Chapter 1. Vector valued functions

This material is covered in Thomas (chapters 12 & 13 in the 11th edition, or chapter 10 in the 10th edition). The approach here will be only slightly different and we will not cover every detail in the book. On the web site `http://www.maths.tcd.ie/~richardt/2E1` there is a more detailed explanation of the sections of those chapters we are covering.

One aspect of this course is advanced or multivariable calculus. This means generalising what was covered in 1E1/1E2 dealing with single variable calculus, developing similar but different (and somewhat more complicated) ideas in several variables. So, while we dealt with function $f \colon \mathbb{R} \to \mathbb{R}$ in single variable calculus (and also functions which were defined on domains like intervals that are smaller than all of the real line $\mathbb{R}$), we will now start to deal with functions $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}^m$ where both the variables and the values can be vectors (or tuples of several numbers).

This is not totally new as you have dealt with the method of Gaussian elimination to solve systems of linear equations. One can view a system of m linear equations in n unknowns as a special kind of vector equation $\mathbf{f}(\mathbf{x}) = \mathbf{y}$, where $\mathbf{f}(\mathbf{x})$ represents the left hand sides of the equations and $\mathbf{y}$ the right hand sides. What we had in that case is a *linear function* $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}^m$, and a linear equation. We will progress by studying functions that are not necessarily linear and later we will look at the topic of linear equations (linear algebra) in more depth than we did before.

Why might an engineer be interested in these topics? The answer is that the world of an engineer is rarely 1-dimensional. You could say that the world around us is 3-dimensional (which it is) but many engineering problems involve complicated machines or structures where there are many different variables — inputs, measurements or outputs and controls. Consider a flying plane, where you can easily imagine several variables in addition to the position of (the center of mass of) the plane in space. Direction, banking angle, flap position, rudder angle, engine power, wind speed, ... A car would be maybe simpler, but still there are many variables. For the structural engineer, there could be many stresses and strains in different parts of a building (or structure like a bridge). These are in fact varying depending on the conditions (or the usage) of the structure. I guess you all realise that even buildings are not just static — quite a lot of research is considering the dynamics of (say) buildings, how this affects the deterioration of the components of the structure, and so on. A valid mathematical model needs to have lots of variables or parameters. And you know that most engineers end up working things out by calculation (with computers), which means using mathematical models, though they will eventually build something to check that the theory really works in practice.

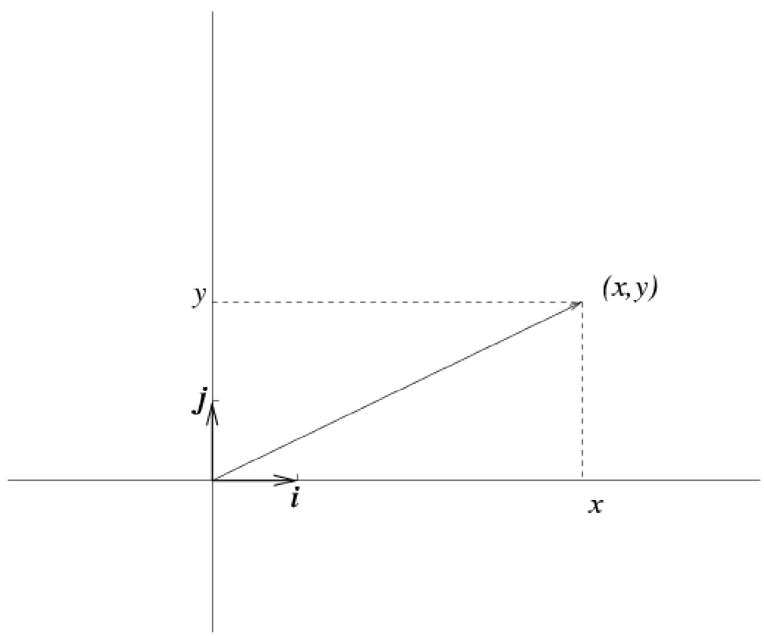
In this chapter we will start with functions of one variable, with vector values.

1.1 Remark. Recall that $\mathbb{R}^n$ means the set of all n-tuples $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of real numbers $x_1, x_2, \ldots, x_n$.

When $n = 1$, $\mathbb{R}^1 = \mathbb{R}$ can be pictured as points on an axis. (One coordinate or one variable.)

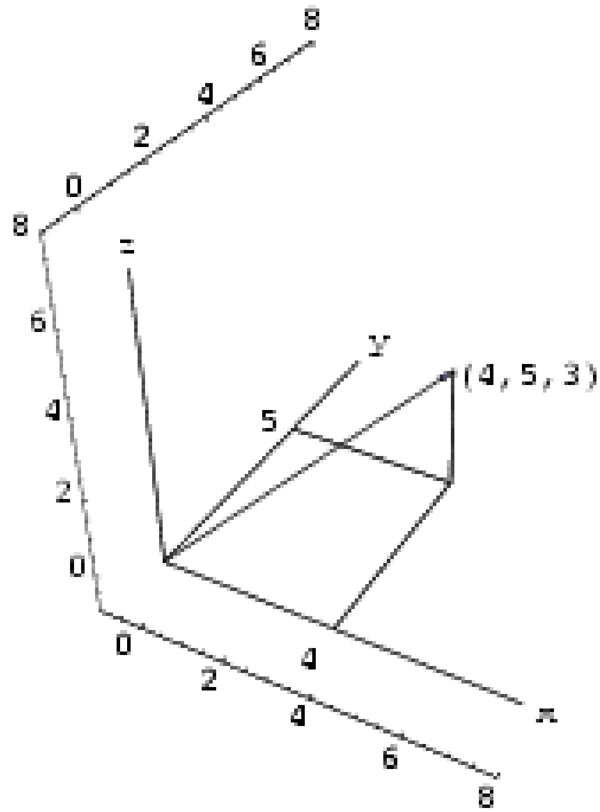
For $n = 2$, we also have a geometrical view of $\mathbb{R}^2$ as the points in a plane where we have selected two axes (perpendicular axes) and points $\mathbf{x} = (x_1, x_2)$ in the plane are labeled by their coordinates with respect to the two axes. Alternatively, we can think in terms of vectors. We can think of the position vector of $\mathbf{x} = (x_1, x_2)$ as the arrow (or directed line segment) from the origin $(0, 0)$ to $\mathbf{x}$. Using the rules of vector addition and multiplication of vectors by scalars we can also write $\mathbf{x} = x_1\mathbf{i} + x_2\mathbf{j}$ where $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$ are 'unit' vectors in the directions of the two axes.

