Differential Geometry

Lecture Notes

Dr. Silvio Fanzon

2 Nov 2023

Academic Year 2023/24

Department of Mathematics

University of Hull

Table of contents

Welcome

These are the Lecture Notes of **Differential Geometry 661955** for T1 2023/24 at the University of Hull. We will study curves and surfaces in $\mathbb{R}^3$. I will follow these lecture notes during the course. If you have any question or find any typo, please email me at

S.Fanzon@hull.ac.uk

Up to date information about the course, Tutorials and Homework will be published on the University of Hull **Canvas Website**

canvas.hull.ac.uk/courses/67594

and on the **Course Webpage** hosted on my website

silviofanzon.com/blog/2023/Differential-Geometry

Digital Notes

Digital version of these notes available at

silviofanzon.com/2023-Differential-Geometry-Notes

Readings

Main textbooks:

- Pressley [6] for differential geometry,
- Manetti [5] for general topology.

Other interesting readings are the books by do Carmo [2] and Abate, Tovena [1]. I will assume some knowledge from Analysis and Linear Algebra. A good place to revise these topics are the books by Zorich [7, 8].

Visualization

It is important to visualize the geometrical objects and concepts we are going to talk about in this course. I will show basic Python code to plot curves and surfaces. This part of the course is `not required` for the final examination. If you want to have fun plotting with Pyhton, I recommend installation through Anaconda or Miniconda. The actual coding can then be done through Jupyter Notebook. Good references for scientific Python programming are [3, 4].
If you do not want to mess around with Python, you can still visualize pretty much everything we will do in this course using the excellent online 3D grapher tool CalcPlot3D. To understand how it works, please refer to the help manual or to the short video introduction. Another nice tool is Desmos.

! You are not expected to purchase any of the above books. These lecture notes will cover 100% of the topics you are expected to known in order to excel in the final exam.

1 Curves

Curves are, intuitively speaking, 1D objects in the 2D or 3D space. For example in two dimensions one could think of a straight line, a hyperbole or a circle. These can be all described by an equation in the x and y coordinates: respectively

$$y = 2x + 1\,, \quad y = e^x\,, \quad x^2 + y^2 = 1\,.$$

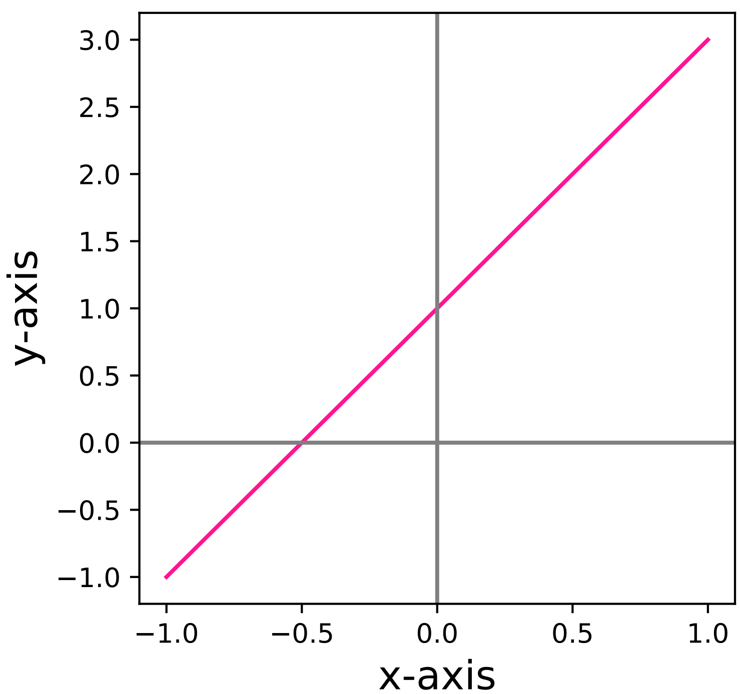


Figure 1.1: Plotting straight line $y = 2x + 1$

Goal

The aim of this course is to study curves by differentiating them.

Question

In what sense do we differentiate the above curves?

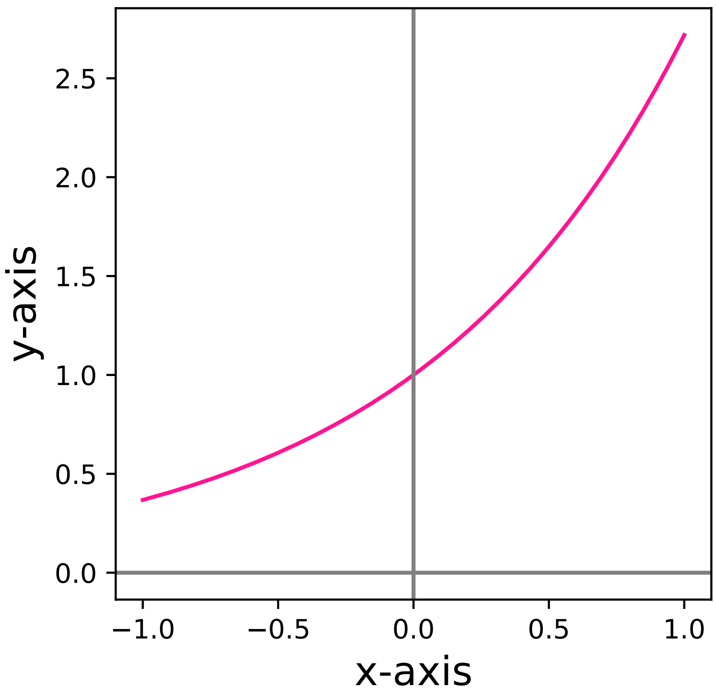


Figure 1.2: Plot of hyperbole $y = e^x$

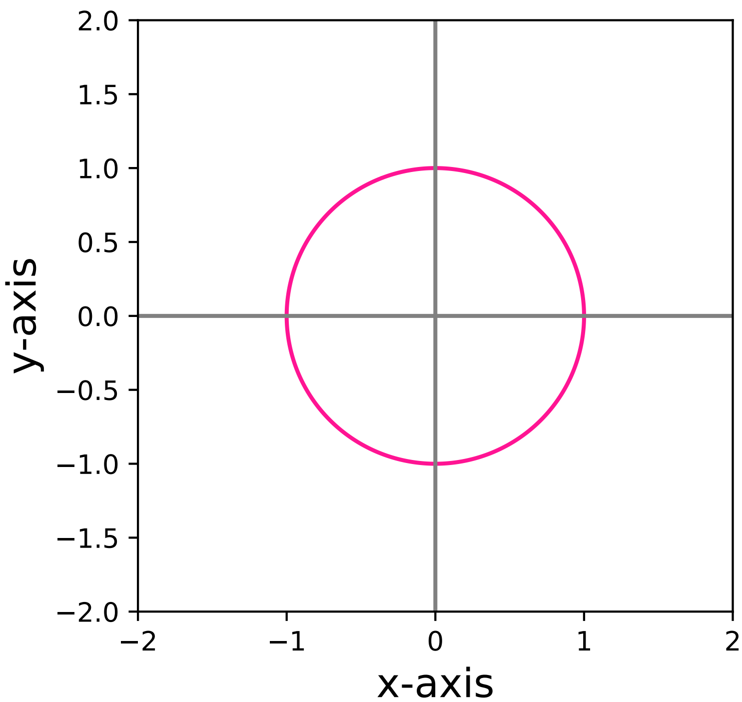


Figure 1.3: Plot of unit circle of equation $x^2 + y^2 = 1$

It is clear that we need a way to mathematically describe the curves. One way of doing it is by means of Cartesian equations. This means that the curve is described as the set of points $(x, y) \in \mathbb{R}^2$ where the equation

$$f(x, y) = c\,,$$

is satisfied, where

$$f : \mathbb{R}^2 \to \mathbb{R}\,.$$

is some given function, and

$$c \in \mathbb{R}$$

some given value. In other words, the curve is identified with the subset of $\mathbb{R}^2$ given by

$$C = \{(x, y) \in \mathbb{R}^2 \,:\, f(x, y) = c\}\,.$$

For example, in the case of the straight line, we would have

$$f(x, y) = y - 2x\,, \quad c = 1\,.$$

while for the circle

$$f(x, y) = x^2 + y^2\,, c = 1\,.$$

But what about for example a helix in 3 dimensions? It would be more difficult to find an equation of the form

$$f(x, y, z) = 0$$

to describe such object.

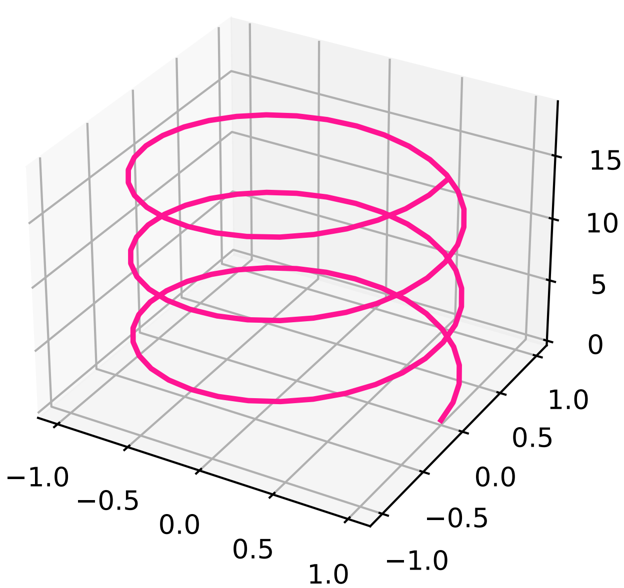


Figure 1.4: Plot of a 3D Helix

Problem

We need a unified way to describe curves.

1.1 Parametrized curves

Rather than Cartesian equations, a more useful way of thinking about curves is viewing them as the *path traced out by a moving point*. If $\boldsymbol{\gamma}(t)$ represents the position a point in $\mathbb{R}^n$ at time t, the whole curve can be identified by the function

$$\boldsymbol{\gamma} \,:\, \mathbb{R} \to \mathbb{R}^n \,, \;\; \boldsymbol{\gamma} = \boldsymbol{\gamma}(t) \,.$$

This motivates the following definition of **parametrized curve**, which will be our **main** definition of curve.

Definition 1.1: Parametrized curve

A **parametrized curve** in $\mathbb{R}^n$ is a function

$$\boldsymbol{\gamma} \,:\, (a,b) \to \mathbb{R}^n \,.$$

where

$$-\infty \leq a < b \leq \infty \,.$$

A few remarks:

- The symbol (a,b) denotes an **open** interval

$$(a,b) = \{t \in \mathbb{R} \,:\, \; a < t < b\} \,.$$

- The requirement that

$$-\infty \leq a < b \leq \infty$$

  means that the interval (a,b) is possibly unbounded.
- For each $t \in (a,b)$ the quantity $\boldsymbol{\gamma}(t)$ is a vector in $\mathbb{R}^n$.
- The **components** of $\boldsymbol{\gamma}(t)$ are denoted by

$$\boldsymbol{\gamma}(t) = (\gamma_1(t), \dots, \gamma_n(t)) \,,$$

  where the components are functions

$$\gamma_i \,:\, (a,b) \to \mathbb{R} \,,$$

  for all $i = 1, \dots, n$.

1.2 Parametrizing Cartesian curves

At the start we said that examples of curves in $\mathbb{R}^2$ were the straight line, the hyperbole and the circle, with equations

$$y = 2x + 1 \,, \quad y = e^x \,, \quad x^2 + y^2 = 1 \,.$$

We saw that these can be represented by Cartesian equations

$$f(x, y) = c$$

for some function $f : \mathbb{R}^2 \to \mathbb{R}$ and value $c \in \mathbb{R}$. Curves that can be represented in this way are called **level curves**. Let us give a precise definition.

Definition 1.2: Level curve

A **level curve** in $\mathbb{R}^n$ is a set $C \subset \mathbb{R}^n$ which can be described as

$$C = \{(x_1, \dots, x_n) \in \mathbb{R}^n \;:\; f(x_1, \dots, x_n) = c\}$$

for some given function

$$f : \mathbb{R}^n \to \mathbb{R}$$

and value

$$c \in \mathbb{R} \,.$$

We now want to represent level curves by means of parametrizations.

Definition 1.3

Suppose given a level curve $C \subset \mathbb{R}^n$. We say that a curve

$$\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^n$$

parametrizes C if

$$C = \{(\gamma_1(t), \dots, \gamma_n(t)) \;:\; t \in (a, b)\} \,.$$

Question

Can we **represent** the level curves we saw above by means of a parametrization $\boldsymbol{\gamma}$?

The answer is YES, as shown in the following examples.

Example 1.4: Parametrizing the straight line

The straight line
$$y = 2x + 1$$
is a **level curve** with
$$C = \{(x, y) \in \mathbb{R}^2 \ : \ f(x, y) = c\} ,$$
where
$$f(x, y) := y - 2x , \quad c := 1 .$$
How do we represent C as a **parametrized curve** $\boldsymbol{\gamma}$? We know that the curve is 2D, therefore we need to find a function
$$\boldsymbol{\gamma} \ : \ (a, b) \to \mathbb{R}^2$$
with componenets
$$\boldsymbol{\gamma}(t) = (\gamma_1(t), \gamma_2(t)) .$$
The curve $\boldsymbol{\gamma}$ needs to be chosen so that it parametrizes the set C, in the sense that
$$C = \{(\gamma_1(t), \gamma_2(t)) \ : \ t \in (a, b)\} . \tag{1.1}$$
Thus we need to have
$$(x, y) = (\gamma_1, \gamma_2) . \tag{1.2}$$
How do we define such $\boldsymbol{\gamma}$? Note that the points (x, y) in C satisfy
$$(x, y) \in C \iff y = 2x + 1 .$$
Therefore, using (1.2), we have that
$$\gamma_1 = x , \quad \gamma_2 = y = 2x + 1$$
from which we deduce that $\boldsymbol{\gamma}$ must satisfy
$$\gamma_2(t) = 2\gamma_1(t) + 1 \tag{1.3}$$
for all $t \in (a, b)$. We can then choose
$$\gamma_1(t) := t ,$$
and from (1.3) we deduce that
$$\gamma_2(t) = 2t + 1 .$$
This choice of $\boldsymbol{\gamma}$ works:
$$\begin{aligned} C &= \{(x, 2x + 1) \ : \ x \in \mathbb{R}\} && (1.4) \\ &= \{(t, 2t + 1) \ : \ -\infty < t < \infty\} && (1.5) \\ &= \{(\gamma_1(t), \gamma_2(t)) \ : \ -\infty < t < \infty\} , && (1.6) \end{aligned}$$
where in the second line we just swapped the symbol x with the symbol t. In this case we have to choose the time interval as
$$(a, b) = (-\infty, \infty) .$$
In this way $\boldsymbol{\gamma}$ satisfies (1.1) and we have successfully parametrized the straight line C.

Remark 1.5: Parametrization is not unique

Let us consider again the straight line

$$C = \{(x, y) \in \mathbb{R}^2 \ : \ 2x + 1 = y\}.$$

We saw that $\boldsymbol{\gamma} : (-\infty, \infty) \to \mathbb{R}^2$ defined by

$$\boldsymbol{\gamma}(t) := (t, 2t + 1)$$

is a parametrization of C. But of course any $\boldsymbol{\gamma}$ satisfying

$$\gamma_2(t) = 2\gamma_1(t) + 1$$

would yield a parametrization of C. For example one could choose

$$\gamma_1(t) = 2t\,, \quad \gamma_2(t) = 2\gamma_1(t) + 1 = 4t + 1\,.$$

In general, any time rescaling would work: the curve $\boldsymbol{\gamma}$ defined by

$$\gamma_1(t) = nt\,, \quad \gamma_2(t) = 2\gamma_1(t) + 1 = 2nt + 1$$

parametrizes C for all $n \in \mathbb{N}$. Hence there are **infinitely many** parametrizations of C.

Example 1.6: Parametrizing the circle

The circle C is described by all the points $(x, y) \in \mathbb{R}^2$ such that

$$x^2 + y^2 = 1\,.$$

Therefore if we want to find a curve

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$$

which parametrizes C, this has to satisfy

$$\gamma_1(t)^2 + \gamma_2(t)^2 = 1 \tag{1.7}$$

for all $t \in (a, b)$.
How to find such curve? We could proceed as in the previous example, and set

$$\gamma_1(t) := t\,.$$

Then (1.7) implies

$$\gamma_2(t) = \sqrt{1 - t^2}\,,$$

from which we also deduce that

$$-1 \leq t \leq 1$$

are the only admissible values of t. However this curve does not represent the full circle C, but only the upper half, as seen in the plot below.
Simlarly, another solution to (1.7) would be $\boldsymbol{\gamma}$ with

$$\gamma_1(t) = t\,, \quad \gamma_2(t) = -\sqrt{1-t^2}\,,$$

for $t \in [-1,1]$. However this choice does not parametrize the full circle C either, but only the bottom half, as seen in the plot below.
How to represent the whole circle? Recall the trigonometric identity

$$\cos(t)^2 + \sin(t)^2 = 1$$

for all $t \in \mathbb{R}$. This suggests to choose $\boldsymbol{\gamma}$ as

$$\gamma_1(t) := \cos(t)\,, \quad \gamma_2(t) := \sin(t)$$

for $t \in [0, 2\pi)$. This way $\boldsymbol{\gamma}$ satisfies (1.7), and actually parametrizes C, as shown below.
Note the following:

- If we had chosen $t \in [0, 4\pi]$ then $\boldsymbol{\gamma}$ would have covered C twice.
- If we had chosen $t \in [0, \pi]$, then $\boldsymbol{\gamma}$ would have covered the upper semi-circle
- If we had chosen $t \in [\pi, 2\pi]$, then $\boldsymbol{\gamma}$ would have covered the lower semi-circle
- Similarly, we can choose $t \in [\pi/6, \pi/2]$ to cover just a portion of C, as shown below.

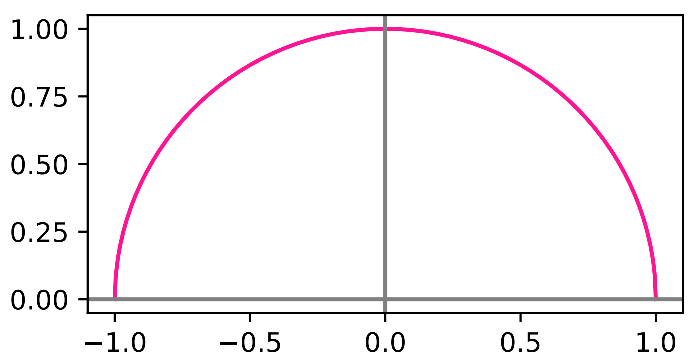


Figure 1.5: Upper semi-circle

Finally we are also able to give a mathematical description of the 3D Helix.

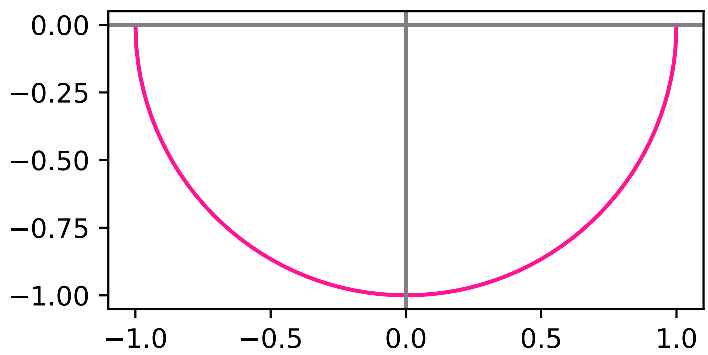


Figure 1.6: Lower semi-circle

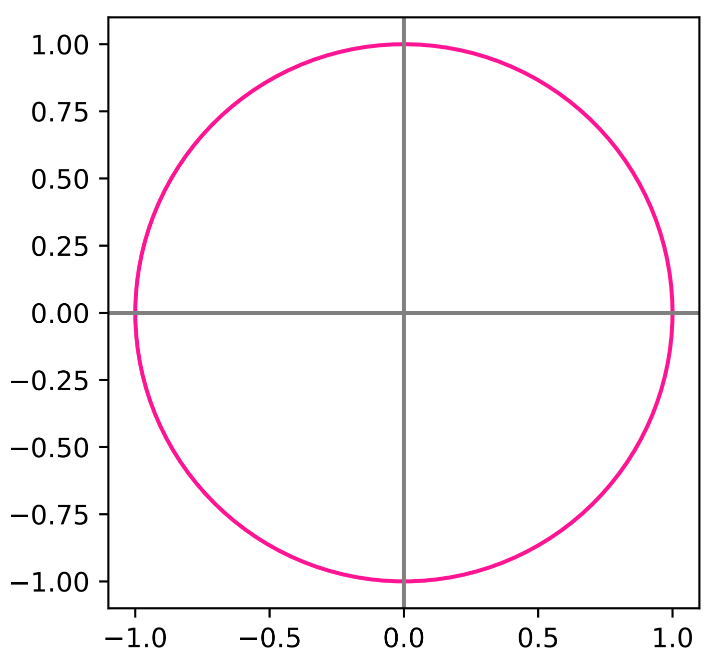


Figure 1.7: Lower semi-circle

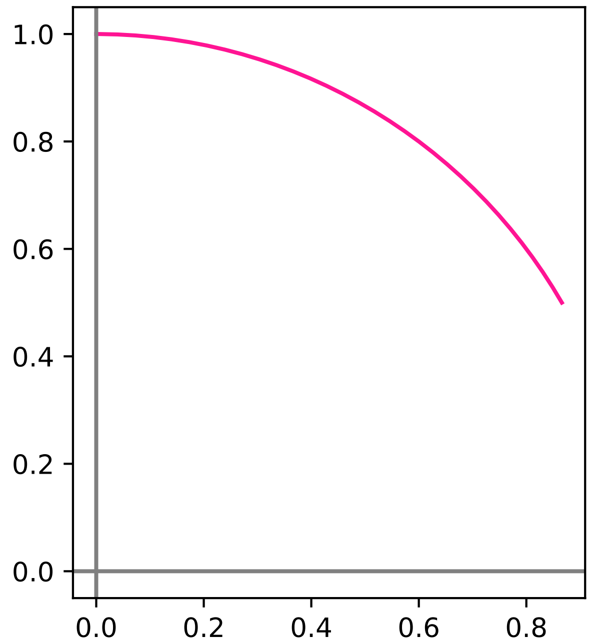


Figure 1.8: Plotting a portion of C

Example 1.7: Parametrizing the helix

The Helix plotted above can be parametrized by

$$\boldsymbol{\gamma} \,:\, (-\infty, \infty) \to \mathbb{R}^3$$

defined by

$$\gamma_1(t) = \cos(t)\,, \;\; \gamma_2(t) = \sin(t)\,, \;\; \gamma_3(t) = t\,.$$

The above equations are in line with our intuition: the helix can be drawn by *tracing a circle while at the same time lifting the pencil.*

1.3 Smooth curves

Let us recall the definition of **parametrized curve**.

Definition 1.8: Parametrized curve

A **parametrized curve** in $\mathbb{R}^n$ is a function

$$\boldsymbol{\gamma} \,:\, (a,b) \to \mathbb{R}^n\,.$$

where

$$(a, b) = \{t \in \mathbb{R} \ : \ a < t < b\},$$

with

$$-\infty \leq a < b \leq \infty .$$

The **components** of $\boldsymbol{\gamma}(t) \in \mathbb{R}^n$ are denoted by

$$\boldsymbol{\gamma}(t) = (\gamma_1(t), \dots, \gamma_n(t)),$$

where the components are functions

$$\gamma_i \ : \ (a, b) \to \mathbb{R},$$

for all $i = 1, \dots, n$.

As we already mentioned, the aim of the course is to study curves by **differentiating** them. Let us see what that means for curves.

Definition 1.9: Smooth functions

A scalar function $f \ : \ (a, b) \to \mathbb{R}$ is called **smooth** if the derivative

$$\frac{d^n f}{dt^n}$$

exists for all $n \geq 1$ and $t \in (a, b)$.

We will denote the first and second derivatives of f as follows:

$$\dot{f} := \frac{df}{dt}, \quad \ddot{f} := \frac{d^2 f}{dt^2}.$$

Example 1.10

The function $f(x) = x^4$ is smooth, with

$$\begin{gathered}
\frac{df}{dt} = 4x^3, \ \frac{d^2 f}{dt^2} = 12x^2, \\
\frac{d^3 f}{dt^3} = 24x, \ \frac{d^4 f}{dt^4} = 24, \\
\frac{d^n f}{dt^n} = 0 \ \text{ for all } \ n \geq 5.
\end{gathered}$$

Other examples smooth functions are polynomials, as well as

$$f(t) = \cos(t), \ f(t) = \sin(t), \ f(t) = e^t.$$

Definition 1.11

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^n$ with

$$\boldsymbol{\gamma}(t) = (\boldsymbol{\gamma}_1(t), \dots, \boldsymbol{\gamma}_n(t))$$

be a parametrized curve. We say that $\boldsymbol{\gamma}$ is **smooth** if the components

$$\boldsymbol{\gamma}_i : (a, b) \to \mathbb{R}$$

are smooth for all $i = 1, \dots, n$. The derivatives of $\boldsymbol{\gamma}$ are

$$\frac{d^k \boldsymbol{\gamma}}{dt^k} := \left(\frac{d^k \gamma_1}{dt^k}, \dots, \frac{d^k \gamma_n}{dt^k} \right)$$

for all $k \in \mathbb{N}$. As a shorthand, we will denote the first derivative of $\boldsymbol{\gamma}$ as

$$\dot{\boldsymbol{\gamma}} := \frac{d\boldsymbol{\gamma}}{dt} = \left(\frac{d\gamma_1}{dt}, \dots, \frac{d\gamma_n}{dt} \right)$$

and the second by

$$\ddot{\boldsymbol{\gamma}} := \frac{d^2 \boldsymbol{\gamma}}{dt^2} = \left(\frac{d^2 \gamma_1}{dt^2}, \dots, \frac{d^2 \gamma_n}{dt^2} \right) .$$

In Figure 1.9 we skectch a smooth and a non-smooth curve. Notice that the curve on the right is smooth, except for the point x.

We will work under the following assumption.

Assumption

All the parametrized curves in this lecture notes are assumed to be **smooth**.

