $p \in \mathbb{R}^{3}$, the tangent vectors

$$
E_{1}(p):=(1,0,0)_{p}, \quad E_{2}(p):=(0,1,0)_{p}, \quad E_{3}(p):=(0,0,1)_{p}
$$

constitute a frame, called the natural frame (at $p$ ).
If $v$ is a tangent vector to $\mathbb{R}^{3}$ at some point $p$, then

$$
v=\left(v_{1}, v_{2}, v_{3}\right)_{p}=v_{1} E_{1}(p)+v_{2} E_{2}(p)+v_{3} E_{3}(p)
$$

Exercise 55 Let $v \in T_{p} \mathbb{R}^{3}$ and let $\left(u_{1}, u_{2}, u_{3}\right)$ be a frame (at $p$ ). Show that

$$
v=\left(v_{1}, v_{2}, v_{3}\right)_{p}=\left(v \bullet u_{1}\right) u_{1}+\left(v \bullet u_{2}\right) u_{2}+\left(v \bullet u_{3}\right) u_{3} .
$$

The numbers $v \bullet u_{i}(i=1,2,3)$ are the coordinates of the tangent vector $v$ with respect to the frame $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$.
2.1.8 DEFINITION. The cross product of tangent vectors $v_{p}$ and $w_{p}$ at the same point $p \in \mathbb{R}^{3}$ is the tangent vector (at $p$ )

$$
\begin{aligned}
v_{p} \times w_{p} & :=\left|\begin{array}{ccc}
E_{1}(p) & E_{2}(p) & E_{3}(p) \\
v_{1} & v_{2} & v_{3} \\
w_{1} & w_{2} & w_{3}
\end{array}\right| \\
& =\left(v_{2} w_{3}-v_{3} w_{2}\right) E_{1}(p)+\left(v_{3} w_{1}-v_{1} w_{3}\right) E_{2}(p)+\left(v_{1} w_{2}-v_{2} w_{1}\right) E_{3}(p)
\end{aligned}
$$

Note : Likewise, this definition provides a cross product on each tangent space $T_{p} \mathbb{R}^{3}$ with the same properties as the original cross product on $\mathbb{R}^{3}$. In particular, two tangent vectors $v_{p}$ and $w_{p}$ are collinear if and only if $v_{p} \times w_{p}=0$.
$\diamond$ Exercise 56 If $\left(u_{1}, u_{2}, u_{3}\right)$ is a frame, show that

$$
u_{1} \bullet u_{2} \times u_{3}= \pm 1
$$

Let $F, x \mapsto A x+c$ be an isometry on $\mathbb{R}^{3}$. Then its orthogonal part $A$ defines a mapping $F_{*}$ that carries each tangent vector at $p$ to a tangent vector at $F(p)$. The mapping

$$
F_{*}=F_{*, p}: T_{p} \mathbb{R}^{3} \rightarrow T_{F(p)} \mathbb{R}^{3}, \quad v_{p} \mapsto(A v)_{F(p)}
$$

is called the tangent mapping of $F$ (at $p)$. In terms of Euclidean coordinates, we have

$$
\begin{aligned}
F_{*}\left(v_{1} E_{1}(p)+v_{2} E_{2}(p)+v_{3} E_{3}(p)\right)= & \left(a_{11} v_{1}+a_{12} v_{2}+a_{13} v_{3}\right) E_{1}(F(p))+ \\
& \left(a_{21} v_{1}+a_{22} v_{2}+a_{23} v_{3}\right) E_{2}(F(p))+ \\
& \left(a_{31} v_{1}+a_{32} v_{2}+a_{33} v_{3}\right) E_{3}(F(p)) \\
= & \sum_{i, j=1}^{3}\left(a_{i j} v_{j}\right) E_{i}(F(p)) .
\end{aligned}
$$

Exercise 57 If $T$ is a translation on $\mathbb{R}^{3}$, then for every tangent vector $v \in$ $T_{p} \mathbb{R}^{3}$ show that $T_{*}(v)$ is parallel to $v$.
$\diamond$ Exercise 58 If $F$ and $G$ are two isometries on $\mathbb{R}^{3}$, show that

$$
(G F)_{*}=G_{*} F_{*} \quad \text { and } \quad\left(F^{-1}\right)_{*}=\left(F_{*}\right)^{-1}
$$

Exercise 59 Given an isometry $F$ on $\mathbb{R}^{3}$, show that its tangent mapping $F_{*}$ preserves the dot product of any two (tangent) vectors.

Since dot products are preserved, it follows automatically that derived concepts such as norm and orthogonality are preserved. Explicitly, if $F$ is an isometry, then $\left\|F_{*}(v)\right\|=\|v\|$, and if $v$ and $w$ are orthogonal, so are $F_{*}(v)$ and $F_{*}(w)$. Thus frames are also preserved.
$\diamond$ Exercise 60 If $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ is a frame at some point $p \in \mathbb{R}^{3}$ and $F$ is an isometry on $\mathbb{R}^{3}$, show that $F_{*}(\underline{u})=\left(F_{*}\left(u_{1}\right), F_{*}\left(u_{2}\right), F_{*}\left(u_{3}\right)\right)$ is a frame at $F(p)$.

Recall that two points uniquely determine a translation. We now show that two frames uniquely determine an isometry.
2.1.9 Theorem. Given any two frames on $\mathbb{R}^{3}$, say $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ at the point $p$ and $\underline{w}=\left(w_{1}, w_{2}, w_{3}\right)$ at the point $q$, there exists a unique isometry $F$ on $\mathbb{R}^{3}$ such that

$$
F_{*}\left(u_{i}\right)=w_{i}, \quad i=1,2,3
$$

Proof : First we show that there is such an isometry. Let $u_{1}, u_{2}, u_{3}$ and $w_{1}, w_{2}, w_{3}$ be the points of $\mathbb{R}^{3}$ corresponding to (the vector parts of) the elements in the two frames. Let $A$ be the unique linear transformation on $\mathbb{R}^{3}$ such that $A\left(u_{i}\right)=w_{i}, i=1,2,3$.
$\diamond$ Exercise 61 Check that the transformation (matrix) $A$ is orthogonal.
Let $T$ be the translation by (the vector) $q-A(p)$. We claim that the isometry $F=T A$ carries the frame $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ to the frame $\underline{w}=\left(w_{1}, w_{2}, w_{3}\right)$. First observe that

$$
F(p)=T A(p)=q-A(p)+A(p)=q .
$$

Then we get

$$
F_{*}\left(u_{i}\right)=\left(A u_{i}\right)_{F(p)}=\left(w_{i}\right)_{F(p)}=\left(w_{i}\right)_{q}=w_{i}, \quad i=1,2,3
$$

To prove uniqueness, we observe that the choice of $A$ is the only possibility for the orthogonal part of the required isometry. The translation part is then completely determined also, since it must carry $p$ to $q$. Hence the isometry $F=T A$ is uniquely determined.

The isometry $F=T A$ (that carries the frame $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ to the frame $\left.\underline{w}=\left(w_{1}, w_{2}, w_{3}\right)\right)$ can be computed explicitly as follows. Let

$$
u_{i}=\left(u_{1 i}, u_{2 i}, u_{3 i}\right)_{p} \quad \text { and } \quad w_{i}=\left(w_{1 i}, w_{2 i}, w_{3 i}\right)_{q}, \quad i=1,2,3
$$

Then we form the $3 \times 3$ matrices (called the attitude matrices of the frames)

$$
U:=\left[\begin{array}{lll}
u_{1} & u_{2} & u_{3}
\end{array}\right]=\left[u_{i j}\right] \quad \text { and } \quad W:=\left[\begin{array}{lll}
w_{1} & w_{2} & w_{3}
\end{array}\right]=\left[w_{i j}\right] .
$$

$\diamond$ Exercise 62 Show that the attitude matrix of a frame is orthogonal.
We claim that (the orthogonal matrix) $A$ is $W U^{T}$. To verify this it suffices to check that

$$
W U^{\top}\left(u_{i}\right)=w_{i}, \quad i=1,2,3
$$

since this uniquely characterizes $A$. For $i=1$, we have

$$
W U^{\top}\left(u_{1}\right)=W U^{\top}\left[\begin{array}{l}
u_{11} \\
u_{21} \\
u_{31}
\end{array}\right]=W\left[\begin{array}{l}
1 \\
0 \\
0
\end{array}\right]=\left[\begin{array}{l}
w_{11} \\
w_{21} \\
w_{31}
\end{array}\right]=w_{1}
$$

That is, $W U^{\top}\left(u_{1}\right)=w_{1}$. The cases $i=2,3$ are similar; hence

$$
A=W U^{\top}\left(=W U^{-1}\right)
$$

As noted above, $T$ is then necessarily the translation by $q-A(p)$.
$\diamond$ Exercise 63 In each case decide whether $F$ is an isometry on $\mathbb{R}^{3}$. If isometry exists, find the translation and orthogonal parts.
(a) $F(x)=-x$;
(b) $F(x)=(x \bullet a) a$ where $\|a\|=1$;
(c) $F(x)=\left(x_{3}-3, x_{2}-2, x_{1}+1\right)$;
(d) $F(x)=\left(x_{1}, x_{2}, 2\right)$.
$\diamond$ Exercise 64 Identify the isometry $F, x \mapsto-x$ on $\mathbb{R}^{3}$.

Exercise 65 Show that the matrix

$$
A=\left[\begin{array}{lll}
0 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 0
\end{array}\right]
$$

represents a reflection in a plane. Find the plane.
$\diamond$ Exercise 66 Given the frame

$$
u_{1}=\frac{1}{3}(2,2,1)_{p}, \quad u_{2}=\frac{1}{3}(-2,1,2)_{p}, \quad u_{3}=\frac{1}{3}(1,-2,2)_{p}
$$

at $p=(0,1,0)$ and the frame

$$
w_{1}=\frac{1}{\sqrt{2}}(1,0,1)_{q}, \quad w_{2}=(0,1,0)_{q}, \quad w_{3}=\frac{1}{\sqrt{2}}(1,0,-1)_{q}
$$

at $q=(3,-1,1)$, find $c$ and $A$ such that the isometry $F=T_{c} A$ carries the frame $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ to the frame $\underline{w}=\left(w_{1}, w_{2}, w_{3}\right)$.

## Frame fields

2.1.10 Definition. Vector fields $U_{1}, U_{2}, U_{3}$ on $\mathbb{R}^{3}$ constitute a frame field on $\mathbb{R}^{3}$ provided

$$
U_{i} \bullet U_{j}=\delta_{i j}, \quad i, j=1,2,3
$$

Thus at each point $p \in \mathbb{R}^{3}$ the (tangent) vectors $U_{1}(p), U_{2}(p), U_{3}(p)$ form a frame.
$\diamond$ Exercise 67 If $X$ and $Y$ are vector fields on $\mathbb{R}^{3}$ that are linearly independent at each point, show that

$$
U_{1}=\frac{X}{\|X\|}, \quad U_{2}=\frac{\tilde{Y}}{\|\widetilde{Y}\|}, \quad U_{3}=U_{1} \times U_{2}
$$

is a frame field, where $\tilde{Y}=Y-\left(Y \bullet U_{1}\right) U_{1}$.

Let $\left(U_{1}, U_{2}, U_{3}\right)$ be a frame field on $\mathbb{R}^{3}$. If $X$ is a vector field on $\mathbb{R}^{3}$, then

$$
X=f_{1} U_{1}+f_{2} U_{2}+f_{3} U_{3}
$$

where the (differentible) functions $f_{i}=X \bullet U_{i}$ are called the coordinate functions of $X$ with respect to the frame $\left(U_{1}, U_{2}, U_{3}\right)$. If

$$
X=f_{1} U_{1}+f_{2} U_{2}+f_{3} U_{3} \quad \text { and } \quad Y=g_{1} U_{1}+g_{2} U_{2}+g_{3} U_{3}
$$

are two vector fields on $\mathbb{R}^{3}$, then

$$
X \bullet Y=f_{1} g_{1}+f_{2} g_{2}+f_{3} g_{3}
$$

In particular,

$$
\|X\|=\sqrt{f_{1}^{2}+f_{2}^{2}+f_{3}^{2}} .
$$

Note : A given vector field $X$ has a different set of coordinates functions with respect to each choice of a frame field $\left(U_{1}, U_{2}, U_{3}\right)$. The Euclidean coordinate functions, of course, come from the natural frame field ( $E_{1}, E_{2}, E_{3}$ ). In studying curves in $\mathbb{R}^{3}$ we shall be able to choose a frame field specifically adapted to the problem at hand. Not only does this simplify computations, but it gives a clearer understanding of geometry than if we had insisted on using the same frame field in every situation.

## Orientation

Let $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ be a frame at a point $p \in \mathbb{R}^{3}$. Recall that associated with each frame $\underline{u}$ is its attitude matrix $U$.

Note : $u_{1} \bullet u_{2} \times u_{3}=\operatorname{det} U= \pm 1$.
We make the following definition.
2.1.11 Definition. The frame $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ is said to be

- positively-oriented (or right-handed) provided $u_{1} \bullet u_{2} \times u_{3}=+1$;
- negatively-oriented (or left-handed) provided $u_{1} \bullet u_{2} \times u_{3}=-1$.

At each point $p$ of $\mathbb{R}^{3}$, the natural frame $\left(e_{1}, e_{2}, e_{3}\right)$ is positively-oriented.
$\diamond$ Exercise 68 Show that a frame $\left(u_{1}, u_{2}, u_{3}\right)$ is positively-oriented if and only if $u_{1} \times u_{2}=u_{3}$. Thus the orientation of a frame can be determined, for practical purposes, by the "right-hand rule".

We know that the tangent mapping of an isometry carries frames to frames. The following result tells what happens to their orientations.
2.1.12 Proposition. If $\underline{u}=\left(u_{1}, u_{2}, u_{3}\right)$ is a frame at a point $p \in \mathbb{R}^{3}$ and $F$ is an isometry on $\mathbb{R}^{3}$, then

$$
F_{*}\left(u_{1}\right) \bullet F_{*}\left(u_{2}\right) \times F_{*}\left(u_{3}\right)=(\operatorname{sgn} F) u_{1} \bullet u_{2} \times u_{3} .
$$

Proof: If

$$
u_{j}=u_{1 j} E_{1}(p)+u_{2 j} E_{2}(p)+u_{3 j} E_{3}(p), \quad j=1,2,3
$$

then we have

$$
F_{*}\left(u_{j}\right)=\sum_{i, k=1}^{3} a_{i k} u_{k j} E_{i}(F(p)),
$$

where $A=\left[a_{i k}\right]$ is the orthogonal part of $F$. Thus the attitude matrix of the frame $F_{*}(\underline{u})=\left(F_{*}\left(u_{1}\right), F_{*}\left(u_{2}\right), F_{*}\left(u_{3}\right)\right)$ is the matrix

$$
\left[\sum_{k=1}^{3} a_{i k} u_{k j}\right]=A U .
$$

But the triple scalar product of a frame is the determinant of its attitude matrix, and hence

$$
\begin{aligned}
F_{*}\left(u_{1}\right) \bullet F_{*}\left(u_{2}\right) \times F_{*}\left(u_{3}\right) & =\operatorname{det}(A U) \\
& =\operatorname{det} A \cdot \operatorname{det} U \\
& =(\operatorname{sgn} F) u_{1} \bullet u_{2} \times u_{3} .
\end{aligned}
$$

This result shows that if the isometry $F$ is direct (i.e., $\operatorname{sgn} F=+1$ ), then $F_{*}$ carries positively-oriented frames to positively-oriented frames and carries negatively-oriented frames to negatively-oriented frames. On the other hand, if the isometry $F$ is opposite (i.e., $\operatorname{sgn} F=-1$ ), then positive goes to negative and negative to positive.

Note : Direct isometries preserve orientation (of frames), and opposite isometries reverse it. For this very reason, direct isometries are also called orientation-preserving isometries, whereas opposite isometries are called orientation-reversing isometries.