We could use $x\mathbf{i} + y\mathbf{j} = (x, y)$ instead of (x_1, x_2) in 2-dimensions, but the reason for using the $x_1, x_2, \ldots$ notation is partly to make it easier to have enough letters for the different points (or vectors) we might want to consider at once.



When we go the $\mathbb{R}^3$ (or $n = 3$ dimensions) we also have a geometrical view (the space around us) we can use. We need to fix an origin and 3 perpendicular axes (the x_1, x_2 and x_3 axes, or the x, y and z axes if you prefer) meeting at the origin. Once we have the axes we can identify every point in space by 3 coordinates (or we can identify the position vector of the point as the vector with those 3 components).

Perhaps a handy way to think is as follows. Think of the first two axes as lying along the floor, pointing east and north say. (We mean a flat floor here, extending infinitely far.) Using two coordinates with respect to these two axes we can identify any point we want on the floor. Then we can use a third number to give a height (altitude) above the floor. If this third number is negative it corresponds to going down below the floor level. This way, using 3 coordinates we can identify any point in space (with reference to the 3 fixed axes).



We should recall the way to add points (or vectors) in $\mathbb{R}^n$, and to multiply by scalars. There is also the *dot product* or scalar product, the distance formula and the length or *magnitude* of a vector. If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ and $\lambda \in \mathbb{R}$, then

$$\begin{aligned}
\mathbf{x} + \mathbf{y} &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \\
\lambda\mathbf{x} &= (\lambda x_1, \lambda x_2, \ldots, \lambda x_n) \\
\mathbf{x} \cdot \mathbf{y} &= x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \\
&= \sum_{i=1}^{n} x_i y_i \\
\mathrm{dist}(\mathbf{x}, \mathbf{y}) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \\
\|\mathbf{x}\| &= \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \\
&= \sqrt{\mathbf{x} \cdot \mathbf{x}} \\
&= \mathrm{dist}(\mathbf{x}, (0, 0, \ldots, 0)).
\end{aligned}$$

In 2 or 3 dimensions we can think of arrows representing vectors and visualise $\mathbf{x} + \mathbf{y}$ as the resultant arrow, got by putting $\mathbf{x}$ and $\mathbf{y}$ nose to tail (or from the parallelogram law, the diagonal of the parallelogram with $\mathbf{x}$ and $\mathbf{y}$ as two side). We can think of $\lambda\mathbf{x}$ as got by stretching $\mathbf{x}$ by a factor λ if $\lambda > 0$, and if $\lambda < 0$ by first reversing the arrow and then stretching by a factor $|\lambda|$.

This latter is consistent with the following general fact:

$$\|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\| \quad (\lambda \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n).$$

We cannot visualise vector algebra in dimensions 4 or higher — but we usually think by analogy with the 3 dimensional setting.

In 2 or 3 dimensions we can prove that the distance formula is correct (relying on a picture and Pythagoras' theorem), but in n dimensions we take the formula as a definition of something we still refer to as 'distance'. In 2 or 3 dimensions we can prove that

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta$$

where θ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. In higher dimensions we take this formula as a definition of the angle θ. (We need to keep θ in the range 0 to π or it would be ambiguously defined. We also need to have nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ for $\theta = \cos^{-1}(\mathbf{x} \cdot \mathbf{y}/(\|\mathbf{x}\| \|\mathbf{y}\|)$ to make sense — to avoid having any division by zero.)

We say that $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are *perpendicular* or *orthogonal* if $\mathbf{x} \cdot \mathbf{y} = 0$. This is the same as having the angle between $\mathbf{x}$ and $\mathbf{y}$ being $\theta = \pi/2$, except that we define the zero vector $\mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{R}^n$ to be perpendicular to every vector.

1.2 Definition. For a vector valued function $\mathbf{x} \colon \mathbb{R} \to \mathbb{R}^n$ of a single variable, we write

$$\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$$

where $x_1(t), x_2(t), \ldots, x_n(t)$ are n ordinary functions (scalar valued functions of a single variable). We can allow $\mathbf{x}(t)$ to be defined only for some subset of $t \in \mathbb{R}$ (say for t in some interval).

We define the *derivative* of $\mathbf{x}(t)$ to be

$$\frac{d\mathbf{x}}{dt} = \left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots \frac{dx_n}{dt} \right)$$

(so the vector-valued function got by differentiating the components or coordinates of $\mathbf{x}(t)$ individually).

1.3 Remarks. There may be an advantage in using other notations like the prime notation $\mathbf{x}'(t)$ or the dot notation $\dot{\mathbf{x}}(t)$ (which some people like to use for a time derivative). Perhaps

$$\mathbf{x}'(t) = (x_1'(t), x_2'(t), \ldots, x_n'(t))$$

looks more tidy than the version with d/dt notation above.

We often think of $\mathbf{x}(t)$ as a point moving is space, where t represents time. (This is not the only way to think of it, but it is often helpful.) We might imagine the case $n = 2$ where we draw with a pen on a flat paper, and $\mathbf{x}(t)$ records where the tip of the pen is at time t. When we are finished drawing, the pen will have followed some track or path around the page and we will see the resulting curve drawn. The function $\mathbf{x}(t)$ describes the curve you see in that picture, but describes more in fact, because it says how you drew the curve, how the pen moved, which direction the pen went along the curve.

In $n = 3$ dimensions, we can have a similar view of $\mathbf{x}(t)$ as tracing out a curve in space (though perhaps we could find it harder to imagine drawing in space). This idea is more or less exactly the idea of a parametric curve, or parametric equations for a curve.

1.4 Examples. (i) Consider $n = 2$ and the function

$$\mathbf{x}(t) = (5\cos t, 5\sin t),$$

that is the case where $x_1(t) = 5\cos t$ and $x_2(t) = 5\sin t$. If you recall $\cos^2 t + \sin^2 t = 1$, you can quickly see that the points $(x, y) = \mathbf{x}(t) = (5\cos t, 5\sin t)$ all satisfy

$$x^2 + y^2 = 5^2\cos^2 t + 5^2\sin^2 t = 5^2(\cos^2 t + \sin^2 t) = 5^2$$

and so they are all on the circle of radius 5 about the origin in the plane $\mathbb{R}^2$.