Example 1.12

The circle

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

is a smooth parametrized curve, since both $\cos(t)$ and $\sin(t)$ are smooth functions. We have

$$\dot{\boldsymbol{\gamma}} = (-\sin(t), \cos(t)) .$$

For example the derivative of $\boldsymbol{\gamma}$ at the point $(0, 1)$ is given by

$$\dot{\boldsymbol{\gamma}}(\pi/2) = (-\sin(\pi/2), \cos(\pi/2)) = (-1, 0) .$$

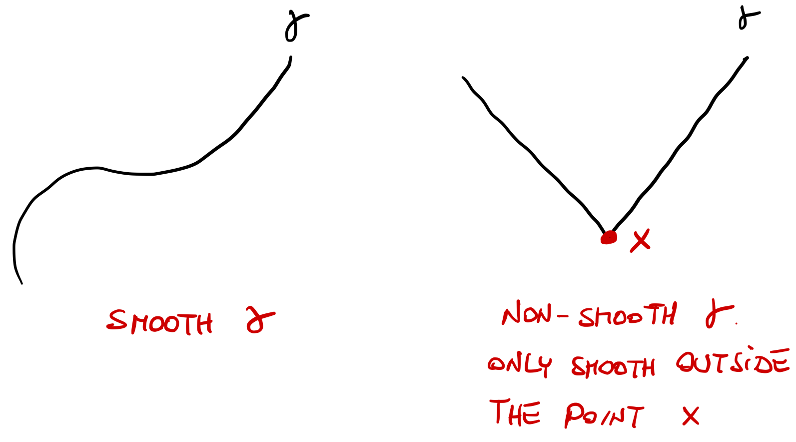


Figure 1.9: Example of smooth and non-smooth curves

The plot of the circle and the derivative vector at $(-1, 0)$ can be seen in Figure 1.10.

1.4 Tangent vectors

Looking at Figure 1.10, it seems like the vector

$$\dot{\boldsymbol{\gamma}}(\pi/2) = (-1, 0)$$

is **tangent** to the circle at the point

$$\boldsymbol{\gamma}(\pi/2) = (0, 1) .$$

Is this a coincidence? Not that all. Let us look at the definition of derivative at a point:

$$\dot{\boldsymbol{\gamma}}(t) := \lim_{\delta \to 0} \frac{\boldsymbol{\gamma}(t+\delta) - \boldsymbol{\gamma}(t)}{\delta} .$$

If we just look at the quantity

$$\frac{\boldsymbol{\gamma}(t+\delta) - \boldsymbol{\gamma}(t)}{\delta}$$

for non-negative δ, we see that this vector is parallel to the chord joining $\boldsymbol{\gamma}(t)$ to $\boldsymbol{\gamma}(t+\delta)$, as shown in Figure 1.11 below. As $\delta \to 0$, the length of the chord tends to zero. However the **direction** of the chord becomes **parallel**

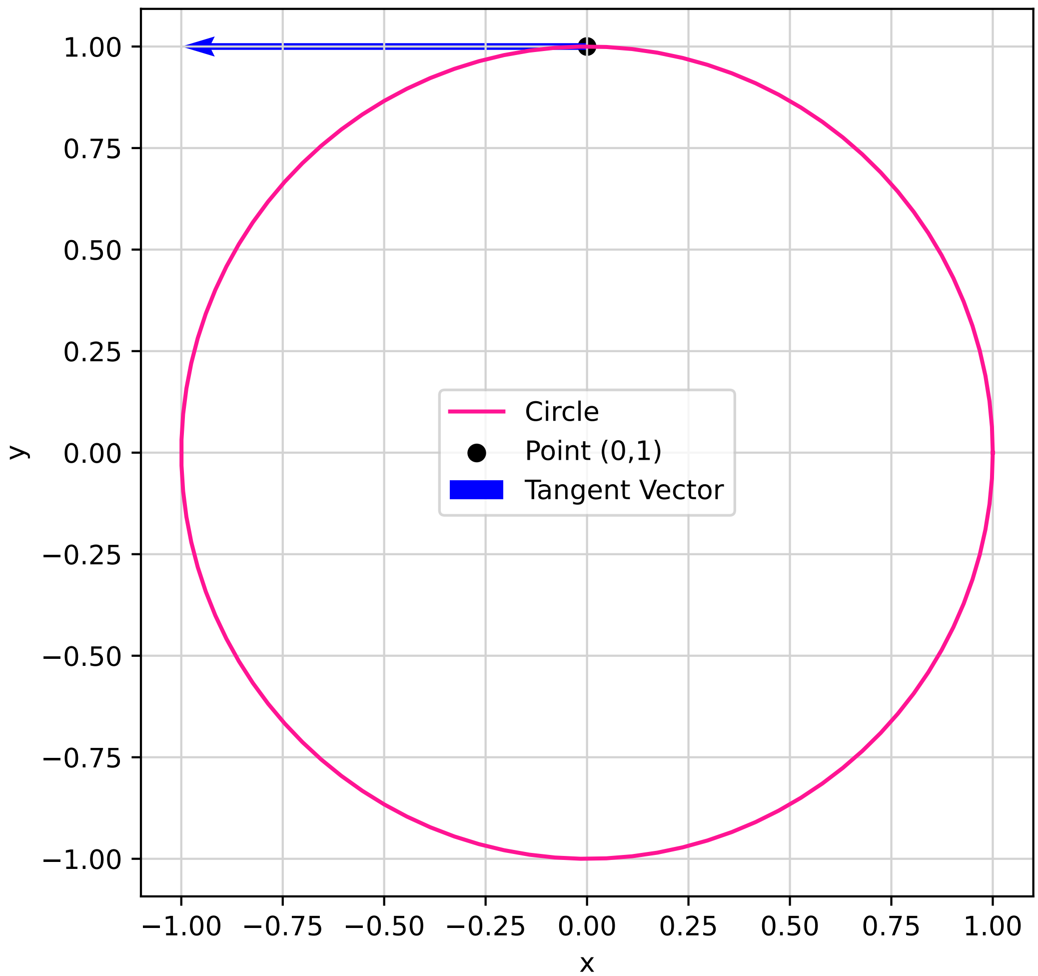


Figure 1.10: Plot of Circle and Tangent Vector at $(0, 1)$

to that of the tangent vector of the curve $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$. Since

$$\frac{\boldsymbol{\gamma}(t+\delta)-\boldsymbol{\gamma}(t)}{\delta} \to \dot{\boldsymbol{\gamma}}(t)$$

as $\delta \to 0$, we see that $\dot{\boldsymbol{\gamma}}(t)$ is **parallel** to the tangent of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$, as showin in Figure 1.11.

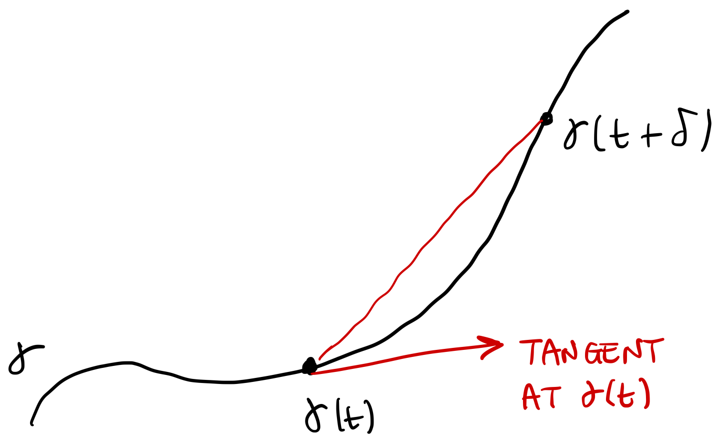


Figure 1.11: Approximating the tangent vector

The above remark motivates the following definition.

Definition 1.13: Tangent vector

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a parametrized curve. The tangent vector to $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(t)$ is defined as

$$\tau := \dot{\boldsymbol{\gamma}}(t)\,.$$

Example 1.14: Tangent vector to helix

The helix is described by the parametric curve

$$\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^3$$

with

$$\boldsymbol{\gamma}_1(t) = \cos(t)\,, \;\; \boldsymbol{\gamma}_2(t) = \sin(t)\,, \;\; \boldsymbol{\gamma}_3(t) = t.$$

This is plotted in Figure 1.12 below. The tangent vector at point $\boldsymbol{\gamma}(t)$ is given by

$$\dot{\boldsymbol{\gamma}}(t) = (-\sin(t), \cos(t), 1) .$$

For example in Figure 1.12 we plot the tangent vector at time $t = \pi/2$, that is,

$$\dot{\boldsymbol{\gamma}}(\pi/2) = (-1, 0, \pi/2) .$$

The above looks very similar to the tangent vector to the circle. Except that there is a z component, and that component is constant and equal to 1. Intuitively this means that the helix is *lifting* from the plane xy with constant speed with respect to the z-axis. We will soon give a name to this concept.

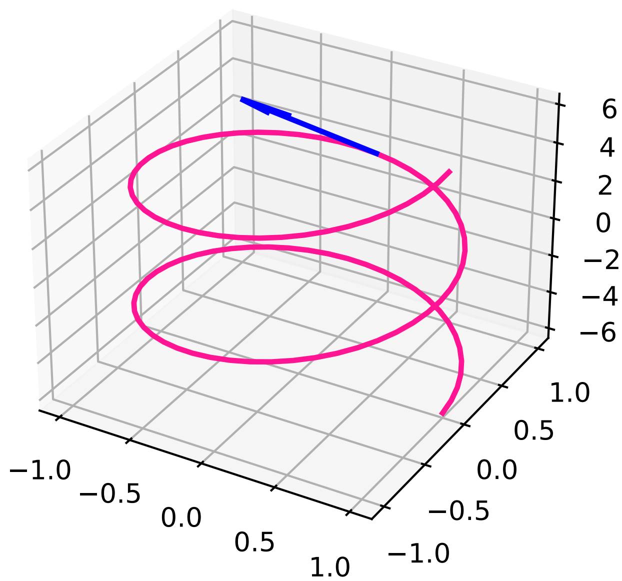


Figure 1.12: Plot of Helix with tangent vector

Remark 1.15: Avoiding potential ambiguities

Sometimes it will happen that a curve self intersects, meaning that there are two time instants t_1 and t_2 and a point $p \in \mathbb{R}^n$ such that

$$p = \boldsymbol{\gamma}(t_1) = \boldsymbol{\gamma}(t_2) .$$

In this case there is ambiguity in talking about the tangent vector at the point p: in principle there are two tangent vectors $\dot{\boldsymbol{\gamma}}(t_1)$ and $\dot{\boldsymbol{\gamma}}(t_2)$, and it could happen that

$$\dot{\boldsymbol{\gamma}}(t_1) \neq \dot{\boldsymbol{\gamma}}(t_1) .$$

Thus the concept of tangent at p is not well-defined. We need then to be more precise and talk about tangent at a certain **time-step** t, rather than at some **point** p. We however do not amend Definition 1.13, but you should keep this potential ambiguity in mind.

Example 1.16: The Lemniscate, a self intersecting curve

For example consider $\boldsymbol{\gamma} : [0, 2\pi] \to \mathbb{R}^2$ defined as

$$\boldsymbol{\gamma}_1(t) = \sin(t)\,, \ \boldsymbol{\gamma}_2(t) = \sin(t)\cos(t)\,.$$

Such curve is called **Lemniscate**, see Wikipedia page, and is plotted in Figure 1.13 below. The orgin $(0, 0)$ is a point of self-intersection, meaning that

$$\boldsymbol{\gamma}(0) = \boldsymbol{\gamma}(\pi) = (0, 0)\,.$$

The tangent vector at point $\boldsymbol{\gamma}(t)$ is given by

$$\dot{\boldsymbol{\gamma}}(t) = (\cos(t), \cos^2(t) - \sin^2(t))$$

and therefore we have two tangents at $(0, 0)$, that is,

$$\tau_1 = \dot{\boldsymbol{\gamma}}(0) = (1, 1)\,, \ \tau_2 = \dot{\boldsymbol{\gamma}}(\pi) = (-1, 1)\,.$$

1.5 Length of curves

For a vector $v \in \mathbb{R}^n$ with components

$$v = (v_1, \dots, v_n),$$

its **length** is defined by

$$\|v\| := \sqrt{\sum_{i=1}^n v_i^2}\,.$$

The above is just an extension of the Pythagoras theorem to $\mathbb{R}^n$, and the length of v is computed from the origin.

If we have a second vector $u \in \mathbb{R}^n$, then the quantity

$$\|u - v\| := \sqrt{\sum_{i=1}^n (u_i - v_i)^2}$$

measures the length of the difference between u and v.

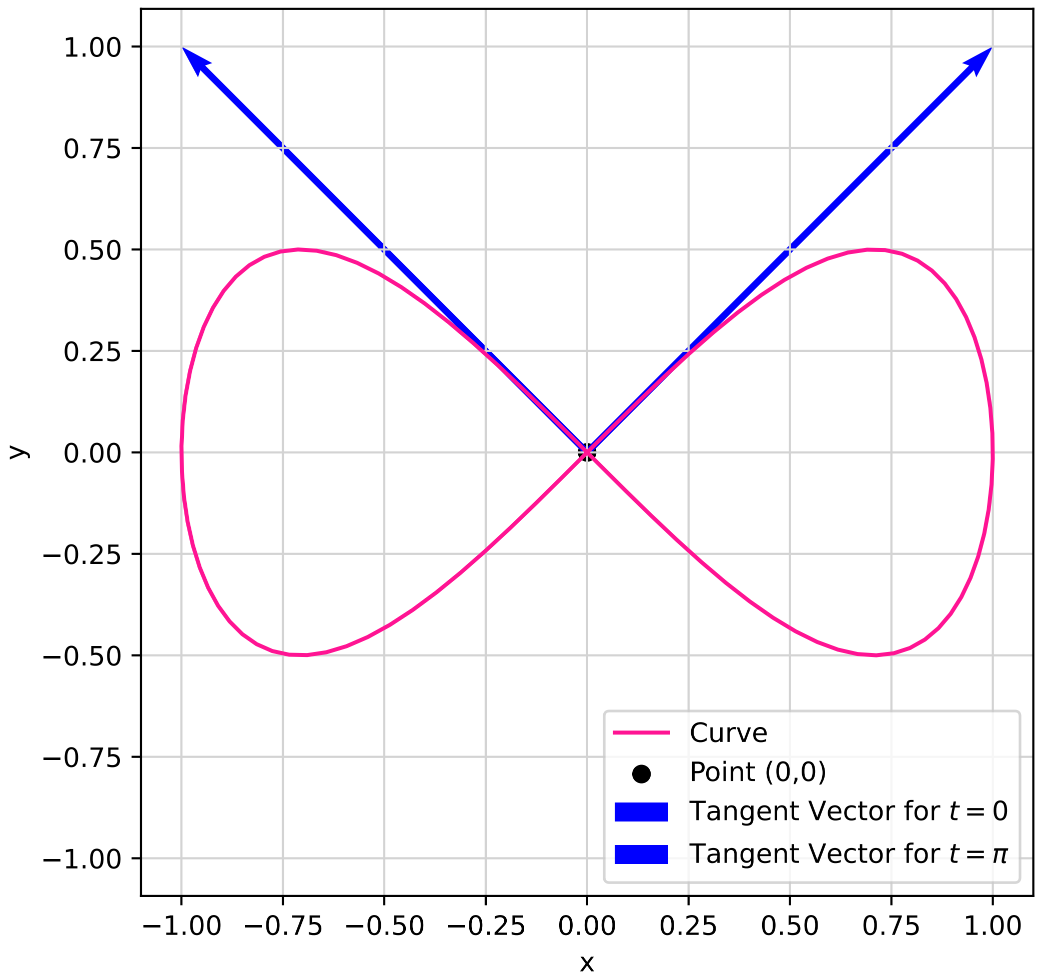


Figure 1.13: The Lemniscate curve

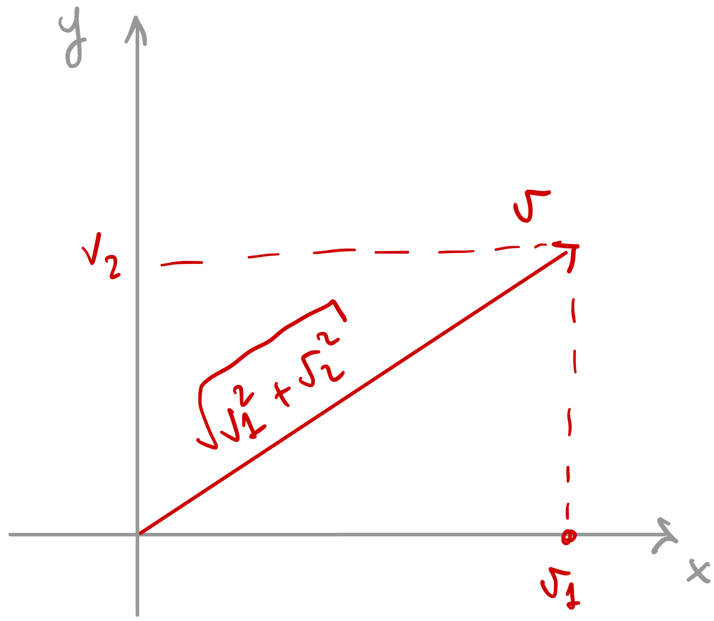


Figure 1.14: Interpretation of $\|v\|$ in $\mathbb{R}^2$

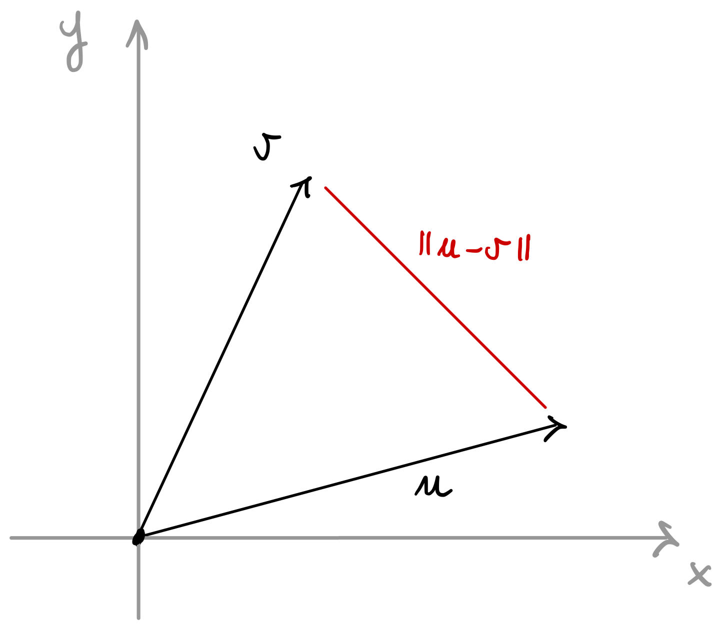


Figure 1.15: Interpretation of $\|u - v\|$ in $\mathbb{R}^2$

We would like to define the concept of **length** of a curve. Intuitively, one could proceed by approximation as in the figure below.

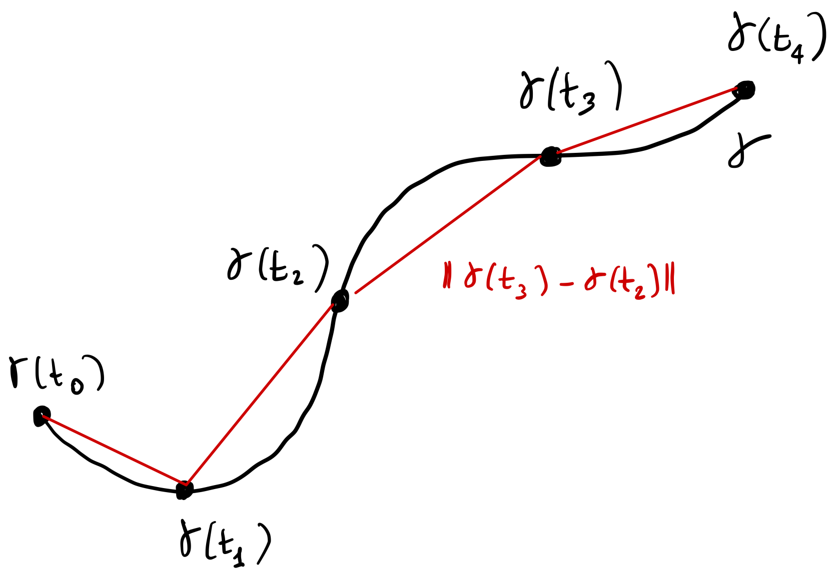


Figure 1.16: Approximating the length of $\boldsymbol{\gamma}$

In formulae, this means choosing some time instants

$$t_0, \dots, t_m \in (a,b)\,.$$

The length of the segment connecting $\boldsymbol{\gamma}(t_{i-1})$ to $\boldsymbol{\gamma}(t_i)$ is given by

$$\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\|\,.$$

Thus

$$L(\boldsymbol{\gamma}) \approx \sum_{i=1}^{m} \|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\|\,. \tag{1.8}$$

Intuitively, if we increase the number of points t_i, the quantity on the RHS of (1.8) should approximate $L(\boldsymbol{\gamma})$ better and better. Let us make this precise.

Definition 1.17: Partition

Let (a,b) be an interval. A partition $\mathscr{P}$ of $[a,b]$ is a vector of time instants

$$\mathscr{P} = (t_0, \dots, t_k) \in [a,b]^{m+1}$$

with

$$t_0 = a < t_1 < \ldots < t_{m-1} < t_m = b\,.$$

If $\mathscr{P}$ is a partition of $[a,b]$, we define its maximum length as

$$\|\mathscr{P}\| := \max_{1 \le i \le m} |t_i - t_{i-1}|\,.$$

Note that $\|\mathscr{P}\|$ measures how fine the partition $\mathscr{P}$ is.

Definition 1.18: Length of approximating polygonal curve

Suppose $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ is a parametrized curve and $\mathscr{P}$ a partition of $[a,b]$. We define the length of the polygonal curve connecting the points

$$\boldsymbol{\gamma}(t_0)\,,\ \boldsymbol{\gamma}(t_1)\,,\ \ldots,\ \boldsymbol{\gamma}(t_m)$$

as

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

If $\|\mathscr{P}\|$ becomes smaller and smaller, that is, the partition $\mathscr{P}$ is finer and finer, it is reasonable to say that

$$L(\boldsymbol{\gamma}, \mathscr{P})$$

is approximating the length of $\boldsymbol{\gamma}$. We take this as definition of length.

Definition 1.19: Rectifiable curve and length

Suppose $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ is a parametrized curve. We say that $\boldsymbol{\gamma}$ is **rectifiable** if the limit

$$L(\boldsymbol{\gamma}) = \lim_{\|P\| \to 0} L(\boldsymbol{\gamma}, \mathscr{P})$$

exists finite. In such case we call $L(\boldsymbol{\gamma})$ the **length** of $\boldsymbol{\gamma}$.

This definition definitely corresponds to our geometrical intuition of length of a curve.

Question 1.20

How do we use such definition in practice to compute the length of a given curve $\boldsymbol{\gamma}$?

Thankfully, when $\boldsymbol{\gamma}$ is smooth, the length $L(\boldsymbol{\gamma})$ can be characterized in terms of $\dot{\boldsymbol{\gamma}}$. Indeed, when δ is small, then the quantity

$$\|\boldsymbol{\gamma}(t+\delta) - \boldsymbol{\gamma}(t)\|$$

is approximating the length of $\boldsymbol{\gamma}$ between $\boldsymbol{\gamma}(t)$ and $\boldsymbol{\gamma}(t+\delta)$. Multiplying and dividing by δ we obtain

$$\frac{\|\boldsymbol{\gamma}(t+\delta) - \boldsymbol{\gamma}(t)\|}{\delta}\,\delta$$

which for small δ is close to

$$\|\dot{\boldsymbol{\gamma}}(t)\|\ \delta\,.$$

We can now divide the time interval (a,b) in steps $t_0,\dots,t_m$ with $|t_i - t_{i-1}| < \delta$ and obtain

$$\begin{aligned}\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| &= \frac{\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\|}{|t_i - t_{i-1}|}|t_i - t_{i-1}| \\ &\approx \|\dot{\boldsymbol{\gamma}}(t_i)\|\,\delta\end{aligned}$$

since δ is small. Therefore

$$L(\boldsymbol{\gamma}) \approx \sum_{i=1}^{m} \|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| \approx \sum_{i=1}^{m} \|\dot{\boldsymbol{\gamma}}(t_i)\|\ \delta\,.$$

The RHS is a Riemann sum, therefore

$$L(\boldsymbol{\gamma}) \approx \int_a^b \|\dot{\boldsymbol{\gamma}}(t)\|\ dt\,.$$

The above argument can be made rigorous, as we see in the next theorem.