Both dot and cross product were originally defined in terms of Euclidean coordinates. It is easy to see that the dot product is given by the same formula, no matter what frame $\left(u_{1}, u_{2}, u_{3}\right)$ is used to get coordinates. Indeed, we have

$$
\begin{aligned}
v \bullet w & =\left(v_{1} u_{1}+v_{2} u_{2}+v_{3} u_{3}\right) \bullet\left(w_{1} u_{1}+w_{2} u_{2}+w_{3} u_{3}\right) \\
& =\sum_{i, j=1}^{3}\left(v_{i} w_{j}\right) u_{i} \bullet u_{j} \\
& =\sum_{i, j=1}^{3} \delta_{i j} v_{i} w_{j} \\
& =v_{1} w_{1}+v_{2} w_{2}+v_{3} w_{3} .
\end{aligned}
$$

Almost the same result holds for cross products, but orientation is now involved.
$\diamond$ Exercise 69 Let $\left(u_{1}, u_{2}, u_{3}\right)$ be a frame at a point $p \in \mathbb{R}^{3}$. If $v=\sum v_{i} u_{i}$ and $w=\sum w_{i} u_{i}$, show that

$$
v \times w=\epsilon\left|\begin{array}{ccc}
e_{1} & e_{2} & e_{3} \\
v_{1} & v_{2} & v_{3} \\
w_{1} & w_{2} & w_{3}
\end{array}\right|,
$$

where $\epsilon=u_{1} \bullet u_{2} \times u_{3}= \pm 1$.
It follows immediately that the effect of an isometry on cross products also involves orientation. Explicitly, if $v$ and $w$ are tangent vectors to $\mathbb{R}^{3}$ at $p$, and $F$ is an isometry on $\mathbb{R}^{3}$, then

$$
F_{*}(v \times w)=(\operatorname{sgn} F) F_{*}(v) \times F_{*}(w)
$$

### 2.2 Directional Derivatives

Associated with each tangent vector $v_{p}$ to $\mathbb{R}^{3}$ is the line $t \mapsto p+t v$. If $f$ is a differentiable function on $\mathbb{R}^{3}$, then

$$
t \mapsto f(p+t v)
$$

is an ordinary differentiable function $\mathbb{R} \rightarrow \mathbb{R}$.
Note : The derivative of this function at $t=0$ tells the initial rate of change of $f$ as $p$ moves in the $v$ direction.

We make the following definition.
2.2.1 Definition. Given a differentiable function $f: \mathbb{R}^{3} \rightarrow \mathbb{R}$ and a tangent vector $v_{p} \in T_{p} \mathbb{R}^{3}$, the number

$$
v_{p}[f]:=\left.\frac{d}{d t} f(p+t v)\right|_{t=0}
$$

is called the directional derivative of $f$ with respect to $v_{p}$.

Note : This definition appears in elementary calculus with the additional restriction that $v_{p}$ be a unit vector. Even though we do not impose this restriction, we shall nevertheless refer to $v_{p}[f]$ as a directional derivative.
2.2.2 Example. We compute (the directional derivative) $v_{p}[f]$ for the function $f\left(x_{1}, x_{2}, x_{3}\right)=x_{1}^{2} x_{2} x_{3}$ with $p=(1,1,0)$ and $v=(1,0,-3)$. Then

$$
p+t v=(1,1,1)+t(1,0,-3)=(1+t, 1,-3 t)
$$

describes the line through $p$ in the $v$ direction. Evaluating $f$ along this line, we get

$$
f(p+t v)=(1+t)^{2} \cdot 1 \cdot(-3 t)=-3 t-6 t^{2}-3 t^{3}
$$

Now

$$
\frac{d}{d t} f(p+t v)=-3-12 t-9 t^{2}
$$

and hence, at $t=0$, we find $v_{p}[f]=-3$. Thus, in particular, the function $f$ is initially decreasing as $p$ moves in the $v$ direction.
$\diamond$ Exercise 70 Compute the directional derivative of the function $f\left(x_{1}, x_{2}, x_{3}\right)=$ $x_{1} x_{2}+x_{3}$ with respect to $v_{p}=(1,-4,2)_{p}$, where $p=(1,1,0)$.

The following result shows how to compute $v_{p}[f]$ in general, in terms of the partial derivatives of $f$ at the point $p$.
2.2.3 Proposition. If $v_{p}=\left(v_{1}, v_{2}, v_{3}\right)_{p}$ is a tangent vector to $\mathbb{R}^{3}$, then

$$
v_{p}[f]=v_{1} \frac{\partial f}{\partial x_{1}}(p)+v_{2} \frac{\partial f}{\partial x_{2}}(p)+v_{3} \frac{\partial f}{\partial x_{3}}(p) .
$$

Proof: Let $p=\left(p_{1}, p_{2}, p_{3}\right)$. Then

$$
p+t v=\left(p_{1}+t v_{1}, p_{2}+t v_{2}, p_{3}+t v_{3}\right)
$$

We use the chain rule to compute the derivative at $t=0$ of the function

$$
f(p+t v)=f\left(p_{1}+t v_{1}, p_{2}+t v_{2}, p_{3}+t v_{3}\right) .
$$

We obtain

$$
\begin{aligned}
v_{p}[f] & =\left.\frac{d}{d t} f(p+t v)\right|_{t=0} \\
& =\sum_{i=1}^{3} \frac{d}{d t}\left(p_{i}+t v_{i}\right) \frac{\partial f}{\partial x_{i}}(p) \\
& =v_{1} \frac{\partial f}{\partial x_{1}}(p)+v_{2} \frac{\partial f}{\partial x_{2}}(p)+v_{3} \frac{\partial f}{\partial x_{3}}(p) .
\end{aligned}
$$

Note: We can write the directional derivative $v_{p}[f]$ in matrix form:

$$
v_{p}[f]=\left[\begin{array}{lll}
\frac{\partial f}{\partial x_{1}}(p) & \frac{\partial f}{\partial x_{2}}(p) & \frac{\partial f}{\partial x_{3}}(p)
\end{array}\right]\left[\begin{array}{l}
v_{1} \\
v_{2} \\
v_{3}
\end{array}\right] .
$$

The main properties of this notion of derivative are as follows.
$\diamond$ Exercise 71 Let $f$ and $g$ be differentiable functions on $\mathbb{R}^{3}, v_{p}$ and $w_{p}$ tangent vectors at a point $p \in \mathbb{R}^{3}$, and $\lambda$ and $\mu$ real numbers. Show that :
(a) $\left(\lambda v_{p}+\mu w_{p}\right)[f]=\lambda v_{p}[f]+\mu w_{p}[f]$.
(b) $v_{p}[\lambda f+\mu g]=\lambda v_{p}[f]+\mu v_{p}[f]$.
(c) $v_{p}[f g]=v_{p}[f] g(p)+f(p) v_{p}[g]$.

The first two properties may be summarized by saying that (the mapping) $\left(v_{p}, f\right) \mapsto v_{p}[f]$ is linear in $v_{p}$ and $f$. The third property is essentially just the usual Leibniz rule for differentiation of a product.

Note : No matter what form differentiation may take, it will always have suitable linear and Leibnizian properties.
$\diamond$ Exercise 72 Given two tangent vectors $v_{p}, w_{p}$ to $\mathbb{R}^{3}$, show that if

$$
v_{p}[f]=w_{p}[f]
$$

for every differentiable function $f$ on $\mathbb{R}^{3}$, then $v_{p}=w_{p}$.

We now use the pointwise principle to define the operation of a vector field on a function. Let $X$ be a vector field and $f$ a differentiable function on $\mathbb{R}^{3}$. Then we define the function $X[f]$ (or simply $X f$ ) by

$$
X[f](p):=X(p)[f] .
$$

That is, the value of $X[f]$ at the point $p$ is the directional derivative of $f$ with respect to the tangent vector $X(p)$ at $p$.
$\diamond$ Exercise 73 If $X$ and $Y$ are vector fields, and $f, g, h$ are differentiable functions on $\mathbb{R}^{3}$, then show that (for $\lambda, \mu \in \mathbb{R}$ ):
(a) $(f X+g Y)[h]=f X[h]+g Y[h]$.
(b) $X[\lambda f+\mu g]=\lambda X[f]+\mu X[g]$.
(c) $X[f g]=X[f] \cdot g+f \cdot X[g]$.

In particular, if $\left(E_{1}, E_{2}, E_{3}\right)$ is the natural frame field on $\mathbb{R}^{3}$, then

$$
E_{i}[f]=\frac{\partial f}{\partial x_{i}} \quad(i=1,2,3) .
$$

This is an immediate consequence of Proposition 2.2.3. For example, $E_{1}(p)=(1,0,0)_{p}$ and hence (for all points $\left.p=\left(p_{1}, p_{2}, p_{3}\right)\right)$

$$
E_{1}(p)[f]=\left.\frac{d}{d t} f\left(p_{1}+t, p_{2}, p_{3}\right)\right|_{t=0}=\frac{\partial f}{\partial x_{1}}(p) .
$$

If $X=\left(X_{1}, X_{2}, X_{3}\right)$ is a vector field on $\mathbb{R}^{3}$ we can write

$$
\begin{aligned}
X & =X_{1} E_{1}+X_{2} E_{2}+X_{3} E_{3} \\
& =X_{1} \frac{\partial}{\partial x_{1}}+X_{2} \frac{\partial}{\partial x_{2}}+X_{3} \frac{\partial}{\partial x_{3}}
\end{aligned}
$$

This notation makes it a simple matter to carry out explicit computations.
2.2.4 Example. For (the vector field)

$$
X=x_{1} \frac{\partial}{\partial x_{1}}-x_{2}^{2} \frac{\partial}{\partial x_{3}}
$$

and (the differentiable function) $f=x_{1}^{2} x_{2}+x_{3}^{3}$ we compute

$$
\begin{aligned}
X[f] & =x_{1} \frac{\partial}{\partial x_{1}}\left(x_{1}^{2} x_{2}+x_{3}^{3}\right)-x_{2}^{2} \frac{\partial}{\partial x_{3}}\left(x_{1}^{2} x_{2}+x_{3}^{3}\right) \\
& =x_{1}\left(2 x_{1} x_{2}\right)-x_{2}^{2}\left(3 x_{3}^{2}\right) \\
& =2 x_{1}^{2} x_{2}-3 x_{2}^{2} x_{3}^{2}
\end{aligned}
$$

$\diamond$ Exercise 74 Given a vector field $X$, show that

$$
X=X\left[x_{1}\right] \frac{\partial}{\partial x_{1}}+X\left[x_{2}\right] \frac{\partial}{\partial x_{2}}+X\left[x_{3}\right] \frac{\partial}{\partial x_{3}}
$$

where $x \mapsto x_{i}, i=1,2,3$ are the natural coordinate functions.
$\diamond$ Exercise 75 Let

$$
\begin{aligned}
X & =\frac{\partial}{\partial x_{1}}+\frac{\partial}{\partial x_{2}} \\
Y & =x_{1} \frac{\partial}{\partial x_{1}}-\frac{\partial}{\partial x_{3}} \\
f & =x_{1} x_{2} .
\end{aligned}
$$

Compute $X[f], Y[f], X\left[f^{2}\right], X[X[f]]$, and $X[Y[f]]-Y[X[f]]$.

### 2.3 Curves in Euclidean 3-Space $\mathbb{R}^{3}$

## Parametrized curves

We want to characterize certain subsets of Euclidean 3 -space $\mathbb{R}^{3}$ (to be called curves) that are, in a certain sense, one-dimensional and to which the methods of calculus can be applied.

Note : There are various notions of a curve in $\mathbb{R}^{3}$. We shall deal here with only one such notion. The definition is not entirely satisfactory but sufficient for our purposes.

A convenient way of defining such subsets is through differentiable maps. Let $J$ be an interval on the real line (the interval may be open or closed, finite, semi-infinite or the entire real line). One can picture a curve in $\mathbb{R}^{3}$ as a trip taken by a moving particle $\alpha$. At each "time" $t, \alpha$ is located at the point $\alpha(t)=\left(\alpha_{1}(t), \alpha_{2}(t), \alpha_{3}(t)\right)$ in $\mathbb{R}^{3}$. We make the following definition.
2.3.1 Definition. A (parametrized) curve in $\mathbb{R}^{3}$ is a smooth map

$$
\alpha: J \rightarrow \mathbb{R}^{3}, \quad t \mapsto\left(\alpha_{1}(t), \alpha_{2}(t), \alpha_{3}(t)\right) .
$$

The curve is regular if $\dot{\alpha}(t)=\frac{d \alpha}{d t}(t) \neq 0$ for all $t \in J$.
The functions $\alpha_{1}, \alpha_{2}, \alpha_{3}$ are called the Euclidean coordinate functions and $t$ is called the parameter of $\alpha$. We write $\alpha=\left(\alpha_{1}, \alpha_{2}, \alpha_{3}\right)$.

Note : More generally, we can speak of differentiability of order $k$ (or class $C^{k}$ ). One then requires the appropriate order of differentiability in each definition and theorem. To focus more on the geometry than the analysis we have ignored this subtlety by assuming curves to be smooth (i.e., of class $C^{\infty}$ ).

The image set $\alpha(J) \subset \mathbb{R}^{3}$ is called the trace of $\alpha$, which is the geometric object of interest. One should carefully distinguish a parametrized curve, which is a map, from its trace, which is a subset of $\mathbb{R}^{3}$.

Note : A given trace may be the image set (or route) of many (parametrized) curves. In this setting, it may be appropriate to call the common trace a geometric curve and refer to the curves as parametrizations (or parametric representations).
2.3.2 Example. (The line) A (straight) line is the simplest type of geometric curve in $\mathbb{R}^{3}$. We know that two points determine a line. For two points $p, q \in \mathbb{R}^{3}$, the line $\overleftrightarrow{p q}$ may be described as follows. To attain the line, add
the vector $p$. To travel along the line, use the direction vector $q-p$ since this is the direction from $p$ to $q$. A parameter $t$ tells exactly how far along $q-p$ to go. Putting these steps together produces

$$
\alpha: \mathbb{R} \rightarrow \mathbb{R}^{3}, \quad t \mapsto p+t(q-p), \quad q \neq p
$$

which gives a parametrization of the line through the points $p$ and $q$ (or, if one prefers, the line through the point $p$ with direction vector $q-p$ ).
$\diamond$ Exercise 76 Find a parametrization of the line through the points $(-1,0,5)$ and $(3,-1,-2)$.
2.3.3 EXAMPLE. (The circle) The circle of radius $a$ with centre $p=$ $\left(p_{1}, p_{2}, 0\right) \in \mathbb{R}^{2}$ is the set (locus) of points $x$ in the plane $\mathbb{R}^{2}$ (i.e., the $x_{1} x_{2^{-}}$ plane of $\mathbb{R}^{3}$ ) such that

$$
\|x-p\|=a
$$

(the distance between $x$ and $p$ is the fixed positive real number $a$ ). A natural parametric representation is

$$
\alpha: \mathbb{R} \rightarrow \mathbb{R}^{3}, \quad t \mapsto\left(p_{1}+a \cos t, p_{2}+a \sin t, 0\right)
$$

$\diamond$ Exercise 77 Find a curve $\alpha: \mathbb{R} \rightarrow \mathbb{R}^{3}$ whose trace is the unit circle $x_{1}^{2}+x_{2}^{2}=$ $1, x_{3}=0$ and such that $\alpha(t)$ runs clockwise around the circle with $\alpha(0)=(0,1)$.
2.3.4 Example. (The helix) A (circular) helix is a geometric curve represented (given) parametrically by

$$
\alpha: \mathbb{R} \rightarrow \mathbb{R}^{3}, \quad t \mapsto(a \cos t, a \sin t, b t) ; \quad a>0, b \neq 0
$$

It rises (when $b>0$ ) or falls (when $b<0$ ) at a constant rate on the (circular) cylinder $x_{1}^{2}+x_{2}^{2}=a^{2}$. The $x_{3}$-axis is called the axis, and $2 \pi b$ the pitch of the helix.
2.3.5 Example. The curve

$$
\alpha: \mathbb{R} \rightarrow \mathbb{R}^{3}, \quad t \mapsto\left(t^{3}, t^{2}, 0\right)
$$

is not regular because $\dot{\alpha}(0)=(0,0,0)$. The trace has a cusp at the origin.
2.3.6 Example. The curve

$$
\alpha: \mathbb{R} \rightarrow \mathbb{R}^{3}, \quad t \mapsto(\cosh t, \sinh t, t)
$$

is known as the hyperbolic helix. (Recall that the hyperbolic trigonometric functions are defined by the formulas

$$
\cosh t=\frac{e^{t}+e^{-t}}{2}, \quad \sinh t=\frac{e^{t}-e^{-t}}{2}, \quad \tanh t=\frac{e^{t}-e^{-t}}{e^{t}+e^{-t}}
$$

We have the fundamental identity $\cosh ^{2} t-\sinh ^{2} t=1$.)
If we visualize a (parametrized) curve $\alpha$ in $\mathbb{R}^{3}$ as a moving particle, then at every time $t$ there is a tangent vector at the point $\alpha(t)$ which gives the instantaneous velocity of $\alpha$ at that time.
2.3.7 Definition. Let $\alpha: J \rightarrow \mathbb{E}^{3}$ be a curve in $\mathbb{R}^{3}$ with $\alpha=\left(\alpha_{1}, \alpha_{2}, \alpha_{3}\right)$. For each number $t \in J$, the velocity vector of $\alpha$ at $t$ is the tangent vector

$$
\dot{\alpha}(t):=\left(\frac{d \alpha_{1}}{d t}(t), \frac{d \alpha_{2}}{d t}(t), \frac{d \alpha_{3}}{d t}(t)\right)_{\alpha(t)}
$$

at the point $\alpha(t) \in \mathbb{R}^{3}$.
If $\alpha$ is a regular curve, all its velocity vectors are different from zero. A regular curve can have no corners or cusps.
2.3.8 Example. The velocity vector of the (straight) line $\alpha(t)=p+t(q-$ $p$ ) is

$$
\dot{\alpha}(t)=\left(q_{1}-p_{1}, q_{2}-p_{2}, q_{3}-p_{3}\right)_{\alpha(t)}=(q-p)_{\alpha(t)} .
$$

The fact that $\alpha$ is "straight" is reflected in the fact that all its velocity vectors are parallel; only the point of application changes as $t$ changes.
2.3.9 Example. For the helix represented by $\alpha(t)=(a \cos t, a \sin t, b t)$, the velocity vector at $t$ is

$$
\dot{\alpha}(t)=(-a \sin t, a \cos t, b)_{\alpha(t)} .
$$

The fact that the helix "rises" constantly is shown by the constancy of the $x_{3}$-coordinate of $\dot{\alpha}(t)$.