We might express this by saying that

$$\begin{cases} x = 5\cos t \\ y = 5\sin t \end{cases}$$

are parametric equations for a curve in the circle or radius 5, but in fact we can say more precisely that $\mathbf{x}(t)$ is the point on the circle of radius 5 around $(0, 0)$ that is an angle t radians[1] anticlockwise around the circle from $(5, 0)$.

In this case we see that once t reaches 2π (from 0) we get back around the circle and we just keep going around more an more times as t increases.

(ii) A somewhat similar example is

$$\mathbf{x}(t) = (3\cos t, 7\sin t).$$

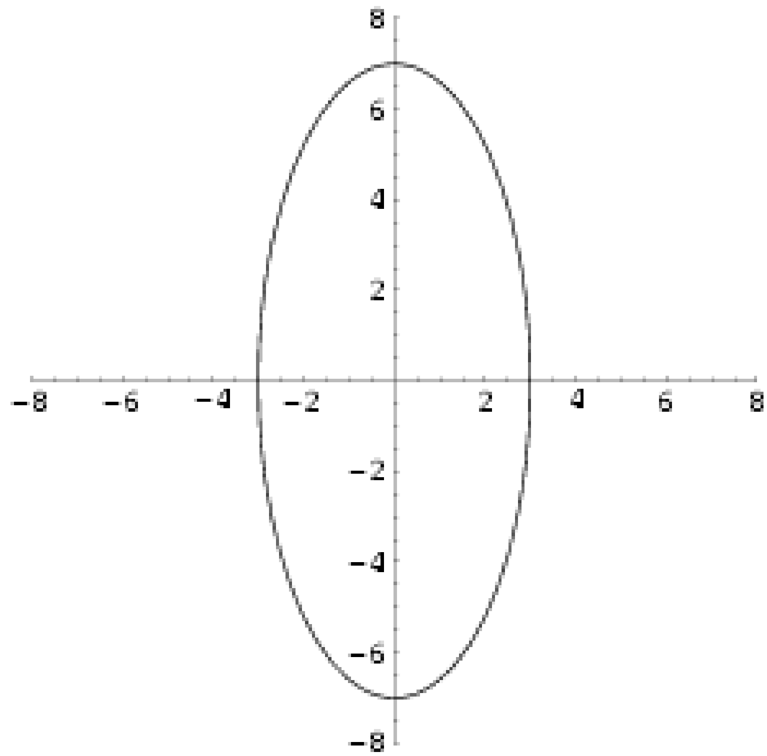
The points $(x, y) = \mathbf{x}(t) = (3\cos t, 7\sin t)$ all satisfy

$$\frac{x^2}{3^2} + \frac{y^2}{7^2} = \cos^2 t + \sin^2 t = 1$$

and so they are all on the curve with equation $\frac{x^2}{3^2} + \frac{y^2}{7^2} = 1$.

This curve is called an ellipse and looks like this:

[1] This is a good place to remind you that calculus with trigonometric functions always needs to be using radians. If we used degrees all the differentiation formulae would have factors of $\pi/180$ in them. The answers you get if you use the trigonometric functions with degrees will usually be irredeemably wrong — can't be related to the correct answer except by starting over again and doing the calculation in radians. One way to say it is that radians are dimensionless, but degrees are not.



It is again the case that after t runs from 0 to 2π, the point $\mathbf{x}(t)$ will have travelled all the way around the ellipse. But you cannot say where $\mathbf{x}(t)$ is on the ellipse using angles.

(iii) If $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ and $\mathbf{b} \neq \mathbf{0}$, then we can quite easily see that the points

$$\mathbf{x}(t) = \mathbf{a} + t\mathbf{b}$$

lie on the line through the point $\mathbf{a}$ in the direction parallel to $\mathbf{b}$. As t varies through all of $\mathbb{R}$ (positive, zero and negative values) we get all the points on that line (once).

If $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$, then what we have is that

$$\begin{cases} x &= a_1 + b_1 t \\ y &= a_2 + b_2 t \\ z &= a_3 + b_3 t \end{cases}$$

are parametric equations for the line in $\mathbb{R}^3$.

In this case $d\mathbf{x}/dt = \mathbf{b}$.

1.5 Remark. We could have defined the derivative $d\mathbf{x}/dt = \mathbf{x}'(t)$ using first principles as

$$\frac{d\mathbf{x}}{dt} = \mathbf{x}'(t) = \lim_{\Delta t \to 0} \frac{\mathbf{x}(t + \Delta t) - \mathbf{x}(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{x}}{\Delta t}$$

(where $\Delta \mathbf{x}$ just means $\mathbf{x}(t + \Delta t) - \mathbf{x}(t)$, the change in $\mathbf{x}$ that happens when t changes from t to $t + \Delta t$.

The justification for this limit formula (which we will not go into) is not very complicated. Basically the limits pass through to the coordinates or components — and we get in component number i the limit that defines $dx_i/dt = x_i'(t)$.

From this limit definition, we can argue that in dimensions $n = 2$ and $n = 3$, the limiting vector should have a direction tangent to the curve described by $\mathbf{x}(t)$. For Δt small, $\mathbf{x}(t+\Delta t) - \mathbf{x}(t)$ is a short chord of the curve, and dividing by Δt makes it longer. We can see that the vector will be almost tangent to the curve when Δt is small. In the limit, it should be exactly tangent.

In case there is any doubt about this argument, we take the derivative vector as our definition of tangent vector.

1.6 Definition. If $\mathbf{x}\colon \mathbb{R} \to \mathbb{R}^n$ is a vector function of a single variable $t \in \mathbb{R}$, thus describing a parametric curve in $\mathbb{R}^n$, we define the *tangent vector* to the curve at $t = t_0$ to be

$$\frac{d\mathbf{x}}{dt} \mid_{t=t_0} = \mathbf{x}'(t_0)$$

and we define the *unit tangent vector to the curve* to be

$$\mathbf{T} = \mathbf{T}(t) = \frac{\frac{d\mathbf{x}}{dt}}{\left\|\frac{d\mathbf{x}}{dt}\right\|} = \frac{\mathbf{x}'(t)}{\|\mathbf{x}'(t)\|}.$$

1.7 Remark. We can't define the unit tangent vector if $d\mathbf{x}/dt$ is the zero vector. We have to exclude this case, that is exclude the points where $d\mathbf{x}/dt = 0$.