Theorem 1.21: Characterizing the length of $\boldsymbol{\gamma}$

Assume $\boldsymbol{\gamma} : [a,b] \to \mathbb{R}^n$ is a parametrized curve, with $[a,b]$ bounded. Then $\boldsymbol{\gamma}$ is rectifiable and

$$L(\boldsymbol{\gamma}) = \int_a^b \|\dot{\boldsymbol{\gamma}}(t)\|\ dt\,. \tag{1.9}$$

Proof

Step 1. The integral in (1.9) is bounded.
Since $\boldsymbol{\gamma}$ is smooth, in particular $\dot{\boldsymbol{\gamma}}$ is continuous. Since $[a,b]$ is bounded, then $\dot{\boldsymbol{\gamma}}$ is bounded, that is

$$\sup_{t\in[a,b]} \|\dot{\boldsymbol{\gamma}}(t)\| \leq C$$

for some constant $C \geq 0$. Therefore

$$\int_a^b \|\dot{\boldsymbol{\gamma}}(t)\|\ dt \leq C(b-a) < \infty\,.$$

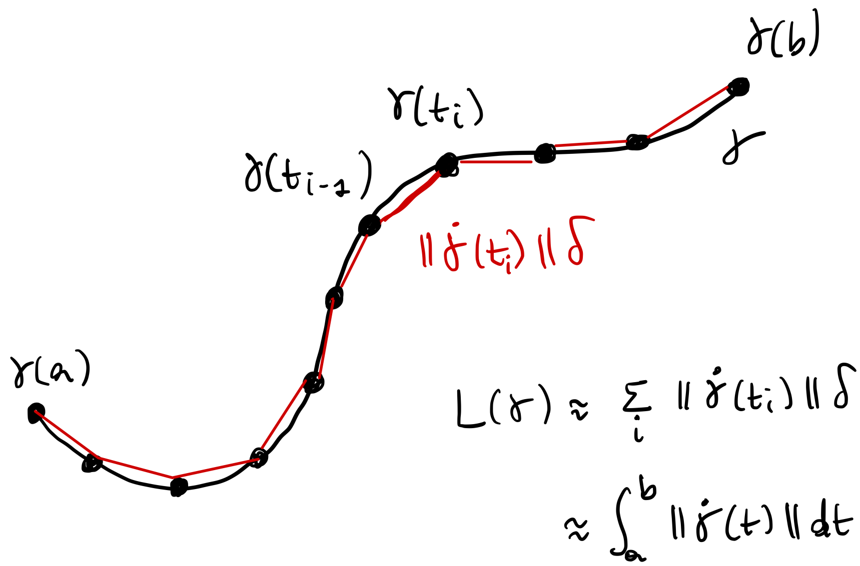


Figure 1.17: Approximating $L(\boldsymbol{\gamma})$ via $\dot{\boldsymbol{\gamma}}$

Step 2. Writing (1.9) as limit.
Recalling that

$$L(\boldsymbol{\gamma}) = \lim_{\|\mathscr{P}\| \to 0} L(\boldsymbol{\gamma}, \mathscr{P}),$$

whenever the limit is finite, in order to show (1.9) we then need to prove

$$L(\boldsymbol{\gamma}, \mathscr{P}) \to \int_a^b \|\dot{\boldsymbol{\gamma}}(t)\| \, dt$$

as $\|\mathscr{P}\| \to 0$. Showing the above means proving that: for every $\varepsilon > 0$ there exists a $\delta > 0$ such that, if $\mathscr{P}$ is a partition of $[a, b]$ such that $\|\mathscr{P}\| < \delta$, then

$$\left| \int_a^b \|\dot{\boldsymbol{\gamma}}(t)\| \, dt - L(\boldsymbol{\gamma}, \mathscr{P}) \right| < \varepsilon \,. \tag{1.10}$$

Step 3. First estimate in (1.10).
This first estimate is easy, and only relies on the Fundamental Theorem of Calculus. To be more precise, we will show that each polygonal has shorter length than $\int_a^b \|\dot{\boldsymbol{\gamma}}(t)\| \, dt$. To this end, take an arbitrary partition $\mathscr{P} = (t_0, \dots, t_m)$ of $[a, b]$. Then for each $i = 1, \dots, m$ we have

$$\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| = \left\| \int_{t_{i-1}}^{t_i} \dot{\boldsymbol{\gamma}}(t) \, dt \right\| \leq \int_{t_{i-1}}^{t_i} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt$$

where we used the Fundamental Theorem of calculus, and usual integral properties. Therefore by definition

$$\begin{aligned} L(\boldsymbol{\gamma}, \mathscr{P}) &= \sum_{i=1}^{m} \|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| \\ &\leq \sum_{i=1}^{m} \int_{t_{i-1}}^{t_i} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt \\ &= \int_a^b \|\dot{\boldsymbol{\gamma}}(t)\| \, dt \,. \end{aligned}$$

We have then shown

$$L(\boldsymbol{\gamma}, \mathscr{P}) \leq \int_a^b \|\dot{\boldsymbol{\gamma}}(t)\| \, dt \tag{1.11}$$

for all partitions $\mathscr{P}$.

Step 4. Second estimate in (1.10).

The second estimate is more delicate. We need to carefully construct a polygonal so that its length is close to $\int_a^b \|\dot{\boldsymbol{\gamma}}\| \, dt$. This will be possible by uniform continuity of $\dot{\boldsymbol{\gamma}}$. Indeed, note that $\dot{\boldsymbol{\gamma}}$ is continuous on the compact set $[a, b]$. Therefore it is uniformly continuous by the Heine-Borel Theorem. Fix $\varepsilon > 0$. By uniform continuity of $\dot{\boldsymbol{\gamma}}$ there exists $\delta > 0$ such that

$$|t - s| < \delta \implies \|\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s)\| < \frac{\varepsilon}{b - a} \,. \tag{1.12}$$

for all $t, s \in [a, b]$. Let $\mathscr{P} = (t_0, \dots, t_m)$ be a partition of $[a, b]$ with $\|\mathscr{P}\| < \delta$. Recall that

$$\|\mathscr{P}\| = \max_{i=1,\dots,m} |t_i - t_{i-1}| \,.$$

Therefore the condition $\|\mathscr{P}\| < \delta$ implies

$$|t_i - t_{i-1}| < \delta \tag{1.13}$$

for each $i = 1, \dots, m$. For all $i = 1, \dots, m$ and $s \in [t_{i-1}, t_i]$ we have

$$\begin{aligned} \boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1}) &= \int_{t_{i-1}}^{t_i} \dot{\boldsymbol{\gamma}}(t) \, dt \\ &= \int_{t_{i-1}}^{t_i} \dot{\boldsymbol{\gamma}}(s) + (\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s)) \, dt \\ &= (t_i - t_{i-1})\dot{\boldsymbol{\gamma}}(s) + \int_{t_{i-1}}^{t_i} (\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s)) \, dt \end{aligned}$$

Therefore

$$\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| = \left\| (t_i - t_{i-1})\dot{\boldsymbol{\gamma}}(s) + \int_{t_{i-1}}^{t_i} (\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s)) \, dt \right\| \tag{1.14}$$

We can now use the reverse triangle inequality

$$\left| \|x\| - \|y\| \right| \leq \|x - y\| \,,$$

for all $x, y \in \mathbb{R}^n$, which implies

$$\|x + y\| = \|x - (-y)\| \geq \|x\| - \|y\|$$

for all $x, y \in \mathbb{R}^n$. Applying the above to (1.14) we get

$$\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| \geq (t_i - t_{i-1}) \|\dot{\boldsymbol{\gamma}}(s)\| - \left\| \int_{t_{i-1}}^{t_i} (\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s))\, dt \right\| \tag{1.15}$$

By standard properties of integral we also have

$$\left\| \int_{t_{i-1}}^{t_i} (\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s))\, dt \right\| \leq \int_{t_{i-1}}^{t_i} \|\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s)\| \, dt \,,$$

so that (1.15) implies

$$\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| \geq (t_i - t_{i-1}) \|\dot{\boldsymbol{\gamma}}(s)\| - \int_{t_{i-1}}^{t_i} \|\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s)\| \, dt \,. \tag{1.16}$$

Since $t, s \in [t_{i-1}, t_i]$, then

$$|t - s| \leq |t_i - t_{i-1}| < \delta$$

where the last inequality follows by (1.13). Thus by uniform continuity (1.12) we get

$$\|\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s)\| < \frac{\varepsilon}{b - a} \,.$$

We can therefore further estimate (1.16) and obtain

$$\begin{aligned} \|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| &\geq (t_i - t_{i-1}) \|\dot{\boldsymbol{\gamma}}(s)\| - \int_{t_{i-1}}^{t_i} \|\dot{\boldsymbol{\gamma}}(t) - \dot{\boldsymbol{\gamma}}(s)\| \, dt \\ &\geq (t_i - t_{i-1}) \|\dot{\boldsymbol{\gamma}}(s)\| - (t_i - t_{i-1}) \frac{\varepsilon}{b - a} \, dt \,. \end{aligned}$$

Dividing the above by $t_i - t_{i-1}$ we get

$$\frac{\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\|}{t_i - t_{i-1}} \geq \|\dot{\boldsymbol{\gamma}}(s)\| - \frac{\varepsilon}{b - a} \,.$$

Integrating the above over s in the interval $[t_{i-1}, t_i]$ we get

$$\|\boldsymbol{\gamma}(t_i) - \boldsymbol{\gamma}(t_{i-1})\| \geq \int_{t_{i-1}}^{t_i} \|\dot{\boldsymbol{\gamma}}(s)\| \, ds - \frac{\varepsilon}{b - a} (t_i - t_{i-1}) \,.$$

Summing over $i = 1, \dots, m$ we get

$$L(\mathscr{P}, \boldsymbol{\gamma}) \geq \int_a^b \|\dot{\boldsymbol{\gamma}}(s)\| \, ds - \varepsilon \tag{1.17}$$

since

$$\sum_{i=1}^{m} (t_i - t_{i-1}) = t_m - t_0 = b - a \, .$$

Conclusion.
Putting together (1.11) and (1.17) we get

$$\int_a^b \|\dot{\boldsymbol{\gamma}}(s)\| \, ds - \varepsilon \leq L(\mathscr{P}, \boldsymbol{\gamma}) \leq \int_a^b \|\dot{\boldsymbol{\gamma}}(s)\| \, ds$$

which implies (1.10), concluding the proof.

Thanks to the above theorem we have now a way to compute $L(\boldsymbol{\gamma})$. Let us check that we have given a meaningful definition of length by computing $L(\boldsymbol{\gamma})$ on known examples.

Example 1.22: Length of Circle

The circle of radius R is parametrized by $\boldsymbol{\gamma} \, : \, [0, 2\pi] \to \mathbb{R}^2$ defined by

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t)) \, .$$

Then

$$\dot{\boldsymbol{\gamma}}(t) = (-R\sin(t), R\cos(t))$$

and

$$\begin{aligned} \|\dot{\boldsymbol{\gamma}}(t)\| &= \sqrt{\dot{\gamma}_1^2(t) + \dot{\gamma}_2^2(t)} \\ &= R\sqrt{\sin^2(t) + \cos^2(t)} = R \, . \end{aligned}$$

Therefore

$$L(\boldsymbol{\gamma}) = \int_0^{2\pi} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt = \int_0^{2\pi} R \, dt = 2\pi R$$

as expected.

Example 1.23: Length of helix

Let us consider one full turn of the Helix of radius R and rise H. This is parametrized by

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t), Ht)$$

for $t \in [0, 2\pi]$. Then

$$\dot{\boldsymbol{\gamma}}(t) = (-R\sin(t), R\cos(t), H)\,,$$

and

$$\begin{aligned}\|\dot{\boldsymbol{\gamma}}(t)\| &= \sqrt{\dot{\gamma}_1^2 + \dot{\gamma}_2^2 + \dot{\gamma}_3^2} \\ &= \sqrt{R^2\sin^2(t) + R^2\cos^2(t) + H^2} = \sqrt{R^2 + H^2}\,.\end{aligned}$$

Therefore

$$L(\boldsymbol{\gamma}) = \int_0^{2\pi} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt = 2\pi\sqrt{R^2 + H^2}\,.$$

1.6 Arc-length

We have just shown in Theorem 1.21 that the length of a regular curve $\boldsymbol{\gamma} : [a,b] \to \mathbb{R}^n$ with $[a,b]$ bounded is given by

$$L(\boldsymbol{\gamma}) = \int_a^b \|\dot{\boldsymbol{\gamma}}(t)\| \, dt\,.$$

Using this formula, we introduce the notion of length of a portion of $\boldsymbol{\gamma}$.

Definition 1.24: Arc-length

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a curve, with (a,b) possibly unbounded. We define the **arc-length** of $\boldsymbol{\gamma}$ starting at the point $\boldsymbol{\gamma}(t_0)$ as the function $s : \mathbb{R} \to \mathbb{R}$ defined by

$$s(t) := \int_{t_0}^t \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau\,.$$

Remark 1.25

A few remarks:

- Arc-length is well-defined

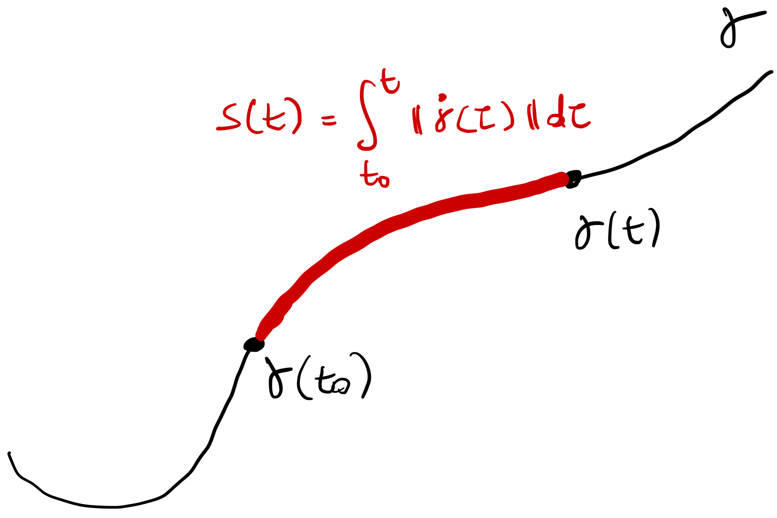


Figure 1.18: Arc-length of $\boldsymbol{\gamma}$ starting at $\boldsymbol{\gamma}(t_0)$

Indeed, $\boldsymbol{\gamma}$ is smooth, and so $\dot{\boldsymbol{\gamma}}$ is continuous. WLOG assume $t \geq t_0$. Then

$$s(t) = \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \leq (t - t_0) \max_{\tau \in [t_0, t]} \|\dot{\boldsymbol{\gamma}}(\tau)\| < \infty \,.$$

- We always have
$$s(t_0) = 0 \,.$$
- We have
$$t > t_0 \implies s(t) \geq 0$$
and
$$t < t_0 \implies s(t) \leq 0 \,.$$
- Choosing a different starting point changes the arc-length by a **constant**:

  For example define $\tilde{s}$ as the arc-length starting from $\tilde{t}_0$
$$\tilde{s}(t) := \int_{\tilde{t}_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \,.$$

Then by the properties of integral

$$\begin{aligned} s(t) &= \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \\ &= \int_{t_0}^{\tilde{t}_0} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau + \int_{\tilde{t}_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \\ &= \int_{t_0}^{\tilde{t}_0} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau + \tilde{s}(t) \, . \end{aligned}$$

Hence

$$s = c + \tilde{s}$$

with

$$c := \int_{t_0}^{\tilde{t}_0} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \, .$$

Note that c is the arc-length of $\boldsymbol{\gamma}$ between the starting points $\boldsymbol{\gamma}(t_0)$ and $\boldsymbol{\gamma}(\tilde{t}_0)$.

- The arc-length is a differentiable function, with

$$\dot{s}(t) = \frac{d}{dt} \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau = \|\dot{\boldsymbol{\gamma}}(t)\| \, .$$

Since $\dot{\boldsymbol{\gamma}}$ is continuous, the above follows by the Fundamental Theorem of Calculus.

Example 1.26: Circle

The circle of radius R is parametrized by $\boldsymbol{\gamma} \, : \, [0, 2\pi] \to \mathbb{R}^2$ defined by

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t)) \, .$$

Then

$$\dot{\boldsymbol{\gamma}}(t) = (-R\sin(t), R\cos(t)) \, , \quad \|\dot{\boldsymbol{\gamma}}(t)\| = R \, .$$

Therefore, for any fixed $t_0 \in [0, 2\pi]$ we have

$$s(t) = \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau = \int_{t_0}^{t} R \, d\tau = (t - t_0)R \, .$$

In particular we see that $\dot{s} = R$ is constant.

Example 1.27: Logarithmic spiral

The Logarithmic spiral is defined by $\boldsymbol{\gamma} \colon [0, 2\pi] \to \mathbb{R}^2$ with

$$\boldsymbol{\gamma}(t) = (e^{kt} \cos(t), e^{kt} \sin(t)) ,$$

where $k \in \mathbb{R}$, $k \neq 0$, is called the **growth factor**. Then

$$\dot{\gamma}_1(t) = e^{kt}(k \cos(t) - \sin(t))$$

$$\dot{\gamma}_2(t) = e^{kt}(k \sin(t) + \cos(t))$$

and so, after some calculations,

$$\left\|\dot{\boldsymbol{\gamma}}(t)\right\|^2 = \dot{\gamma}_1^2 + \dot{\gamma}_2^2 = (k^2 + 1)e^{2kt} .$$

The arc-length starting from t_0 is

$$\begin{aligned} s(t) &= \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \\ &= \sqrt{k^2+1} \int_{t_0}^{t} e^{k\tau} \, d\tau \\ &= \frac{\sqrt{k^2+1}}{k} (e^{kt} - e^{kt_0}) . \end{aligned}$$

1.7 Scalar product in $\mathbb{R}^n$

Let us start by defining the scalar product in $\mathbb{R}^2$.

Definition 1.28: Scalar product in $\mathbb{R}^2$

Let $u, v \in \mathbb{R}^2$ and denote by $\theta \in [0, \pi]$ the angle formed by u and v. The *scalar product* between u and v is defined by

$$u \cdot v := |u||v| \cos(\theta) .$$

Remark 1.29

The scalar product is maximized for $\theta = 0$, for which we have

$$u \cdot v = |u||v| \cos(\theta) = |u||v| .$$

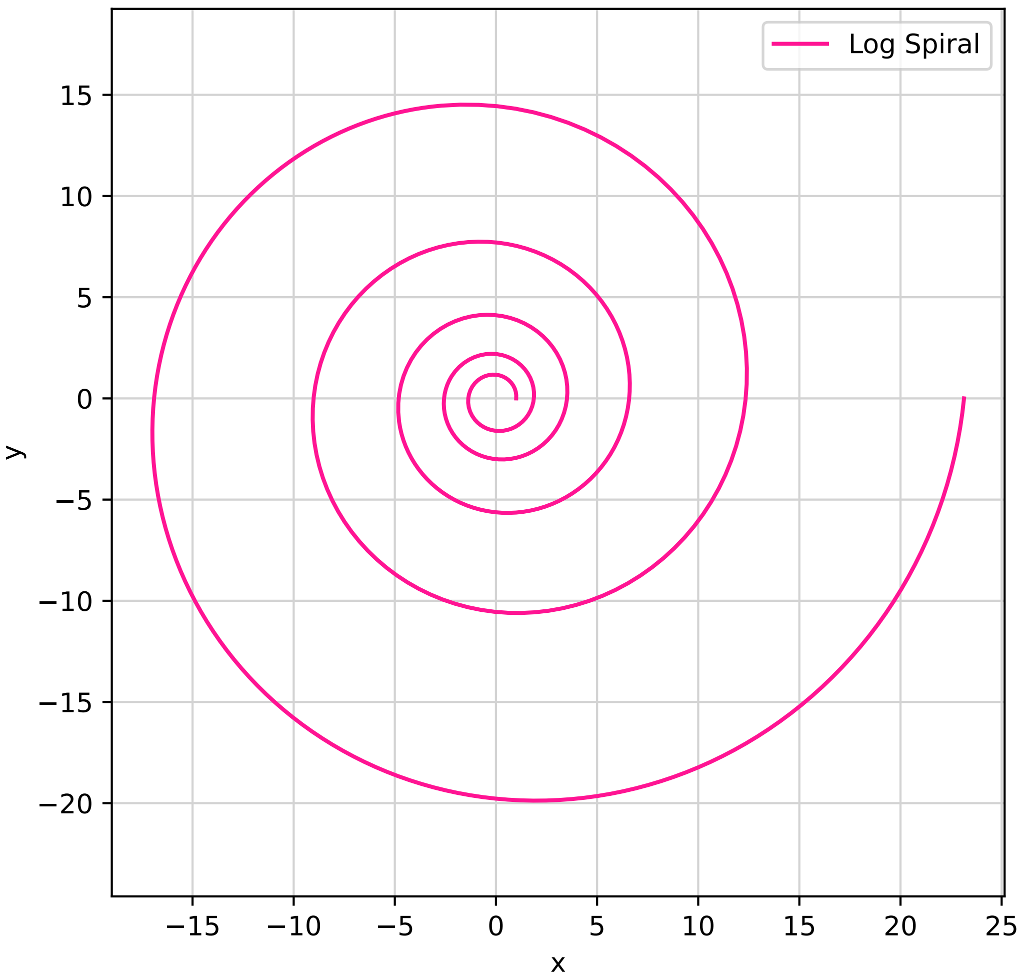


Figure 1.19: Plot of Logarithmic Spiral with $k = 0.1$

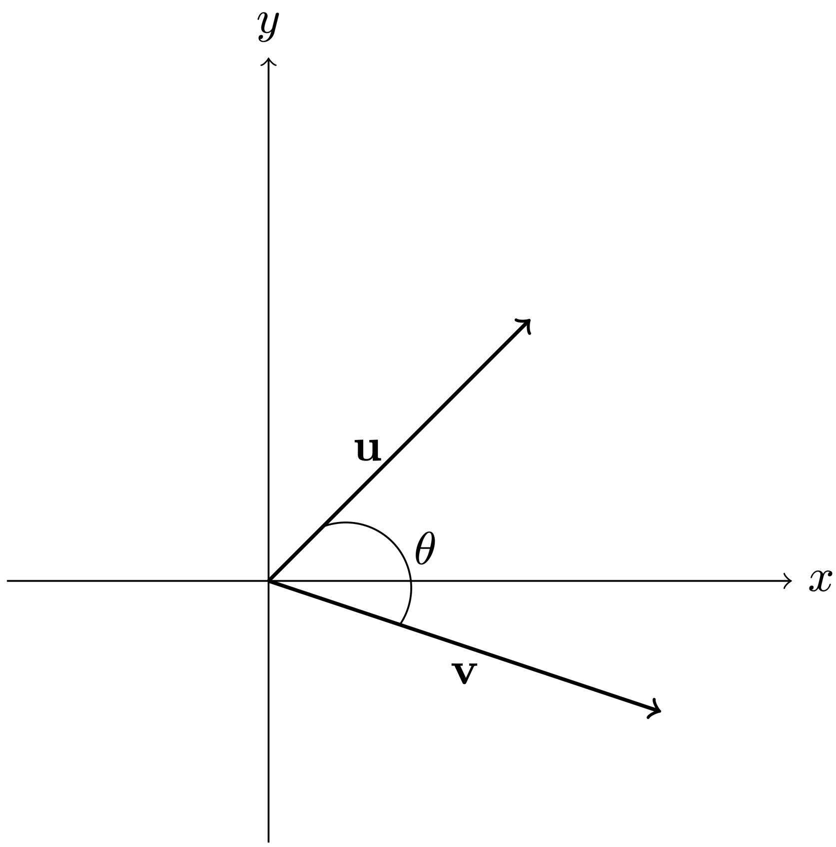


Figure 1.20: Vectors u and v in $\mathbb{R}^2$ forming angle θ

It is instead minimized for $\theta = \pi$, for which

$$u \cdot v = |u||v| \cos(\theta) = -|u||v| \,.$$

Definition 1.30: Orthogonal vectors

Let $u, v \in \mathbb{R}^2$. If

$$u \cdot v = 0$$

we say that u and v are **orthogonal**.

Proposition 1.31: Bilinearity and symmetry of scalar product

Let $u, v, w \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$. Then

- **Symmetry**: $u \cdot v = v \cdot u$
- **Bilinearity**: It holds

$$\lambda(u \cdot v) = (\lambda u) \cdot v = u \cdot (\lambda v) \,,$$

$$u \cdot (v + w) = u \cdot v + u \cdot w \,.$$

We leave the proof to the reader. The above proposition is saying that the scalar product is **bilinear** and **symmetric**.

Proposition 1.32: Scalar products written wrt euclidean coordinates

Denote by

$$e_1 = (1, 0) \,, \quad e_2 = (0, 1)$$

the euclidean basis of $\mathbb{R}^2$. Let $u, v \in \mathbb{R}^2$ and denote by

$$u = (u_1, u_2) = u_1 e_1 + u_2 e_2$$

$$v = (v_1, v_2) = v_1 e_1 + v_2 e_2$$

their coordinates with respect to e_1, e_2. Then

$$u \cdot v = u_1 v_2 + u_2 v_2 \,.$$

Proof

Note that

$$e_1 \cdot e_1 = 1\,, \quad e_2 \cdot e_2 = 1\,, \quad e_1 \cdot e_2 = e_2 \cdot e_1 = 0\,.$$

Using the bilinearity of scalar product we have

$$\begin{aligned} u \cdot v &= (u_1 e_1 + u_2 e_2) \cdot (v_1 e_1 + v_2 e_2) \\ &= u_1 v_1 e_1 \cdot e_1 + u_1 v_2 e_1 \cdot e_2 + u_2 v_1 e_2 \cdot e_1 + u_2 v_2 e_2 \cdot e_2 \\ &= u_1 v_1 + u_2 v_2\,. \end{aligned}$$

The above proposition provides a way to generalize of the scalar product to $\mathbb{R}^n$..

Definition 1.33: Scalar product in $\mathbb{R}^n$

Let $u, v \in \mathbb{R}^n$ and denote their coordinates by

$$u = (u_1, \dots, u_n)\,, \quad u = (v_1, \dots, v_n)\,.$$

We define the scalar product between u and v by

$$u \cdot v := \sum_{i=1}^n u_i v_i\,.$$

With the above definition we still have that the scalar product is bilinear and symmetric, as detailed in the following proposition:

Proposition 1.34: Bilinearity and symmetry of scalar product in $\mathbb{R}^n$

Let $u, v, w \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Then

- **Symmetry**: $u \cdot v = v \cdot u$
- **Bilinearity**: It holds

$$\lambda(u \cdot v) = (\lambda u) \cdot v = u \cdot (\lambda v)\,,$$

$$u \cdot (v + w) = u \cdot v + u \cdot w\,.$$

The proof of the above proposition is an easy check, and is left to the reader for exercise.

Definition 1.35

Let $u, v \in \mathbb{R}^n$. We say that u and v are **orthogonal** if

$$u \cdot v = 0\,.$$

Proposition 1.36: Differentiating scalar product

Let $\boldsymbol{\gamma}, \boldsymbol{\eta} : (a,b) \to \mathbb{R}^n$ be parametrized curves. Then the scalar map

$$\boldsymbol{\gamma} \cdot \boldsymbol{\eta} : (a,b) \to \mathbb{R}$$

is smooth, and

$$\frac{d}{dt}(\boldsymbol{\gamma} \cdot \boldsymbol{\eta}) = \dot{\boldsymbol{\gamma}} \cdot \boldsymbol{\eta} + \boldsymbol{\gamma} \cdot \dot{\boldsymbol{\eta}}$$

for all $t \in (a,b)$.