Note : The line, the circle, the ellipse, and the helix (circular or hyperbolic) are all regular curves.
$\diamond$ Exercise 78 For a fixed $t$, the tangent line to a regular curve $\alpha: J \rightarrow \mathbb{R}^{3}$ at the point $\alpha(t)$ is the line $u \mapsto \alpha(t)+u \dot{\alpha}(t)$. Find the tangent line to the helix

$$
\alpha(t)=(2 \cos t, 2 \sin t, t)
$$

at the points $p=\alpha(0)$ and $q=\alpha\left(\frac{\pi}{4}\right)$.
$\diamond$ Exercise 79 Find the curve $\alpha: \mathbb{R} \rightarrow \mathbb{R}^{3}$ such that

$$
\alpha(0)=(2,3,0) \quad \text { and } \quad \dot{\alpha}(t)=\left(e^{t},-2 t, t^{2}\right) .
$$

Let $\alpha: J \rightarrow \mathbb{R}^{3}$ be a curve. The norm of the velocity vector $\dot{\alpha}(t)$ of $\alpha$ at $t$

$$
\|\dot{\alpha}(t)\|:=\sqrt{\dot{\alpha}(t) \bullet \dot{\alpha}(t)}=\sqrt{\left(\frac{d \alpha_{1}}{d t}(t)\right)^{2}+\left(\frac{d \alpha_{2}}{d t}(t)\right)^{2}+\left(\frac{d \alpha_{3}}{d t}(t)\right)^{2}}
$$

is called the speed of $\alpha$ at $t$. Again, thinking of $\alpha$ as the path of a moving particle and $t$ as time, we see that the length of the velocity vector is precisely the speed of the particle at the given time.

Note : A regular curve has speed always greater than zero.
$\diamond$ Exercise 80 If $\alpha: J \rightarrow \mathbb{R}^{3}$ is a curve, its acceleration vector at $t$ is given by

$$
\ddot{\alpha}(t):=\left(\frac{d^{2} \alpha_{1}}{d t^{2}}(t), \frac{d^{2} \alpha_{2}}{d t^{2}}(t), \frac{d^{2} \alpha_{3}}{d t^{2}}(t)\right)_{\alpha(t)} .
$$

What can be said about $\alpha$ if its acceleration is identically zero?
$\diamond$ Exercise 81 Verify that the curve $\alpha(t)=(\cos t, \sin t, 1)$ has constant speed, but nonzero acceleration.
$\diamond$ Exercise 82 For the curve $\alpha(t)=\left(2 t, t^{2}, \frac{t^{3}}{3}\right)$, find the velocity, speed, and acceleration for arbitrary $t$, and at $t=1$.
$\diamond$ Exercise 83 Show that (the trace of) the curve $\alpha(t)=(t \cos t, t \sin t, t)$ lies on a cone in $\mathbb{R}^{3}$. Find the velocity, speed, and acceleration of $\alpha$ at the vertex of the cone.

## Other examples of curves

The following plane parametrized curves arise naturally throughout the physical sciences and mathematics.
2.3.10 Example. (The catenary) Let $f: J \rightarrow \mathbb{R}$ be any smooth function. The graph of $f$ is the set of all points $(t, f(t)) \in \mathbb{R}^{2}$ with $t \in J$, so is the trace of the (regular) curve

$$
\alpha: J \rightarrow \mathbb{R}^{2}, \quad t \mapsto(t, f(t))
$$

In particular, for $f(t)=a \cosh \frac{t}{a}$, we get the catenary (from the Latin for "chain").

Note : The catenary is of historical interest, representing the form (shape) adopted by a perfect inextensible chain of uniform density suspended by its ends and acted upon by gravity. It was studied first by Galileo Galilei (1564-1642), who mistook it for a parabola, and later by Gottfried Leibniz (1646-1716), Christiaan Huygens (1629-1695), and Johann Bernoulli (1667-1748). (They were responding to the challenge put out by Jakob (Jacques) Bernoulli (1654-1705) to find the equation of the "chain-curve".) It is also of contemporary mathematics interest, being a plane section of the minimal surface (a soap film catenoid) spanning two circular discs, the only minimal surface of revolution.
2.3.11 Example. (The cycloid) Suppose a circle of radius $a$ sits on the $x_{1^{-}}$ axis making contact at the origin. Let the circle to roll (without slipping) along the positive $x_{1}$-axis. The figure (path) described by the point on the circle, originally in contact with the $x_{1}$-axis, is a geometric curve called cycloid. It can be shown that a parametric reprezentation of the cycloid is

$$
\alpha: \mathbb{R} \rightarrow \mathbb{E}^{2}, \quad t \mapsto(a(t-\sin t), a(1-\cos t))
$$

where (the parameter) $t$ is the angle formed by the (new) point of contact with the axis, the centre of the circle, and the original point of contact.
$\diamond$ Exercise 84 Draw a picture (i.e., sketch the graph) of the cycloid.

We can see that the cycloid has infinitely many cusps (corresponding to $t=$ $2 k \pi, k \in \mathbb{Z}$ ) : the arc of the cycloid between any two consecutive cusps is called an arch. Generally, when a curve rolls (without slipping) along another fixed curve, any point which moves with the moving curve describes a curve, called a roulette (from the French for "small wheel"). Consider now the roulette of a tracing point carried by a (moving) circle of radius a rolling along a line (the $x_{1}$-axis, say). It is assumed that in the initial configuration the (moving) circle is tangent to the $x_{1}$-axis at the origin and that the tracing point is the point on the $x_{2}$-axis distance $h a$ from the centre of the circle. The resulting roulette, also known as a cycloid, has the parametrization

$$
x_{1}(t)=a(t-h \sin t), \quad x_{2}(t)=a(1-h \cos t) .
$$

The form of the cycloid depends on whether the tracing point is inside ( $h<1$ ), on $(h=1)$ or outside $(h>1)$ the moving circle. For $h<1$ we obtain a "shortened" cycloid reminiscent of the sine curve. For $h>1$ we obtain an "extended" cycloid with infinitely many self crossings. Finally, for $h=1$ we get the (standard or "cuspidal") cycloid as introduced above.

Note : The cycloid has two additional names and a lot of interesting history. The other two names are the tautochrone and the brachistochrone. Christiaan Huygens (1629-1695) discovered a remarkable property of the cycloid : it is the only curve such that a body falling under its own weight is guided by this curve so as to oscillate with a period that is independent of the initial point where the body is released. Therefore, he called this curve (i.e. the cycloid) the tautochrone (from the Greek for "same time"; $\tau \alpha v \tau o ́ \varsigma: ~ e q u a l, ~ a n d ~ \chi \rho o ́ v o \varsigma: ~ t i m e) . ~$

In 1696 Johann Bernoulli (1667-1784) posed a question (problem) and invited his fellow mathematicians to solve it. The problem (which he had solved and which he considered very beautiful and very difficult), called the brachistochrone problem, is the following : Given two points $p$ and $q$ in a vertical plane (with $q$ below and to the right of $p$ ) find, among all (smooth) curves with endpoints $p$ and $q$, the curve such that a particle which slides without friction along the curve under the influence of gravity will travel from the one point to the other in the least possible time. (Johann Bernoulli solved the problem ingeniously by employing Fermat's principle that light travels to minimize time together with Snell's law of refraction. The other solvers included Johann's brother, Jakob Bernoulli, as well as Gottfried

Leibniz (1646-1716), Isaac Newton (1642-1727), and L'Hôpital (1661-1704).) This problem is important because it led to the systematic consideration of similar problems; the new discipline which developed thereby is called the calculus of variations. Moreover, Bernoulli's problem is a true minimum time problem of the kind that is studied today in optimal control theory. Bernoulli called the "fastest path" the brachistochrone (from the Greek for "shortest time"; $\beta \rho \alpha ́ \chi \iota \sigma \tau o \varsigma: ~ s h o r t e s t, ~ a n d$ $\chi$ оо́vos: time).
2.3.12 EXAMPLE. (The astroid) A parametric reprezentation of the curve called the astroid is

$$
\alpha:[0,2 \pi] \rightarrow \mathbb{E}^{2}, \quad t \mapsto\left(a \cos ^{3} t, a \sin ^{3} t\right)
$$

The definition of the astroid is very similar to that of the cycloid. For the astroid, however, a circle is rolled (without slipping), not on a line, but inside another (fixed) circle. More precisely, let a circle of radius $\frac{a}{4}$ roll inside a large circle of radius $a$ (and centered at the origin say). For concreteness, suppose we start the little circle at $(a, 0)$ and follow the path of the point originally in contact with $(a, 0)$ as the circle rolls up. Let $t$ denote the angle from the centre of the large circle to the new contact point. One can show that, with respect to the origin, the rolling point moves to

$$
\alpha(t)=\left(\frac{3 a}{4} \cos t+\frac{a}{4} \cos 3 t, \frac{3 a}{4} \sin t-\frac{a}{4} \sin 3 t\right) .
$$

$\diamond$ Exercise 85 Show that the formula for the astroid may be reduced to

$$
\alpha(t)=\left(a \cos ^{3} t, a \sin ^{3} t\right)
$$

with implicit form

$$
x_{1}^{\frac{2}{3}}+x_{2}^{\frac{2}{3}}=a^{\frac{2}{3}} .
$$

Note : Recall that when one curve rolls along another fixed curve, any point which moves with the moving curve describes a curve, called a roulette. When the curves are circles the resulting roulette is called a trochoid. Trochoids occur naturally in the physical sciences. Assume that the fixed circle has centre $o$ at the origin and radius $a>0$, and that the moving circle has centre $o^{\prime}$ and "radius" $a^{\prime} \neq 0$. The case $a^{\prime}>0$ is interpreted as the moving circle rolling on the outside of the fixed one
(an epitrochoid), while $a^{\prime}<0$ is interpreted as the moving circle rolling on the inside of it (an hypotrochoid). Suppose the tracing point is distant $a^{\prime} h$ from the centre $a^{\prime}$ of the moving circle. Write $t$ for the angle between the line $o o^{\prime}$ and the $x_{1}$-axis, and assume (without loss of generality) that when $t=0$ the tracing point lies on the $x_{1}$-axis. It can be shown that the following parametric reprezentation results

$$
t \mapsto((\lambda+1) \cos t-h \cos (\lambda+1) t,(\lambda+1) \sin t-h \sin (\lambda+1) t)
$$

where $\lambda:=\frac{a}{a^{\prime}} \neq 0$ and $h>0$. The case $h=1$ (i.e. when the tracing point lies on the moving circle) is of special significance. Various names have been assigned traditionally to the curves (trochoids) arising by taking certain values of $\lambda$; for example, for $\lambda=1$ : the cardioid (the "heart shaped" curve), for $\lambda=-3$ : the deltoid, or for $\lambda=-4$ : the astroid. Ellipses are special case of trochoids. (Consider the special case when $\lambda=-2$ : the moving circle rolls inside the fixed circle, and has half the radius. For $0<h<1$ this is an ellipse; for $h=0$ the ellipse becomes a circle, concentric with the fixed circle, and of half its radius.)

Another special case is obtained when $\lambda=1$ : the moving circle rolls outside the fixed circle, and has the same radius. The (epi)trochoid is then a limancon with parametric reprezentation

$$
\left(x_{1}(t), x_{2}(t)\right)=(2 \cos t-h \cos 2 t, 2 \sin t-h \sin 2 t) .
$$

The form of the limancon depends on the value of $h$. (When $h=1$ we get the cardioid.)
2.3.13 Example. (The tractrix) The trace of the parametrized curve

$$
\alpha: \mathbb{R} \rightarrow \mathbb{E}^{2}, \quad t \mapsto\left(t-\tanh t, \frac{1}{\cosh t}\right)
$$

is called the tractrix.
$\diamond$ Exercise 86 Show that the tractrix has the following remarkable property : the length of the line segment of the tangent of the tractrix between the point of tangency and the $x_{1}$-axis is constantly equal to 1 .

There is another way of saying this : the circle of unit radius centered at the point $(t, 0)$ passes through the point $x(t)$ (on the tractrix), and the tangent line to the circle at $x(t)$ is orthogonal to the tangent line to the tractrix at that point. Thus the tractrix has the property that it meets all circles of unit
radius centered on the $x_{1}$-axis orthogonally. (For that reason the tractrix is described as an "orthogonal trajectory" of that family of circles.)

Note : The tractrix gives rise to an interesting example in the elementary geometry of surfaces : the surface of revolution obtained by rotating it about the $x_{1}$-axis is the pseudosphere, distinguished by the property of having constant negative (Gaussian) curvature. (Intuitively, the curvature of a surface is a number $\kappa$ that measures the extent to which the surface "bends". In general, the curvature $\kappa$ varies from point to point, being close to zero at points where the surface is rather flat, large at points where the surface bends sharply. For some surfaces the curvature is the same at all points, so naturally these are called surfaces of constant curvature к.)
2.3.14 Example. (The standard conics) A general conic is a set of points defined by the vanishing of a polynomial of degree two in two variables :

$$
A x_{1}^{2}+2 B x_{1} x_{2}+C x_{2}^{2}+2 D x_{1}+2 E x_{2}+F=0,
$$

where $A, B, C, D, E, F \in \mathbb{R}$ and not all of $A, B$, and $C$ are zero. A class of conics arises from the following classical construction. One is given a line $\mathcal{D}$ (the directrix), a point $f$ (the focus) not on $\mathcal{D}$, and a variable point $p$ subject to the constraint that its distance from $f$ is proportional to its distance from $\mathcal{L}$. Write $p^{\prime}$ for the (orthogonal) projection of $p$ onto $\mathcal{D}$. Then the constraint reads

$$
\|p-f\|=\epsilon\left\|p-p^{\prime}\right\|
$$

for some positive constant of proportionality $\epsilon$, known as the eccentricity. (The line through $f$ and its projection $f^{\prime}$ onto $\mathcal{D}$ is an axis of symmetry.) The locus of $p$ is a parabola when $\epsilon=1$, an ellipse when $\epsilon<1$, or a hyperbola when $\epsilon>1$. The "standard conics" arise from the special case when $\mathcal{D}$ is parallel to one of the coordinate axes (the $x_{2}$-axis say), and the focus lies on the other coordinate axis (the $x_{1}$-axis say). Then we get an equation of the form

$$
\left(1-\epsilon^{2}\right) x_{1}^{2}+2 \beta x_{1}+x_{2}^{2}+\gamma=0 .
$$

(Circles cannot be constructed in this way since the eccentricity is positive.) Convenient forms for the equations of the three "standard conics" can be now obtained easily.
(i) Consider first the case $\epsilon=1$ of a parabola. The equation of the conic reduces to that of a standard parabola

$$
x_{2}^{2}=4 a x_{1}
$$

with directrix the line $x_{1}=-a$ and focus the point $f=(a, 0)$. The $x_{1}$-axis is the axis of symmetry of the parabola, and the point where it meets the parabola (in this case, the origin) is the vertex. A standard parametrization of this parabola is

$$
x_{1}=a t^{2}, \quad x_{2}=2 a t
$$

(ii) Consider next the case $\epsilon<1$ of an ellipse. The equation of the conic reduces to that of a standard ellipse

$$
\frac{x_{1}^{2}}{a^{2}}+\frac{x_{2}^{2}}{b^{2}}=1
$$

where $b^{2}=a^{2}\left(1-\epsilon^{2}\right)$ and $0<b<a$. The coordinate axes are axes of symmetry of a standard ellipse. The points $(0, \pm b),( \pm a, 0)$ where the axes meet the ellipse provide the four vertices. It is traditional to refer to $a$ as the major semiaxis and $b$ as the minor semiaxis. The directrix $\mathcal{D}^{-}$is the line $x_{1}=-\frac{a}{\epsilon}$, and the focus is the point $f^{-}=(-a \epsilon, 0)$. The symmetry of the equation shows that there is a second directrix $\mathcal{D}^{+}$ with equation $x_{1}=\frac{a}{\epsilon}$ having a corresponding focus $f^{+}=(a \epsilon, 0)$. The centre of a standard ellipse is the mid-point of the line segment joining the two foci (in this case, the origin).