1.8 Remark. When $n = 2$ or $n = 3$ (or even for $n = 1$) we can think of a vector function $\mathbf{x}\colon \mathbb{R} \to \mathbb{R}^n$ as representing the position $\mathbf{x}(t)$ at time t of a moving particle. Then the derivative vector $d\mathbf{x}/dt$ is the rate of change of position $\mathbf{x}$ with respect to time t, and we **define** the (instantaneous) *velocity* of the particle at time t to be

$$\mathbf{v} = \mathbf{v}(t) = \frac{d\mathbf{x}}{dt}.$$

A rationale for this definition can be given, similar to what you know from one dimensional motion, based on $d\mathbf{x}/dt = \lim_{\Delta t \to 0} \Delta\mathbf{x}/\Delta t$. $\Delta\mathbf{x} = \mathbf{x}(t+\Delta t) - \mathbf{x}(t)$ is the displacement (or change of position) of the particle over the time interval from t to $t + \Delta t$ and $\Delta\mathbf{x}/\Delta t$ is the average velocity over the time interval (displacement divided my elapsed time). The limit is the (definition of) the velocity at the instant t.

The *speed* is the magnitude $\|\mathbf{v}(t)\|$ of the velocity.

1.9 Theorem (Chain rule)**.** *Suppose given a vector function $\mathbf{x}(t)$ where $t = t(u)$ is a function of another real variable u. Then ultimately $\mathbf{x}$ depends on u and we have two possible derivatives of $\mathbf{x}$ (with respect to t and to u). Then*

$$\frac{d\mathbf{x}}{du} = \frac{d\mathbf{x}}{du}\frac{du}{dt}$$

(In other words the chain rule says it is ok to cancel the dt's. We could alternatively, more precisely, write $\mathbf{y}(u) = \mathbf{x}(t(u))$ for the composition of the functions $u \mapsto t(u)$ and $t \mapsto \mathbf{x}(t)$. The Chain Rule expresses the derivative of the composition $\mathbf{y}'(u) = \mathbf{x}'(t(u))t'(u)$ as the product of the two derivatives, the vector $\mathbf{x}'$ and the scalar $t'(u) = dt/du$.)

Proof. The proof is just to write out

$$\begin{aligned}
\mathbf{x}(t) &= (x_1(t), x_2(t), \ldots, x_n(t)) \\
\mathbf{x}(t(u)) &= (x_1(t(u)), x_2(t(u)), \ldots, x_n(t(u))) \\
\frac{d\mathbf{x}}{du} &= \left(\frac{dx_1}{du}, \frac{dx_2}{du}, \ldots, \frac{dx_n}{du}\right).
\end{aligned}$$

Then, using the ordinary chain rule n times we can say

$$\frac{dx_i}{du} = \frac{dx_i}{dt}\frac{dt}{du} \qquad (i = 1, 2 \ldots, n).$$

Thus

$$\begin{aligned}
\frac{d\mathbf{x}}{du} &= \left(\frac{dx_1}{du}, \frac{dx_2}{du}, \ldots, \frac{dx_n}{du}\right) \\
&= \left(\frac{dx_1}{dt}\frac{dt}{du}, \frac{dx_2}{dt}\frac{dt}{du}, \ldots, \frac{dx_n}{dt}\frac{dt}{du}\right) \\
&= \left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_n}{dt}\right)\frac{dt}{du} \\
&= \frac{d\mathbf{x}}{dt}\frac{dt}{du}. \qquad \square
\end{aligned}$$

(We will not emphasise proofs much, but this proof amounts to working out the two sides of the equation to be proved, and then invoking something we know to show the two sides are always equal.)

1.10 Remark. If we consider the situation of the chain rule again for a moment, we see that we have two parametric curves $t \mapsto \mathbf{x}(t)$ and $u \mapsto \mathbf{y}(u) = \mathbf{x}(t(u))$, which describe the same points in $\mathbb{R}^n$. To be more precise, the curve which is the image of the map $u \mapsto \mathbf{y}(u) = \mathbf{x}(t(u))$ is at least contained in the image curve of the map $t \mapsto \mathbf{x}(t)$.

The chain rules implies that the tangent vectors will be in the same direction if $dt/du > 0$ and opposite directions if $dt/du < 0$. We have

$$\begin{aligned}
\frac{d\mathbf{y}}{du} &= \frac{dt}{du}\frac{d\mathbf{x}}{dt} \\
\left\|\frac{d\mathbf{y}}{du}\right\| &= \left|\frac{dt}{du}\right|\left\|\frac{d\mathbf{x}}{dt}\right\|
\end{aligned}$$

and so if we compute the unit tangent vector for $\mathbf{y}$ we get

$$\mathbf{T} = \mathbf{T}(u) = \frac{\frac{d\mathbf{y}}{du}}{\left\|\frac{d\mathbf{y}}{du}\right\|} = \frac{\frac{dt}{du}}{\left|\frac{dt}{du}\right|}\frac{\frac{d\mathbf{x}}{dt}}{\left\|\frac{d\mathbf{x}}{dt}\right\|}$$

and will be the same as the unit tangent for $\mathbf{x}$ at t if $dt/du > 0$.

This tells us that the unit tangent vector depends on the curve, rather than on the way it is parametrised, except that the direction travelled along the curve is important.

1.11 Theorem (Product Rules). *Let $\mathbf{x}\colon \mathbb{R} \to \mathbb{R}^n$ and $\mathbf{y}\colon \mathbb{R} \to \mathbb{R}^n$ be two (differentiable) functions and $\alpha\colon \mathbb{R} \to \mathbb{R}$ a scalar-valued (also differentiable) function. Then*

(i)

$$\frac{d}{dt}(\mathbf{x}(t)\cdot\mathbf{y}(t)) = \frac{d\mathbf{x}}{dt}\cdot\mathbf{y}(t) + \mathbf{x}(t)\cdot\frac{d\mathbf{y}}{dt}$$

(ii)

$$\frac{d}{dt}(\alpha(t)\mathbf{x}(t)) = \frac{d\alpha}{dt}\mathbf{x}(t) + \alpha(t)\frac{d\mathbf{x}}{dt}$$

Proof. We won't give the proof, but it is not really complicated. If we write out the two sides of the equations to be proved using coordinates (of $\mathbf{x}$ and $\mathbf{y}$), they follow from the usual product rule for scale functions (applied n times). □

1.12 Remark. In $n = 3$ dimensions there is another product we could consider, the cross product of vectors. There is also a product rule for that, but we need to be careful because the cross product depends on the order of the vectors and in the product rule you need to keep the vector functions in order:

$$\frac{d}{dt}(\mathbf{x}(t)\times\mathbf{y}(t)) = \frac{d\mathbf{x}}{dt}\times\mathbf{y}(t) + \mathbf{x}(t)\times\frac{d\mathbf{y}}{dt}.$$