Proof

Denote by

$$\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_n) , \quad \boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$$

the coordinates of $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$. Clearly the map

$$t \mapsto \boldsymbol{\gamma} \cdot \boldsymbol{\eta} = \sum_{i=1}^{n} \boldsymbol{\gamma}_i \eta_i$$

is smooth, being sum and product of smooth functions.
Concerning the formula, by definition of scalar product and linearity of the derivative we have

$$\begin{aligned}
\frac{d}{dt}(\boldsymbol{\gamma} \cdot \boldsymbol{\eta}) &= \frac{d}{dt}\left(\sum_{i=1}^{n} \boldsymbol{\gamma}_i \eta_i\right) \\
&= \sum_{i=1}^{n} \frac{d}{dt}(\boldsymbol{\gamma}_i \eta_i) \\
&= \sum_{i=1}^{n} \dot{\gamma}_i \eta_i + \dot{\gamma}_i \dot{\eta}_i \\
&= \dot{\boldsymbol{\gamma}} \cdot \boldsymbol{\eta} + \boldsymbol{\gamma} \cdot \dot{\boldsymbol{\eta}} ,
\end{aligned}$$

where in the second to last equality we used the product rule of differentiation.

1.8 Speed of a curve

Given a curve $\boldsymbol{\gamma}$ we defined the **tangent** vector at $\boldsymbol{\gamma}(t)$ to be

$$\dot{\boldsymbol{\gamma}}(t) .$$

The tangent vector measures the change of direction of the curve. Therefore the magnitude of $\dot{\boldsymbol{\gamma}}$ can be interpreted as the **speed** of the curve.

Definition 1.37

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a curve. We define the speed of $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(t)$ by

$$\|\dot{\boldsymbol{\gamma}}(t)\| \,.$$

We say that $\boldsymbol{\gamma}$ is a **unit-speed** curve if

$$\|\dot{\boldsymbol{\gamma}}(t)\| = 1 \,, \quad \forall\, t \in (a,b) \,.$$

Remark 1.38

The derivative of the arc-length s gives the speed of $\boldsymbol{\gamma}$:

$$s(t) := \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \quad \Longrightarrow \quad \dot{s}(t) = \|\dot{\boldsymbol{\gamma}}(t)\| \,.$$

The reason why we introduce unit speed curves is because they make calculations easy. This is essentially because of the next proposition.

Proposition 1.39

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a unit speed curve. Then

$$\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0$$

for all $t \in (a,b)$.

Proof

Let us consider the identity

$$\dot{\boldsymbol{\gamma}}(t) \cdot \dot{\boldsymbol{\gamma}}(t) = \sum_{i=1}^{n} \dot{\boldsymbol{\gamma}}_i^2(t) = \|\dot{\boldsymbol{\gamma}}(t)\|^2 \,. \tag{1.18}$$

Since $\boldsymbol{\gamma}$ is unit speed we have

$$\|\dot{\boldsymbol{\gamma}}(t)\|^2 = 1 \quad \forall\, t \in (a,b) \,.$$

and therefore

$$\frac{d}{dt}\left(\|\dot{\boldsymbol{\gamma}}(t)\|^2\right) = 0 \quad \forall\, t \in (a,b)\,. \tag{1.19}$$

We can differentiate the LHS of (1.18) to get

$$\frac{d}{dt}(\dot{\boldsymbol{\gamma}}\cdot\dot{\boldsymbol{\gamma}}) = \ddot{\boldsymbol{\gamma}}\cdot\dot{\boldsymbol{\gamma}} + \dot{\boldsymbol{\gamma}}\cdot\ddot{\boldsymbol{\gamma}} = 2\dot{\boldsymbol{\gamma}}\cdot\ddot{\boldsymbol{\gamma}}\,. \tag{1.20}$$

where we used Proposition 1.36 and symmetry of the scalar product. Differentiating (1.18) and using (1.19)-(1.20) we conclude

$$2\dot{\boldsymbol{\gamma}}\cdot\ddot{\boldsymbol{\gamma}} = 0 \quad \forall\, t \in (a,b)\,.$$

Remark 1.40

Proposition 1.39 is saying that if $\boldsymbol{\gamma}$ is unit speed, then its tangent vector $\dot{\boldsymbol{\gamma}}$ is always orthogonal to the second derivative $\ddot{\boldsymbol{\gamma}}$. This will be very useful in the future.

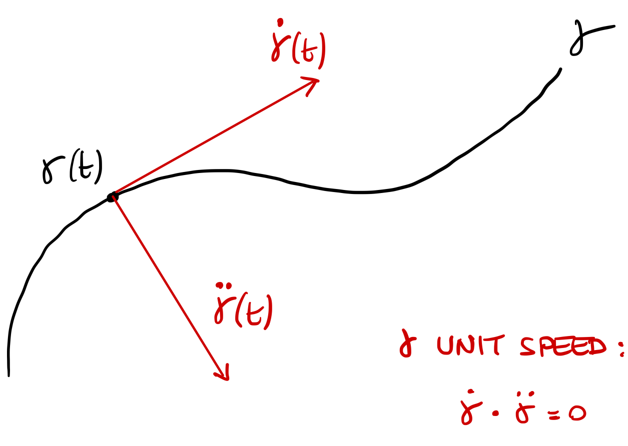


Figure 1.21: If $\boldsymbol{\gamma}$ is unit speed then $\dot{\boldsymbol{\gamma}}$ and $\ddot{\boldsymbol{\gamma}}$ are orthogonal

1.9 Reparametrization

As we have observed in the Examples of Chapter 1, there is in general no unique way to parametrize a curve. However we would like to understand when two parametrizations are related. In other words, we want to clarify the concept of **equivalence** of two parametrizations.

Definition 1.41: Diffeomorphism

Let $\phi : (a,b) \to (\tilde{a},\tilde{b})$. We say that ϕ is a **diffeomorphism** if the following conditions are satisfied:

1. ϕ is invertible, with inverse $\phi^{-1} : (\tilde{a},\tilde{b}) \to (a,b)$. Thus
$$\phi^{-1} \circ \phi = \phi \circ \phi^{-1} = \mathrm{Id}\,,$$
where $\mathrm{Id} : \mathbb{R} \to \mathbb{R}$ is the identity map on $\mathbb{R}$, that is,
$$\mathrm{Id}(t) = t\,, \quad \forall\, t \in \mathbb{R}\,.$$
2. ϕ is smooth,
3. ϕ^{-1} is smooth.

Definition 1.42: Reparametrization

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a parametrized curve. A **reparametrization** of $\boldsymbol{\gamma}$ is another parametrized curve $\tilde{\boldsymbol{\gamma}} : (\tilde{a},\tilde{b}) \to \mathbb{R}^n$ such that
$$\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(\phi(t)) \quad \forall\, t \in (\tilde{a},\tilde{b})\,, \tag{1.21}$$
where
$$\phi : (\tilde{a},\tilde{b}) \to (a,b)$$
is a diffeomerphism. We call both ϕ and ϕ^{-1} **reparametrization maps**.

Remark 1.43

A comment about the above definition. Given a parametrized curve $\boldsymbol{\gamma}$, this identifies a 1D shape $\Gamma \subset \mathbb{R}^n$. A reparametrization $\tilde{\boldsymbol{\gamma}}$ is just an equivalent way to describe Γ. For $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ to be reparametrizations of each other, there must exist a smooth rule ϕ to switch from one to another, according to formula (1.21)

Example 1.44: Change of orientation

The map $\phi : (\tilde{a},\tilde{b}) \to (a,b)$ defined by
$$\phi(t) := -t$$
is a diffeomoprhism. The inverse of ϕ is given by $\phi^{-1} : (a,b) \to (\tilde{a},\tilde{b})$ defined by
$$\phi^{-1}(t) = -t\,.$$

Note that ϕ can be used to **reverse the orientation** of a curve.

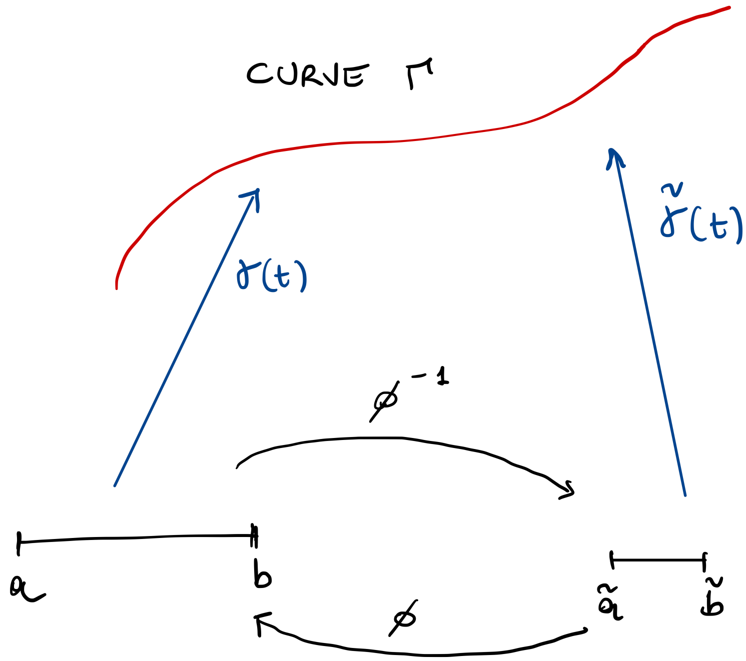


Figure 1.22: Sketch of 1D shaper □parametrized by $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$

Example 1.45: Reversing orientation of circle

Consider the unit circle parametrized as usual by $\boldsymbol{\gamma} : [0, 2\pi] \to \mathbb{R}^2$ defined as

$$\boldsymbol{\gamma}(t) := (\cos(t), \sin(t)) .$$

To reverse the orientation we can reparametrize $\boldsymbol{\gamma}$ by using the diffeomorphism

$$\phi(t) := -t .$$

This way we obtain $\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ \phi : [0, 2\pi] \to [0, 2\pi]$,

$$\begin{aligned} \tilde{\boldsymbol{\gamma}}(t) &= \boldsymbol{\gamma}(\phi(t)) \\ &= (\cos(-t), \sin(-t)) \\ &= (\cos(t), -\sin(t)) , \end{aligned}$$

where in the last identity we used the properties of cos and sin. Notice that in this way, for example,

$$\boldsymbol{\gamma}(\pi/2) = (0, 1) , \quad \boldsymbol{\gamma}(\pi/2) = (0, -1) .$$

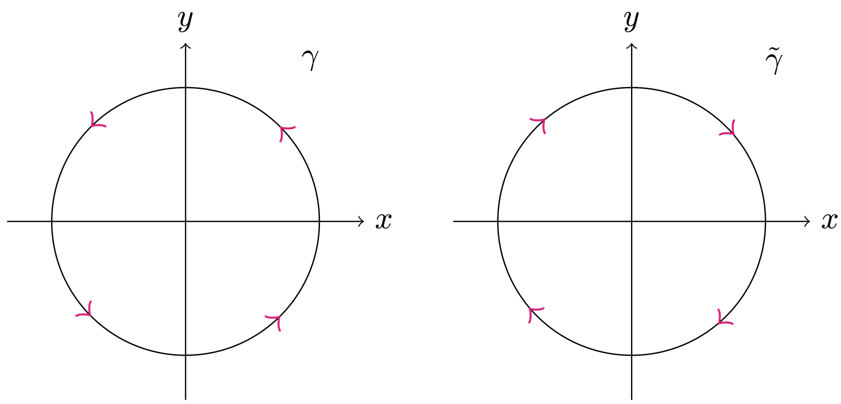


Figure 1.23: Unit circle with usual parametrization $\boldsymbol{\gamma}$, and with reversed orientation $\tilde{\boldsymbol{\gamma}}$

Example 1.46: Change of speed

Let $k > 0$. The map $\phi : (\tilde{a}, \tilde{b}) \to (a, b)$ defined by

$$\phi(t) := kt$$

is a diffeomoprhism. The inverse of ϕ is given by $\phi^{-1} : (a, b) \to (\tilde{a}, \tilde{b})$ defined by

$$\phi^{-1}(t) = \frac{t}{k}\,.$$

Note that ϕ can be used to **change the speed** of a curve:

- If $k > 1$ the speed increases ,
- If $0 < k < 1$ the speed decreases.

Example 1.47: Doubling the speed of Lemniscate

Recall the Lemniscate

$$\boldsymbol{\gamma}(t) := (\sin(t), \sin(t)\cos(t))\,, \quad t \in [0, 2\pi]\,.$$

We can double the speed of the Lemniscate by using the Using the diffeomorphism

$$\phi(t) := 2t\,.$$

This way we obtain $\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ \phi : [0, \pi] \to [0, 2\pi]$ with

$$\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(\phi(t)) = (\sin(2t), \sin(2t)\cos(2t))\,.$$

In this case we have that

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) = 2\dot{\boldsymbol{\gamma}}(\phi(t))\,.$$

The above follows by chain rule. Indeed, $\dot{\phi} = 2$, so that

$$\dot{\tilde{\boldsymbol{\gamma}}} = \frac{d}{dt}(\boldsymbol{\gamma}(\phi(t))) = \dot{\phi}(t)\dot{\boldsymbol{\gamma}}(\phi(t)) = 2\dot{\boldsymbol{\gamma}}(\phi(t))\,.$$

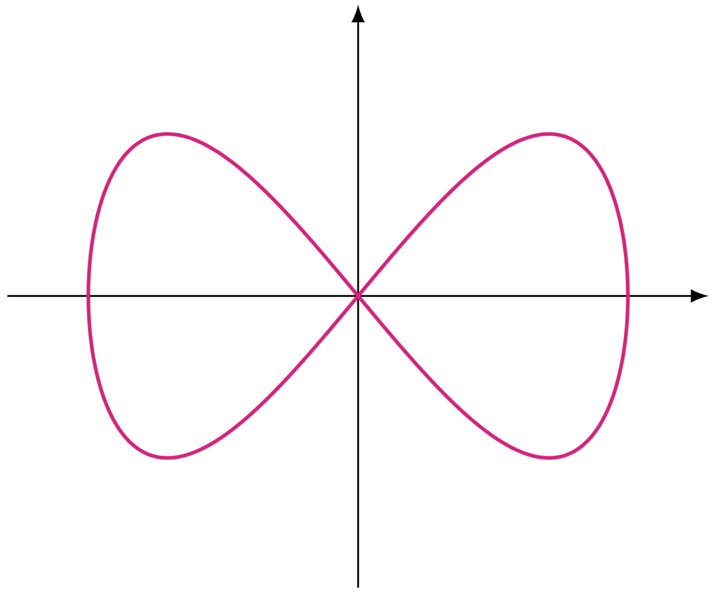


Figure 1.24: Lemniscate curve

Important

The main reason we are interested in reparametrizations is because we want to parametrize curves by **arc-lenght**: This means that, for a curve $\boldsymbol{\gamma}$, we want to find a reparametrization $\tilde{\boldsymbol{\gamma}}$ such that $\tilde{\boldsymbol{\gamma}}$ is unit speed:

$$\|\dot{\tilde{\boldsymbol{\gamma}}}\| = 1\,, \quad \forall t \in (a,b)\,.$$

We will see that this is not always possible.

Definition 1.48: Regular points

Let $\boldsymbol{\gamma} \,:\, (a,b) \to \mathbb{R}^n$ be a parametrized curve. We say that:

- $\boldsymbol{\gamma}(t_0)$ is a **regular point** if
$$\dot{\boldsymbol{\gamma}}(t_0) \neq 0\,.$$
- A point $\boldsymbol{\gamma}(t_0)$ is **singular** if it is not regular.
- The curve $\boldsymbol{\gamma}$ is regular if every point of $\boldsymbol{\gamma}$ is regular, that is,
$$\dot{\boldsymbol{\gamma}}(t) \neq 0\,, \quad \forall\, t \in (a,b)\,.$$

Note that when $\dot{\boldsymbol{\gamma}}(t_0) = 0$, this means the curve is *stopping* at time t_0. Before making an example, let us prove a useful lemma about diffeomorphisms.

Lemma 1.49

Let $\phi : (a,b) \to (\tilde{a},\tilde{b})$ be a diffeomorphism. Then

$$\dot{\phi}(t) \neq 0 \quad \forall \, t \in (a,b) \, .$$

Proof

We know that ϕ is smooth with smooth inverse

$$\psi := \phi^{-1} : (\tilde{a},\tilde{b}) \to (a,b) \, .$$

In particular it holds

$$\psi(\phi(t)) = t \, , \quad \forall \, t \in (a,b) \, .$$

We can differentiate both sides of the above expression to get

$$\frac{d}{dt} \left(\psi(\phi(t)) \right) = 1 \, . \tag{1.22}$$

We can differentiate the LHS by chain rule

$$\frac{d}{dt} \left(\psi(\phi(t)) \right) = \dot{\psi}(\phi(t)) \, \dot{\phi}(t) \, .$$

From (1.22) we then get

$$\dot{\psi}(\phi(t)) \, \dot{\phi}(t) = 1 \, , \quad \forall \, t \in (a,b) \, .$$

Since on the LHS we have a product, this means that none of the LHS terms vanishes, so that

$$\dot{\phi}(t) \neq 0 \, , \quad \forall \, t \in (a,b) \, .$$

Example 1.50: A curve with one singular point

Consider the parabola

$$\Gamma := \{(x,y) \in \mathbb{R}^2 \, : \, y = x^2, \, -1 \leq x \leq 1\} \, .$$

This can be parametrized in two ways by $\boldsymbol{\gamma}, \boldsymbol{\eta} : [-1,1] \to \mathbb{R}^2$ defined as

$$\boldsymbol{\gamma}(t) = (t, t^2) \, , \quad \boldsymbol{\eta}(t) = (t^3, t^6) \, .$$

We will see that the above parametrizations are **not** equivalent. This is intuitively clear, since the change of variables map should be

$$\phi(t) = t^3 \, .$$

This is smooth and invertible, with inverse

$$\phi^{-1}(t) = \sqrt[3]{x} \, .$$

However ϕ^{-1} is not smooth at $t = 0$, and thus ϕ is not a diffeomorphism. Alternatively we could have just noticed that

$$\dot{\phi}(t) = 3t^2 \quad \Longrightarrow \quad \dot{\phi}(0) = 0\,,$$

and therefore ϕ cannot be a diffeomorphism due to Lemma 1.49.
Let us look at the derivatives:

$$\dot{\boldsymbol{\gamma}}(t) = (1, 2t)\,, \quad \dot{\boldsymbol{\eta}}(t) = (3t^2, 6t^5)\,.$$

We notice a difference:

- $\boldsymbol{\gamma}$ is a regular parametrization,
- $\boldsymbol{\eta}(t)$ is regular only for $t \neq 0$.

Indeed if we animate the plots of the above parametrizations, we see that:

- The point $\boldsymbol{\gamma}(t)$ moves with constant horizontal speed
- The point $\boldsymbol{\eta}(t)$ is decelerating for $t < 0$, it stops at $t = 0$, and then accelerates again for $t > 0$.

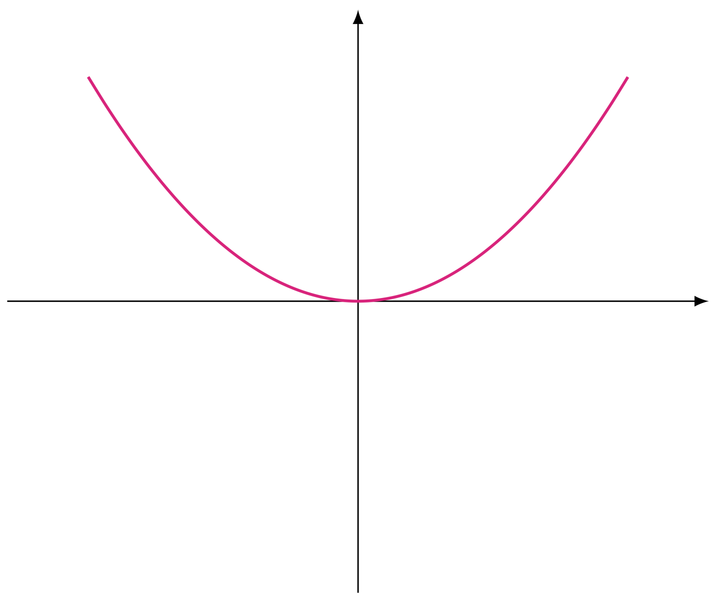


Figure 1.25: Parabola Γ

Proposition 1.51: Regularity is invariant for reparametrization

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^n$ be a parametrized curve and suppose that $\boldsymbol{\gamma}$ is regular, that is,

$$\dot{\boldsymbol{\gamma}}(t) \neq 0\,, \quad \forall\, t \in (a, b)\,.$$

Then every reparametrization of $\boldsymbol{\gamma}$ is also regular.

Proof

Let $\tilde{\boldsymbol{\gamma}} : (\tilde{a}, \tilde{b}) \to \mathbb{R}^n$ be a reparametrization of $\boldsymbol{\gamma}$. Then there exist $\phi : (\tilde{a}, \tilde{b}) \to (a, b)$ diffeomorphism such that

$$\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(\phi(t)), \quad \forall\, t \in (\tilde{a}, \tilde{b}).$$

By the chain rule we have

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) = \frac{d}{dt}(\boldsymbol{\gamma}(\phi(t))) = \dot{\boldsymbol{\gamma}}(\phi(t))\dot{\phi}(t).$$

Therefore

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) \neq 0 \quad \iff \quad \dot{\boldsymbol{\gamma}}(\phi(t))\dot{\phi}(t) \neq 0. \tag{1.23}$$

But we are assuming that $\boldsymbol{\gamma}$ is regular, so that

$$\dot{\boldsymbol{\gamma}}(\phi(t)) \neq 0, \quad \forall\, t \in (\tilde{a}, \tilde{b}).$$

Thus (1.23) is equivalent to

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) \neq 0 \quad \iff \quad \dot{\phi}(t) \neq 0. \tag{1.24}$$

Since ϕ is a diffeomorphism, by Lemma 1.49 we have that

$$\dot{\phi}(t) \neq 0, \quad \forall\, t \in (\tilde{a}, \tilde{b}).$$

By (1.24) we conclude that

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) \neq 0, \quad \forall\, t \in (\tilde{a}, \tilde{b}),$$

proving that $\tilde{\boldsymbol{\gamma}}$ is regular.

Example 1.52

Let us go back to the parabola

$$\Gamma := \{(x, y) \in \mathbb{R}^2 : \ y = x^2, \ -1 \leq x \leq 1\},$$

with the two parametrizations $\boldsymbol{\gamma}, \boldsymbol{\eta} : [-1, 1] \to \mathbb{R}^2$ with

$$\boldsymbol{\gamma}(t) = (t, t^2), \quad \boldsymbol{\eta}(t) = (t^3, t^6).$$

We have that

$$\dot{\boldsymbol{\gamma}}(t) = (1, 2t), \quad \dot{\boldsymbol{\eta}}(t) = (3t^2, 6t^5).$$

Therefore

- $\boldsymbol{\gamma}$ is a regular parametrization,
- $\boldsymbol{\eta}(t)$ is regular only for $t \neq 0$.

Proposition 1.51 implies that $\boldsymbol{\eta}$ is **NOT** a reparametrization of $\boldsymbol{\gamma}$.

Definition 1.53: Unit speed reparametrization

Let $\boldsymbol{\gamma}$ be a parametrized curve. A **unit speed reparametrization** of $\boldsymbol{\gamma}$ is a reparametrization $\tilde{\boldsymbol{\gamma}}$ such that $\tilde{\boldsymbol{\gamma}}$ is unit speed.

The next theorem states that a curve is regular if and only if it has a unit speed reparametrization. For the proof, it is crucial to recall the definition of arc-length of a curve $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$, which is given by

$$s(t) := \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \,,$$

for some arbitrary $t_0 \in (a,b)$ fixed. Indeed, we will see that for ϕ regular the unit speed parametrization map can be taken as

$$\phi = s^{-1} \,.$$

Theorem 1.54: Existence of unit speed reparametrization

Let $\boldsymbol{\gamma}$ be a parametrized curve. They are equivalent:

- $\boldsymbol{\gamma}$ is regular,
- $\boldsymbol{\gamma}$ has a unit speed reparametrization.