Note : Despite the fact that the circle does not appear as a standard conic, it is profitable to think of a circle (centered at the origin) as the limiting case of standard ellipses as $b \rightarrow a$ (which corresponds to $\epsilon \rightarrow 0$ ).

A standard parametrization of this ellipse is

$$
\left(x_{1}(t), x_{2}(t)\right)=(a \cos t, b \sin t)
$$

(iii) Finally, consider the case $\epsilon>1$ of a hyperbola. The equation of the conic reduces to that of a standard hyperbola

$$
\frac{x_{1}^{2}}{a^{2}}-\frac{x_{2}^{2}}{b^{2}}=1
$$

where $b^{2}=a^{2}\left(\epsilon^{2}-1\right)$ and $0<a, b$. The coordinate axes are axes of symmetry of a standard hyperbola. Only the $x_{1}$-axis meets the hyperbola, at the vertices $( \pm a, 0)$. Again we have directrix lines $x_{1}=-\frac{a}{\epsilon}, x_{1}=\frac{a}{\epsilon}$ with corresponding foci ( $\pm a \epsilon, 0$ ). The centre of a standard hyperbola is the mid-point of the line segment joining the two foci (i.e. the origin). The lines $x_{2}= \pm \frac{b}{a} x$ are the asymptotes of the hyperbola. (The asymptotes are orthogonal if and only if $a=b$; this corresponds to the case when the eccentricity $\epsilon=\sqrt{2}$.)

NOTE : A point $p=\left(x_{1}, x_{2}\right)$ satisfying the equation of a standard hyperbola is subject only to the constraint that $x_{1} \geq a$ or $x_{1} \leq-a$. Thus the key feature of a (standard) hyperbola is that it splits into two "branches" : the positive branch (defined for $x_{1} \geq a$ ) and the negative branch (defined for $x_{1} \leq-a$ ).
$\diamond$ Exercise 87 Find parametrizations for each of the two branches of this hyperbola.

Note: A general conic

$$
A x_{1}^{2}+2 B x_{1} x_{2}+C x_{2}^{2}+D x_{1}+E x_{2}+F=0
$$

(where not all of $A, B$, and $C$ are zero) represents one of the following eight types of loci (geometric curves) : an ellipse, a hyperbola, a parabola, a pair of intersecting lines, a pair of parallel lines, a line "counted twice", a single point, or the empty set. Moreover, the cases that can occur are governed by the sign of the expression $A C-B^{2}$ as follows :

- If $A C-B^{2}>0$, the possibilities are an ellipse, a single point, or the empty set.
- If $A C-B^{2}=0$, the possibilities are a parabola, two parallel lines, a single line, or the empty set.
- If $A C-B^{2}<0$, the possibilities are a hyperbola or two intersecting lines.
2.3.15 Example. (General algebraic curves)


## The shortest distance between two points

Let's consider the following question : What is the shortest distance between two points $p, q \in \mathbb{R}^{3}$ ? We have been taught since childhood that the answer is the (straight) line, but now we can see why our intuition is correct.

Let $\alpha: J \rightarrow \mathbb{R}^{3}$ be a curve. Let $\left[t_{0}, t_{1}\right] \subset J$ and set $\alpha\left(t_{0}\right)=p$ and $\alpha\left(t_{1}\right)=q$. We defined the speed of $\alpha$ at $t$ as the length of the velocity vector $\alpha^{\prime}(t)$. Thus speed is a real-valued (continuous) function on the interval $J$. In physics, the distance traveled by a moving particle is determined by integrating its speed with respect to time. Thus we define the arc length of $\alpha$, from $p$ to $q$, to be the number

$$
L(\alpha):=\int_{t_{0}}^{t_{1}}\|\dot{\alpha}(t)\| d t
$$

We are able to answer the question of which route between two given points gives the shortest distance.
2.3.16 Theorem. The line is the curve of least arc length between two points.

Proof : Consider two points $p, q \in \mathbb{R}^{3}$. The line segment between them may be parametrized by

$$
\lambda:[0,1] \rightarrow \mathbb{E}^{3}, \quad t \mapsto p+t(q-p)
$$

where $q-p$ gives the direction. Then

$$
\dot{\lambda}(t)=q-p \quad \text { and } \quad\|\dot{\lambda}(t)\|=\|q-p\| .
$$

Therefore

$$
L(\lambda)=\int_{0}^{1}\|\dot{\lambda}(t)\| d t=\|q-p\| \int_{0}^{1} d t=\|q-p\|
$$

and the length of the line segment (or direction vector) from $p$ to $q$ is the distance from $p$ to $q$ (as of course expected). Now we consider another curve
segment $\alpha:\left[t_{0}, t_{1}\right] \rightarrow \mathbb{R}^{3}$ which joins $p$ and $q$; that is, $\alpha\left(t_{0}\right)=p$ and $\alpha\left(t_{1}\right)=q$.

We want to show that $L(\alpha)>L(\lambda)$ and, since $\alpha$ is arbitrary, this will say that the straight line minimizes distance.

Now, why should $\alpha$ be longer than $\lambda$ ? One intuitive explanation is to say that $\alpha$ starts off in the wrong direction. That is, $\dot{\alpha}\left(t_{0}\right)$ is not "pointing toward" $q$. How can we measure this deviation? The angle between the unit vector $u$ in the direction of $q-p$ and the velocity vector $\dot{\alpha}\left(t_{0}\right)$ of $\alpha$ at $p$ may be calculated by taking the dot product $\dot{\alpha}(t) \bullet u$. The total deviation may be added up by integration to give us an idea of why $L(\alpha)>L(\lambda)$ should hold.

Let $u=\frac{1}{\|q-p\|}(q-p)$. We have

$$
\frac{d}{d t}(\alpha(t) \bullet u)=\dot{\alpha}(t) \bullet u+\alpha(t) \bullet \dot{u}=\dot{\alpha}(t) \bullet u .
$$

Now we compute the integral

$$
\int_{t_{0}}^{t_{1}} \dot{\alpha}(t) \bullet u d t
$$

in two different ways to obtain the inequality. On the one hand, we have

$$
\begin{aligned}
\int_{t_{0}}^{t_{1}} \dot{\alpha}(t) \bullet u d t & =\int_{t_{0}}^{t_{1}} \frac{d}{d t}(\alpha(t) \bullet u) d t=\alpha\left(t_{1}\right) \bullet u-\alpha\left(t_{0}\right) \bullet u \\
& =q \bullet u-p \bullet u \\
& =(q-p) \bullet u \\
& =\frac{(q-p) \bullet(q-p)}{\|q-p\|} \\
& =\|q-p\| \\
& =L(\lambda) .
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
\int_{t_{0}}^{t_{1}} \dot{\alpha}(t) \bullet u d t & \leq \int_{t_{0}}^{t_{1}}\|\dot{\alpha}(t)\|\|u\| d t \quad \text { (by the Cauchy-Schwarz inequality) } \\
& =\int_{t_{0}}^{t_{1}}\|\dot{\alpha}(t)\| d t \\
& =L(\alpha)
\end{aligned}
$$

Consequently,

$$
L(\lambda)=\int_{t_{0}}^{t_{1}} \dot{\alpha}(t) \bullet u d t \leq L(\alpha) .
$$

Observe that

$$
\dot{\alpha}(t) \bullet u=\|\dot{\alpha}(t)\|\|u\|
$$

only when $\cos \theta=1$, or $\theta=0$. That is, the vector $\dot{\alpha}(t)$ must be collinear with $q-p$ for all $t$. In this case $\alpha$ is (a parametrization of) the line segment from $p$ to $q$. Therefore we have strict inequality $L(\lambda)<L(\alpha)$ unless $\alpha$ is a line segment.

## Exercise 88

(a) Find the arc length of the catenary $t \mapsto(t, \cosh t)$ from $t=0$ to $t=t_{1}$.
(b) Show that the curve $t \mapsto\left(3 t^{2}, t-3 t^{3}\right)$ has a unique self crossing, determine the corresponding parameters $a$ and $b$, and then find the arc length from $t=a$ to $t=b$.
(c) Find the arc length of the astroid $t \mapsto\left(\cos ^{3} t, \sin ^{3} t\right)$ from $t=0$ to $t=\frac{\pi}{2}$ and then from $t=0$ to $t=\pi$. Compare the results.
(d) Show that the arc length of the parabola $t \mapsto\left(t^{2}, 2 t\right)$ from $t=0$ to $t=t_{1}$ is given by

$$
L=t_{1} \sqrt{1+t_{1}^{2}}+\ln \left(t_{1}+\sqrt{1+t_{1}^{2}}\right) .
$$

$\diamond$ Exercise 89 Find an expression for the arc length of the cycloid $t \mapsto(a(t-\sin t), a(1-\cos t))$ from $t=0$ to $t=t_{0}$, where $0 \leq t_{0} \leq 2 \pi$. Deduce that the arc length from $t=0$ to $t=2 \pi$ is 8 .

## Arc length parametrizations

Given a (parametrized) curve $\alpha$, we can construct many new (parametrized) curves that follow the same route (i.e., have the same trace) as $\alpha$ but travel at different speeds. Any such alteration is called a reparametrization. More precisely, we have the following definition.
2.3.17 Definition. Let $J$ and $J^{\prime}$ be intervals on the real line. Let $\alpha$ : $J \rightarrow \mathbb{R}^{3}$ be a curve and let $h: J^{\prime} \rightarrow J$ be a smooth function (usually with smooth inverse). Then the composite function

$$
\beta=\alpha(h): J^{\prime} \rightarrow \mathbb{E}^{3}, \quad \beta(s)=\alpha(h(s))
$$

is a curve called the reparametrization of $\alpha$ by $h$. (The function $h$ is the change of parameter.)

The curves $\beta$ and $\alpha$ pass through the same points in $\mathbb{R}^{3}$, but they reach any of these points in different "times" ( $s$ and $t$ ).
$\diamond$ Exercise 90 A smooth function

$$
h: J_{\theta} \rightarrow J_{t}, \quad \theta \mapsto t=h(\theta)
$$

is said to be an allowable change of parameter on (the interval) $J_{\theta}$ if it is onto and $h^{\prime}(\theta) \neq 0$ on $J_{\theta}$. Show that if $h$ is an allowable change of parameter, then $h$ is invertible and its inverse $h^{-1}$ is also an allowable change of parameter (on $J_{t}$ ).
2.3.18 EXAMPLE. The (smooth) function $s \mapsto s+s^{3}$ defines an allowable change of parameter on $\mathbb{R}$. On the other hand, the function $s \mapsto s^{3}$ does not (since its derivative vanishes at $s=0$ ).
$\diamond$ Exercise 91 Check that
(a) The function $\theta \mapsto t=\frac{\theta-a}{b-a}$ is an allowable change of parameter which takes the interval $[a, b]$ onto $[0,1]$.
(b) The function $\theta \mapsto t=\frac{1}{\pi}\left(\frac{\pi}{2}+\arctan \theta\right)$ is an allowable change of parameter which takes the interval $(-\infty, \infty)$ onto $(0,1)$.
(c) The function $\theta \mapsto t=\frac{\arctan \theta-\arctan a}{\frac{\pi}{2}-\arctan a}$ is an allowable change of parameter which takes the interval $[a, \infty)$ onto $[0,1)$.
2.3.19 Example. The reparametrization of the curve

$$
\alpha:(0,4) \rightarrow \mathbb{R}^{3}, \quad t \mapsto(\sqrt{t}, t \sqrt{t}, 1-t)
$$

by (the change of parameter) $h:(0,2) \rightarrow(0,4), \quad s \mapsto s^{2}$ is

$$
\beta(s)=\alpha(h(s))=\alpha\left(s^{2}\right)=\left(s, s^{3}, 1-s^{2}\right), \quad s \in(0,2)
$$

Exercise 92 Reparametrize
(a) the circle

$$
t \mapsto(a \cos t, a \sin t), \quad t \in[-\pi, \pi]
$$

by $h:[-1,1] \rightarrow[-\pi, \pi], \quad \theta \mapsto 4 \arctan \theta$.
(b) the positive branch of the (standard) hyperbola

$$
t \mapsto(a \cosh t, b \sinh t), \quad t \in \mathbb{R}
$$

by $h:(0, \infty) \rightarrow \mathbb{R}, \quad \theta \mapsto \ln \theta$.
(c) the tractrix

$$
t \mapsto\left(t-\tanh t, \frac{1}{\cosh t}\right), \quad t \in \mathbb{R}
$$

$$
\text { by } h:(0, \pi) \rightarrow \mathbb{R}, \quad \theta \mapsto \ln \tan \frac{\theta}{2} .
$$

The following result relates the velocities of a curve and of a reparametrization.
2.3.20 Proposition. If $\beta$ is a reparametrization of $\alpha$ by $h$, then

$$
\dot{\beta}(s)=\frac{d h}{d s}(s) \cdot \dot{\alpha}(h(s)) .
$$

Proof : If $\alpha=\left(\alpha_{1}, \alpha_{2}, \alpha_{3}\right)$, then

$$
\beta(s)=\alpha(h(s))=\left(\alpha_{1}(h(s)), \alpha_{2}(h(s)), \alpha_{3}(h(s))\right) .
$$

By the chain-rule we obtain

$$
\begin{aligned}
\dot{\beta}(s) & =\frac{d \beta}{d s}(s)=\frac{d \alpha(h)}{d s}(s) \\
& =\left(\frac{d \alpha_{1}}{d s}(h(s)) \cdot \frac{d h}{d s}(s), \frac{d \alpha_{2}}{d s}(h(s)) \cdot \frac{d h}{d s}(s), \frac{d \alpha_{3}}{d s}(h(s)) \cdot \frac{d h}{d s}(s)\right) \\
& =\frac{d h}{d s}(s) \cdot \dot{\alpha}(h(s))
\end{aligned}
$$

$\diamond$ Exercise 93 Recall that the arc length of a curve $\alpha:[a, b] \rightarrow \mathbb{R}^{3}$ is given by $L(\alpha)=\int_{a}^{b}\|\dot{\alpha}(t)\| d t$. Let $\beta:[c, d] \rightarrow \mathbb{R}^{3}$ be a reparametrization of $a$ by (the change of parameter) $h:[c, d] \rightarrow[a, b]$. Show that the arc length does not change under reparametrization.

Exercise 94 Let $\alpha: J \rightarrow \mathbb{R}^{3}$ be a curve and $f: \mathbb{R}^{3} \rightarrow \mathbb{R}$ a (smooth) function on $\mathbb{R}^{3}$. Show that (for $t \in J$ )

$$
\dot{\alpha}(t)[f]=\frac{d}{d t} f(\alpha(t))
$$

This simple result shows that the rate of change of $f$ along the line through the point $\alpha(t)$ in the direction $\alpha^{\prime}(t)$ is the same as along the curve $\alpha$ itself.

Sometimes one is interested only in the trace of a curve and not in the particular speed at which it is covered. One way to ignore the speed of a curve $\alpha$ is to reparametrize to a curve $\beta$ which has unit speed. Then $\beta$ represents a "standard trip" along the trace of $\alpha$.
2.3.21 THEOREM. If $\alpha: J \rightarrow \mathbb{R}^{3}$ is a regular curve, then there exists $a$ reparametrization $\beta$ of $\alpha$ such that $\beta$ has unit speed.

Proof: Fix a number $t_{0}$ in $J$ and consider the arc length function

$$
s(t):=\int_{t_{0}}^{t}\|\dot{\alpha}(u)\| d u
$$

Since $\alpha$ is regular, the Fundamental Theorem of Calculus implies

$$
\frac{d s}{d t}=\|\dot{\alpha}(t)\|>0
$$

By the Mean Value Theorem, $s$ is strictly increasing on $J$ and so is one-to-one. Therefore $s$ has an inverse function $t=t(s)$ and the respective derivatives are inversely related :

$$
\frac{d t}{d s}(s)=\frac{1}{\frac{d s}{d t}(t(s))}>0
$$

Let $\beta(s)=\alpha(t(s))$ be the reparametrization of $\alpha$. We claim that $\beta$ has unit speed. Indeed, we have

$$
\dot{\beta}(s)=\frac{d t}{d s}(s) \cdot \dot{\alpha}(t(s))
$$

and hence

$$
\begin{aligned}
\|\dot{\beta}(s)\| & =\left|\frac{d t}{d s}(s)\right| \cdot\|\dot{\alpha}(t(s))\| \\
& =\frac{d t}{d s}(s) \cdot \frac{d s}{d t}(t(s)) \\
& =1
\end{aligned}
$$

Without loss of generality, suppose $\beta$ is defined on the interval $[0,1]$. Consider the arc length of the reparametrization $\beta$ out to a certain parameter value $s_{0}$

$$
L(\beta)=\int_{0}^{s_{0}}\|\dot{\beta}(s)\| d s=\int_{0}^{s_{0}} d s=s_{0} .
$$

Thus $\beta$ has its arc length as parameter. We sometimes call the unit speed curve $\beta$ the arc length parametrization of $\alpha$.