This might be a good place to recall the definition of the cross product $\mathbf{x}\times\mathbf{y}$ of two 3-dimensional vectors $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$. One way to recall the formula is to write $\mathbf{x}\times\mathbf{y}$ as a kind of determinant to be worked out by cofactor expansions

$$\mathbf{x}\times\mathbf{y} = \det\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix} = \mathbf{i}(x_2y_3 - x_3y_2) - \mathbf{j}(x_1y_3 - x_3y_1) + \mathbf{k}(x_1y_2 - x_2y_1)$$

or

$$\mathbf{x}\times\mathbf{y} = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1).$$

The properties of the cross product are that it is perpendicular to both vectors $\mathbf{x}$ and $\mathbf{y}$, $\|\mathbf{x}\times\mathbf{y}\| = \|\mathbf{x}\|\|\mathbf{y}\|\sin\theta$ where θ is the angle between the vectors.

If one uses a right-handed coordinate system (which means that a right-handed screw placed along the x_3 axis and turned from the x_1 towards the x_2 axis will go in the positive direction along the x_3 axis) then the direction of $\mathbf{x}\times\mathbf{y}$ is also given by a right-handed rule.

We have $\mathbf{x}\times\mathbf{y} = -(\mathbf{y}\times\mathbf{x})$, so that the order of the vectors makes a difference.

1.13 Proposition. *Let $\mathbf{x}\colon \mathbb{R} \to \mathbb{R}^n$ be a (differentiable) vector function and $\mathbf{T}(t) = \mathbf{x}'(t)/\|\mathbf{x}'(t)\|$ be the unit tangent vector for $\mathbf{x}$ (defined as long as $\mathbf{x}'(t) \neq \mathbf{0}$). Then*

$$\frac{d\mathbf{T}}{dt}\cdot\mathbf{T} = 0$$

Proof. Recall that, as $\mathbf{T} = \mathbf{T}(t)$ is a unit vector

$$\mathbf{T} \cdot \mathbf{T} = \|\mathbf{T}\|^2 = 1.$$

Differentiating both sides with respect to t and using the product rule we get

$$\begin{aligned} \frac{d}{dt}(\mathbf{T} \cdot \mathbf{T}) &= \frac{d}{dt}1 \\ \frac{d\mathbf{T}}{dt} \cdot \mathbf{T} + \mathbf{T} \cdot \frac{d\mathbf{T}}{dt} &= 0 \\ 2\frac{d\mathbf{T}}{dt} \cdot \mathbf{T} &= 0 \\ \frac{d\mathbf{T}}{dt} \cdot \mathbf{T} &= 0 \end{aligned}$$

Thus $d\mathbf{T}/dt$ is perpendicular to $\mathbf{T}$ (for all t). □

1.14 Definition. The unit vector in the direction of $d\mathbf{T}/dt$ is called the *unit normal vector* and is denoted $\mathbf{N}$.

We have

$$\mathbf{N} = \frac{\frac{d\mathbf{T}}{dt}}{\left\| \frac{d\mathbf{T}}{dt} \right\|}.$$

(At some stage we should point out that we can't always do this. If $d\mathbf{T}/dt$ is the zero vector, we can't find the unit normal $\mathbf{N}$. In fact, a similar problem applies to the unit tangent vector. At some points on a curve, any points where $\mathbf{x}'(t) = \mathbf{0}$, $\mathbf{T}$ will not be defined.)

1.15 Remark. We've been considering vector-valued functions $\mathbf{x} \colon \mathbb{R} \to \mathbb{R}^n$, which we may also think of as describing a parametric curve in $\mathbb{R}^n$. Sometimes we want to concentrate on the curve that is traced out, the image curve of the function $\mathbf{x}$, without much reference to the way it is parametrised (or precisely how the point $\mathbf{x}(t)$ moves as t varies).

We've seen that the velocity vector $\mathbf{v}(t) = \mathbf{x}'(t) = d\mathbf{x}/dt$ is tangent to the curve and the unit tangent vector is something that depends only on the curve — rather than how it is paramatrised. To be precise the direction we travel along the curve matters. If the curve is reparametrised

$$\mathbf{y}(u) = \mathbf{x}(t(u)),$$

then $\mathbf{y}'(u) = \mathbf{x}'(t(u))\frac{dt}{du}$ will have exactly the opposite direction to $\mathbf{x}(t)$ if $dt/du < 0$.

One way to get away from ideas that depend on how a curve is parametrised and to get to something that depends only on the curve itself is to reparametrise in a standard way — by the arc length parameter, measuring how far we have travelled along the curve from some starting point.

To explain this, we start by working out a formula for the length of a curve.

1.16 Remark. Consider a finite section of a curve $\mathbf{x} = \mathbf{x}(t)$ (in $\mathbb{R}^n$) where t varies $a \leq t \leq b$. We are going to motivate the formula for the length of the curve. For the motivation, we stick to $n = 2$ and $n = 3$ where we can see things geometrically.

What we want is a formula for the distance travelled by a particle that follows the trajectory $\mathbf{x} = \mathbf{x}(t)$ (as time varies $a \leq t \leq b$). If you think of a path in the plane, it is the distance you travel if you walk along the path. Or if you unrolled a piece of thread (or string) along the path (no slipping, so that at the end the shape of the thread is the path) then we want a formula for the total length of thread you would need.

What we do is divide the path into many very short sections $\mathbf{x}(t)$ to $\mathbf{x}(t + \Delta t)$ and use the approximation that over such a short distance we can take the length of the curve to be the straight line distance between the end points. That is

$$\|\mathbf{x}(t + \Delta t) - \mathbf{x}(t)\| = \Delta\mathbf{x} \cong \|\mathbf{x}'(t)\,\Delta t\| = \|\mathbf{x}'(t)\|\,\Delta t$$

We add all these small lengths up to get the total length. To do it more accurately we should take Δt very small (and so need correspondingly more little lengths in the sum). In the limit, what we get is an integral.

In the end, we take this as only a motivation and define the length as follows.

1.17 Definition. If $\mathbf{x} = \mathbf{x}(t)$ $(a \leq t \leq b)$ is a (differentiable) $\mathbb{R}^n$-valued function (parametric path) then the *arc length* is defined as

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

(Note that it may help to remember this as the integral of the speed.)