Proof

Step 1. Direct implication.
Assume $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ is regular, that is,

$$\dot{\boldsymbol{\gamma}}(t) \neq 0 \,, \quad \forall \, t \in (a,b) \,.$$

Let $s : (a,b) \to \mathbb{R}$ be the arc-length of $\boldsymbol{\gamma}$ starting at any point $t_0 \in (a,b)$. By the Fundamental Theorem of Calculus we have

$$\dot{s}(t) = \|\dot{\boldsymbol{\gamma}}(t)\| \tag{1.25}$$

so that

$$\dot{s}(t) > 0 \,, \quad \forall \, t \in (a,b) \,.$$

Since s is a scalar function, the above condition and the Inverse Function Theorem guarantee the exsistence of a smooth inverse

$$s^{-1} : (\tilde{a}, \tilde{b}) \to (a,b)$$

for some $\tilde{\alpha} < \tilde{\beta}$. Define the reparametrization map ϕ as

$$\phi := s^{-1}$$

and the corresponding reparametrization of $\boldsymbol{\gamma}$ given by the curve

$$\tilde{\boldsymbol{\gamma}} : (\tilde{a}, \tilde{b}) \to \mathbb{R}^n \,, \quad \tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ \phi \,.$$

We claim that $\tilde{\boldsymbol{\gamma}}$ is unit speed. Indeed, by definition

$$\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ \phi \quad \Longrightarrow \quad \boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi^{-1} = \tilde{\boldsymbol{\gamma}} \circ s \,,$$

or in other words

$$\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\gamma}}(s(t)) \,, \quad \forall t \in (a,b) \,.$$

Differentiating the above expression and using the chain rule we get

$$\dot{\boldsymbol{\gamma}}(t) = \dot{\tilde{\boldsymbol{\gamma}}}(s(t))\, \dot{s}(t) = \dot{\tilde{\boldsymbol{\gamma}}}(s(t))\, \|\dot{\boldsymbol{\gamma}}(t)\|$$

where in the last equality we used (1.25). Taking the absolute value of the above yileds

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \left\|\dot{\tilde{\boldsymbol{\gamma}}}(s(t))\right\| \, \|\dot{\boldsymbol{\gamma}}(t)\| \,. \tag{1.26}$$

Since $\boldsymbol{\gamma}$ is regular, we have

$$\|\dot{\boldsymbol{\gamma}}(t)\| \neq 0 \,, \quad \forall\, t \in (a,b) \,.$$

Therefore we can divide (1.26) by $\|\dot{\boldsymbol{\gamma}}(t)\|$ and obtain

$$\left\|\dot{\tilde{\boldsymbol{\gamma}}}(s(t))\right\| = 1 \,, \quad \forall\, t \in (a,b) \,.$$

By invertibility of s, the above holds if and only if

$$\left\|\dot{\tilde{\boldsymbol{\gamma}}}(t)\right\| = 1 \,, \quad \forall\, t \in (\tilde{a},\tilde{b}) \,,$$

showing that $\tilde{\boldsymbol{\gamma}}$ is a unit speed reparametrization of $\boldsymbol{\gamma}$.
Step 2. Reverse implication.
Suppose there exists a unit speed reparametrization of $\boldsymbol{\gamma}$ denoted by

$$\tilde{\boldsymbol{\gamma}} \,:\, (\tilde{a},\tilde{b}) \to \mathbb{R}^n \,, \quad \tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} \circ \phi$$

for some reparametrization map $\phi \,:\, (\tilde{a},\tilde{b}) \to (a,b)$. Differentiating $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} \circ \phi$ and using the chain rule we get

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) = \dot{\boldsymbol{\gamma}}(\phi(t))\, \dot{\phi}(t) \,.$$

Taking the norm

$$\left\|\dot{\tilde{\boldsymbol{\gamma}}}(t)\right\| = \|\dot{\boldsymbol{\gamma}}(\phi(t))\| \, |\dot{\phi}(t)| \,.$$

Since $\tilde{\boldsymbol{\gamma}}$ is unit speed we obtain

$$\|\dot{\boldsymbol{\gamma}}(\phi(t))\| \, |\dot{\phi}(t)| = 1 \,, \quad \forall\, t \in (\tilde{a},\tilde{b}) \,. \tag{1.27}$$

Since ϕ is a diffeomorphism from $(\tilde{a},\tilde{b})$ into (a,b), Lemma 1.49 guarantees that

$$\dot{\phi}(t) \neq 0 \,, \quad \forall\, t \in (a,b) \,.$$

In particular (1.27) implies

$$\dot{\boldsymbol{\gamma}}(\phi(t)) \neq 0 \,, \quad \forall\, t \in (\tilde{a},\tilde{b}) \,.$$

As ϕ is invertible, we also have

$$\dot{\boldsymbol{\gamma}}(t) \neq 0 \,, \quad \forall\, t \in (a,b) \,,$$

proving that $\boldsymbol{\gamma}$ is regular.

The proof of Theorem 1.54 told us that, if $\boldsymbol{\gamma}$ is regular, then

$$\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} \circ s^{-1}$$

is a unit speed reparametrization of $\boldsymbol{\gamma}$. In the next proposition we show that the arc-length s is essentially the only unit-speed reparametrization of a regular curve.

Proposition 1.55: Arc-length and unit speed reparametrization

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a regular curve. Let $\tilde{\boldsymbol{\gamma}} : (\tilde{a}, \tilde{b}) \to \mathbb{R}^n$ be reparametrization of $\boldsymbol{\gamma}$, so that

$$\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\gamma}}(\phi(t)), \quad \forall\, t \in (a,b)\,.$$

for some diffeomorphism $\phi : (a,b) \to (\tilde{a}, \tilde{b})$. Denote by

$$s(t) := \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau\,, \quad t \in (a,b)$$

the arc-length of $\boldsymbol{\gamma}$ starting at any point $t_0 \in (a,b)$. We have:

1. If $\tilde{\boldsymbol{\gamma}}$ is unit speed, then there exists $c \in \mathbb{R}$ such that
$$\phi(t) = \pm s(t) + c\,, \quad \forall\, t \in (a,b)\,. \tag{1.28}$$
2. If ϕ is given by (1.28) for some $c \in \mathbb{R}$, then $\tilde{\boldsymbol{\gamma}}$ is unit speed.

Proof

Step 1. First Point.
First note that a unit speed reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ exists by Theorem 1.54, since $\boldsymbol{\gamma}$ is assumed to be regular. Thus assume $\tilde{\boldsymbol{\gamma}}$ is unit speed reparametrization of $\boldsymbol{\gamma}$. By differentiating both sides of

$$\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\gamma}}(\phi(t)), \quad \forall\, t \in (a,b)\,,$$

we obtain

$$\dot{\boldsymbol{\gamma}}(t) = \frac{d}{dt}\tilde{\boldsymbol{\gamma}}(\phi(t)) = \dot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\, \dot{\phi}(t)\,.$$

Taking the norms we then have

$$\begin{aligned} \|\dot{\boldsymbol{\gamma}}(t)\| &= \left\|\dot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\, \dot{\phi}(t)\right\| \\ &= \left\|\dot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\right\| \, |\dot{\phi}(t)| \\ &= |\dot{\phi}(t)|\,, \end{aligned}$$

where in the last equality we used that $\tilde{\boldsymbol{\gamma}}$ is unit speed, and so

$$\left\|\dot{\tilde{\boldsymbol{\gamma}}}\right\| \equiv 1\,.$$

To summarize, so far we have proven that

$$\|\dot{\boldsymbol{\gamma}}(t)\| = |\dot{\phi}(t)| \,, \quad \forall \, t \in (a,b) \,.$$

Therefore

$$s(t) = \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau = \int_{t_0}^{t} |\dot{\phi}(\tau)| \, d\tau \,.$$

By the Fundamental Theorem of Calculus we get

$$\dot{s}(t) = |\dot{\phi}(t)|$$

and therefore

$$\phi = \pm s + c$$

for some $c \in \mathbb{R}$, concluding the proof.
Step 2. Second Point.
Suppose that

$$\phi := \pm s + c$$

for some $c \in \mathbb{R}$, so that $\phi : (a,b) \to (\tilde{a}, \tilde{b})$. We have

$$\dot{\phi}(t) = \pm \dot{s}(t) = \pm \|\dot{\boldsymbol{\gamma}}(t)\| \neq 0 \tag{1.29}$$

where the last term is non-zero since $\boldsymbol{\gamma}$ is regular. Therefore, due to the Inverse Function Theorem, ϕ is invertible with smooth inverse. This proves that $\tilde{\boldsymbol{\gamma}}$ defined by

$$\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ \psi \,, \quad \psi := \phi^{-1} \,,$$

is a reparametrization of $\boldsymbol{\gamma}$. In particular

$$\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi \,.$$

Differentiating the above, and recalling (1.29), we get

$$\dot{\boldsymbol{\gamma}}(t) = \dot{\tilde{\boldsymbol{\gamma}}}(\phi(t)) \, \dot{\phi}(t) = \dot{\tilde{\boldsymbol{\gamma}}}(\phi(t)) \, (\pm \|\dot{\boldsymbol{\gamma}}(t)\|) \,.$$

Taking the absolute value of the above yields

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \left\|\dot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\right\| \, \|\dot{\boldsymbol{\gamma}}(t)\| \,.$$

Since $\boldsymbol{\gamma}$ is regular we can divide by $\|\dot{\boldsymbol{\gamma}}(t)\|$ to get

$$\left\|\dot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\right\| = 1 \quad \forall \, t \in (a,b) \,.$$

Since ϕ is invertible, the above is equivalent to

$$\left\|\dot{\tilde{\boldsymbol{\gamma}}}(t)\right\| = 1 \quad \forall \, t \in (\tilde{a}, \tilde{b}) \,,$$

proving that $\tilde{\boldsymbol{\gamma}}$ is a unit speed reparametrization.

Remark 1.56

Let $\boldsymbol{\gamma}$ be regular. The above proposition tells us that they are equivalent:

1. Computing a unit speed reparametrization of $\boldsymbol{\gamma}$,
2. Computing s the arc-length of $\boldsymbol{\gamma}$.

In some cases however, unit speed reparametrization and arc-length are impossible to characterize in terms of elementary functions, even for very simple curves.

Example 1.57: Twisted cubic

Define the **twisted cubic** $\boldsymbol{\gamma} \,:\, \mathbb{R} \to \mathbb{R}^3$ by

$$\boldsymbol{\gamma}(t) = (t, t^2, t^3)\,.$$

Therefore

$$\dot{\boldsymbol{\gamma}}(t) = (1, 2t, 3t^2)\,,$$

so that

$$\dot{\boldsymbol{\gamma}}(t) \neq 0\,, \quad \forall\, t \in \mathbb{R}\,,$$

meaning that $\boldsymbol{\gamma}$ is regular. In particular we have

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \sqrt{1 + 4t^2 + 9t^4}$$

so that the arc-length of $\boldsymbol{\gamma}$ is

$$s(t) = \int_{t_0}^{t} \sqrt{1 + 4\tau^2 + 9\tau^4}\, d\tau\,.$$

Since $\boldsymbol{\gamma}$ is regular, by Proposition 1.55 we know that $\boldsymbol{\gamma}$ admits a unit speed reparametrization $\tilde{\boldsymbol{\gamma}}$ such that

$$\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi$$

with the diffeomorphism ϕ given by

$$\phi(t) = \pm s(t) + c = \pm \int_{t_0}^{t} \sqrt{1 + 4\tau^2 + 9\tau^4}\, d\tau + c$$

for some $c \in \mathbb{R}$. It can be shown that the above integral does not have a closed form in terms of elementary functions. Therefore the unit speed parametrization $\tilde{\boldsymbol{\gamma}}$ cannot be computed explicitly.

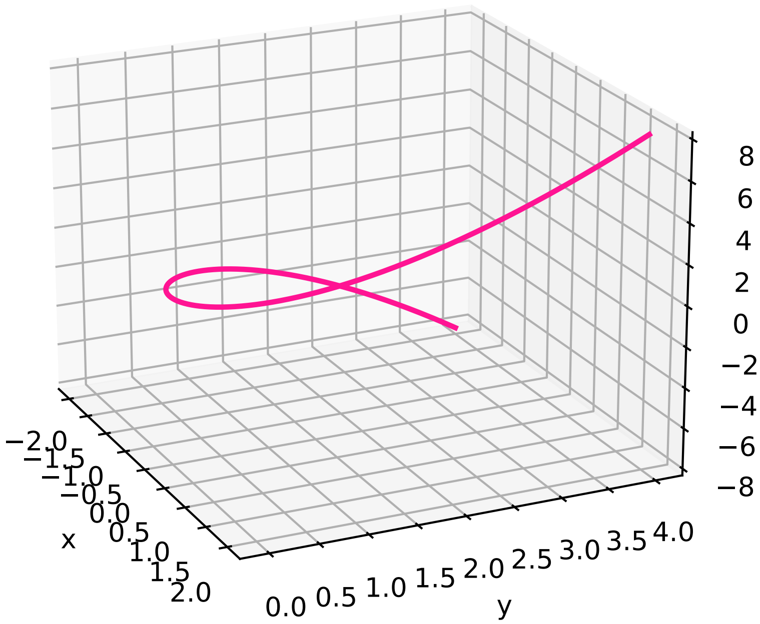


Figure 1.26: Plot of Twisted Cubic for t between -2 and 2

1.10 Closed curves

So far we have seen examples of:

- Curves which are infinite, or **unbounded**. This is for example the parabola

$$\boldsymbol{\gamma}(t) := (t, t^2), \quad \forall t \in \mathbb{R},$$

- Curves which are finite and have end-points, such as the semi-circle

$$\boldsymbol{\gamma}(t) := (\cos(t), \sin(t)), \quad \forall t \in [0, \pi],$$

- Curves which form **loops**, such as the circle

$$\boldsymbol{\gamma}(t) := (\cos(t), \sin(t)), \quad \forall t \in [0, 2\pi].$$

However there are examples of curves which are in between the above types.

Example 1.58

For example consider the curve $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^2$

$$\boldsymbol{\gamma}(t) := (t^2 - 1, t^3 - t) \quad \forall t \in \mathbb{R}.$$

This curve has two main properties:

- $\boldsymbol{\gamma}$ is unbounded: If define $\tilde{\boldsymbol{\gamma}}$ as the restriction of $\boldsymbol{\gamma}$ to the time interval $[1, \infty)$, then $\tilde{\boldsymbol{\gamma}}$ is unbounded. A point which starts at $\boldsymbol{\gamma}(1) = (0, 0)$ goes towards infinity.
- $\boldsymbol{\gamma}$ contains a loop: If we define $\tilde{\boldsymbol{\gamma}}$ as the restriction of $\boldsymbol{\gamma}$ to the time interval $[-1, 1]$, then $\tilde{\boldsymbol{\gamma}}$ is a closed loop starting at $\boldsymbol{\gamma}(-1) = (0, 0)$ and returnning at $\boldsymbol{\gamma}(1) = (0, 0)$.

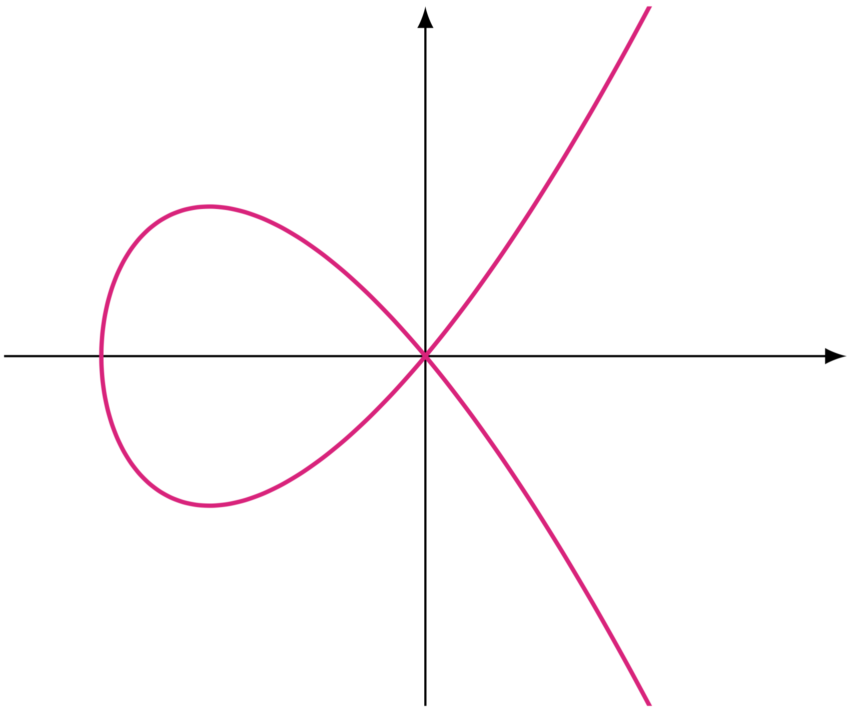


Figure 1.27: Plot of curve $\boldsymbol{\gamma}(t) = (t^2 - 1, t^3 - 1)$ for $t \in [-2, 2]$

The aim of this section is to make precise the concept of **looping curve**. To do that, we need to define **periodic curves**.

Definition 1.59: Periodic curve

Let $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ be a parametrized curve, and let $T \in \mathbb{R}$. We say that $\boldsymbol{\gamma}$ is **T-periodic** if

$$\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(t + T), \quad \forall\, t \in \mathbb{R}.$$

Note that every curve is 0-periodic. Therefore to define a closed curve we need to rule out this case.

Definition 1.60: Closed curve

Let $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ be a parametrized curve. We say that $\boldsymbol{\gamma}$ is **closed** if:

- $\boldsymbol{\gamma}$ is not constant,
- $\boldsymbol{\gamma}$ is T-periodic for some $T \neq 0$.

Remark 1.61

We have the following basic facts:

1. If $\boldsymbol{\gamma}$ is T-periodic, then a point moving around $\boldsymbol{\gamma}$ returns to its starting point after time T.

   This is exactly the definition of T-periodicity. Indeed let $p = \boldsymbol{\gamma}(a)$ be the point in question, then
   $$\boldsymbol{\gamma}(a+T) = \boldsymbol{\gamma}(a) = p$$
   by periodicity. Thus $\boldsymbol{\gamma}$ returns to p after time T.

2. If $\boldsymbol{\gamma}$ is T-periodic, then $\boldsymbol{\gamma}$ is determined by its restriction to any interval of length $|T|$.

3. Conversely, suppose that $\boldsymbol{\gamma} : [a,b] \to \mathbb{R}^n$ satisfies
   $$\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b)\,, \qquad \frac{d^k \boldsymbol{\gamma}}{dt^k}(a) = \frac{d^k \boldsymbol{\gamma}}{dt^k}(b)$$
   for all $k \in \mathbb{N}$. Set
   $$T := b - a\,.$$
   Then $\boldsymbol{\gamma}$ can be extended to a T-periodic curve $\tilde{\boldsymbol{\gamma}} : \mathbb{R} \to \mathbb{R}^n$ defined by
   $$\tilde{\boldsymbol{\gamma}}(t) := \boldsymbol{\gamma}(\tilde{t})\,, \quad \tilde{t} := t - \left\lfloor \frac{t-a}{b-a} \right\rfloor (b-a)\,, \quad \forall\, t \in \mathbb{R}\,.$$
   The above means that $\tilde{\boldsymbol{\gamma}}(t)$ is defined by $\boldsymbol{\gamma}(\tilde{t})$ where $\tilde{t}$ is the unique point in $[a,b]$ such that
   $$t = \tilde{t} + k(b-a)$$
   with $k \in \mathbb{Z}$ defined by
   $$k := \left\lfloor \frac{t-a}{b-a} \right\rfloor,$$
   see figure below. In this way $\tilde{\boldsymbol{\gamma}}$ is T-periodic.

4. If $\boldsymbol{\gamma}$ is T-periodic, then it is also $(-T)$-periodic.

   Because if $\boldsymbol{\gamma}$ is T-periodic then
   $$\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}((t-T)+T) = \boldsymbol{\gamma}(t-T)$$
   where in the first equality we used the trivial identity $t = (t-T)+T$, while in the second equality we used T-periodicity of $\boldsymbol{\gamma}$.

5. If $\boldsymbol{\gamma}$ is T-periodic for some $T \neq 0$, then it is T-periodic for some $T > 0$.

   This is an immediate consequence of Point 4.

6. If $\boldsymbol{\gamma}$ is T-periodic the $\boldsymbol{\gamma}$ is (kT)-periodic, for all $k \in \mathbb{Z}$.

   By point 4 we can assume WLOG that $k \geq 0$. We proceed by induction:

   - The statement is true for $k = 1$, since $\boldsymbol{\gamma}$ is T-periodic.
   - Assume now that $\boldsymbol{\gamma}$ is kT-periodic. Then

   $$\begin{aligned} \boldsymbol{\gamma}(t + (k+1)T) &= \boldsymbol{\gamma}((t+T) + kT) \\ &= \boldsymbol{\gamma}(t+T) && \text{(by } kT\text{-periodicity)} \\ &= \boldsymbol{\gamma}(t) && \text{(by } T\text{-periodicity)} \end{aligned}$$

   showing that $\boldsymbol{\gamma}$ is $(k+1)T$-periodic.

   By induction we conclude that $\boldsymbol{\gamma}$ is (kT)-periodic for all $k \in \mathbb{N}$.

7. If $\boldsymbol{\gamma}$ is T_1-periodic and T_2-periodic then $\boldsymbol{\gamma}$ is $(k_1T_1 + k_2T_2)$-periodic, for all $k_1, k_2 \in \mathbb{Z}$.

   By Point 6 we know that $\boldsymbol{\gamma}$ is k_1T_1-periodic and k_2T_2-periodic. Set $T := k_1T_1 + k_2T_2$. We have

   $$\begin{aligned} \boldsymbol{\gamma}(t + T) &= \boldsymbol{\gamma}((t + k_1T_1) + k_2T_2) \\ &= \boldsymbol{\gamma}(t + k_1T_1) && \text{(by } k_2T_2\text{-periodicity)} \\ &= \boldsymbol{\gamma}(t) && \text{(by } k_1T_1\text{-periodicity)} \end{aligned}$$

   showing that $\boldsymbol{\gamma}$ is $(k_1T_1 + k_2T_2)$-periodic.

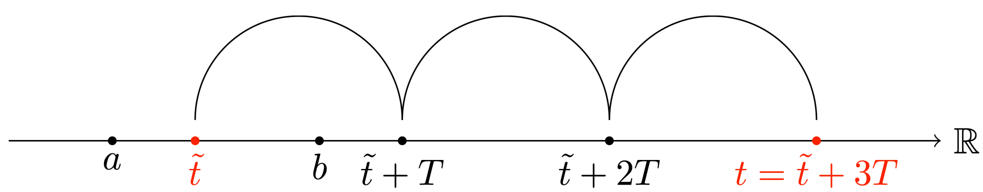


Figure 1.28: The points $t \in \mathbb{R}$ and $\tilde{t} \in [a, b]$ from Point 3 in Remark 1.61. In this skecth $t = \tilde{t} + 3T$, with $T = b - a$.

Definition 1.62

Let $\boldsymbol{\gamma}$ be a closed curve. The **period** of $\boldsymbol{\gamma}$ is the smallest $T > 0$ such that $\boldsymbol{\gamma}$ is T-periodic, that is

$$\text{Period of } \boldsymbol{\gamma} := \min\{T \ : \ T > 0 \,,\ \boldsymbol{\gamma} \ \text{ is T-periodic}\} \,.$$

We need to show that the above definition is well-posed, i.e., that there exists such smallest $T > 0$.

Proposition 1.63

Let γ be a closed curve. Then there exists a smallest $T > 0$ such that γ is T-periodic. In other words, the set

$$S := \{T \; : \; T > 0 \, , \; \gamma \; \text{is T-periodic}\} \, .$$

admits positive minumum

$$P = \min S \, , \quad P > 0 \, .$$

Proof

We make 2 observations about the set S:

- Since γ is closed, we have that γ is T-periodic for some $T \neq 0$. By Remark 1.61 Point 5, we know that T can be chosen such that $T > 0$. Therefore

$$S \neq \emptyset \, .$$

- S is bounded below by 0. This is by definition of S.

Thus, by the Axiom of Completeness of the Real Numbers, the set S admits an infimum

$$P = \inf S \, .$$

The proof is concluded if we show that:
Claim. We have

$$P = \min S \, .$$

This is equivalent to saying that

$$P \in S \, .$$

Proof of claim.
To see that $P \in S$ we need to show that

1. γ is P-periodic,
2. $P > 0$.

Since P is the infimum of S, there exists an infimizing sequence $\{T_n\}_{n \in \mathbb{N}} \subset S$ such that

$$T_n \to P \, .$$

WLOG we can choose T_n decreasing, that is, such that

$$T_1 > T_2 > \ldots > T_n > \ldots > 0 \, .$$

Proof of Point 1. As $T_n \in S$, we have that $\boldsymbol{\gamma}$ is T_n-periodic. Then

$$\boldsymbol{\gamma}(t + T_n) = \boldsymbol{\gamma}(t) \,, \quad \forall\, t \in \mathbb{R} \,,\; n \in \mathbb{N} \,.$$

Since $T_n \to P$, we can take the limit as $n \to \infty$ and use the continuity of $\boldsymbol{\gamma}$ to obtain

$$\boldsymbol{\gamma}(t) = \lim_{n \to \infty} \boldsymbol{\gamma}(t + T_n) = \boldsymbol{\gamma}(t + P) \,, \quad \forall\, t \in \mathbb{R} \,,$$

showing that $\boldsymbol{\gamma}$ is P-periodic.
Proof of Point 2. Suppose by contradiction that

$$P = 0 \,.$$

Fix $t \in \mathbb{R}$. Since $T_n > 0$, we can find unique

$$t_n \in [0, T_n] \,, \quad k_n \in \mathbb{Z} \,,$$

such that

$$t = t_n + k_n T_n \,,$$

as shown in the figure below. Indeed, it is sufficient to define

$$k_n := \left\lfloor \frac{t}{T_n} \right\rfloor \in \mathbb{Z} \,, \quad t_n := t - k_n T_n \,.$$

Since $T_n \in S$, we know that $\boldsymbol{\gamma}$ is T_n-periodic. Remark 1.61 Point 6 implies that $\boldsymbol{\gamma}$ is also $k_n T_n$-periodic, since $k_n \in \mathbb{Z}$. Thus

$$\begin{aligned} \boldsymbol{\gamma}(t) &= \boldsymbol{\gamma}(t_n + k_n T_n) && \text{(definition of } t_n\text{)} \\ &= \boldsymbol{\gamma}(t_n) && \text{(by } k_n T_n\text{-periodicity)} \,. \end{aligned}$$

Therefore

$$\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(t_n) \,, \quad \forall\, n \in \mathbb{N} \,. \tag{1.30}$$

Also notice that

$$0 \leq t_n \leq T_n \,, \quad \forall\, n \in \mathbb{N} \,.$$

by construction. Since $T_n \to 0$, by the Squeeze Theorem we conclude that

$$t_n \to 0 \quad \text{as} \;\; n \to \infty \,.$$

Using the continuity of $\boldsymbol{\gamma}$, we can pass to the limit in (1.30) and obtain

$$\boldsymbol{\gamma}(t) = \lim_{n \to \infty} \boldsymbol{\gamma}(t_n) = \boldsymbol{\gamma}(0) \,.$$

Since $t \in \mathbb{R}$ was arbitrary, we have shown that

$$\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(0) \,, \quad \forall\, t \in \mathbb{R} \,.$$

Therefore $\boldsymbol{\gamma}$ is constant. This is a contradiction, as we were assuming that $\boldsymbol{\gamma}$ is closed, and, in particular, not constant.

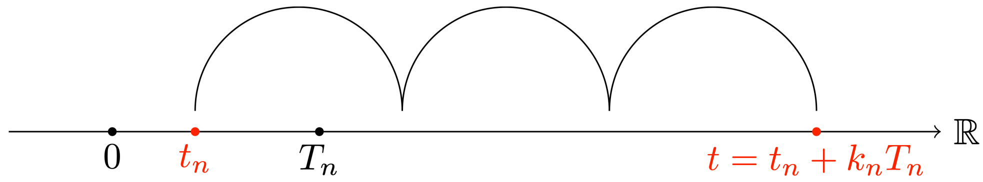


Figure 1.29: For each $t \in \mathbb{R}$ there exist unique $k_n \in \mathbb{Z}$ and $\tilde{t}_n \in [0, T_n]$ such that $t = \tilde{t} + k_n T_n$. In this skecth $k_n = 3$.