Note: Observe that a curve $\beta$ parametrized by arc length has speed given by

$$
\|\dot{\beta}(s)\|=\left|\frac{d t}{d s}\right| \cdot\|\dot{\alpha}(t)\|=\frac{\|\dot{\alpha}(t)\|}{\|\dot{\alpha}(t)\|}=1 .
$$

Hence we speak of a unit speed curve as a curve parametrized by arc length. We reserve the variable $s$ for the arc length parameter when it is convenient and $t$ for an arbitrary parameter.
2.3.22 EXAMPLE. If $\alpha(t)=p+t(q-p)$, then $\dot{\alpha}(t)=(q-p)_{\alpha(t)}$ and hence

$$
\|\dot{\alpha}(t)\|=\|q-p\| .
$$

Then

$$
s(t)=\int_{0}^{t}\|\dot{\alpha}(t)\| d t=\int_{0}^{t}\|q-p\| d t=\|q-p\| t
$$

and the inverse function is $t(s)=\frac{1}{\|q-p\|} s$. So an arc length parametrization is given by

$$
\beta(s)=p+s \frac{q-p}{\|q-p\|} .
$$

Note that $\|\dot{\beta}(s)\|=1$.
$\diamond$ Exercise 95 Find an arc length parametrization of the circle $x_{1}^{2}+x_{2}^{2}=a^{2}$.
2.3.23 Example. Consider the helix $\alpha(t)=(a \cos t, a \sin t, b t)$ with $\dot{\alpha}(t)=$ $(-a \sin t, a \cos t, b)_{\alpha(t)}$. We have

$$
\|\dot{\alpha}(t)\|^{2}=\dot{\alpha}(t) \bullet \dot{\alpha}(t)=a^{2} \sin ^{2} t+a^{2} \cos ^{2} t+b^{2}=a^{2}+b^{2} .
$$

Thus $\alpha$ has constant speed $k=\sqrt{a^{2}+b^{2}}$. Then

$$
s(t)=\int_{0}^{t}\|\dot{\alpha}(t)\| d t=\int_{0}^{t} k d t=k t
$$

Hence $t(s)=\frac{s}{k}$. Substituting in the formula for $\alpha$, we get the unit speed reparametrization

$$
\beta(s)=\alpha\left(\frac{s}{k}\right)=\left(a \cos \frac{s}{k}, a \sin \frac{s}{k}, \frac{b s}{k}\right) .
$$

It is easy to check directly that $\|\dot{\beta}(s)\|=1$.

Note : If a curve $\alpha$ has constant speed, then it may be parametrized by arc length explicitly. For general curves, however, the integral defining $s$ may be impossible to compute in closed form.
2.3.24 Example. The curve $\alpha(t)=(a \cos t, b \sin t)$ gives an ellipse in (the $x_{1} x_{2}$-plane of $\mathbb{R}^{3}$, identified with) $\mathbb{R}^{2}$. Furthermore, $\dot{\alpha}(t)=(-a \sin t, b \cos t)_{\alpha(t)}$ and

$$
\|\dot{\alpha}(t)\|=\sqrt{a^{2} \sin ^{2} t+b^{2} \cos ^{2} t}=\sqrt{a^{2}+\left(b^{2}-a^{2}\right) \cos ^{2} t}
$$

The resulting length-function

$$
s(t)=\int_{0}^{t} \sqrt{a^{2}+\left(b^{2}-a^{2}\right) \cos t} d t
$$

is not generally expressible in terms of elementary functions (it is an example of an elliptic integral).

## Vector fields (on curves)

The general notion of vector field can be adapted to curves as follows.
2.3.25 Definition. A vector field $X$ on a curve $\alpha: J \rightarrow \mathbb{R}^{3}$ is a (differentible) mapping

$$
t \in J \mapsto X(t) \in T_{\alpha(t)} \mathbb{R}^{3}
$$

We have already met such a vector field : for any curve $\alpha$, its velocity $\dot{\alpha}$ clearly satisfies this definition.

Note : It is important to realize that unlike $\dot{\alpha}$, arbitrary vector fields on $\alpha$ need not be tangent to (the trace of) the curve $a$, but may point in any direction.

A vector field $X$ on a curve $\alpha$ is a unit vector field if each vector $\alpha^{\prime}(t)$ (which is a tangent vector to $\mathbb{E}^{3}$ at $\alpha(t)$ ) is a unit vector.

The properties of vector fields on curves are analogous to those of vector fields on $\mathbb{R}^{3}$. For example, if $X$ is a vector field on the curve $\alpha: J \rightarrow \mathbb{R}^{3}$, then for each $t \in J$ we can write

$$
\begin{aligned}
X(t) & =\left(X_{1}(t), X_{2}(t), X_{3}(t)\right)_{\alpha(t)} \\
& =X_{1}(t) E_{1}(\alpha(t))+X_{2}(t) E_{2}(\alpha(t))+X_{3}(t) E_{3}(\alpha(t)) \\
& =\left.X_{1}(t) \frac{\partial}{\partial x_{1}}\right|_{\alpha(t)}+\left.X_{2}(t) \frac{\partial}{\partial x_{2}}\right|_{\alpha(t)}+\left.X_{3}(t) \frac{\partial}{\partial x_{3}}\right|_{\alpha(t)}
\end{aligned}
$$

We have thus defined real-valued functions $X_{1}, X_{2}, X_{3}$ on $J$, called the $E u$ clidean coordinate functions of $X$. These will always be assumed to be differentiable (in fact, smooth).

Note : The composite function $t \mapsto E_{i}(\alpha(t))=\left.\frac{\partial}{\partial x_{i}}\right|_{\alpha(t)}$ is a vector field on $\alpha$. Where it seems to be safe to do so, we shall often write merely $E_{i}$ (or $\frac{\partial}{\partial x_{i}}$ ) insted of $E_{i}(\alpha(t))$.

The operations of addition, scalar multiplication, dot product, and cross product of vector fields (on the same curve) are all defined in the usual pointwise fashion.
2.3.26 Example. Given

$$
X(t)=t^{2} \frac{\partial}{\partial x_{1}}-t \frac{\partial}{\partial x_{3}}, \quad Y(t)=\left(1-t^{2}\right) \frac{\partial}{\partial x_{2}}+t \frac{\partial}{\partial x_{3}}, \quad \text { and } \quad f(t)=\frac{1+t}{t}
$$

we obtain the vector fields

$$
\begin{aligned}
(X+Y)(t) & =t^{2} \frac{\partial}{\partial x_{1}}+\left(1-t^{2}\right) \frac{\partial}{\partial x_{2}} \\
(f X)(t) & =t(t+1) \frac{\partial}{\partial x_{1}}-(t+1) \frac{\partial}{\partial x_{3}} \\
(X \times Y)(t) & =\left|\begin{array}{ccc}
E_{1} & E_{2} & E_{3} \\
t^{2} & 0 & -t \\
0 & 1-t^{2} & t
\end{array}\right| \\
& =t\left(1-t^{2}\right) \frac{\partial}{\partial x_{1}}-t^{3} \frac{\partial}{\partial x_{2}}+t^{2}\left(1-t^{2}\right) \frac{\partial}{\partial x_{3}}
\end{aligned}
$$

and the real-valued function

$$
(X \bullet Y)(t)=-t^{2} .
$$

To differentiate a vector field on a one simply differentiate its Euclidean coordinate functions, thus obtaining a new vector field on $\alpha$. Explicitly, if

$$
X=X_{1} \frac{\partial}{\partial x_{1}}+X_{2} \frac{\partial}{\partial x_{2}}+X_{3} \frac{\partial}{\partial x_{3}}
$$

then

$$
\dot{X}=\dot{X}_{1} \frac{\partial}{\partial x_{1}}+\dot{X}_{2} \frac{\partial}{\partial x_{2}}+\dot{X}_{3} \frac{\partial}{\partial x_{3}} .
$$

In particular, the derivative $\ddot{\alpha}$ of the velocity vector field $\dot{\alpha}$ is called the acceleration of $\alpha$.

Note : By contrast with velocity, acceleration is generally not tangent to the curve.
The following basic differentiation rules hold (for $X$ and $Y$ vector fields on $\mathbb{R}^{3}$, $f$ a real-valued (differentiable) function on $\mathbb{R}^{3}$, and $\lambda$ and $\mu$ real numbers) :

$$
\begin{aligned}
(\lambda X+\mu Y)^{\cdot} & =\lambda \dot{X}+\mu \dot{Y} \\
(f X)^{\cdot} & =\dot{f} X+f \dot{X} \\
(X \bullet Y)^{\cdot} & =\dot{X} \bullet Y+X \bullet \dot{Y} .
\end{aligned}
$$

If the function $X \bullet Y$ is constant, the last formula shows that

$$
\dot{X} \bullet Y+X \bullet \dot{Y}=0
$$

$\diamond$ Exercise 96 Show that a curve has constant speed if and only if its acceleration is everywhere orthogonal to its velocity.
$\diamond$ Exercise 97 Let $X$ be a vector field on the helix $\alpha(t)=(\cos t, \sin t, t)$. In each of the following cases, express $X$ in the form $\sum X_{i} \frac{\partial}{\partial x_{i}}$.
(a) $X(t)$ is the vector from $\alpha(t)$ to the origin of $\mathbb{R}^{3}$;
(b) $X(t)=\dot{\alpha}(t)-\ddot{\alpha}(t)$;
(c) $X(t)$ has unit length and is orthogonal to both $\dot{\alpha}(t)$ and $\ddot{\alpha}(t)$;
(d) $X(t)$ is the vector from $\alpha(t)$ to $\alpha(t+\pi)$.

Recall that tangent vectors are parallel if they have the same vector parts. We say that a vector field $X$ on a curve is parallel provided all its (tangent vector) values are parallel. In this case, if the common vector part is $\left(c_{1}, c_{2}, c_{3}\right)$, then

$$
X(t)=\left(c_{1}, c_{2}, c_{3}\right)_{\alpha(t)}=c_{1} \frac{\partial}{\partial x_{1}}+c_{2} \frac{\partial}{\partial x_{2}}+c_{3} \frac{\partial}{\partial x_{3}}
$$

for all $t$. The parallelism for a vector field is equivalent to the constancy of its Euclidean coordinate functions.
$\diamond$ Exercise 98 Let $\alpha, \bar{\alpha}: J \rightarrow \mathbb{R}^{3}$ be two curves such that $\dot{\alpha}(t)$ and $\dot{\bar{\alpha}}(t)$ are parallel (same Euclidean coordinates) at each $t$. Show that $\alpha$ and $\bar{\alpha}$ are parallel in the sense that there is a point $p \in \mathbb{R}^{3}$ such that $\bar{\alpha}(t)=\alpha(t)+p$ for all $t$.

### 2.4 Serret-Frenet Formulas

The geometry of a curve (i.e., its turning and twisting) may be (completely) described by attaching a "moving trihedron" (or moving frame) along the curve. The variation of its elements is described by the so-called Serret-Frenet formulas, which are fundamental in the study of (differential) geometry of curves in $\mathbb{R}^{3}$. We start by deriving mathematical measurements of the turning and twisting of a curve.

## The Serret-Frenet frame

Let $\beta: J \rightarrow \mathbb{R}^{3}$ be a unit speed curve, so $\|\dot{\beta}(s)\|=1$ for all $s \in J$.
2.4.1 Definition. The vector field $T:=\dot{\beta}$ is called the unit tangent vector field on $\beta$.

Since $T$ has constant length 1 , its derivative $\dot{T}=\ddot{\beta}$ measures only the rate of change of $T^{\prime} s$ direction (i.e., measures the way the curve is turning in $\mathbb{R}^{3}$ ). Hence $\dot{T}$ is a good choice to detect some of the geometry of $\beta$.
2.4.2 Definition. The vector field $\dot{T}$ is called the curvature vector field on $\beta$.

Differentiation of $T \bullet T=1$ gives

$$
0=(T \bullet T)^{\cdot}=\dot{T} \bullet T+T \bullet \dot{T}=2 T \bullet \dot{T}
$$

Hence $T \bullet \dot{T}=0$ and, therefore, $\dot{T}$ is orthogonal to $T$. We say that $\dot{T}$ is normal to $\beta$. The length of the curvature vector field $\dot{T}$ gives a numerical measurement of the turning of $\beta$.
2.4.3 Definition. The real-valued function $\kappa: I \rightarrow \mathbb{R}$ given by

$$
\kappa(s):=\|\dot{T}(s)\|
$$

is called the curvature function of $\beta$.
Of course $\kappa \geq 0$ and $\kappa$ increases as $\beta$ turns more sharply.
Note : If $\kappa=0$, then (as we will see in Theorem 2.4.10 below) we know everything about the curve $\beta$ already.

We assume that $\kappa$ is never zero, so $\kappa>0$. Then the vector field $N=\frac{1}{\kappa} \dot{T}$ on $\beta$ tells the direction in which $\beta$ is turning at each point.
2.4.4 Definition. The vector field

$$
N:=\frac{1}{\kappa} \dot{T}
$$

is called the principal normal vector field on $\beta$.
We need to introduce a third vector field on $\beta$ as part of our "moving trihedron" along the curve and this vector field should be orthogonal to both $T$ and $N$ (just as $T$ and $N$ are to each other).
2.4.5 Definition. The vector field

$$
B:=T \times N
$$

is called the binormal vector field on $\beta$.
2.4.6 Proposition. Let $\beta: J \rightarrow \mathbb{R}^{3}$ be a unit speed curve with nonzero curvature. Then the three vector fields $T, N$, and $B$ on $\beta$ are unit vector fields which are mutually orthogonal at each point.

Proof: By definition, $\|T\|=1$. Since $\kappa=\|\dot{T}\|>0$, we have

$$
\|N\|=\frac{1}{\kappa}\|\dot{T}\|=\frac{\|\dot{T}\|}{\|\dot{T}\|}=1 .
$$

We saw that $T$ and $N$ are orthogonal; that is, $T \bullet N=0$. Now $B=T \times N$ is orthogonal to both $T$ and $N$, and we have

$$
\|B\|=\|T \times N\|=\sqrt{\|T\|^{2}\|N\|^{2}-(T \bullet N)^{2}}=\sqrt{1-0}=1 .
$$

The ordered set $(T, N, B)$ is a frame field, called the Serret-Frenet frame field, on (the unit speed curve) $\beta$. The Serret-Frenet frame field on $\beta$ is full of information about $\beta$.

Note : The moving trihedron (with its curvature and torsion functions) was introduced in 1847 by Jean-Frédéric Frenet (1816-1900) and independently by Joseph Serret (1819-1885) in 1851.

## The Serret-Frenet formulas

Let $\beta: J \rightarrow \mathbb{R}^{3}$ be a unit speed curve with nonzero curvature (i.e., $\kappa>0$ ) and consider the associated Serret-Frenet frame field ( $T, N, B$ ). The measurement of how $T, N$, and $B$ vary as we move along (the trace of) the curve $\beta$ will tell us how the curve itself turns and twists through space. The variation of $T, N$, and $B$ will be determined by calculating the derivatives $\dot{T}, \dot{N}$, and $\dot{B}$. We already know

$$
\dot{T}=\kappa N
$$

by definition of $N$. So the curvature $\kappa$ describes $T^{\prime} s$ variation in direction.
$\diamond$ Exercise 99 Show that $\dot{B} \bullet B=0$ and $\dot{B} \bullet T=0$.
Because $\dot{B}$ is orthogonal to both $B$ and $T$, it follows that, at each point, $\dot{B}$ is a scalar multiple of $N$.
2.4.7 Definition. The real-valued function $\tau: J \rightarrow \mathbb{R}^{3}$ given by

$$
\dot{B}=-\tau N
$$

is called the torsion function of $\beta$. The minus sign is traditional.
Note: By contrast with curvature, there is no restriction on the values of $\tau$ : it can be positive, negative, or zero at various points of $I$. We shall show that the torsion function $\tau$ does measure the twisting (or torsion) of the curve $\beta$.

For a unit speed curve $\beta: J \rightarrow \mathbb{R}^{3}$, the associated collection

$$
\{\kappa, \tau, T, N, B\}
$$

is called the Serret-Frenet apparatus of $\beta$.
2.4.8 Example. Consider the arc length parametrization of a circle of radius $a$

$$
\beta(s)=\left(a \cos \frac{s}{a}, a \sin \frac{s}{a}, 0\right) .
$$

The unit tangent vector field is given by

$$
T(s)=\dot{\beta}(s)=\left(-\sin \frac{s}{a}, \cos \frac{s}{a}, 0\right)
$$

and

$$
\dot{T}(s)=\ddot{\beta}(s)=-\frac{1}{a}\left(\cos \frac{s}{a}, \sin \frac{s}{a}, 0\right) .
$$

Hence

$$
\kappa(s)=\|\dot{T}(s)\|=\frac{1}{a} .
$$

It follows that

$$
N(s)=\frac{1}{\kappa(s)} \dot{T}(s)=\left(-\cos \frac{s}{a},-\sin \frac{c}{a}, 0\right) .
$$

To compute the binormal vector field $B$, we take the cross product :

$$
B(s)=T(s) \times N(s)=e_{3}=(0,0,1) .
$$

Hence $-\tau(s) N(s)=\dot{B}(s)=0$, and therefore $\tau=0$.