1.18 Example. Find the length of the curve

$$\mathbf{x}(t) = (\cos t, \sin t, t) \qquad (0 \leq t \leq 2\pi)$$

We apply our formula, which involves the speed. So we need to work out the velocity $\mathbf{x}'(t)$ and the speed $\|\mathbf{x}'(t)\|$ first.

$$\begin{aligned}
\mathbf{x}'(t) &= (-\sin t, \cos t, 1)\\
\|\mathbf{x}'(t)\| &= \sqrt{\sin^2 t + \cos^2 t + 1}\\
&= \sqrt{2}\\
\text{length} &= \int_0^{2\pi} \|\mathbf{x}'(t)\|\,dt\\
&= \int_0^{2\pi} \sqrt{2}\,dt\\
&= \sqrt{2}(2\pi) = 2\sqrt{2}\pi
\end{aligned}$$

1.19 Arc Length Parametrisation. We now explain the idea of reparametrising a curve in terms of distance travelled rather than by the original parameter t (which we may think of as time). You could maybe imagine a boat that sailed around the coast and logged its position at every instant. In this way we have a recording $\mathbf{x}(t)$ of the position at all times t. Or think of a car journey where the car has some kind of GPS system recording its position as it drove from Dublin to Belfast. We are going to record the position as a function of the distance s travelled (since some starting point).

In this way we get to a description of the course followed by the boat, or the route followed by the car, which is independent of the speed the boot (or car) was traveling at along the way. If we want to, we can remember that the distance travelled s on the particular journey did depend on time and so we get $s = s(t)$ is a function of time.

We can use our arc length formula to say how s depends on t, if we measure distance from time $t = 0$. It gives

$$s(t) = \int_0^t \|\mathbf{x}'(u)\|\,du. \tag{1}$$

(The justification for using the parameter u in this length formula is, first, the name of the running variable in a definite integral is immaterial. The definite integral has a numerical value and does not depend on the name given to the variable of integration. But secondly, and more crucially, the variable t is not available to us here as a running variable as t is already in use as the upper limit of the integral.)

Now we can (usually) solve this for $t = t(s)$. We then can express $\mathbf{x} = \mathbf{x}(t) = \mathbf{x}(t(s))$ as a function of s. This is what we mean by reparametrisation by arc length.

For us it is important to know ds/dt. There is a theorem called the Fundamental theorem of integral calculus[2] that tells us

$$\frac{ds}{dt} = \|\mathbf{x}'(t)\|. \tag{2}$$

A consequence of this is that $ds/dt \geq 0$ always, and so $s(t)$ is an increasing function of t. If $\|\mathbf{x}'(t)\| > 0$ for all t, then $s(t)$ is a strictly increasing function of t, and that allows us to justify the claim that we can solve for t in terms of s.

We can rephrase (2) as saying that

$$\frac{ds}{dt} = \text{ speed}$$

or that the rate of change of distance travelled is speed. (The speedometer in a car know how far the car has travelled, but does not know the direction. So it does not know the velocity vector of the car, but counts distance travelled in terms of the number of turns of the wheels. Working by the rate distance travelled is increasing, it figures out ds/dt, or speed.)

[2] A bad explanation may convince you. Imagine working out the integral for $s(t)$ by funding an antiderivative for the integrand. So we find $g(u)$ so that $g'(u) = \|\mathbf{x}'(u)\|$. Then $s(t) = [g(u)]_{u=0}^t = g(t) - g(0)$. If we now differentiate this, we get $ds/dt = g'(t) - 0 = \|\mathbf{x}'(t)\|$. The reason that this is a bad explanation is that the Fundamental theorem is what allows you to evaluate definite integrals in terms of antiderivatives.

We can also work out velocity in the new arclength paramatrisation of the curve. We find, using the Chain rule.

$$\frac{d\mathbf{x}}{ds} = \frac{d\mathbf{x}}{dt}\frac{dt}{ds} = \frac{d\mathbf{x}}{dt}\Big/\frac{ds}{dt} = \frac{\frac{d\mathbf{x}}{dt}}{\text{speed}} = \frac{\mathbf{x}'(t)}{\|\mathbf{x}'(t)\|} = \mathbf{T}.$$

Thus, in the arclength parametrisation, the velocity is a unit vector, and the speed as measured with respect to s (rather than t) is $\|d\mathbf{x}/ds\| = \|\mathbf{T}\| = 1$. In other words the arclength parametrisation is a *unit speed* parametrisation.

[Though it is a small point, we should maybe point out that s is not quite arc length. For $t > 0$, $s(t)$ is the distance travelled along the curve $\mathbf{x} = \mathbf{x}(t)$ since $t = 0$. But for $t < 0$, the integral (1) is going backwards from 0, and so we have $s(t) < 0$ for $t < 0$. We can say that $s(t)$ is minus the length of the path from time t to 0 when $t < 0$. So s has a sign.]

1.20 Curvature. We now consider the meaning of the rate of change of the unit tangent vector. If $\mathbf{T}$ is constant, it means that the velocity vector always has the same directions, and it is not hard to believe that this can only happen if the curve is a straight line. So if $\mathbf{T}$ is changing, then the path $\mathbf{x} = \mathbf{x}(t)$ is changing direction, and we might believe that the rate of change of the unit tangent vector has something to do with how fast the path is changing direction or how curved it is.

In fact $d\mathbf{T}/dt$ is a vector, and it is always perpendicular to $\mathbf{T}$ according to Proposition 1.13. The magnitude of the vector $d\mathbf{T}/dt$ depends on how we parametrise the curve as well as on the intrinsic shape of the curve. To get to something that depends on the curve only we consider the derivative of $\mathbf{T}$ with respect to arclength s.,

$$\frac{d\mathbf{T}}{ds} = \frac{d\mathbf{T}}{dt}\frac{dt}{ds} = \frac{d\mathbf{T}}{dt}\Big/\frac{ds}{dt} = \frac{\mathbf{T}'(t)}{\|\mathbf{x}'(t)\|}$$

We call the magnitude of the the curvature and denote it by κ. In summary

$$\text{curvature } = \kappa = \left\|\frac{d\mathbf{T}}{ds}\right\| = \frac{\left\|\frac{d\mathbf{T}}{dt}\right\|}{\|\mathbf{x}'(t)\|}.$$