Example 1.64

Some examples of closed curves:

- The circumference
$$\gamma(t) = (\cos(t), \sin(t)), \quad t \in \mathbb{R}$$
is not costant and is 2π-periodic. Thus γ is closed. The period of γ is 2π.
- The Lemniscate
$$\gamma(t) = (\sin(t), \sin(t)\cos(t)), \quad t \in \mathbb{R}$$
is not costant and is 2π-periodic. Thus γ is closed. The period of γ is 2π.
- Consider again the curve from Example 1.58
$$\gamma(t) := (t^2 - 1, t^3 - t), \quad t \in \mathbb{R}.$$
According to our definition, γ is not periodic. Therefore γ is not closed. However there is a point of **self-intersection** on γ, namely
$$p := (0, 0),$$
for which we have
$$p = \gamma(-1) = \gamma(1).$$

The last curve in the above example motivates the definition of **self-intersecting** curve.

Definition 1.65: Self-intersecting curve

Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be a parametrized curve. We say that γ is **self-intersecting** at a point p on the curve if

1. There exist two times $a \neq b$ such that
$$p = \gamma(a) = \gamma(b),$$
2. If γ is closed with period T, then $b - a$ is not an integer multiple of T.

Remark 1.66

The second condition in the above definition is important: if we did not require it, then any closed curve would be self-intersecting. Indeed consider a closed curve $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^n$ and let T be its period. Then by Point 6 in Remark 1.61 we have

$$\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(a + kT) , \quad \forall\, a \in \mathbb{R},\, k \in \mathbb{Z} .$$

Therefore every point $\boldsymbol{\gamma}(a)$ would be of self-intersection. Point 2 in the above definition rules this example out. Indeed set $b := a + kT$, then

$$b - a = kT ,$$

meaning that $b - a$ is an integer multiple of T.

Example 1.67

Let us go back to the curve of Example 1.58, that is,

$$\boldsymbol{\gamma}(t) := (t^2 - 1, t^3 - t) , \quad t \in \mathbb{R} .$$

We have that $\boldsymbol{\gamma}$ is not periodic, and therefore not closed. However $p = (0, 0)$ is a point of **self-intersection** on $\boldsymbol{\gamma}$, since we have

$$p = \boldsymbol{\gamma}(-1) = \boldsymbol{\gamma}(1) .$$

Example 1.68: The Limaçon

Define the parametrized curve $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^2$ by

$$\boldsymbol{\gamma}(t) = ((1 + 2\cos(t))\cos(t), (1 + 2\cos(t))\sin(t)) , \quad \forall\, t \in \mathbb{R} .$$

Such curve, plotted bolow, is called limaçon (French for snail). This curve is non constant and 2π-periodic. Therefore it is closed. The period of $\boldsymbol{\gamma}$ is 2π. Moreover we have

$$\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b) = (0, 0) .$$

with $a = 2\pi/3$ and $b = 4\pi/3$. Note that

$$b - a = \frac{4\pi}{3} - \frac{2\pi}{3} = \frac{2\pi}{3}$$

which is not an integer multiple of the period 2π. Therefore $\boldsymbol{\gamma}$ is **self-intersecting** at $(0, 0)$.

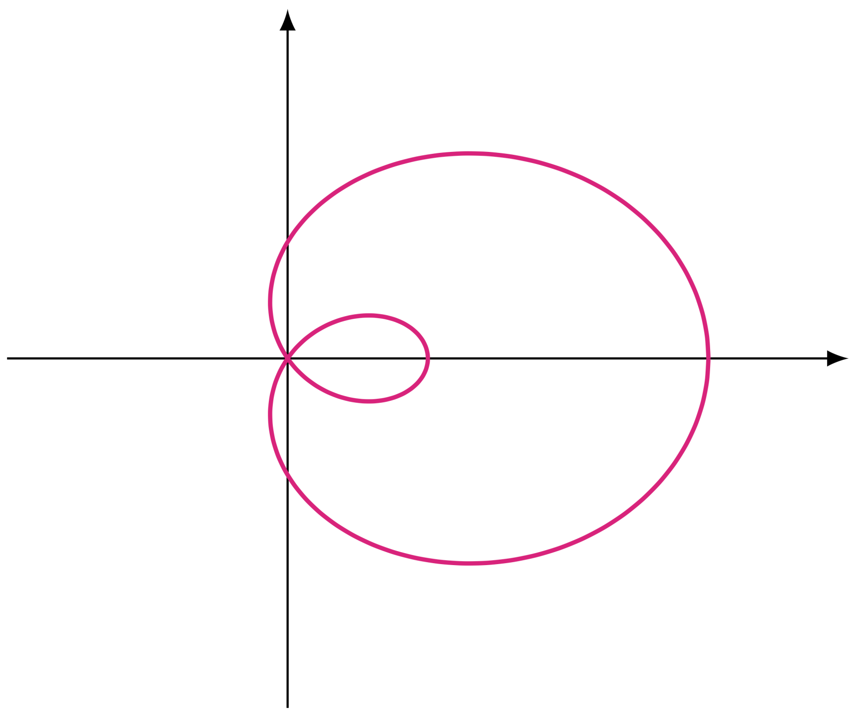


Figure 1.30: Limaçon curve

2 Curvature and Torsion

We have seen how to describe curves and reparametrized them. Now we want to look at local properties of curves:

- How much does a curve twist?
- How much does a curve bend?

We will measure two quantities:

- **Curvature**: measures how much a curve $\boldsymbol{\gamma}$ deviates from a straight line.
- **Torsion**: measures how much a curve $\boldsymbol{\gamma}$ fails to lie on a plane.

For example a 2D spiral is curved, but still lies in a plane. Instead the Helix both deviates from a straight line and *pulls away* from any fixed plane.

2.1 Curvature

We start with an informal discussion. Suppose $\boldsymbol{\gamma}$ is a straight line

$$\boldsymbol{\gamma}(t) = \mathbf{a} + t\mathbf{v}$$

with $\mathbf{a}, \mathbf{v} \in \mathbb{R}^3$. The tangent vector to $\boldsymbol{\gamma}$ is constant

$$\dot{\boldsymbol{\gamma}}(t) = \mathbf{v}\,.$$

Whatever the definition of curvature will be, it has to hold that $\boldsymbol{\gamma}$ has zero curvature in this case. If we further derive the tangent vector, we obtain

$$\ddot{\boldsymbol{\gamma}}(t) = \mathbf{0}\,.$$

Thus $\ddot{\boldsymbol{\gamma}}$ seems to be a good candidate for the definition of curvature of $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}(t)$.

Suppose now that $\boldsymbol{\gamma}$ is a curve in $\mathbb{R}^2$ with unit speed. We have proven that in this case

$$\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0\,,$$

that is, the vector $\ddot{\boldsymbol{\gamma}}$ is orthogonal to the tangent $\dot{\boldsymbol{\gamma}}$ at all times. Now let $\mathbf{n}(t)$ be the unit vector orthogonal to $\dot{\boldsymbol{\gamma}}(t)$ at the point $\boldsymbol{\gamma}(t)$. The amount that the curve $\boldsymbol{\gamma}$ deviates from its tangent at $\boldsymbol{\gamma}(t)$ after time t_0 is

$$(\boldsymbol{\gamma}(t + t_0) - \boldsymbol{\gamma}(t)) \cdot \mathbf{n}(t)\,, \tag{2.1}$$

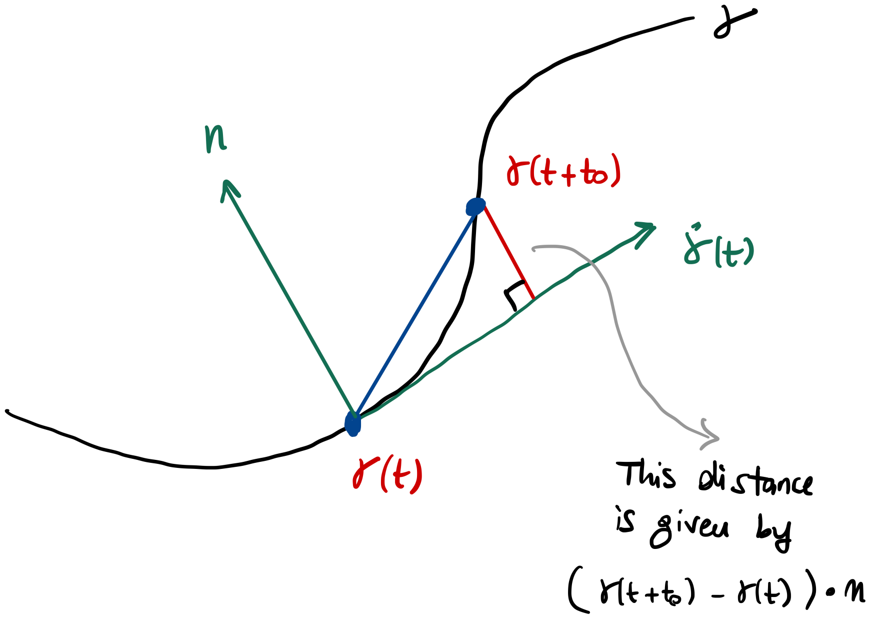


Figure 2.1: Amount that $\boldsymbol{\gamma}$ deviates from tangent is $(\boldsymbol{\gamma}(t+t_0) - \boldsymbol{\gamma}(t)) \cdot \mathbf{n}(t)$

as seen in the figure below.

Equation (2.1) is what we take as measure of curvature. Since

$$\dot{\boldsymbol{\gamma}}(t) \cdot \ddot{\boldsymbol{\gamma}}(t) = 0 \quad \text{and} \quad \dot{\boldsymbol{\gamma}}(t) \cdot \mathbf{n}(t) = 0\,,$$

we conclude that $\ddot{\boldsymbol{\gamma}}(t)$ is parallel to $\mathbf{n}(t)$. Since $\mathbf{n}(t)$ is a unit vector, there exists a scalar $\kappa(t)$ such that

$$\ddot{\boldsymbol{\gamma}}(t) = \kappa(t)\,\mathbf{n}(t)\,.$$

Note that, since $\mathbf{n}$ is unitary, we have

$$\kappa(t) = \|\ddot{\boldsymbol{\gamma}}(t)\|$$

Now, approximate $\boldsymbol{\gamma}$ at t with its second order Taylor polynomial:

$$\boldsymbol{\gamma}(t+t_0) = \boldsymbol{\gamma}(t) + \dot{\boldsymbol{\gamma}}(t)t_0 + \frac{\ddot{\boldsymbol{\gamma}}(t)}{2}t_0^2 + o(t_0)$$

with the remainder $o(t_0)$ is such that

$$\lim_{t_0 \to 0} \frac{o(t_0)}{t_0^2} = 0\,.$$

Therefore, forgetting about the remainder,

$$\boldsymbol{\gamma}(t+t_0) - \boldsymbol{\gamma}(t) \approx \dot{\boldsymbol{\gamma}}(t)t_0 + \frac{\ddot{\boldsymbol{\gamma}}(t)}{2}t_0^2\,.$$

Multiplying by $\mathbf{n}(t)$ we get

$$(\boldsymbol{\gamma}(t+t_0) - \boldsymbol{\gamma}(t)) \cdot \mathbf{n}(t) \approx \dot{\boldsymbol{\gamma}}(t) \cdot \mathbf{n}(t) t_0 + \frac{\ddot{\boldsymbol{\gamma}}(t) \cdot \mathbf{n}(t)}{2} t_0^2 \,.$$

Recalling that

$$\dot{\boldsymbol{\gamma}}(t) \cdot \mathbf{n}(t) = 0 \,, \quad \ddot{\boldsymbol{\gamma}}(t) \cdot \mathbf{n}(t) = \kappa(t) \,,$$

we then obtain

$$(\boldsymbol{\gamma}(t+t_0) - \boldsymbol{\gamma}(t)) \cdot \mathbf{n}(t) \approx \frac{1}{2} \kappa(t) \, t_0^2$$

Important

The amount that $\boldsymbol{\gamma}$ deviates from a straight line is proportional to

$$\kappa(t) = \|\ddot{\boldsymbol{\gamma}}(t)\| \,.$$

We take this as definition of curvature for a general unit speed curve in $\mathbb{R}^n$.

Definition 2.1

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a unit speed curve. The **curvature** of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is

$$\kappa^{\boldsymbol{\gamma}}(t) := \|\ddot{\boldsymbol{\gamma}}(t)\| \,.$$

Note that $\kappa(t)$ is a function of time. Therefore the curvature of $\boldsymbol{\gamma}$ can change from point to point.

We now define curvature for curves which are regular, but not necessarily unit speed.

Definition 2.2

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a regular. The **curvature** of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is

$$\kappa^{\boldsymbol{\gamma}}(t) := \left\|\ddot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\right\| \,, \quad \forall \, t \in (a,b) \,,$$

where $\tilde{\boldsymbol{\gamma}}$ is a unit speed reparametrization of $\boldsymbol{\gamma}$, with $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi$.

Remark 2.3

The above definition is well posed:

- Since $\boldsymbol{\gamma}$ is regular, there exist a unit speed reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$.
- If $\hat{\boldsymbol{\gamma}}$ is another unit speed reaprametrization of $\boldsymbol{\gamma}$, with $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}} \circ \hat{\phi}$, then

$$\kappa^{\boldsymbol{\gamma}}(t) = \left\|\ddot{\hat{\boldsymbol{\gamma}}}(\hat{\phi}(t))\right\| \,,$$

showing that there is no ambiguity in the definition of $\kappa^{\boldsymbol{\gamma}}$.

Indeed, since $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ are both reparametrizations of $\boldsymbol{\gamma}$, then

$$\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\gamma}}(\tilde{\phi}(t)), \quad \boldsymbol{\gamma}(t) = \hat{\boldsymbol{\gamma}}(\hat{\phi}(t))$$

for some diffeomorphisms $\tilde{\phi}, \hat{\phi}$. Hence

$$\tilde{\boldsymbol{\gamma}}(t) = \hat{\boldsymbol{\gamma}}(\phi(t)), \quad \phi := \hat{\phi} \circ (\tilde{\phi})^{-1}, \tag{2.2}$$

where ϕ is a diffeomorphism, since it is composition of diffeomorphisms. Differentiating (2.2) we get

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) = \dot{\hat{\boldsymbol{\gamma}}}(\phi(t))\dot{\phi}(t). \tag{2.3}$$

Taking the norms of the above, and recalling that $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ are unit speed, we get

$$|\dot{\phi}(t)| = 1, \quad \forall\, t. \tag{2.4}$$

Since ϕ is a diffeomorphism, we already know that $|\dot{\phi}| \neq 0$. As $\dot{\phi}$ is continuous, this means that the sign of $\dot{\phi}$ is constant. Thus (2.4) implies

$$\dot{\phi}(t) \equiv 1 \quad \text{or} \quad \dot{\phi}(t) \equiv -1.$$

In both cases, we have

$$\ddot{\phi} \equiv 0.$$

Differentiating (2.3) we then obtain

$$\begin{aligned}
\ddot{\tilde{\boldsymbol{\gamma}}}(t) &= \ddot{\hat{\boldsymbol{\gamma}}}(\phi(t))\dot{\phi}^2(t) + \dot{\hat{\boldsymbol{\gamma}}}(\phi(t))\ddot{\phi}(t) \\
&= \ddot{\hat{\boldsymbol{\gamma}}}(\phi(t))\dot{\phi}^2(t).
\end{aligned}$$

Taking the norms and using again that $|\dot{\phi}| \equiv 1$, we get that

$$\left\|\ddot{\tilde{\boldsymbol{\gamma}}}(t)\right\| = \left\|\ddot{\hat{\boldsymbol{\gamma}}}(\phi(t))\right\|.$$

Recalling that $\phi = \hat{\phi} \circ (\tilde{\phi})^{-1}$ we get

$$\left\|\ddot{\tilde{\boldsymbol{\gamma}}}(\tilde{\phi}(t))\right\| = \left\|\ddot{\hat{\boldsymbol{\gamma}}}(\hat{\phi}(t))\right\|, \quad \forall\, t \in (a,b).$$

Therefore

$$\kappa^{\boldsymbol{\gamma}}(t) = \left\|\ddot{\tilde{\boldsymbol{\gamma}}}(\tilde{\phi}(t))\right\| = \left\|\ddot{\hat{\boldsymbol{\gamma}}}(\hat{\phi}(t))\right\|.$$

Remark 2.4: Methods for computing curvature

In summary, the curvature of a regular curve

$$\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$$

is defined via unit speed reparametrizations of $\boldsymbol{\gamma}$. To compute κ we do the following:

- We find a unit speed reparametrization $\tilde{\boldsymbol{\gamma}}$ of the regular curve $\boldsymbol{\gamma}$
- This can be done by computing s the arc-length of $\boldsymbol{\gamma}$, and then defining

$$\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ \psi \,, \quad \psi := s^{-1}$$

- Then we compute

$$\kappa^{\tilde{\boldsymbol{\gamma}}}(t) = \|\ddot{\tilde{\boldsymbol{\gamma}}}\|(t)$$

- We obtain the curvature of $\boldsymbol{\gamma}$ by

$$\kappa^{\boldsymbol{\gamma}}(t) = \kappa^{\tilde{\boldsymbol{\gamma}}}(t)$$

When $\boldsymbol{\gamma}$ is regular and has values in $\mathbb{R}^3$, there is a way to compute κ without reparametrizing. To do this, we will need the notion of **cross product**, or **vector product**. We will see this in the following sections.

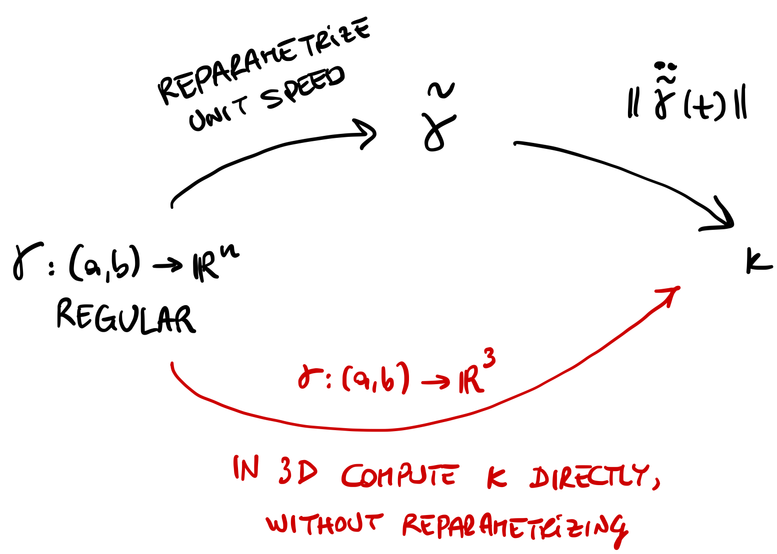


Figure 2.2: Procedure for computing curvature κ

We conclude with two examples in which we compute the curvature κ using unit speed reparametrizations.

Example 2.5

Consider the circle of radius $R > 0$:

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t)) \,, \quad t \in [0, 2\pi] \,.$$

To compute the curvature of $\boldsymbol{\gamma}$ we need to find a unit speed reparametrization. We have shown that:

$$\boldsymbol{\gamma} \text{ regular} \quad \Longrightarrow \quad \phi = s^{-1} \text{ unit speed reparametrization}$$

where s is the arc length of $\boldsymbol{\gamma}$:

$$s(t) := \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau \,.$$

In our case

$$\dot{\boldsymbol{\gamma}}(t) = (-R\sin(t), R\cos(t)) \quad \Longrightarrow \quad \|\dot{\boldsymbol{\gamma}}(t)\| = R$$

and so $\boldsymbol{\gamma}$ is regular. However $\boldsymbol{\gamma}$ is not unit speed, therefore we need to find a unit speed reparametrization. The arc length starting at $t_0 = 0$ is

$$s(t) = \int_0^t R d\tau = tR \,.$$

The inverse of s is

$$\phi(t) := s^{-1}(t) = \frac{t}{R} \,.$$

Therefore a unit speed reparametrization of $\boldsymbol{\gamma}$ is

$$\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ \phi$$

which reads

$$\tilde{\boldsymbol{\gamma}}(t) := \left(R\cos\left(\frac{t}{R}\right), R\sin\left(\frac{t}{R}\right) \right) \,.$$

We have

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) = \left(-\sin\left(\frac{t}{R}\right), \cos\left(\frac{t}{R}\right) \right)$$
$$\ddot{\tilde{\boldsymbol{\gamma}}}(t) = \left(-\frac{1}{R}\cos\left(\frac{t}{R}\right), -\frac{1}{R}\sin\left(\frac{t}{R}\right) \right)$$

Therefore the curvature of $\boldsymbol{\gamma}$ is

$$\kappa(t) = \|\ddot{\tilde{\boldsymbol{\gamma}}}(t)\| = \frac{1}{R} \,.$$

In this case $\kappa(t)$ is constant. The curvature also tells us that the smaller the circle, the higher the curvature. For a large circle, like the Earth, the curvature is barely noticeable.

Before proceeding with the next example, let us give a short overview of the **Hyperbolic functions**.

Remark 2.6: Hyperbolic functions

The Hyperbloic functions are the analogous of the trigonometric functions, but defined using the hyperbola rather than the circle. Their formulas can be obtained by means of the exponential function e^t. We have:

- Hyperbolic cosine: The **even part** of the function e^t, that is,

$$\cosh(t) = \frac{e^t + e^{-t}}{2} = \frac{e^{2t} + 1}{2e^t} = \frac{1 + e^{-2t}}{2e^{-t}} .$$

- Hyperbolic sine: The **odd part** of the function e^t, that is,

$$\sinh(t) = \frac{e^t - e^{-t}}{2} = \frac{e^{2t} - 1}{2e^t} = \frac{1 - e^{-2t}}{2e^{-t}} .$$

- Hyperbolic tangent: Defined by

$$\tanh(t) = \frac{\sinh t}{\cosh t} = \frac{e^t - e^{-t}}{e^t + e^{-t}} = \frac{e^{2t} - 1}{e^{2t} + 1} .$$

- Hyperbolic cotangent: The reciprocal of tanh for $t \neq 0$,

$$\coth t = \frac{\cosh t}{\sinh t} = \frac{e^t + e^{-t}}{e^t - e^{-t}} = \frac{e^{2t} + 1}{e^{2t} - 1} .$$

- Hyperbolic secant: The reciprocal of cosh

$$\operatorname{sech}(t) = \frac{1}{\cosh t} = \frac{2}{e^t + e^{-t}} = \frac{2e^t}{e^{2t} + 1} .$$

- Hyperbolic cosecant: The reciprocal of sinh for $t \neq 0$,

$$\operatorname{csch}(t) = \frac{1}{\sinh t} = \frac{2}{e^t - e^{-t}} = \frac{2e^t}{e^{2t} - 1} .$$

For a plot cosh, sinh, tanh see Figure 2.3 below. The properties of the hyperbolic functions which are of interest to us are:

1. Identities:

$$\begin{aligned}
\cosh(t) + \sinh(t) &= e^t \\
\cosh(t) - \sinh(t) &= e^{-t} \\
\cosh^2(t) - \sinh^2(t) &= 1 \\
\operatorname{sech}^2(t) - \tanh^2(t) &= 1
\end{aligned}$$

2. Derivatives:

$$\frac{d}{dt}[\sinh(t)] = \cosh(t)$$
$$\frac{d}{dt}[\cosh(t)] = \sinh(t)$$
$$\frac{d}{dt}[\tanh(t)] = 1 - \tanh^2(t) = -\operatorname{csch}^2(t)$$

3. Integrals:

$$\int_{t_0}^{t} \sinh(u)\, du = \cosh(t) - \cosh(t_0)$$
$$\int_{t_0}^{t} \cosh(u)\, du = \sinh(t) - \sinh(t_0)$$
$$\int_{t_0}^{t} \tanh(u)\, du = \log(\cosh(t)) - \log(\cosh(t_0))$$

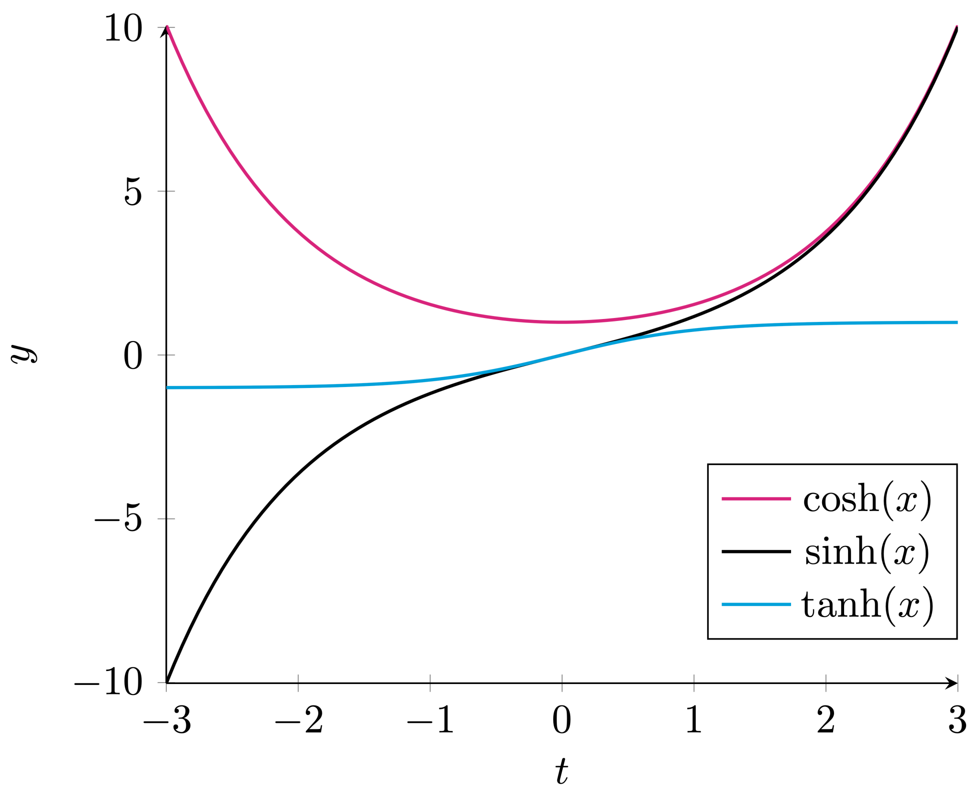


Figure 2.3: Plot of cosh, sinh, tanh.