Note : For a circle of radius $a$, the curvature function is constant and is equal to $\frac{1}{a}$. This makes sense intuitively since, as $a$ increases, the circle becomes less curved. The limit $\kappa=\frac{1}{a} \rightarrow 0$ reflects this. Moreover, the circle has zero torsion. We shall see a general reason for this fact shortly.
$\diamond$ Exercise 100 Compute the Serret-Frenet apparatus of the unit speed curve (the helix)

$$
\beta(s)=\left(a \cos \frac{s}{k}, a \sin \frac{s}{k}, \frac{b s}{k}\right) \quad \text { with } \quad k=\sqrt{a^{2}+b^{2}} .
$$

2.4.9 Theorem. (The Serret-Frenet Theorem) If $\beta: J \rightarrow \mathbb{R}^{3}$ is a unit speed curve with nonzero curvature, then

$$
\begin{aligned}
\dot{T} & =\kappa N \\
\dot{N} & =-\kappa T+\tau B \\
\dot{B} & =-\tau N .
\end{aligned}
$$

Proof : The first and the third formulas are essentially just the definitions of curvature and torsion. To prove the second formula, we express $\dot{N}$ in terms of $T, N$, and $B$ :

$$
\dot{N}=(\dot{N} \bullet T) T+(\dot{N} \bullet N) N+(\dot{N} \bullet B) B .
$$

These coefficients are easily found. Differentiating $N \bullet T=0$, we get $\dot{N} \bullet T+$ $N \bullet \dot{T}=0$, and hence

$$
\dot{N} \bullet T=-N \bullet \dot{T}=-N \bullet(\kappa N)=-\kappa .
$$

As usual, $\dot{N} \bullet N=0$, since $N$ is a unit vector field. Finally,

$$
\dot{N} \bullet B==-N \bullet \dot{B}=-N \bullet(-\tau N)=\tau .
$$

Note : We can record the Serret-Frenet formulas more succinctly in the matrix expression

$$
\left[\begin{array}{c}
\dot{T} \\
\dot{N} \\
\dot{B}
\end{array}\right]=\left[\begin{array}{ccc}
0 & \kappa & 0 \\
-\kappa & 0 & \tau \\
0 & -\tau & 0
\end{array}\right]\left[\begin{array}{l}
T \\
N \\
B
\end{array}\right]
$$

or, equivalently,

$$
\left[\begin{array}{ccc}
\dot{T} & \dot{N} & \dot{B}
\end{array}\right]=\left[\begin{array}{lll}
T & N & B
\end{array}\right]\left[\begin{array}{ccc}
0 & -\kappa & 0 \\
\kappa & 0 & -\tau \\
0 & \tau & 0
\end{array}\right]
$$

$\diamond$ Exercise 101 If a rigid body moves along a (unit speed) curve $\beta$, then the motion of the body consists of translation along (the trace of) $\beta$ and rotation about (the trace of) $\beta$. The rotation is determined by an angular velocity vector $\omega$ which satisfies

$$
\dot{T}=\omega \times T, \quad \dot{N}=\omega \times N, \quad \text { and } \quad \dot{B}=\omega \times B
$$

The vector $\omega$ is called the Darboux vector. Show that $\omega$, in terms of $T, N$, and $B$, is given by $\omega=\tau T+\kappa B$. (Hint : Write $\omega=a T+b N+c B$ and take cross products with $T, N$, and $B$ to determine $a, b$, and $c$.)
$\diamond$ Exercise 102 Show that

$$
\dot{T} \times \ddot{T}=\kappa^{2} \omega
$$

where $\omega$ is the Darboux vector.

## Constraints on curvature and torsion

Constraints on curvature and torsion produce constraints on the geometry of the curve. The simplest constraints are contained in the following two results.
2.4.10 Theorem. Let $\beta: J \rightarrow \mathbb{R}^{3}$ be a unit speed curve. Then

$$
\kappa=0 \quad \text { if and only if } \beta \text { is a (part of a) line. }
$$

Proof: $(\Rightarrow)$ Suppose $\kappa=0$. Then $\dot{T}=0$ by the Serret-Frenet formulas, and so $T=v$ is a constant (with $\|v\|=1$ since $\beta$ has unit speed). But

$$
\dot{\beta}(s)=T=v \Rightarrow \beta(s)=p+s v
$$

with $p$ a constant of integration. Hence $\beta$ is (the arc length parametrization of) a line.
$(\Leftarrow)$ Suppose $\beta$ is (the arc length parametrization of) a line. Then $\beta(s)=$ $p+s v$ with $\|v\|=1$ (so $\beta$ has unit speed). It follows that

$$
T(s)=\dot{\beta}(s)=v=\mathrm{constant}
$$

and so $\dot{T}=0=\kappa N$, and hence $\kappa=0$.
A plane curve in $\mathbb{R}^{3}$ is a curve that lies in a single plane of $\mathbb{R}^{3}$. That is, the trace of the curve is a subset of a certain plane of $\mathbb{R}^{3}$. Clearly, the straight line and the circle are plane curves.
2.4.11 Theorem. Let $\beta: J \rightarrow \mathbb{R}^{3}$ be a unit speed curve with nonzero curvature. Then

$$
\tau=0 \quad \text { if and only if } \beta \text { is a plane curve. }
$$

Proof: $\quad(\Rightarrow) \quad$ Suppose $\tau=0$. Then, by the Serret-Frenet formulas, $\dot{B}=0$ and so $B$ is constant (parallel). But this means that $\beta(s)$ should always lie in the plane through $\beta(0)$ orthogonal to $B$. We show this.

Take the plane determined by the point $\beta(0)$ and the normal vector $B$. Recall that a point $p$ is in this plane if $(p-\beta(0)) \bullet B=0$. Consider the real-valued function

$$
f(s):=(\beta(s)-\beta(0)) \bullet B
$$

for all $s$. Then

$$
\dot{f}(s)=(\beta(s)-\beta(0)) \bullet B+(\beta(s)-\beta(0)) \bullet \dot{B}=\dot{\beta}(s) \bullet B=T \bullet B=0 .
$$

Hence $f(s)=$ constant. To identify the constant, evaluate

$$
f(0)=(\beta(0)-\beta(0))=0 .
$$

Then (for all $s$ ) $(\beta(s)-\beta(0)) \bullet B=0$ and hence $\beta(s)$ is in the plane determined by $\beta(0)$ and the (constant) vector $B$.
$(\Leftarrow)$ Suppose $\beta$ lies in a plane. Then the plane is determined by a point $p$ and a normal vector $n \neq 0$. Since $\beta$ lies in the plane, we have (for all $s$ )

$$
(\beta(s)-p) \bullet n=0 .
$$

By differentiating, we get

$$
\dot{\beta}(s) \bullet n=\ddot{\beta}(s) \bullet n=0 .
$$

That is, $T \bullet n$ and $\kappa N \bullet n=0$. These equations say that $n$ is orthogonal to both $T$ and $N$. Thus $n$ is collinear to $B$ and

$$
B= \pm \frac{1}{\|n\|} n .
$$

Hence $\dot{B}=0$ and the Serret-Frenet formulas then give $\tau=0$.
We now see that curvature measures the deviation of a curve from being a (straight) line and torsion the deviation of a curve from being contained in a plane. We know that the standard circle of radius $a$ in the $x_{1} x_{2}$-plane in $\mathbb{R}^{3}$ has $\kappa=\frac{1}{a}$ and $\tau=0$. To see that a circle located anywhere in $\mathbb{R}^{3}$ has these properties we have two choices. We could give a parametrization for an arbitrary circle in $\mathbb{R}^{3}$ or we could use the familiar definition of a circle as the locus of points in a plane equally distanced from a fixed point (in the plane). In order to emphasize geometry, we take the latter approach.
2.4.12 Theorem. Let $\beta: J \rightarrow \mathbb{R}^{3}$ be a unit speed curve. Then (the trace of) $\beta$ is a part of a circle if and only if $\kappa>0$ is constant and $\tau=0$.

Proof: $(\Rightarrow)$ Suppose (the trace of) $\beta$ is part of a circle. By definition, $\beta$ is a plane curve, so $\tau=0$. Also by definition (for all $s$ ) $\|\beta(s)-p\|=r$. Squaring both sides gives $(\beta(s)-p) \bullet(\beta(s)-p)=a^{2}$. If we differentiate this expression, we get (for $T=\dot{\beta}$ )

$$
2 T \bullet(\beta(s)-p)=0 \quad \text { or } \quad T \bullet(\beta(s)-p)=0
$$

If we differentiate again, then we obtain

$$
\begin{align*}
\dot{T} \bullet(\beta(s)-p)+T \bullet T & =0 \\
\kappa N \bullet(\beta(s)-p)+1 & =0  \tag{*}\\
\kappa N \bullet(\beta(s)-p) & =-1
\end{align*}
$$

This means, in particular, that $\kappa>0$ and $N \bullet(\beta(s)-p) \neq 0$. Now differentiating (*) produces

$$
\begin{array}{r}
\frac{d \kappa}{d s} N \bullet(\beta(s)-p)+\kappa \dot{N} \bullet(\beta(s)-p)+\kappa N \bullet T
\end{array}=0 .
$$

Since $\tau=0$ and $T \bullet(\beta(s)-p)=0$ by above, we have

$$
\frac{d \kappa}{d s} N \bullet(\beta(s)-p)=0
$$

Also $N \bullet(\beta(s)-p) \neq 0$ by above, and so $\frac{d \kappa}{d s}=0$. This means, of course, that $\kappa>0$ is constant.
$(\Leftarrow)$ Suppose now that $\tau=0$ and $\kappa>0$ is constant. To show $\beta(s)$ is part of a circle we must show that each $\beta(s)$ is a fixed distance from a fixed point.

For the standard circle, from any point on the circle to the center we proceed in the normal direction a distance equal to the radius. That is, we go $a N=\frac{1}{\kappa} N$. We do the same here.

Let $\gamma$ denote the curve

$$
\gamma(s):=\beta(s)+\frac{1}{\kappa} N .
$$

Since we want $\gamma$ to be a single point (the center of the desired circle), we must have $\dot{\gamma}(s)=0$. Computing, we obtain

$$
\begin{aligned}
\dot{\gamma}(s) & =\dot{\beta}(s)+\frac{1}{\kappa} \dot{N} \\
& =T+\frac{1}{\kappa}(-\kappa T+\tau B) \\
& =T-T \\
& =0 .
\end{aligned}
$$

Hence $\gamma(s)$ is a constant $p$. Then we have

$$
\|\beta(s)-p\|=\left\|-\frac{1}{\kappa} N\right\|=\frac{1}{\kappa}
$$

so $p$ is the center of a circle $\beta(s)$ of radius $\frac{1}{\kappa}$.

Exercise 103 Compute the Serret-Frenet apparatus of the unit speed curve

$$
\beta(s)=\left(\frac{4}{5} \cos s, 1-\sin s,-\frac{3}{5} \cos s\right) .
$$

$\diamond$ Exercise 104 Let $\beta$ be a unit speed curve which lies entirely on the sphere of radius $a$ centered at the origin. Show that the curvature $\kappa$ is such that $\kappa \geq \frac{1}{a}$. (Hint : Differentiate $\beta \bullet \beta=a^{2}$ and use the Serret-Frenet formulas to get $\kappa \beta \bullet N=-1$.)
$\diamond$ Exercise 105 Let $\beta$ be a unit speed curve which lies entirely on the sphere of center $p$ and radius $a$. Show that, if $\tau \neq 0$, then

$$
\beta(s)-p=-\frac{1}{\kappa} N-\left(\frac{1}{\kappa}\right) \frac{1}{\tau} B \quad \text { and } \quad a^{2}=\left(\frac{1}{\kappa}\right)^{2}+\left(\left(\frac{1}{\kappa}\right) \frac{1}{\tau}\right)^{2}
$$

$\diamond$ Exercise 106 Show that, if

$$
\left(\frac{1}{\kappa}\right) \neq 0 \quad \text { and } \quad\left(\frac{1}{\kappa}\right)^{2}+\left(\left(\frac{1}{\kappa}\right) \frac{1}{\tau}\right)^{2} \text { is a constant }
$$

then the unit speed curve $\beta$ lies entirely on a sphere. (Hint : Show that the "center curve" $\gamma(s):=\beta(s)+\left(\frac{1}{\kappa}\right)^{\circ} \frac{1}{\tau} B$ is constant.)
$\diamond$ Exercise 107 Find the curvature $\kappa$ and torsion $\tau$ for the curve

$$
\beta(s)=\left(\frac{1}{\sqrt{2}} \cos s, \sin s, \frac{1}{\sqrt{2}} \cos s\right) .
$$

Identify the curve.

### 2.5 The Fundamental Theorem for Curves

Recall the notion of a vector field on a curve. If $X$ is a vector field on $\alpha: J \rightarrow \mathbb{R}^{3}$ and $F$ is an isometry, then $\widetilde{X}=F_{*}(X)$ is a vector field on the image curve $\widetilde{\alpha}=F(\alpha)$. In fact, for each $t \in J, X(t)$ is a tangent vector to $\mathbb{E}^{3}$ at the point $\alpha(t)$. But then $\widetilde{X}(t)=F_{*}(X(t))$ is a tangent vector to $\mathbb{E}^{3}$ at the point $F(\alpha(t))=\widetilde{\alpha}(t)$.

Isometries preserve the derivatives of such vector fields.
$\diamond$ Exercise 108 If $X$ is a vector field on a curve $\alpha$ in $\mathbb{R}^{3}$ and $F$ is an isometry on $\mathbb{R}^{3}$, then $\widetilde{X}=F_{*}(X)$ is a vector field on $\widetilde{\alpha}=F(\alpha)$. Show that

$$
\dot{\tilde{X}}=F_{*}(\dot{X})
$$

It follows immediately that (if we set $X=\dot{\alpha}$ )

$$
\ddot{\widetilde{\alpha}}=\dot{\widetilde{X}}=F_{*}(\dot{X})=F_{*}(\ddot{\alpha})
$$

That is, isometries preserve acceleration. Now we show that the Serret-Frenet apparatus of a curve is preserved by isometries.

Note : This is certainly to be expected on intuitive grounds, since a rigid motion ought to carry one curve into another that turns and twists in exactly the same way. And this is what happens when the isometry is orientation-preserving.
2.5.1 Proposition. Let $\beta$ be a unit speed curve on $\mathbb{R}^{3}$ with nonzero curvature and let $\widetilde{\beta}=F(\beta)$ be the image curve of $\beta$ under the isometry $F$ on $\mathbb{R}^{3}$. Then

$$
\begin{aligned}
\widetilde{\kappa} & =\kappa, \quad \widetilde{T}=F_{*}(T) \\
\widetilde{\tau} & =(\operatorname{sgn} F) \tau, \quad \widetilde{N}=F_{*}(N) \\
\widetilde{B} & =(\operatorname{sgn} F) F_{*}(B)
\end{aligned}
$$

Proof : Observe first that $\widetilde{\beta}$ is also a unit speed curve, since

$$
\|\dot{\widetilde{\beta}}\|=\left\|F_{*}(\dot{\beta})\right\|=\|\dot{\beta}\|=1
$$

Thus

$$
\widetilde{T}=\dot{\widetilde{\beta}}=F_{*}(\dot{\beta})=F_{*}(T)
$$

Since $F_{*}$ preserves both acceleration and norms, it follows from the definition of curvature that

$$
\widetilde{\kappa}=\|\ddot{\widetilde{\beta}}\|=\left\|F_{*}(\ddot{\beta})\right\|=\|\ddot{\beta}\|=\kappa .
$$

To get the full Serret-Frenet frame, we now use the hypothesis $\kappa>0$ (which implies $\widetilde{\kappa}>0$, since $\widetilde{\kappa}=\kappa$ ). By definition, $N=\frac{1}{\kappa} \ddot{\beta}$ and hence

$$
\widetilde{N}=\frac{\ddot{\tilde{\beta}}}{\widetilde{\kappa}}=\frac{F_{*}(\ddot{\beta})}{\kappa}=F_{*}\left(\frac{\ddot{\beta}}{\kappa}\right)=F_{*}(N) .
$$

It remains only to prove the interesting cases $B$ and $\tau$. We have

$$
\widetilde{B}=\widetilde{T} \times \widetilde{N}=F_{*}(T) \times F_{*}(N)=(\operatorname{sgn} F) F_{*}(T \times N)=(\operatorname{sgn} F) F_{*}(B)
$$

Furthermore,

$$
\widetilde{\tau}=\widetilde{B} \bullet \dot{\tilde{N}}=(\operatorname{sgn} F) F_{*}(B) \bullet F_{*}(\dot{N})=(\operatorname{sgn} F) B \bullet \dot{N}=(\operatorname{sgn} F) \tau .
$$

Note: The presence of $\operatorname{sgn} F$ in the formula for the torsion of $F(\beta)$ shows that the torsion of a curve gives more subtle description of the curve than has been apparent so far. The sign of $\tau$ measures the orientation of the twisting of the curve.