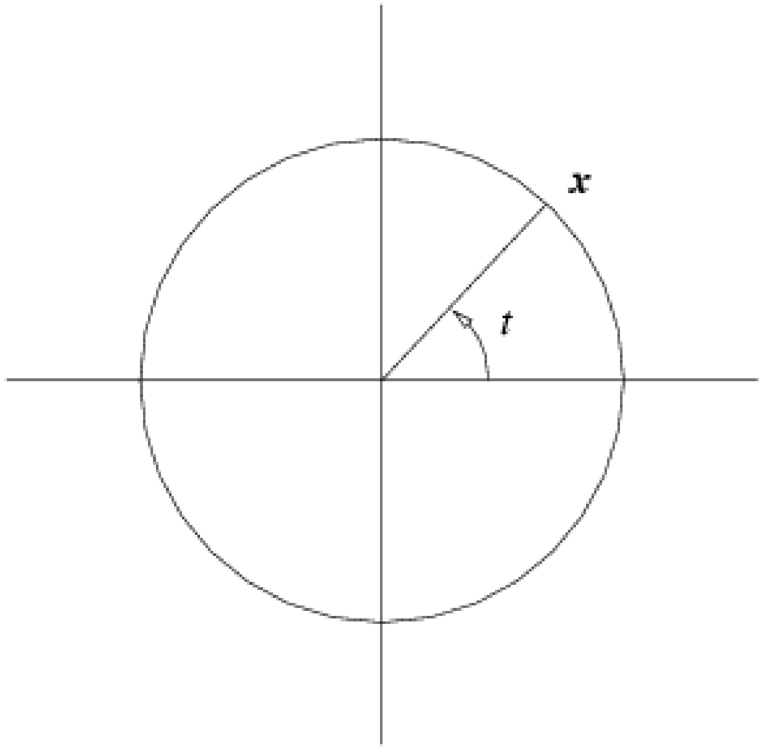
Notice that $d\mathbf{T}/ds$ and $d\mathbf{T}/dt$ have the same direction and so the unit normal vector $\mathbf{N}$ is the unit vector in the direction of $d\mathbf{T}/ds$. We can say

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N} \tag{3}$$

1.21 Example. We consider the curve in $\mathbb{R}^2$ given by

$$\mathbf{x}(t) = (r\cos t, r\sin t)$$

(where $r > 0$ is a constant). It is easy to understand what this curve is. All the points of the curve are on the circle $x_1^2 + x_2^2 = r^2$ of radius r about the origin $(0, 0)$. In fact we can say exactly where $\mathbf{x}(t)$ is.



$\mathbf{x}(t)$ is the point which is t radians around the circle, in the anticlockwise direction if $t > 0$, from the positive x_1-axis. (We saw a similar thing with $r = 5$ in Examples 1.4 (i).)

We can work out all the quantities we have been discussing quite easily in this example (velocity vector $\mathbf{x}'(t)$, arc length parameter $s = s(t)$, unit tangent vector $\mathbf{T}$, unit normal vector $\mathbf{N}$ and curvature κ).

$$\begin{aligned} \mathbf{x}'(t) &= (-r\sin t, r\cos t) \\ \text{speed} &= \|\mathbf{x}'(t)\| \\ &= \sqrt{r^2\sin^2 t + r^2\cos^2 t} \\ &= \sqrt{r^2(\sin^2 t + \cos^2 t)} = \sqrt{r^2} \\ &= r \\ s(t) &= \int_{u=0}^{t} \|\mathbf{x}'(u)\|\,du \\ &= \int_{u=0}^{t} r\,du = rt \end{aligned}$$

So we can (in this case very easily and explicitly) solve for $t = s/r$. The arc length parametrisation gives

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

and we can check that

$$\frac{d\mathbf{x}}{ds} = \left(-r\frac{1}{r}\sin\frac{s}{r}, r\frac{1}{r}\cos\frac{s}{r}\right) = \left(-\sin\frac{s}{r}, \cos\frac{s}{r}\right)$$

is a unit vector (as it is guaranteed to be by the theory above). It is the vector $\mathbf{T}$ expressed in term of s rather than t. We can also see that $\mathbf{T}$ is perpendicular to the radius vector $\mathbf{x}$, as it is supposed to be from the theory, because the tangent to the circle is perpendicular to the radius. We can calculate $d\mathbf{T}/ds$ directly in this case

$$\begin{aligned}\mathbf{T} &= \frac{d\mathbf{x}}{ds} \\ &= \left(-\sin\frac{s}{r}, \cos\frac{s}{r}\right) \\ \frac{d\mathbf{T}}{ds} &= \left(-\frac{1}{r}\cos\frac{s}{r}, -\frac{1}{r}\sin\frac{s}{r}\right) \\ &= -\frac{1}{r}\left(\cos\frac{s}{r}, \sin\frac{s}{r}\right)\end{aligned}$$

We can then compute that the curvature

$$\kappa = \left\|\frac{d\mathbf{T}}{ds}\right\| = \frac{1}{r}$$

is the reciprocal of the radius.

(The idea is that a circle of a big radius seems less curved than a small circle. Think of the equator of the earth, which is almost a circle. It seems hardly curved at all — well you can see it curves by looking at the horizon. Or think of a corner on a road. A gentle corner will be one with a big radius, and the car can easily go around at full speed. A roundabout will often be quite tight and the car will have to slow down to negotiate it. Hence the idea that the reciprocal of the radius represents the amount of curvature.)

In this example we can also see that the the unit normal vector, the unit vector in the direction of $d\mathbf{T}/ds$ (or $d\mathbf{T}/dt$) is

$$\mathbf{N} = -\left(\cos\frac{s}{r}, \sin\frac{s}{r}\right).$$

It is in the opposite direction to $\mathbf{x}$ and so points in along the radius, towards the center from $\mathbf{x}$.

While we could, in this case, explicitly find t in terms of s, and so work out derivatives with respect to s explicitly, usually we would have a hard time finding a formula for $t = t(s)$. We'll repeat the calculation of κ and $\mathbf{N}$ in a way more like what we would do in more complicated examples.

$$
\begin{aligned}
\mathbf{x}'(t) &= (-r\sin t, r\cos t)\\
\|\mathbf{x}'(t)\| &= r\\
\mathbf{T} &= \mathbf{x}'(t)/\|\mathbf{x}'(t)\|\\
&= \mathbf{x}'(t)/r\\
&= (-\sin t, \cos t)\\
\frac{d\mathbf{T}}{dt} &= (-\cos t, -\sin t)\\
\kappa &= \|d\mathbf{T}/ds\|\\
&= \|d\mathbf{T}/dt\|/(ds/dt)\\
&= \|d\mathbf{T}/dt\|/\|\mathbf{x}'(t)\|\\
&= \frac{\sqrt{\cos^2 t+\sin^2 t}}{r} = \frac{1}{r}\\
\mathbf{N} &= d\mathbf{T}/dt/\|d\mathbf{T}/dt\|\\
&= d\mathbf{T}/dt\\
&= (-\cos t, -\sin t)
\end{aligned}
$$

1.22 Circle of curvature. It is possible to relate the fact that the curvature of a circle is $1/$radius to the curvature of a general curve.