Example 2.7: The Catenary

The **catenary** is the shape of a heavy chain suspended at its ends. The chain is only subjected to gravity, see Figure 2.4. This shape looks similar to a parabola, but it is not a parabola. This was first noted by Galilei, see this Wikipedia page. The profile of the hanging chain can be obtained via a minimization problem, and one can show it is of the form

$$\boldsymbol{\gamma}(t) = (t, \cosh(t)) , \quad t \in \mathbb{R} .$$

See Figure 2.5 for a plot of $\boldsymbol{\gamma}$. Let us check if $\boldsymbol{\gamma}$ is regular. We have

$$\dot{\boldsymbol{\gamma}}(t) = (1, \sinh(t))$$

so that

$$\|\dot{\boldsymbol{\gamma}}\|^2 = 1 + \sinh^2(t) = \cosh^2(t) \quad \Longrightarrow \quad \|\dot{\boldsymbol{\gamma}}\| = \cosh(t) .$$

Note that

$$\cosh(t) \geq 1$$

showing that $\boldsymbol{\gamma}$ is regular. However

$$\|\dot{\boldsymbol{\gamma}}(1)\| = \cosh(1) = \frac{e + e^{-1}}{2} \approx 1.54 ,$$

proving that $\boldsymbol{\gamma}$ is not unit speed. Let us then compute the arc length of $\boldsymbol{\gamma}$ starting at $t_0 = 0$

$$s(t) = \int_0^t \|\dot{\boldsymbol{\gamma}}(u)\| \, du = \int_0^t \cosh(u) \, du = \sinh(t)$$

since $\sinh(0) = 0$. We need to invert s. We have

$$s = \sinh(t) \quad \Longleftrightarrow \quad s = \frac{e^t - e^{-t}}{2} \quad \Longleftrightarrow \quad e^{2t} - 2se^t - 1 = 0 ,$$

where the last equation was obtained multuplying both sides by e^t. Now we substitute

$$y = e^t$$

and obtain

$$e^{2t} - 2se^t - 1 = 0 \quad \Longleftrightarrow \quad y^2 - 2sy - 1 = 0 \quad \Longleftrightarrow \quad y = s \pm \sqrt{1 + s^2} .$$

Recalling that $y = e^t$, we only consider the positive solution, and obtain that

$$e^t = s + \sqrt{1 + s^2} \quad \Longrightarrow \quad t = \log\left(s + \sqrt{1 + s^2}\right) .$$

We have proven that the inverse of the arc length $s(t)$ is

$$\psi(t) := s^{-1}(t) = \log\left(t + \sqrt{1 + t^2}\right) .$$

Therefore

$$\tilde{\boldsymbol{\gamma}}(t) := \boldsymbol{\gamma}(\psi(t))$$

is a unit speed reparametrization of $\boldsymbol{\gamma}$. Substituting ψ and using the definition of $\boldsymbol{\gamma}$ we have

$$\tilde{\boldsymbol{\gamma}}(t) = \left(\log\left(t + \sqrt{1+t^2}\right), \sqrt{1+t^2}\right) .$$

We can now compute the curvature. We have:

$$\begin{aligned}
\dot{\tilde{\boldsymbol{\gamma}}}(t) &= \left(\frac{1}{\sqrt{1+t^2}}, \frac{t}{\sqrt{1+t^2}}\right) \\
\ddot{\tilde{\boldsymbol{\gamma}}}(t) &= \left(-\frac{t}{(1+t^2)^{3/2}}, \frac{1}{(1+t^2)^{3/2}}\right)
\end{aligned}$$

Moreover

$$\left\|\ddot{\tilde{\boldsymbol{\gamma}}}(t)\right\|^2 = \frac{t^2}{(1+t^2)^3} + \frac{1}{(1+t^2)^3} = \frac{1}{(1+t^2)^2} .$$

Therefore the curvature is

$$\kappa(t) = \|\ddot{\tilde{\boldsymbol{\gamma}}}(t)\| = \frac{1}{1+t^2} .$$

Figure 2.4: The catenary is the shape of a heavy chain suspended at its ends. Image from Wikipedia.

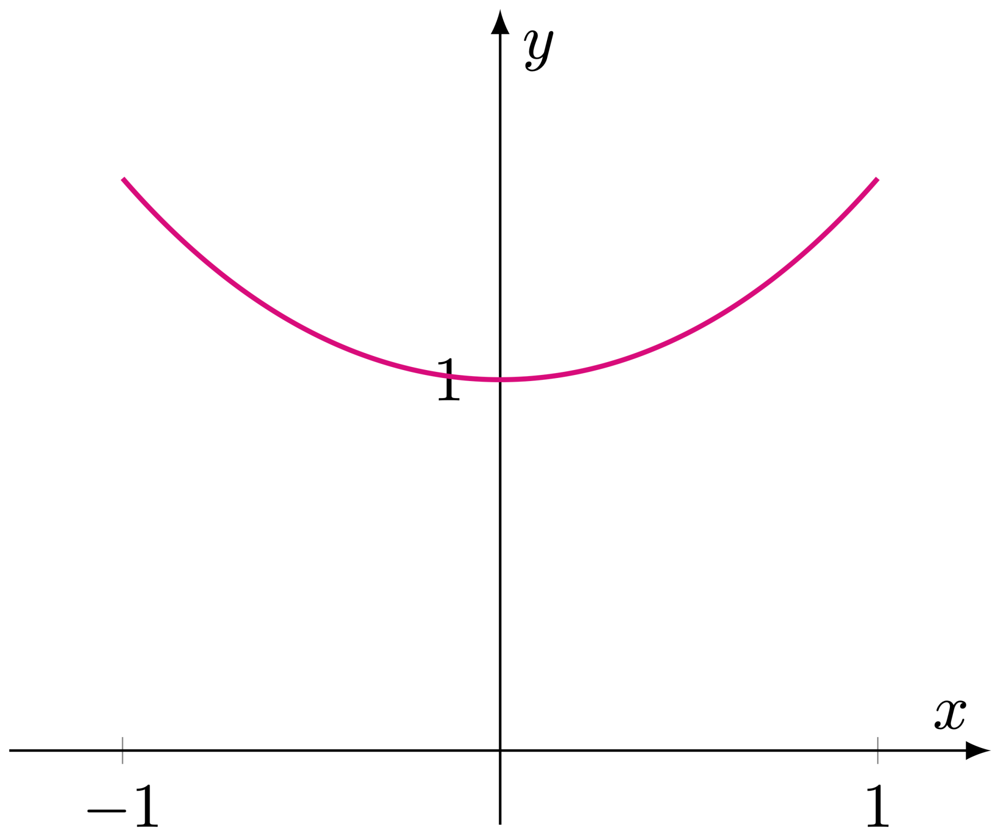


Figure 2.5: Plot of the catenary curve $\boldsymbol{\gamma}(t) = (t, \cosh(t))$.

2.2 Vector product in $\mathbb{R}^3$

The discussion in this section follows [2]. We start by defining **orientation** for a vector space.

Definition 2.8: Same orientation

Consider two ordered basis of $\mathbb{R}^3$

$$B = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3), \quad \tilde{B} = (\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \tilde{\mathbf{b}}_3).$$

We say that B and $\widetilde{B}$ have the same orientation if the matrix of change of basis has positive determinant.

When two basis B and $\widetilde{B}$ have the same orientation, we write

$$\mathbf{b} \sim \tilde{\mathbf{b}}.$$

The above is clearly an equivalence relation on the set of ordered basis. Therefore the set of ordered basis of $\mathbb{R}^3$ can be decomposed into equivalence classes. Since the determinant of the matrix of change of basis can only be positive or negative, there are only two equivalence classes.

Definition 2.9: Orientation

The two equivalence classes determined by $\sim$ on the set of ordered basis are called **orientations**.

Definition 2.10: Positive orientation

Consider the standard basis of $\mathbb{R}^3$

$$E = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$$

where we set

$$\mathbf{e}_1 = (1, 0, 0), \quad \mathbf{e}_2 = (0, 1, 0), \quad \mathbf{e}_3 = (0, 0, 1).$$

Then:

- The orientation corresponding to E is called **positive orientation** of $\mathbb{R}^3$.
- The orientation corresponding to the other equivalence class is called **negative orientation** of $\mathbb{R}^3$.

For a basis B of $\mathbb{R}^3$ we say that:

- B is a **positive basis** if it belongs to the class of e.
- B is a **negative basis** if it does not belong to the class of e.

Example 2.11

Since we are dealing with ordered basis, the order in which vectors appear is fundamental. For example, we defined the equivalence class of

$$E = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)\,,$$

to be the positive orientation of $\mathbb{R}^3$. In particular e is a positive basis.
Consider instead

$$\widetilde{E} = (\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3)\,.$$

The matrix of change of variables between $\widetilde{E}$ and E is

$$(\mathbf{e}_2|\mathbf{e}_1|\mathbf{e}_3) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the latter has negative determinant. Thus $\widetilde{E}$ does not belong to the class of E, and is therefore a negative basis.

We are now ready to define the vector product in $\mathbb{R}^3$.

Definition 2.12: Vector product in $\mathbb{R}^3$

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$. The vector product of $\mathbf{u}$ and $\mathbf{v}$ is the unique vector

$$\mathbf{u} \times \mathbf{v} \in \mathbb{R}^3$$

which satisfies the property:

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}\,, \quad \forall\, \mathbf{w} \in \mathbb{R}^3\,. \tag{2.5}$$

Here $|a_{ij}|$ denotes the determinant of the matrix (a_{ij}), and

$$\mathbf{u} = \sum_{i=1}^{3} u_i \mathbf{e}_i\,, \quad \mathbf{v} = \sum_{i=1}^{3} v_i \mathbf{e}_i\,, \quad \mathbf{w} = \sum_{i=1}^{3} w_i \mathbf{e}_i\,,$$

with $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ standard basis of $\mathbb{R}^3$.

The following proposition gives an explicit formula for computing $\mathbf{u} \times \mathbf{v}$.

Proposition 2.13

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$. Then

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{e}_3 \,. \tag{2.6}$$

Proof

Denote by $(\mathbf{u} \times \mathbf{v})_i$ the i-th component of $\mathbf{u} \times \mathbf{v}$ with respect to the standard basis, that is,

$$\mathbf{u} \times \mathbf{v} = \sum_{i=1}^{3} (\mathbf{u} \times \mathbf{v})_i \, \mathbf{e}_i \,.$$

We can use (2.5) with $\mathbf{w} = \mathbf{e}_1$ to obtain

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{e}_1 = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ 1 & 0 & 0 \end{vmatrix} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}$$

where we used the Laplace expansion for computing the determinant of the 3×3 matrix. As the standard basis is orthonormal, by bilinearity of the scalar product we get

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{e}_1 = \sum_{i=1}^{3} (\mathbf{u} \times \mathbf{v})_i \, \mathbf{e}_i \cdot \mathbf{e}_1 = (\mathbf{u} \times \mathbf{v})_i \,.$$

Therefore we have shown

$$(\mathbf{u} \times \mathbf{v})_1 = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \,.$$

Similarly we obtain

$$(\mathbf{u} \times \mathbf{v})_2 = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ 0 & 1 & 0 \end{vmatrix} = - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}$$

and

$$(\mathbf{u} \times \mathbf{v})_3 = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \,,$$

from which we conclude.

Sometimes we will denote formula (2.6) by

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_2 \\ v_1 & v_2 & v_3 \end{vmatrix} \,.$$

Let us collect some crucial properties of the vector product.

Proposition 2.14

The vector product in $\mathbb{R}^3$ satisfies the following properties: For all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$

1. $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$
2. $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent
3. $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} = 0$, $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = 0$
4. For all $\mathbf{w} \in \mathbb{R}^3$, $a, b \in \mathbb{R}$
$$(a\mathbf{u} + b\mathbf{w}) \times \mathbf{v} = a\mathbf{u} \times \mathbf{v} + b\mathbf{w} \times \mathbf{w}$$

The proof, which is based on the properties of determinants, is omitted.

Remark 2.15: Geometric interpretation of vector product

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ be linearly independent. We make some observations:

1. Property 3 in Proposition 2.14 says that
$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} = 0\,, \quad (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = 0\,.$$
Therefore $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.
2. In particular $\mathbf{u} \times \mathbf{v}$ is orthogonal to the plane generated by $\mathbf{u}$ and $\mathbf{v}$.
3. Since $\mathbf{u}$ and $\mathbf{v}$ are linearly independent, Property 2 in Proposition 2.14 says that
$$\mathbf{u} \times \mathbf{v} \neq \mathbf{0}$$
4. Therefore we have
$$(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{u} \times \mathbf{v}) = \|\mathbf{u} \times \mathbf{v}\|^2 > 0$$
5. On the other hand, using the definition of $\mathbf{u} \times \mathbf{v}$ with $\mathbf{w} = \mathbf{v} \times \mathbf{w}$ yields
$$(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{u} \times \mathbf{v}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ (\mathbf{u} \times \mathbf{v})_1 & (\mathbf{u} \times \mathbf{v})_2 & (\mathbf{u} \times \mathbf{v})_3 \end{vmatrix}$$
6. Therefore the determinant of the matrix
$$(\mathbf{u}|\mathbf{v}|\mathbf{u} \times \mathbf{v})$$
is positive. This shows that
$$(\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v})$$
is a **positive basis** of $\mathbb{R}^3$.

7. For all $\mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ it holds

$$(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{x} \times \mathbf{y}) = \begin{vmatrix} \mathbf{u} \cdot \mathbf{x} & \mathbf{v} \cdot \mathbf{x} \\ \mathbf{u} \cdot \mathbf{y} & \mathbf{v} \cdot \mathbf{y} \end{vmatrix} . \tag{2.7}$$

   Indeed, one can check that the above formula holds for the standard vectors $\mathbf{e}_i$, and thus the general formula follows by linearity.

8. Using (2.7) we get

$$\begin{aligned} \|\mathbf{u} \times \mathbf{v}\|^2 = (\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{u} \times \mathbf{v}) &= \begin{vmatrix} \mathbf{u} \cdot \mathbf{u} & \mathbf{v} \cdot \mathbf{u} \\ \mathbf{u} \cdot \mathbf{v} & \mathbf{v} \cdot \mathbf{v} \end{vmatrix} \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - |\mathbf{u} \cdot \mathbf{v}|^2 \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \cos^2(\theta) \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 (1 - \cos^2(\theta)) \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2(\theta) \\ &= A^2 \end{aligned}$$

   where A is the area of the parallelogram with sides $\mathbf{u}$ and $\mathbf{v}$.

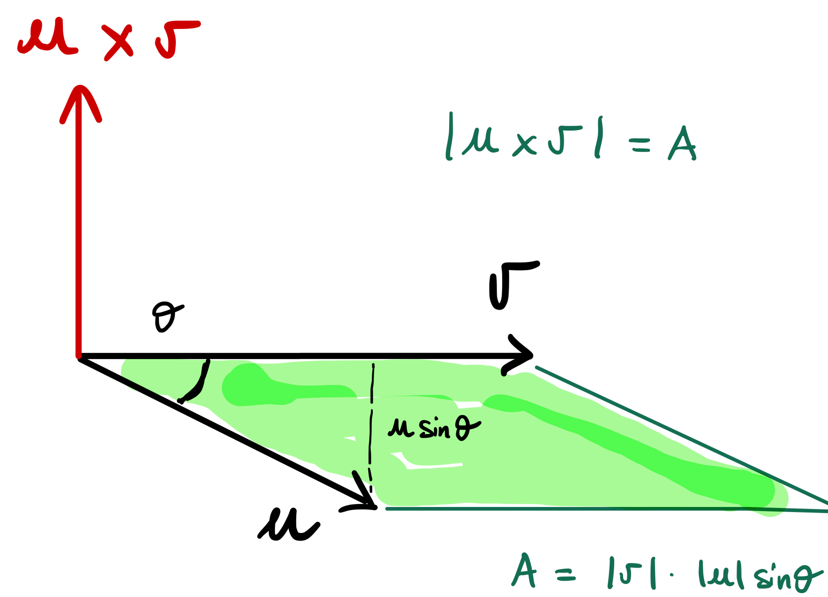


Figure 2.6: For $\mathbf{u}, \mathbf{v}$ linearly independent, $\mathbf{u} \times \mathbf{v}$ is orthogonal to the plane generated by $\mathbf{u}, \mathbf{v}$. Moreover $|\mathbf{u} \times \mathbf{v}|$ is the area of the parallelogram with sides $\mathbf{u}, \mathbf{v}$, and $(\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v})$ is a positive basis of $\mathbb{R}^3$

Let us summarize the above remark.

Remark 2.16: Summary: Properties of $\mathbf{u} \times \mathbf{v}$

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ be linearly independent. Then

- $\mathbf{u} \times \mathbf{v}$ is orthogonal to the plane spanned by $\mathbf{u}, \mathbf{v}$
- $\|\mathbf{u} \times \mathbf{v}\|$ is equal to the area of the parallelogram with sides $\mathbf{u}, \mathbf{v}$
- $\mathbf{u} \times \mathbf{v}$ is such that
$$(\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v})$$
is a positive basis of $\mathbb{R}^3$.

We conclude with noting that the cross product is not associative, and with a useful proposition for differentiating the cross product of curves in $\mathbb{R}^3$.

Proposition 2.17

The vector product is not associative. In particular, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ it holds:

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u}\,. \tag{2.8}$$

The proof is omitted. It follows by observing that both sides of (2.8) are linear in $\mathbf{u}, \mathbf{v}, \mathbf{w}$. Therefore it is sufficient to verify (2.8) for the standard basis vectors $\mathbf{e}_i$. This is left as an exercise.

Proposition 2.18

Suppose $\boldsymbol{\gamma}, \boldsymbol{\eta} : (a,b) \to \mathbb{R}^3$ are parametrized curves. Then the curve

$$\boldsymbol{\gamma} \times \boldsymbol{\eta} : (a,b) \to \mathbb{R}^3$$

is smooth, and

$$\frac{d}{dt}(\boldsymbol{\gamma} \times \boldsymbol{\eta}) = \dot{\boldsymbol{\gamma}} \times \boldsymbol{\eta} + \boldsymbol{\gamma} \times \dot{\boldsymbol{\eta}}\,. \tag{2.9}$$

The proof is omitted. It follows immediately from formula (2.6).

2.3 Curvature formula in $\mathbb{R}^3$

Given a unit speed curve

$$\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$$

we defined its curvature as

$$\kappa(t) = \|\ddot{\boldsymbol{\gamma}}(t)\| \,.$$

If $\boldsymbol{\gamma}$ is not unit speed then the curvature is not defined. However, when $\boldsymbol{\gamma}$ is regular, then we can find a unit-speed reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$, and compute κ as

$$\kappa(t) = \|\ddot{\tilde{\boldsymbol{\gamma}}}(t)\| \,.$$

If $\boldsymbol{\gamma}$ is a regular curve in $\mathbb{R}^3$, there is a way to compute κ without passing through $\tilde{\boldsymbol{\gamma}}$. The formula for computing κ is as follows.

Proposition 2.19: Curvature formula

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a regular curve. The curvature $\kappa(t)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is given by

$$\kappa(t) = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} \,. \tag{2.10}$$

We delay the proof of the above Proposition, as this will get easier when the **Frenet frame** is introduced. For a proof which does not make use of the Frenet frame, see the proof of Proposition 2.1.2 in [6].

For now we use (2.10) the above proposition to compute the curvature on specific curves.

Example 2.20

Consider the straight line

$$\boldsymbol{\gamma}(t) = \mathbf{a} + t\mathbf{v}$$

for some $\mathbf{a}, \mathbf{v} \in \mathbb{R}^3$ fixed, with $\mathbf{v} \neq \mathbf{0}$. Then

$$\dot{\boldsymbol{\gamma}}(t) = \mathbf{v} \,, \quad \ddot{\boldsymbol{\gamma}}(t) = \mathbf{0} \,.$$

Therefore

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \|\mathbf{v}\| \neq 0$$

showing that $\boldsymbol{\gamma}$ is regular. We have

$$\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} = \mathbf{v} \times \mathbf{0} = \mathbf{0} \,.$$

Therefore the curvature is

$$\kappa = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} = 0 \,,$$

as expected.

Example 2.21

Consider the Helix of radius $R > 0$ and rise $H > 0$

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t), Ht)\,, \quad t \in \mathbb{R}\,.$$

Then

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) &= (-R\sin(t), R\cos(t), H) \\ \ddot{\boldsymbol{\gamma}}(t) &= (-R\cos(t), -R\sin(t), 0) \end{aligned}$$

From this we deduce that

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \sqrt{R^2 + H^2}\,,$$

showing that $\boldsymbol{\gamma}$ is regular. Finally

$$\begin{aligned} \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} &= \begin{vmatrix} \dot{\gamma}_2 & \dot{\gamma}_3 \\ \ddot{\gamma}_2 & \ddot{\gamma}_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} \dot{\gamma}_1 & \dot{\gamma}_3 \\ \ddot{\gamma}_1 & \ddot{\gamma}_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} \dot{\gamma}_1 & \dot{\gamma}_2 \\ \ddot{\gamma}_1 & \ddot{\gamma}_2 \end{vmatrix} \mathbf{e}_3 \\ &= \begin{vmatrix} R\cos(t) & H \\ -R\sin(t) & 0 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} -R\sin(t) & H \\ -R\cos(t) & 0 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} -R\sin(t) & R\cos(t) \\ -R\cos(t) & -R\sin(t) \end{vmatrix} \mathbf{e}_3 \\ &= \left(RH\sin(t), -RH\cos(t), R^2\cos^2(t) + R^2\sin^2(t)\right) \\ &= \left(RH\sin(t), -RH\cos(t), R^2\right) \end{aligned}$$

and therefore

$$\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| = R\sqrt{R^2 + H^2}\,.$$

By the general formula we have

$$\kappa = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} = \frac{R(R^2 + H^2)^{\frac{1}{2}}}{(R^2 + H^2)^{\frac{3}{2}}} = \frac{R}{R^2 + H^2}$$

We notice the following:

- If $H = 0$ then the Helix is just a circle of radius R. In this case the curvature is

  $$\kappa = \frac{1}{R}$$

  which agrees with the curvature computed for the circle of radius R.

- If $R = 0$ then the Helix is just parametrizing the z-axis. In this case the curvature is

  $$\kappa = 0\,,$$

  which agrees with the curvature of a straight line.

2.4 Signed curvature of plane curves

In this section we assume to have plane curves, that is, curves with values in $\mathbb{R}^2$. In this case we can give a geometric interpretation for the sign of the curvature. This cannot be done in higher dimension.

Definition 2.22

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^2$ be unit speed. We define the **signed unit normal** to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ as the unit vector $\mathbf{n}(t)$ obtained by rotating $\dot{\boldsymbol{\gamma}}(t)$ anti-clockwise by an angle of $\pi/2$.

Definition 2.23

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^2$ be unit speed. The **signed curvature** of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is the scalar $\kappa_s(t)$ such that

$$\ddot{\boldsymbol{\gamma}}(t) = k_s(t)\mathbf{n}(t)$$

Remark 2.24

Notice that since $\mathbf{n}$ is a unit vector and $\boldsymbol{\gamma}$ is unit speed, then

$$|\kappa_s(t)| = \|\ddot{\boldsymbol{\gamma}}(t)\| = \kappa(t)\,.$$

Thus the signed curvature is related to the curvature by

$$\kappa_s(t) = \pm\kappa(t)\,.$$

Remark 2.25

It can be shown that the signed curvature is the rate at which the tangent vector $\dot{\boldsymbol{\gamma}}$ of the curve $\boldsymbol{\gamma}$ rotates. The signed curvature is:

- positive if $\dot{\boldsymbol{\gamma}}$ is rotating anti-clockwise
- negative if $\dot{\boldsymbol{\gamma}}$ is rotating clockwise

In other words,

- $k_s > 0$ means the curve is turning left,
- $k_s < 0$ means the curve is turning right.

A rigorous justification of the above statement is found in Proposition 2.2.3 in [6].

For curves which are not unit speed, we define the signed curvature as the signed curvature of the unit speed reparametrization.

Definition 2.26

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^2$ be regular and let $\tilde{\boldsymbol{\gamma}}$ be a unit speed reparametrization of $\boldsymbol{\gamma}$. The **signed curvature** of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is the scalar $\kappa_s(t)$ such that

$$\ddot{\tilde{\boldsymbol{\gamma}}}(t) = k_s(t)\mathbf{n}(t)\,,$$

where $\mathbf{n}(t)$ is the unit vector obtained by rotating $\dot{\tilde{\boldsymbol{\gamma}}}(t)$ anti-clockwise by an angle $\pi/2$.

The signed curvature completely characterizes plane curves, in the sense of the following theorem.

Theorem 2.27: Characterization of plane curves

Let $\phi : \mathbb{R} \to \mathbb{R}$ be smooth. Then:

1. There exists a unit speed curve $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^2$ such that its signed curvature κ_s satisfies
$$\kappa_s(t) = \phi(t)\,, \quad \forall\, t \in \mathbb{R}\,.$$
2. Suppose that $\tilde{\boldsymbol{\gamma}} : \mathbb{R} \to \mathbb{R}^2$ is a unit speed curve such that its signed curvature $\tilde{\kappa}_s$ satisfies
$$\tilde{\kappa}_s(t) = \phi(t)\,, \quad \forall\, t \in \mathbb{R}\,.$$
Then
$$\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$$
up to rotations and translations.

We do not prove the above theorem. For a proof, see Theorem 2.2.6 in [6].

2.5 Space curves

In this section we deal with **space curves**, that is, curves with values in $\mathbb{R}^3$. There are several issues compare to the plane case:

- A 3D counterpart of the signed curvature does not exist, since there is no notion of *turning left* or *turning right*.
- We have seen in the previous section that the signed curvature completely characterizes plane curves. In 3D however curvature is not enough to characterize curves: there exist $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ space curves such that
$$\kappa^{\boldsymbol{\gamma}} = \kappa^{\boldsymbol{\eta}}\,, \quad \boldsymbol{\gamma} \neq \boldsymbol{\eta}\,,$$
that is, $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ have same curvature but are different curves.