We have seen that curvature and torsion, individually and in combination, tell us a great deal about the geometry of a curve. In fact, in a very real sense, they tell us everything. Precisely, if two unit speed curves have the same curvature and torsion functions, then there is a rigid motion of $\mathbb{R}^{3}$ taking one curve onto another. Furthermore, given specified curvature and torsion functions, there is a curve which realizes them as its own curvature and torsion. These results are, essentially, theorems about existence and uniqueness of solutions of systems of differential equations. The Serret-Frenet formulas provide the system and the unique solution provides the curve.
2.5.2 Theorem. (The Fundamental Theorem) Let $\kappa, \tau:(a, b) \rightarrow \mathbb{R}$ be (smooth) functions with $\kappa>0$. Then there exists a regular curve $\beta$ : $(a, b) \rightarrow \mathbb{R}^{3}$ parametrized by arc length such that $\kappa$ is the curvature function and $\tau$ is the torsion function of $\beta$. Moreover, any other curve $\bar{\beta}$ satisfying the same conditions, differs from $\beta$ by a proper rigid motion; that is, there exists a direct isometry $F, x \mapsto A x+c$ such that (for all s)

$$
\beta(s)=A \bar{\beta}(s)+c .
$$

Proof : Consider the matrix-valued function

$$
g(s):=\left[\begin{array}{ccc}
0 & \kappa(s) & 0 \\
-\kappa(s) & 0 & \tau(s) \\
0 & -\tau(s) & 0
\end{array}\right]=\left[a_{i j}\right] .
$$

If we write

$$
\xi_{1}=T, \quad \xi_{2}=N, \quad \text { and } \quad \xi_{3}=B
$$

then the Serret-Frenet formulas give us the system of differential equations

$$
\begin{aligned}
\dot{\xi}_{1} & =a_{11} \xi_{1}+a_{12} \xi_{2}+a_{13} \xi_{3} \\
\dot{\xi}_{2} & =a_{21} \xi_{1}+a_{22} \xi_{2}+a_{23} \xi_{3} \\
\dot{\xi}_{3} & =a_{31} \xi_{1}+a_{32} \xi_{2}+a_{33} \xi_{3}
\end{aligned}
$$

or, equivalently, the vector differential equation

$$
\left[\begin{array}{l}
\dot{\xi}_{1} \\
\dot{\xi}_{2} \\
\dot{\xi}_{3}
\end{array}\right]=g(s)\left[\begin{array}{l}
\xi_{1} \\
\xi_{2} \\
\xi_{3}
\end{array}\right] .
$$

It is known that if the matrix-valued function $s \mapsto g(s)$ is continuous, then the diferential equation

$$
\dot{\xi}=g(s) \xi, \quad s \in(a, b)
$$

has solutions $\xi:(a, b) \rightarrow \mathbb{E}^{3}$.
Thus there is a solution $\left(\xi_{1}(s), \xi_{2}(s), \xi_{3}(s)\right)$ dependent upon the initial conditions. For a value $s_{0} \in(a, b)$, we may take $\left(\xi_{1}\left(s_{0}\right), \xi_{2}\left(s_{0}\right), \xi_{3}\left(s_{0}\right)\right)$ to be a choice of a positively-oriented frame on $\mathbb{R}^{3}$. We next show that (for every $s)$ the solution $\left(\xi_{1}(s), \xi_{2}(s), \xi_{3}(s)\right)$ is a frame. Observe that

$$
\left(\xi_{i} \bullet \xi_{j}\right)^{\bullet}=\sum_{k=1}^{3}\left(a_{i k} \xi_{j} \bullet \xi_{k}+a_{j k} \xi_{i} \bullet \xi_{k}\right), \quad i, j=1,2,3 .
$$

Let $\xi_{i j}:=\xi_{i} \bullet \xi_{j}$. We obtain the system of differential equations

$$
\begin{equation*}
\dot{\xi}_{i j}=\sum_{k=1}^{3}\left(a_{i k} \xi_{j k}+a_{j k} \xi_{i k}\right), \quad i, j=1,2,3 \tag{*}
\end{equation*}
$$

with initial conditions

$$
\xi_{i j}=\delta_{i j} \quad \text { (the Kronecker delta function). }
$$

In order to have a frame we need to show that $\xi_{i j}(s)=\delta_{i j}$ holds for all $s$. But notice

$$
\dot{\delta}_{i j}=0=a_{i j}+a_{j i}=\sum_{k=1}^{3}\left(a_{i k} \delta_{j k}+a_{j k} \delta_{i k}\right), \quad i, j=1,2,3
$$

which holds by the skew-symmetry of the matrix $g(s)$. Thus $\delta_{i j}, i, j=1,2,3$ satisfies the differential equation $(*)$ and so, by the uniqueness of solutions to differential equations, we have (for all $s$ ) a frame.

To define the curve $\beta:(a, b) \rightarrow \mathbb{R}^{3}$ we integrate

$$
\beta(s)=\int_{s_{0}}^{s} \xi_{1}(\sigma) d \sigma .
$$

Then $\dot{\beta}(s)=\xi_{1}(s)=T(s)$ and so

$$
\ddot{\beta}=\dot{\xi}_{1}=\kappa \xi_{2}=\kappa N .
$$

Thus $\kappa(s)$ is the curvature of $\beta$ (at $s$ ) and hence $\kappa_{\beta}=\kappa$.
$\diamond$ Exercise 109 Show that if $\beta$ is a unit speed curve with nonzero curvature, then

$$
\tau=\frac{\dot{\beta} \bullet \ddot{\beta} \times \dddot{\beta}}{\kappa^{2}} .
$$

(Hint : Compute $\dot{\beta} \times \ddot{\beta}$ and $\dddot{\beta}$ (in terms of $T, N$, and $B$ ) and then dot them.)
It follows immediately that $\tau(s)$ is the torsion of $\beta$ (at $s$ ) and hence $\tau_{\beta}=\tau$.
Now assume that two (unit speed) curves $\beta$ and $\bar{\beta}$ satisfy the conditions

$$
\kappa_{\beta}=\kappa_{\bar{\beta}}=\kappa \quad \text { and } \quad \tau_{\beta}=\tau_{\bar{\beta}}=\tau .
$$

Let $\left(T_{0}, N_{0}, B_{0}\right)$ and ( $\left.\bar{T}_{0}, \bar{N}_{0}, \bar{B}_{0}\right)$ be the Serret-Frenet frames at $s_{0} \in I=$ $(a, b)$ of $\beta$ and $\bar{\beta}$, respectively. By Theorem 2.1.9, there is a (proper) rigid motion $F, x \mapsto A x+c$ on $\mathbb{E}^{3}$ which takes $\bar{\beta}\left(s_{0}\right)$ into $\beta\left(s_{0}\right)$ and $\left(\bar{T}_{0}, \bar{N}_{0}, \bar{B}_{0}\right)$ into ( $T_{0}, N_{0}, B_{0}$ ).

Denote the Serret-Frenet apparatus of $\widetilde{\beta}=F(\bar{\beta})$ by $\{\widetilde{k}, \widetilde{\tau}, \widetilde{T}, \widetilde{N}, \widetilde{B}\}$. Then (from Proposition 2.5.1 and the information above)

$$
\begin{aligned}
\widetilde{\beta}\left(s_{0}\right) & =\beta\left(s_{0}\right) \\
\widetilde{\kappa} & =\kappa, \quad \widetilde{T}\left(s_{0}\right)=T_{0} \\
\widetilde{\tau} & =\tau, \quad \widetilde{N}\left(s_{0}\right)=N_{0} \\
\widetilde{B}\left(s_{0}\right) & =B_{0} .
\end{aligned}
$$

We need to show that the curves $\beta$ and $\widetilde{\beta}$ coincide; that is (for all $s$ )

$$
\beta(s)=\widetilde{\beta}(s)=F(\bar{\beta}(s))=A \bar{\beta}(s)+c .
$$

We shall show that $T=\widetilde{T}$; that is, the curves $\beta$ and $\widetilde{\beta}$ are parallel.
$\diamond$ Exercise 110 Show that if two curves $\beta, \widetilde{\beta}: J \rightarrow \mathbb{R}^{3}$ are parallel and $\beta\left(s_{0}\right)=$ $\widetilde{\beta}\left(s_{0}\right)$ for some $s_{0} \in I$, then $\beta=\widetilde{\beta}$.

Consider the real-valued function (on the interval $J$ )

$$
f=T \bullet \widetilde{T}+N \bullet \widetilde{N}+B \bullet \widetilde{B}
$$

Since these are unit vector fields, the Cauchy-Schwarz inequality shows that

$$
T \bullet \widetilde{T} \leq 1
$$

Furthermore, $T \bullet \widetilde{T}=1$ if and only if $T=\widetilde{T}$. Similar remarks hold for the other two terms in $f$. Thus it suffices to show that $f$ has constant value 3 .
$\diamond$ Exercise 111 Show that the real-valued function

$$
f=T \bullet \widetilde{T}+N \bullet \widetilde{N}+B \bullet \widetilde{B}
$$

has constant value 3. (Hint : Compute $\dot{f}=0$ and observe that $f\left(s_{0}\right)=3$.)

### 2.6 Some Remarks

## Arbitrary speed curves

Let $\alpha: J \rightarrow \mathbb{R}^{3}$ be a regular curve that does not necessarily have unit speed. We may reparametrize $\alpha$ to get a unit speed curve $\bar{\alpha}$ and then transfer to $\alpha$ the Serret-Frenet apparatus of $\bar{\alpha}$. Explicitly, if $s$ is an arc length function for $\alpha$, then

$$
\alpha(t)=\bar{\alpha}(s(t)) \quad \text { for } t \in J .
$$

Let $\{\bar{\kappa}, \bar{\tau}, \bar{T}, \bar{N}, \bar{B}\}$ be the Serret-Frenet apparatus of $\bar{\alpha}$. We now make the following definition (for $\bar{\kappa}>0$ ).
2.6.1 Definition. We define (for the regular curve $\alpha$ )

- the curvature function : $\kappa(t):=\bar{\kappa}(s(t))$;
- the torsion function : $\tau(t):=\bar{\tau}(s(t))$;
- the unit tangent vector field : $T(t):=\bar{T}(s(t))$;
- the principal normal vector field : $N(t):=\bar{N}(s(t))$;
- the binormal vector field : $B(t):=\bar{B}(s(t))$.

Note : In general, $\kappa$ and $\bar{\kappa}$ are different functions, defined on different intervals. But they give exactly the same description of the turning of the common route of $\alpha$ and $\bar{\alpha}$, since at any point $\alpha(t)=\bar{\alpha}(s(t))$ the numbers $\kappa(t)$ and $\bar{\kappa}(s(t))$ are by definition the same. Similarly with the rest of the Serret-Frenet apparatus; since only the change of parametrization is involved, its fundamental geometric meaning is the same as before.

For purely theoretical work, this simple transference is often all that is needed. Data about $\alpha$ converts into data about the unit speed reparametrization $\bar{\alpha}$; results about $\bar{\alpha}$ convert to results about $\alpha$. However, for explicit numerical computations - and occasionally for the theory as well - this transference is impractical, since it is rarely possible to find explicit formulas for $\bar{\alpha}$.

The Serret-Frenet formulas are valid only for unit speed curves; they tell the rate of change of the frame field $(T, N, B)$ with respect to arc length. However, the speed $\nu$ of the curve is the proper correction factor in the general case.
2.6.2 Proposition. If $\alpha$ is a regular curve with nonzero curvature, then

$$
\begin{aligned}
\dot{T} & =\kappa \nu N \\
\dot{N} & =-\kappa \nu T+\tau \nu B \\
\dot{B} & =-\tau \nu N .
\end{aligned}
$$

Proof : The speed of $\alpha$ is

$$
\nu(t)=\|\dot{\alpha}(t)\|=\frac{d s}{d t} .
$$

Let $\bar{\alpha}$ be a unit speed reparametrization of $\alpha$. Then $T(t)=\bar{T}(s(t))$. The chain rule and the usual Serret-Frenet equations give

$$
\begin{aligned}
\dot{T}=\frac{d T}{d t} & =\frac{d \bar{T}}{d s} \frac{d s}{d t} \\
& =\bar{\kappa} \bar{N} \nu \\
& =\kappa \nu N
\end{aligned}
$$

so the first formula is proved. For the second and third,

$$
\begin{aligned}
\dot{N}=\frac{d N}{d t} & =\frac{d \bar{N}}{d s} \frac{d s}{d t} \\
& =(-\bar{\kappa} \bar{T}+\bar{\tau} \bar{B}) \nu \\
& =-\kappa \nu T+\tau \nu B
\end{aligned}
$$

and

$$
\begin{aligned}
\dot{B}=\frac{d B}{d t} & =\frac{d \bar{B}}{d s} \frac{d s}{d t} \\
& =-\bar{\tau} \bar{N} \nu \\
& =-\tau \nu N .
\end{aligned}
$$

Recall that only for a constant speed curve is acceleration everywhere orthogonal to velocity. In the general case, we analyze velocity and acceleration by expressing them in terms of the Serret-Frenet frame field.
2.6.3 Proposition. If $\alpha$ is a regular curve with speed function $\nu$, then the velocity and acceleration of $\alpha$ are given by

$$
\dot{\alpha}=\nu T \quad \text { and } \quad \ddot{\alpha}=\frac{d \nu}{d t} T+\kappa \nu^{2} N .
$$

Proof : Since $\alpha(t)=\bar{\alpha}(s(t))$, the first calculation is

$$
\begin{aligned}
\dot{\alpha}=\frac{d \alpha}{d t} & =\frac{d \bar{\alpha}}{d s} \frac{d s}{d t} \\
& =\nu \bar{T} \\
& =\nu T
\end{aligned}
$$

while the second is

$$
\begin{aligned}
\ddot{\alpha}=\frac{d \dot{\alpha}}{d t} & =\frac{d \nu}{d t} T+\nu \dot{T} \\
& =\frac{d \nu}{d t} T+\nu \kappa \nu N \\
& =\frac{d \nu}{d t} T+\kappa \nu^{2} N .
\end{aligned}
$$

Note: The formula $\dot{\alpha}=\nu T$ is to be expected since $\dot{\alpha}$ and $T$ are each tangent to the curve and $T$ has unit length, while $\|\dot{\alpha}\|=\nu$. The formula for acceleration is more interesting. By definition, $\ddot{\alpha}$ is the rate of change of the velocity $\dot{\alpha}$, and in general both the length and the direction of $\dot{\alpha}$ are changing. The tangential component $(d \nu / d t) T$ of $\ddot{\alpha}$ measures the rate of change of the length of $\dot{\alpha}$ (i.e. of the speed of $\alpha$ ). The normal component $\kappa \nu^{2} N$ measures the rate of change of the direction of $\dot{\alpha}$. Newton's laws of motion show that these components may be experienced as forces.