Given a (fairly nice) curve $\mathbf{x} = \mathbf{x}(t)$ in $\mathbb{R}^2$ or $\mathbb{R}^3$ (or even in $\mathbb{R}^n$) and a point $\mathbf{x}_0 = \mathbf{x}(t_0)$ on the curve there is exactly one circle that matches the curve more closely than any other at the point $\mathbf{x}_0$. There is a circle through $\mathbf{x}_0$ with the same first derivative and second derivative as $\mathbf{x}$, called the osculating circle of the curve at the point $\mathbf{x}_0$.

The radius of this circle is called the radius of curvature and the curvature of this circle (the reciprocal of its radius) is the curvature κ of the curve.

We will not go into this any more.

1.23 Example. We consider the curve in $\mathbb{R}^3$ given by

$$\mathbf{x}(t) = (r\cos t, r\sin t, t)$$

(where $r > 0$ is fixed).

It is not too hard to understand what this curve looks like. The first two coordinates are the same as the earlier example. So they move around on a circle of radius r, centre the origin, in the horizontal plane. At $t = 0$, we are at $(r, 0, 0)$. As t increases, the first two coordinates move around the circle (t radians around from $(r, 0)$) and then the point we have in $\mathbb{R}^3$ is climbing t units up from the horizontal. Looked at from directly above we will see a circle, but the curve is a *helix*, like a corkscrew or the thread on a screw.

We can calculate

$$\begin{aligned}
\mathbf{x}'(t) &= (-r\sin t, r\cos t, 1)\\
\|\mathbf{x}'(t)\| &= \sqrt{r^2\sin^2 t + r^2\cos^2 t + 1}\\
&= \sqrt{r^2+1}\\
\mathbf{T} &= \mathbf{x}'(t)/\|\mathbf{x}'(t)\|\\
&= \mathbf{x}'(t)/\sqrt{r^2+1}\\
&= \left(-\frac{r}{\sqrt{r^2+1}}\sin t, \frac{r}{\sqrt{r^2+1}}\cos t, \frac{1}{\sqrt{r^2+1}}\right)\\
\frac{d\mathbf{T}}{dt} &= \left(-\frac{r}{\sqrt{r^2+1}}\cos t, -\frac{r}{\sqrt{r^2+1}}\sin t, 0\right)\\
\kappa &= \|d\mathbf{T}/ds\|\\
&= \|d\mathbf{T}/dt\|/(ds/dt)\\
&= \|d\mathbf{T}/dt\|/\|\mathbf{x}'(t)\|\\
&= \frac{\frac{r}{\sqrt{r^2+1}}\sqrt{\cos^2 t+\sin^2 t}}{\sqrt{r^2+1}}\\
&= \frac{r}{r^2+1}\\
\mathbf{N} &= \frac{d\mathbf{T}}{dt}\Big/\left\|\frac{d\mathbf{T}}{dt}\right\|\\
&= \frac{d\mathbf{T}}{dt}\Big/\frac{r}{\sqrt{r^2+1}}\\
&= (-\cos t, -\sin t, 0)
\end{aligned}$$

1.24 Tangential and normal components of acceleration. Given a particle moving in space $\mathbb{R}^3$ (or the plane $\mathbb{R}^2$) so that its position at time t is given by $\mathbf{x} = \mathbf{x}(t)$, we know its velocity vector at time t is

$$\mathbf{v}(t) = \mathbf{x}'(t)$$

and its acceleration is

$$\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{x}''(t).$$

We also know that the speed is $\|\mathbf{v}(t)\|$ and we can write the velocity as speed times the unit tangent vector:

$$\mathbf{v}(t) = \|\mathbf{v}(t)\|\mathbf{T}.$$

If we differentiate both sides of this equation using the product rule (the one for scalar times vector), we get

$$\mathbf{a}(t) = \mathbf{v}'(t) = \frac{d\|\mathbf{v}(t)\|}{dt}\mathbf{T} + \|\mathbf{v}(t)\|\frac{d\mathbf{T}}{dt}.$$

We know that $d\mathbf{T}/dt$ is perpendicular to $\mathbf{T}$, or normal to the tangent direction. So we see that the acceleration is the rate of change of speed times the unit tangent vector plus another contribution in the normal direction.

This normal contribution is to do with the change of direction. Recall that a particle will move at constant speed in a straight line (that is, at constant velocity) if it is not accelerating (not subject to any force). So a particle moving around a circle at a constant speed will have zero rate of change of speed, and so zero tangential contribution to $\mathbf{a}(t)$, but it will still be accelerating because of the changing direction of its velocity.

We can relate the normal part of $\mathbf{a}(t)$ to the unit normal vector $\mathbf{N}$ and the curvature κ. We have

$$\frac{d\mathbf{T}}{dt} = \frac{d\mathbf{T}}{ds}\frac{ds}{dt} = \frac{d\mathbf{T}}{ds}\|v(t)\|$$

and the magnitude of $d\mathbf{T}/ds$ is the curvature κ. Thus

$$\frac{d\mathbf{T}}{dt} = \kappa\mathbf{N}/\|v(t)\|$$

and so we can say

$$\mathbf{a}(t) = \frac{d\|\mathbf{v}(t)\|}{dt}\mathbf{T} + \kappa\|\mathbf{v}(t)\|^2\mathbf{N}.$$

We could write

$$a_T = \frac{d\|\mathbf{v}(t)\|}{dt}, \quad a_N = \kappa\|\mathbf{v}(t)\|^2 \text{ and } \mathbf{a} = a_T\mathbf{T} + a_N\mathbf{N}.$$

Then a simple calculation with dot products via $\|\mathbf{a}\|^2 = \mathbf{a}\cdot\mathbf{a}$ (or a use of Pythagoras' theorem) shows that

$$\|\mathbf{a}\| = \sqrt{a_T^2 + a_N^2}.$$

In particular the magnitude of the acceleration vector will be larger than the magnitude of the rate of change of speed $|a_T| = \left|\frac{d}{dt}\|\mathbf{v}(t)\|\right|$ in general.