We now find effectively computable expressions for the Serret-Frenet apparatus. Clearly we have (for an arbitrary speed curve)

$$
T=\frac{\dot{\alpha}}{\|\dot{\alpha}\|} \quad \text { and } \quad N=B \times T .
$$

We also have
2.6.4 Proposition. For any regular curve $\alpha$ (with positive curvature)

$$
\begin{aligned}
& \text { (1) } \quad B=\frac{\dot{\alpha} \times \ddot{\alpha}}{\|\dot{\alpha} \times \ddot{\alpha}\|} ; \\
& \text { (2) } \quad \kappa=\frac{\|\dot{\alpha} \times \ddot{\alpha}\|}{\|\dot{\alpha}\|^{3}} ; \\
& \text { (3) } \quad \tau=\frac{(\dot{\alpha} \times \ddot{\alpha}) \bullet \dddot{\alpha}}{\|\dot{\alpha} \times \ddot{\alpha}\|^{2}} .
\end{aligned}
$$

Proof : For (1), we use the formulas of Proposition 2.6.3 to get

$$
\begin{aligned}
\dot{\alpha} \times \ddot{\alpha} & =(\nu T) \times\left(\frac{d \nu}{d t} T+\kappa \nu^{2} N\right) \\
& =\nu \frac{d \nu}{d t} T \times T+\kappa \nu^{3} T \times N \\
& =0+\kappa \nu^{3} B .
\end{aligned}
$$

Hence $\|\dot{\alpha} \times \ddot{\alpha}\|=\kappa \nu^{3}$ and so

$$
B=\frac{\dot{\alpha} \times \ddot{\alpha}}{\|\dot{\alpha} \times \ddot{\alpha}\|}
$$

For (2), we use the expression for $\ddot{\alpha}$ in Proposition 2.6.3, take cross product with $T$ and note that $T \times T=0$ to isolate the curvature

$$
\begin{aligned}
T \times \ddot{\alpha} & =0+\kappa \nu^{2} T \times N \\
\frac{\dot{\alpha} \times \ddot{\alpha}}{\|\dot{\alpha}\|} & =\kappa \nu^{2} T \times N
\end{aligned}
$$

We get (by taking norms)

$$
\frac{\|\dot{\alpha} \times \ddot{\alpha}\|}{\nu}=\kappa \nu^{2}\|B\|
$$

and hence

$$
\frac{\|\dot{\alpha} \times \ddot{\alpha}\|}{\nu^{3}}=\kappa .
$$

For (3), we take the third derivative

$$
\begin{aligned}
\dddot{\alpha} & =\left(\frac{d \nu}{d t} T+\kappa \nu^{2} N\right) \\
& =\frac{d^{2} \nu}{d t^{2}} T+\frac{d \nu}{d t} \dot{T}+\frac{d \kappa}{d t} \nu^{2} N+2 \kappa \nu \frac{d \nu}{d t} N+\kappa \nu^{2} \dot{N}
\end{aligned}
$$

Therefore, since $\dot{T}=\kappa N$ and $B$ is othogonal to $T$ and $N$, we get

$$
\begin{aligned}
B \bullet \dddot{\alpha} & =\kappa \nu^{2} B \bullet \dot{N} \\
& =\kappa \nu^{2} B \bullet(-\kappa \nu T+\tau \nu B) \\
& =\kappa \tau \nu^{3} .
\end{aligned}
$$

Now $\dot{\alpha} \times \ddot{\alpha}=\kappa \nu^{3} B$, and so

$$
\begin{aligned}
(\dot{\alpha} \times \ddot{\alpha}) \bullet \dddot{\alpha} & =\kappa \nu^{3} B \bullet \dddot{\alpha} \\
& =\kappa \nu^{3}\left(\kappa \tau \nu^{3}\right) \\
& =\kappa^{2} \nu^{6} \tau
\end{aligned}
$$

Of course $\|\dot{\alpha} \times \ddot{\alpha}\|=\kappa^{2} \nu^{6}$, so we have

$$
\tau=\frac{(\dot{\alpha} \times \ddot{\alpha}) \bullet \dddot{\alpha}}{\kappa^{2} \nu^{6}}=\frac{(\dot{\alpha} \times \ddot{\alpha}) \bullet \dddot{\alpha}}{\|\dot{\alpha} \times \ddot{\alpha}\|^{2}}
$$

Note : Equation (2) shows that for regular curves, $\|\dot{\alpha} \times \ddot{\alpha}\|>0$ is equivalent to the usual condition $\kappa>0$.
2.6.5 Example. We compute the Serret-Frenet apparatus of the regular curve

$$
\alpha(t)=\left(3 t-t^{3}, 3 t^{2}, 3 t+t^{3}\right)
$$

The derivatives are

$$
\begin{aligned}
\dot{\alpha} & =3\left(1-t^{2}, 2 t, 1+t^{2}\right) \\
\ddot{\alpha} & =6(-t, 1, t) \\
\ddot{\alpha} & =6(-1,0,1) .
\end{aligned}
$$

We have $\dot{\alpha} \bullet \dot{\alpha}=18\left(1+2 t^{2}+t^{4}\right)$, and so

$$
\nu=\|\dot{\alpha}\|=3 \sqrt{2}\left(1+t^{2}\right)
$$

Next

$$
\dot{\alpha} \times \ddot{\alpha}=18\left|\begin{array}{ccc}
E_{1} & E_{2} & E_{3} \\
1-t^{2} & 2 t & 1+t^{2} \\
-t & 1 & t
\end{array}\right|=18\left(-1+t^{2},-2 t, 1+t^{2}\right)
$$

and hence

$$
\|\dot{\alpha} \times \ddot{\alpha}\|=18 \sqrt{2}\left(1+t^{2}\right)
$$

The expressions above for $\dot{\alpha} \times \ddot{\alpha}$ and $\dddot{\alpha}$ yield

$$
(\dot{\alpha} \times \ddot{\alpha}) \bullet \dddot{\alpha}=216 .
$$

It remains only to substitute this data into the formulas above, with $N$ being computed by another cross product. The final results are

$$
\begin{aligned}
T & =\frac{\left(1-t^{2}, 2 t, 1+t^{2}\right)}{\sqrt{2}\left(1+t^{2}\right)} \\
N & =\frac{\left(-2 t, 1-t^{2}, 0\right)}{1+t^{2}} \\
B & =\frac{\left(-1+t^{2},-2 t, 1+t^{2}\right)}{\sqrt{2}\left(1+t^{2}\right)} \\
\kappa & =\tau=\frac{1}{3\left(1+t^{2}\right)}
\end{aligned}
$$

$\diamond$ Exercise 112 Compute the Serret-Frenet apparatus for each of the following (regular) curves :
(a) $\alpha(t)=\left(e^{t} \cos t, e^{t} \sin t, e^{t}\right)$.
(b) $\beta(t)=(\cosh t, \sinh t, t) \quad($ the hyperbolic helix $)$.
$\diamond$ Exercise 113 If $\alpha$ is a regular curve with constant speed $c>0$, show that

$$
T=\frac{1}{c} \dot{\alpha} ; \quad N=\frac{\ddot{\alpha}}{\|\ddot{\alpha}\|} ; \quad B=\frac{\dot{\alpha} \times \ddot{\alpha}}{c\|\ddot{\alpha}\|} ; \quad \kappa=\frac{1}{c^{2}}\|\ddot{\alpha}\| ; \quad \tau=\frac{(\dot{\alpha} \times \ddot{\alpha}) \bullet \dddot{\alpha}}{c^{2}\|\ddot{\alpha}\|^{2}} .
$$

$\diamond$ Exercise 114 Consider the unit speed helix

$$
\beta(s)=\left(a \cos \frac{s}{k}, a \sin \frac{s}{k}, \frac{b s}{k}\right), \quad k=\sqrt{a^{2}+b^{2}}
$$

and define the curve $\sigma:=\dot{\beta}$, the spherical image of $\beta$. (For every s , the point $\sigma(s)$ lies on the unit sphere $\mathbb{S}^{2}$, and the motion of $\sigma$ represents the turning of $\beta$.) Show that

$$
\kappa_{\sigma}=\sqrt{1+\left(\frac{\tau_{\beta}}{\kappa_{\beta}}\right)^{2}} \geq 1
$$

(and thus depends only on the ratio of torsion to curvature for the original curve).

## Some implications of curvature and torsion

There are instances in which the ratio of torsion to curvature (for a certain curve) plays an important role (see Exercise 107). This ratio can be used to characterize an entire class of regular curves, called cylindrical helices.

Exercise 115 Consider the standard helix

$$
\alpha(t)=(a \cos t, a \sin t, b t) .
$$

Verify that the angle $\theta$ between the unit tangent vector $T=\frac{\dot{\alpha}}{\|\dot{\alpha}\|}$ of $\alpha$ and the standard unit vector $e_{3}$ is constant.

A cylindrical helix is a generalization of a standard (or circular) helix. We make the following definition.
2.6.6 Definition. A regular curve $\alpha: J \rightarrow \mathbb{R}^{3}$ is called a cylindrical helix provided the unit tangent vector $T$ of $\alpha$ has constant angle $\theta$ with some fixed unit vector $u$; that is, $T(t) \bullet u=\cos \theta$ for all $t \in J$.

The condition is not altered by reparametrization, so without loss of generality we may assume that cylindrical helices have unit speed. Cylindrical helices can be identified by a simple condition on torsion and curvature.
2.6.7 Proposition. Let $\beta: J \rightarrow \mathbb{R}^{3}$ be a unit speed curve with nonzero curvature. Then
$\beta$ is a cylindrical helix if and only if $\frac{\tau}{\kappa}$ is constant.
Proof : $(\Rightarrow)$ If $\beta$ is a cylindrical helix with $T \bullet u=\cos \theta$, then

$$
0=(T \bullet u)^{\cdot}=\dot{T} \bullet u=\kappa N \bullet u
$$

so $N \bullet u=0$ since $\kappa>0$. The unit vector $u$ is orthogonal to $N$ and hence

$$
\begin{aligned}
u & =(u \bullet T) T+(u \bullet B) B \\
& =\cos \theta T+\sin \theta B .
\end{aligned}
$$

By differentiating we obtain

$$
\begin{aligned}
0 & =\cos \theta \dot{T}+\sin \theta \dot{B} \\
& =\kappa \cos \theta N-\tau \sin \theta N \\
& =(\kappa \cos \theta-\tau \sin \theta) N .
\end{aligned}
$$

Thus $\kappa \cos \theta-\tau \sin \theta=0$ which gives

$$
\frac{\tau}{\kappa}=\cot \theta=\text { constant. }
$$

$(\Leftarrow)$ Now suppose $\frac{\tau}{\kappa}$ is constant. Choose an angle $\theta$ such that $\cot \theta=\frac{\tau}{\kappa}$. Define $U:=\cos \theta T+\sin \theta B$ to get

$$
\dot{U}=(\kappa \cos \theta-\tau \sin \theta) N=0 .
$$

This parallel vector field $U$ then determines a unit vector $u$ such that

$$
T \bullet u=\cos \theta .
$$

Thus $\beta$ is a cylindrical helix.

Note : A regular curve with nonzero curvature is a circular helix if and only if both $\tau$ and $\kappa$ are constant. Also, it can be shown that a regular curve is a cylindrical helix if and only if its spherical image is part of a circle.
$\diamond$ Exercise 116 Show that the curve

$$
\alpha(t)=\left(a t, b t^{2}, t^{3}\right)
$$

is a cylindrical helix if and only if $4 b^{4}=9 a^{2}$. In this case, find the vector $u$ such that $T \bullet u=$ constant.

## Plane Curves

Recall that a plane curve in $\mathbb{R}^{3}$ is a curve that lies entirely in a single plane of $\mathbb{R}^{3}$. The theory of plane curves can be viewed as a special case of the theory of curves in $\mathbb{R}^{3}$.

Note : The Euclidean plane $\mathbb{R}^{2}$ can be embedded in $\mathbb{R}^{3}$ and thus identified with a subset (plane) of $\mathbb{R}^{3}$. For instance, we can think of $\mathbb{R}^{2}$ as the $x_{1} x_{2}$-plane of $\mathbb{R}^{3}$; that is, we identify the Euclidean plane $\mathbb{R}^{2}$ with the plane $\left\{x=\left(x_{1}, x_{2}, x_{3}\right) \mid x_{3}=0\right\} \subset \mathbb{R}^{3}$. Another option is to identify $\mathbb{R}^{2}$ with the plane $\left\{x=\left(x_{1}, x_{2}, x_{3}\right) \mid x_{1}=1\right\} \subset \mathbb{R}^{3}$. In this case, it is convenient to represent the point $\left(p_{1}, p_{2}\right)$ of $\mathbb{R}^{2}$ as the column matrix (vector) $\left[\begin{array}{c}1 \\ p_{1} \\ p_{2}\end{array}\right]$.

We can also give an independent treatment of plane curves; this approach has the advantage that the plane $\mathbb{R}^{2}$ can be taken to be oriented. An orientation of $\mathbb{R}^{2}$ may be given by fixing a positive frame at a point $p \in \mathbb{R}^{2}$; an obvious choice is the natural frame $\left(e_{1}, e_{2}\right)$ at the origin.

Let $\beta: J \rightarrow \mathbb{R}^{2}$ an oriented unit speed curve and denote by $T$ the unit tangent vector field on $\beta: T:=\dot{\beta}$. We define the normal vector field (on $\beta$ ) $N$ by requiring the oriented frame field $(T, N)$ to have the "same orientation" as the plane $\mathbb{R}^{2}$.

Then the Serret-Frenet formulas become

$$
\begin{aligned}
\dot{T} & =\kappa N \\
\dot{N} & =-\kappa T
\end{aligned}
$$

where the real-valued function $\kappa: J \rightarrow \mathbb{R}$ is the (signed) curvature of $\beta$.
Note : The curvature $\kappa$ might be either positive or negative. It is clear that $|\kappa|$ agrees with the curvature in the case of space curves and that $\kappa$ changes sign when we change either the orientation of $\beta$ or the orientation of $\mathbb{R}^{2}$.
$\diamond$ Exercise 117 Show that if $\beta=\left(\beta_{1}, \beta_{2}\right)$ is a unit speed curve in $\mathbb{R}^{2}$, then

$$
T(s)=\left(\dot{\beta}_{1}(s), \dot{\beta}_{2}(s)\right) \quad \text { and } \quad N(s)=\left(-\dot{\beta}_{2}(s), \dot{\beta}_{1}(s)\right) .
$$

$\diamond$ Exercise 118 Show that the regular plane curve $\alpha(t)=\left(x_{1}(t), x_{2}(t)\right)$ has curvature

$$
\kappa(t)=\frac{\dot{x}_{1}(t) \ddot{x}_{2}(t)-\ddot{x}_{1}(t) \dot{x}_{2}(t)}{\left(\dot{x}_{1}^{2}(t)+\dot{x}_{2}^{2}(t)\right)^{3 / 2}}
$$

at $\alpha(t)$.
2.6.8 Example. The curvature of the (standard) ellipse $\alpha(t)=(a \cos t, b \sin t)$ is given by

$$
\kappa(t)=\frac{a b}{\left(a^{2} \sin ^{2} t+b^{2} \cos ^{2} t\right)^{3 / 2}}=\frac{a b}{\sqrt{\left(a^{2}+\left(b^{2}-a^{2}\right) \cos ^{2} t\right)^{3}}} .
$$

Recall that $0<b<a$ and hence observe that that the curvature achieves a minimum when $t= \pm \frac{\pi}{2}$ and a maximum when $t=0$ or $\pi$.

Consider the example of a (regular) plane curve given by the graph of a (differentiable) function

$$
\alpha(t)=(t, f(t)) .
$$

Here $\|\dot{\alpha}(t)\|=\sqrt{1+\dot{f}^{2}}$ and so we can compute

$$
\begin{aligned}
T & =\left(\frac{1}{\sqrt{1+\dot{f}^{2}(t)}}, \frac{\dot{f}(t)}{\sqrt{1+\dot{f}^{2}(t)}}\right) \\
N & =\left(\frac{-\dot{f}(t)}{\sqrt{1+\dot{f}^{2}(t)}}, \frac{1}{\sqrt{1+\dot{f}^{2}(t)}}\right) \\
\dot{T} & =\frac{\ddot{f}(t)}{1+\dot{f}^{2}(t)}\left(\frac{-\dot{f}(t)}{\sqrt{1+\dot{f}^{2}(t)}}, \frac{1}{\sqrt{1+\dot{f}^{2}(t)}}\right) .
\end{aligned}
$$

Hence

$$
\kappa(t)=\frac{\ddot{f}(t)}{\left(1+\dot{f}^{2}(t)\right)^{3 / 2}}
$$

Observe that the sign of the curvature is determined by the second derivative $\ddot{f}(t)$, which is positive if $f(t)$ is concave up, negative if $f(t)$ is concave down. Since any curve in $\mathbb{R}^{2}$ is locally the graph of a function, we see that the signed curvature at a point is positive if the curve turns to left of the tangent, negative if to the right.

Exercise 119 Compute the curvature of the semicircle

$$
x_{2}=\sqrt{a^{2}-x_{1}^{2}} .
$$

Exercise 120 Show that the curvature of the (cuspidal) cycloid $t \mapsto(t-$ $\sin t, 1-\cos t$ ) (at a regular value $t$ ) is given by

$$
\kappa(t)=-\frac{1}{4 \sin \frac{t}{2}} .
$$

$\diamond$ Exercise 121 Find a formula for the curvature of the parabola $x_{1}=a t^{2}, x_{2}=$ $2 a t$ (with $a>0$ ). Show that the vertex is the unique point on the parabola where the curvature assumes a maximal value.

In the case of a plane curve, the proof of the Fundamental Theorem is actually very simple.
$\diamond$ Exercise 122 Given a smooth function $\kappa:(a, b) \rightarrow \mathbb{R}$, show that the plane curve $\beta:(a, b) \rightarrow \mathbb{R}^{2}$, parametrized by arc length and having $\kappa$ as (directed) curvature function, is given by

$$
\beta(s)=\left(\int \cos \theta(s) d s+c_{1}, \int \sin \theta(s) d s+c_{2}\right)
$$

where

$$
\theta(s)=\int \kappa(s) d s+\varphi .
$$

Furthermore, any other curve curve $\bar{\beta}$ satisfying the same conditions, differs from $\beta$ by a rotation of angle $\varphi$ followed by a translation by vector $c=\left[\begin{array}{l}c_{1} \\ c_{2}\end{array}\right]$.