Example 2.28

Let $\boldsymbol{\gamma}$ be a circle of radius $R > 0$

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t), 0)\,,$$

and $\boldsymbol{\eta}$ be a helix of radius $S > 0$ and rise $H > 0$

$$\boldsymbol{\eta}(t) = (S\cos(t), S\sin(t), Ht)\,.$$

We have computed that

$$\kappa^{\boldsymbol{\gamma}} = \frac{1}{R}\,, \quad \kappa^{\boldsymbol{\eta}} = \frac{S}{S^2 + H^2}\,.$$

If we now choose $R = 2$ and we impose that $\kappa^{\boldsymbol{\gamma}} = \kappa^{\boldsymbol{\eta}}$ we get

$$\frac{1}{R} = \frac{S}{S^2 + H^2} \qquad \Longrightarrow \qquad H^2 = 2S - S^2$$

Therefore choosing $S = 1$ and $H = 1$ yields

$$\kappa^{\boldsymbol{\gamma}} = \kappa^{\boldsymbol{\eta}}\,, \quad \boldsymbol{\gamma} \neq \boldsymbol{\eta}\,..$$

Therefore curvature is not enough for characterizing space curves, and we need a new quantity. As we did with curvature, we start by considering the simpler case of unit speed curves. We will also need to assume that the curvature is never zero.

Definition 2.29: Principal normal vector

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a unit speed curve with

$$\kappa(t) \neq 0\,, \quad \forall\, t \in (a,b)\,.$$

The **principal normal vector** to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is

$$\mathbf{n}(t) := \frac{1}{\kappa(t)}\ddot{\boldsymbol{\gamma}}(t)\,.$$

Remark 2.30

Since for $\boldsymbol{\gamma}$ unit speed we defined

$$\kappa(t) := \|\ddot{\boldsymbol{\gamma}}(t)\|\,,$$

we have that

$$\|\mathbf{n}(t)\| = 1\,,$$

thus $\mathbf{n}$ is a unit vector. Moreover $\mathbf{n}$ is orthogonal to $\dot{\boldsymbol{\gamma}}$, that is,

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{n} = 0 \, .$$

This is because

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{n} = \frac{1}{\kappa} \dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0 \, ,$$

where the last equality follows from $\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0$, being $\boldsymbol{\gamma}$ unit speed.

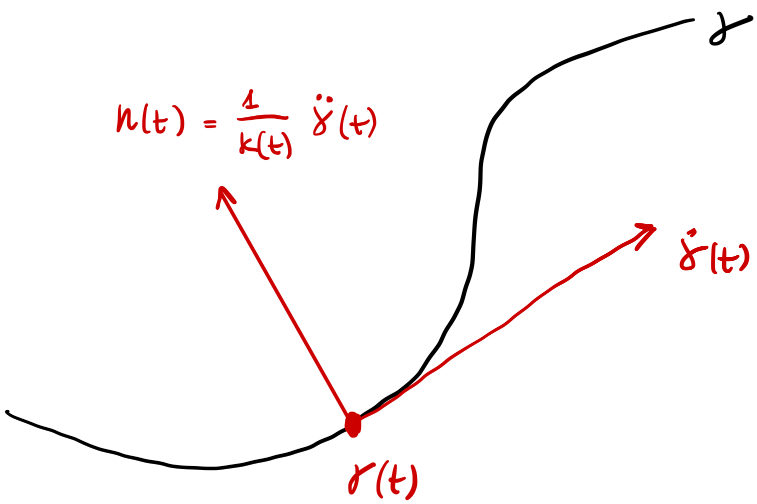


Figure 2.7: Principal normal vector $\mathbf{n}(t)$ to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$.

Question 2.31

Why is the principal normal interesting? Because it can tell the difference beween a plane curve and a space curve. See picture below.

Definition 2.32: Binormal vector

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be a unit speed curve with

$$\kappa(t) \neq 0 \, , \quad \forall \, t \in (a, b) \, .$$

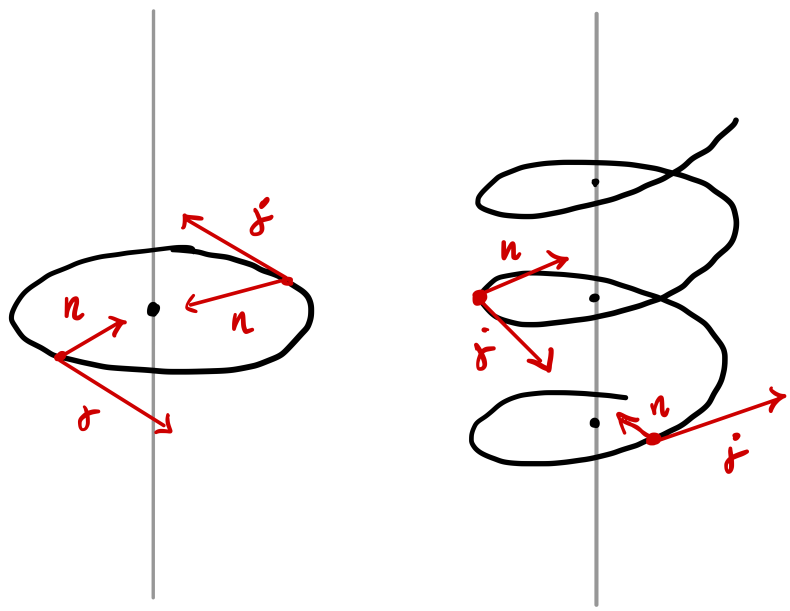


Figure 2.8: Left: Principal normal to a circle. Note that **n** always points towards the origin **0**. Right: Principal normal to a helix. Note that **n** points towards the z-axis, but never towards the same point.

The **binormal vector** to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is

$$\mathbf{b}(t) := \dot{\boldsymbol{\gamma}}(t) \times \mathbf{n}(t) .$$

Definition 2.33: Orthonormal basis

Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ be vectors in $\mathbb{R}^3$. We say that the triple

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

is **orthonormal** if

$$\|v_i\| = 1 , \quad v_i \cdot v_j = 0 , \quad \text{for } i \neq j .$$

Proposition 2.34

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be a unit speed curve with

$$\kappa(t) \neq 0 , \quad \forall\, t \in (a,b) .$$

Then the triple

$$B = (\dot{\boldsymbol{\gamma}}(t), \mathbf{n}(t), \mathbf{b}(t))$$

is a positive orthonormal basis of $\mathbb{R}^3$ for all $t \in (a, b)$.

Proof

Since $\boldsymbol{\gamma}$ is unit speed we have

$$\|\dot{\boldsymbol{\gamma}}(t)\| \equiv 1 \,.$$

Moreover we have already observed that

$$\|\mathbf{n}(t)\| \equiv 1 \,, \quad \dot{\boldsymbol{\gamma}}(t) \cdot \mathbf{n}(t) \equiv 0 \,.$$

As $\mathbf{b}$ is defined by

$$\mathbf{b} := \dot{\boldsymbol{\gamma}} \times \mathbf{n} \,,$$

by the properties of the vector product, see Proposition 2.14, it follows that

$$\mathbf{b} \cdot \dot{\boldsymbol{\gamma}} = 0 \,, \quad \mathbf{b} \cdot \mathbf{n} = 0 \,.$$

By the calculation in Remark 2.15 Point 8, we have that

$$\|\mathbf{b}\|^2 = \|\dot{\boldsymbol{\gamma}}\|^2 \|\mathbf{n}\|^2 - |\dot{\boldsymbol{\gamma}} \cdot \mathbf{n}|^2 = 1 \,.$$

This shows that the vectors

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{n}, \mathbf{b}\}$$

are orthonormal. By the properties of the vector product, see Remark 2.15 Point 6, we also know that

$$(\dot{\boldsymbol{\gamma}}, \mathbf{n}, \mathbf{b})$$

is a positive basis of $\mathbb{R}^3$.

Proposition 2.35

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be a unit speed curve with $\kappa \neq 0$. Then

$$\mathbf{b} = \dot{\boldsymbol{\gamma}} \times \mathbf{n} \,, \quad \mathbf{n} = \mathbf{b} \times \dot{\boldsymbol{\gamma}} \,, \quad \dot{\boldsymbol{\gamma}} = \mathbf{n} \times \mathbf{b} \,. \tag{2.11}$$

Proof

The first equality in (2.11) is true by definition of $\mathbf{b}$. For the other 2 equalities, recall formula (2.8):

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u}\,, \tag{2.12}$$

for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. Applying the above with

$$\mathbf{u} = \dot{\boldsymbol{\gamma}}\,, \quad \mathbf{v} = \mathbf{n}\,, \quad \mathbf{w} = \dot{\boldsymbol{\gamma}}\,,$$

yields

$$\begin{aligned} (\dot{\boldsymbol{\gamma}} \times \mathbf{n}) \times \dot{\boldsymbol{\gamma}} & = (\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}})\mathbf{n} - (\mathbf{n} \cdot \dot{\boldsymbol{\gamma}})\dot{\boldsymbol{\gamma}} \\ & = \|\dot{\boldsymbol{\gamma}}\|^2 \mathbf{n} - 0 \\ & = \mathbf{n}\,, \end{aligned}$$

where we used that $\dot{\boldsymbol{\gamma}}$ is a unit vector and $\mathbf{n} \cdot \dot{\boldsymbol{\gamma}} = 0$. Therefore, by definition of $\mathbf{b}$, we have

$$\mathbf{b} \times \dot{\boldsymbol{\gamma}} = (\dot{\boldsymbol{\gamma}} \times \mathbf{n}) \times \dot{\boldsymbol{\gamma}} = \mathbf{n}$$

showing the second equality in (2.11). For showing the third equality in (2.11), we apply (2.12) with

$$\mathbf{u} = \dot{\boldsymbol{\gamma}}\,, \quad \mathbf{v} = \mathbf{n}\,, \quad \mathbf{w} = \mathbf{n}\,,$$

to get

$$\begin{aligned} (\dot{\boldsymbol{\gamma}} \times \mathbf{n}) \times \mathbf{n} & = (\dot{\boldsymbol{\gamma}} \cdot \mathbf{n})\mathbf{n} - (\mathbf{n} \cdot \mathbf{n})\dot{\boldsymbol{\gamma}} \\ & = 0 - \|\mathbf{n}\|^2 \dot{\boldsymbol{\gamma}} \\ & = -\dot{\boldsymbol{\gamma}} \end{aligned}$$

where we used that $\mathbf{n}$ is a unit vector and $\dot{\boldsymbol{\gamma}} \cdot \mathbf{n} = 0$. Therefore, by definition of $\mathbf{b}$ and anti-commutativity of the vector product, we have

$$\mathbf{n} \times \mathbf{b} = -\mathbf{b} \times \mathbf{n} = -(\dot{\boldsymbol{\gamma}} \times \mathbf{n}) \times \mathbf{n} = \dot{\boldsymbol{\gamma}}\,,$$

showing the last equality in (2.11).

Proposition 2.36

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be a unit speed curve with $\kappa \neq 0$. Then

$$\dot{\mathbf{b}}(t) = -\tau(t)\mathbf{n}(t)\,, \tag{2.13}$$

for some $\tau(t) \in \mathbb{R}$.

Proof

By definition of $\mathbf{b}$ and the formula of derivation of the cross product (2.9) we have

$$\begin{aligned}\dot{\mathbf{b}} &= \frac{d}{dt}(\dot{\boldsymbol{\gamma}} \times \mathbf{n}) \\ &= \ddot{\boldsymbol{\gamma}} \times \mathbf{n} + \dot{\boldsymbol{\gamma}} \times \dot{\mathbf{n}} \\ &= \dot{\boldsymbol{\gamma}} \times \dot{\mathbf{n}}\,,\end{aligned}$$

where we used that

$$\ddot{\boldsymbol{\gamma}} \times \mathbf{n} = 0\,,$$

since $\mathbf{n}$ is defined by $\mathbf{n} := \ddot{\boldsymbol{\gamma}}/\kappa$, and therefore $\mathbf{n}$ and $\ddot{\boldsymbol{\gamma}}$ are parallel. Hence, we have proven that

$$\dot{\mathbf{b}} = \dot{\boldsymbol{\gamma}} \times \dot{\mathbf{n}}\,. \tag{2.14}$$

By the properties of the cross product we have that $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$. Thus (2.14) implies that

$$\dot{\mathbf{b}} \cdot \dot{\boldsymbol{\gamma}} = 0\,.$$

Further, observe that

$$\frac{d}{dt}(\mathbf{b} \cdot \mathbf{b}) = \dot{\mathbf{b}} \cdot \mathbf{b} + \mathbf{b} \cdot \dot{\mathbf{b}} = 2\dot{\mathbf{b}} \cdot \mathbf{b}\,.$$

On the other hand, since $\mathbf{b}$ is a unit vector, we have

$$\frac{d}{dt}(\mathbf{b} \cdot \mathbf{b}) = \frac{d}{dt}(\|\mathbf{b}\|^2) = \frac{d}{dt}(1) = 0$$

Therefore

$$\dot{\mathbf{b}} \cdot \mathbf{b} = 0\,.$$

To summarize, we have shown that $\dot{\mathbf{b}}$ is orthogonal to $\mathbf{b}$ and $\dot{\boldsymbol{\gamma}}$. Since

$$(\dot{\boldsymbol{\gamma}}, \mathbf{n}, \mathbf{b})$$

is an orthonormal basis of $\mathbb{R}^3$ we conclude that $\dot{\mathbf{b}}$ is parallel to $\mathbf{n}$. Therefore there exists $\tau(t) \in \mathbb{R}$ such that

$$\dot{\mathbf{b}} = -\tau(t)\mathbf{n}(t)\,,$$

concluding the proof.

The scalar τ in equation (2.13) is called the torsion of $\boldsymbol{\gamma}$.

Definition 2.37: Torsion of unit speed curve

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a unit speed curve, with $\kappa \neq 0$. The **torsion** of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is the unique scalar

$$\tau(t) \in \mathbb{R}$$

such that

$$\dot{\mathbf{b}}(t) = -\tau(t)\mathbf{n}(t)\,.$$

Remark 2.38

In particular the torsion satisfies:

$$\tau(t) = -\dot{\mathbf{b}}(t) \cdot \mathbf{n}(t)\,.$$

The above can be immediately obtained by multiplying (2.13) by $\mathbf{n}$. Indeed,

$$\dot{\mathbf{b}} = -\tau\mathbf{n} \quad \implies \quad \dot{\mathbf{b}} \cdot \mathbf{n} = -\tau\mathbf{n} \cdot \mathbf{n} = -\tau\,,$$

since $\mathbf{n}$ is a unit vector.

Warning

We defined the torsion only for space curves $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ which are unit speed and have non-vanishing curvature, that is, such that

$$\|\dot{\boldsymbol{\gamma}}(t)\| = 1\,, \quad \kappa(t) = \|\ddot{\boldsymbol{\gamma}}(t)\| \neq 0\,,$$

for all $t \in (a,b)$.

We can extend the definition of torsion to regular curves $\boldsymbol{\gamma}$ with non-vanishing curvature. In this case the torsion of $\boldsymbol{\gamma}$ is defined as the torsion of a unit speed reparametrization of $\boldsymbol{\gamma}$.

Definition 2.39

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a regular curve with non-vanishing curvature. Let $\tilde{\boldsymbol{\gamma}}$ be a unit speed reparametrization of $\boldsymbol{\gamma}$, with

$$\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi\,, \quad \phi \colon (a,b) \to (\tilde{a}, \tilde{b})\,.$$

We define the torsion of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ as

$$\tau^{\boldsymbol{\gamma}}(t) := \tau^{\tilde{\boldsymbol{\gamma}}}(\phi(t))\,,$$

where $\tau^{\tilde{\boldsymbol{\gamma}}}(s)$ denotes the torsion of $\tilde{\boldsymbol{\gamma}}$ at $\tilde{\boldsymbol{\gamma}}(s)$.

As usual, it is possible to check that the above definition of torsion does not depend on the choice of unit speed reparametrization $\tilde{\boldsymbol{\gamma}}$. As with curvature, there is a general formula to compute the torsion without having to reparametrize.

Proposition 2.40: Torsion formula

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a regular curve with non-vanishing curvature. The torsion $\tau(t)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is given by

$$\tau(t) = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} .$$

We delay the proof of the above proposition for a bit. In the meantime, let us look at examples.

Example 2.41: Torsion Helix

Consider the Helix of radius $R > 0$ and rise $H > 0$

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t), Ht) , \quad t \in \mathbb{R} .$$

We have already shown that

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \sqrt{R^2 + H^2} , \quad \kappa = \frac{R}{R^2 + H^2} .$$

Therefore the Helix is regular with non-vanishing curvature. The torsion can be then computed via the formula

$$\tau(t) = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} .$$

Let us compute the quantities appearing in the formula for τ

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) &= (-R\sin(t), R\cos(t), H) \\ \ddot{\boldsymbol{\gamma}}(t) &= (-R\cos(t), -R\sin(t), 0) \\ \dddot{\boldsymbol{\gamma}}(t) &= (R\sin(t), -R\cos(t), 0) \end{aligned}$$

Moreover we had already computed that

$$\begin{aligned} \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} &= \left(RH\sin(t), -RH\cos(t), R^2\right) \\ \|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| &= R\sqrt{R^2 + H^2} . \end{aligned}$$

Finally we compute

$$(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} = R^2 H .$$

We are ready to find the torsion:

$$\tau = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} = \frac{H}{R^2 + H^2} .$$

Example 2.42: Curvature and Torsion of Circle

The Circle of radius $R > 0$ is

$$\boldsymbol{\gamma}(t) := (R\cos(t), R\sin(t), 0)\,.$$

The curvature and torsion of the Helix of radius R and rise $H > 0$ are

$$\kappa = \frac{R}{R^2 + H^2}\,, \quad \tau = \frac{H}{R^2 + H^2}\,.$$

For $H = 0$ the Helix coincides with the Circle $\boldsymbol{\gamma}$. Therefore we can set $H = 0$ in the above formulas to obtain the curvature and torsion of the Circle

$$\kappa = \frac{1}{R}\,, \quad \tau = 0\,.$$

From the above example we notice that the torsion of the circle is 0. This is true in general for space curves which are contained in a plane: we will prove this result in general. For the moment, let us give an example for which this happens, that is, an example of space curve $\boldsymbol{\gamma}$ which is contained in a plane.

Example 2.43

Define the space curve

$$\boldsymbol{\gamma}(t) := \left(\frac{4}{5}\cos(t), 1 - \sin(t), -\frac{3}{5}\cos(t) \right)\,,$$

for $t \in \mathbb{R}$. As seen in the plot below, $\boldsymbol{\gamma}$ is just a Circle which has been rotated an translated. Therefore $\boldsymbol{\gamma}$ is contained in a plane, and we expect curvature and torsion to be

$$\kappa = \frac{1}{R}, \quad \tau = 0\,,$$

for some $R > 0$, radius of the Circle $\boldsymbol{\gamma}$. Let us proceed with the calculations:

$$\dot{\boldsymbol{\gamma}} = \left(-\frac{4}{5}\sin(t), -\cos(t), \frac{3}{5}\sin(t) \right)$$

so that

$$\|\dot{\boldsymbol{\gamma}}\|^2 = \frac{16}{25}\sin^2(t) + \cos^2(t) + \frac{9}{25}\sin^2(t) = 1\,,$$

showing that $\boldsymbol{\gamma}$ is regular and unit speed. Further

$$\ddot{\boldsymbol{\gamma}} = \left(-\frac{4}{5}\cos(t), \sin(t), \frac{3}{5}\cos(t) \right)\,.$$

As $\boldsymbol{\gamma}$ is unit speed, we have

$$\kappa = \|\ddot{\boldsymbol{\gamma}}\| = \frac{16}{25}\cos^2(t) + \sin^2(t) + \frac{9}{25}\cos^2(t) = 1\,.$$

As $\boldsymbol{\gamma}$ is unit speed, the normal vector is

$$\mathbf{n} = \frac{1}{\kappa}\ddot{\boldsymbol{\gamma}} = \left(-\frac{4}{5}\cos(t), \sin(t), \frac{3}{5}\cos(t)\right).$$

We can then compute the binormal

$$\begin{aligned}
\mathbf{b} &= \dot{\boldsymbol{\gamma}} \times \mathbf{n} \\
&= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\frac{4}{5}\sin(t) & -\cos(t) & \frac{3}{5}\sin(t) \\ -\frac{4}{5}\cos(t) & \sin(t) & \frac{3}{5}\cos(t) \end{vmatrix} \\
&= \left(-\frac{3}{5}\cos^2(t) - \frac{3}{5}\sin^2(t), -\frac{12}{25}\cos(t)\sin(t) + \frac{12}{25}\cos(t)\sin(t), -\frac{4}{5}\sin^2(t) - \frac{4}{5}\cos^2(t)\right) \\
&= \left(-\frac{3}{5}, 0, -\frac{4}{5}\right).
\end{aligned}$$

Therefore

$$\dot{\mathbf{b}} = 0\,,$$

and we obtain that the torsion is

$$\tau = -\dot{\mathbf{b}} \cdot \mathbf{n} = 0\,.$$

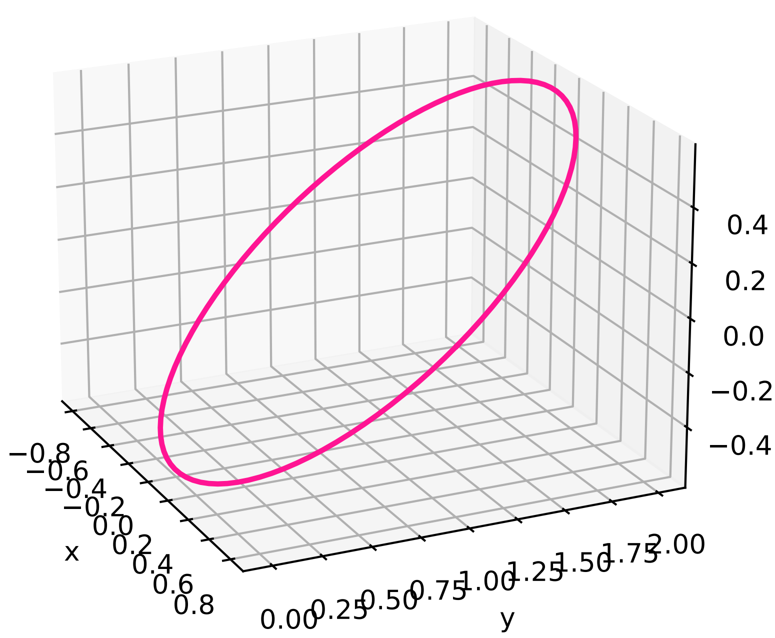


Figure 2.9: Plot of the curve in example above

2.6 Frenet frame

For a unit speed curve $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ with non-vanishing curvature we computed the triple

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{n}, \mathbf{b}\}.$$

We saw that the above is a positive orthonormal basis of $\mathbb{R}^3$. We also used this triple to compute curvature κ and torsion τ of $\boldsymbol{\gamma}$:

$$\kappa = \|\ddot{\boldsymbol{\gamma}}\| , \quad \tau = -\dot{\mathbf{b}} \cdot \mathbf{n} .$$

This triple is so important that it has a name.

Definition 2.44: Frenet frame

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be unit speed with $\kappa \neq 0$. The positive orthonormal basis

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{n}, \mathbf{b}\}$$

is called **Frenet frame** of $\boldsymbol{\gamma}$.

We can also define the Frenet frame for regular curves with non-vanishing curvature.

Definition 2.45

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be regular with $\kappa \neq 0$. The Frenet frame of $\boldsymbol{\gamma}$ is defined as the Frenet frame of a unit speed reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$.

Remark 2.46

We should check that the above definition is well-posed:

- Note that $\tilde{\boldsymbol{\gamma}}$ is unit speed. Moreover the curvature of $\kappa^{\tilde{\boldsymbol{\gamma}}}$ is given by

  $$\kappa^{\tilde{\boldsymbol{\gamma}}}(t) = \kappa^{\boldsymbol{\gamma}}(\phi(t))$$

  for some ϕ diffeomorphism. Therefore $\kappa^{\tilde{\boldsymbol{\gamma}}} \neq 0$ as we are assuming $\kappa^{\boldsymbol{\gamma}} \neq 0$. Therefore the Frenet-Frame of $\tilde{\boldsymbol{\gamma}}$ is well defined.

- If $\hat{\boldsymbol{\gamma}}$ is another unit speed reparametrization of $\boldsymbol{\gamma}$, then the Frenet frame generated by $\hat{\boldsymbol{\gamma}}$ coincides with the one generated by $\tilde{\boldsymbol{\gamma}}$. The proof is left as an exercise.

From the Frenet frame we can define the Frenet-Serret equations.

Theorem 2.47: Frenet-Serret equations

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be unit speed with $\kappa \neq 0$. The **Frenet-Serret** equations are

$$\begin{aligned} \ddot{\boldsymbol{\gamma}} &= \kappa \mathbf{n} \\ \dot{\mathbf{n}} &= -\kappa \dot{\boldsymbol{\gamma}} + \tau \mathbf{b} \\ \dot{\mathbf{b}} &= -\tau \mathbf{n} \end{aligned}$$

Proof

The first Frenet-Serret equation

$$\ddot{\boldsymbol{\gamma}} = \kappa \mathbf{n} \tag{2.15}$$

holds by definition of $\mathbf{n}$ and κ. The third Frenet-Serret equation

$$\dot{\mathbf{b}} = -\tau \mathbf{n} \tag{2.16}$$

holds by Proposition 2.36. Now, recall that in Proposition 2.35 we have proven

$$\mathbf{b} = \dot{\boldsymbol{\gamma}} \times \mathbf{n} \,, \quad \mathbf{n} = \mathbf{b} \times \dot{\boldsymbol{\gamma}} \,, \quad \dot{\boldsymbol{\gamma}} = \mathbf{n} \times \mathbf{b} \,. \tag{2.17}$$

Differentiating the second equation in (2.17) and using (2.15)-(2.16) we get

$$\begin{aligned} \dot{\mathbf{n}} &= \dot{\mathbf{b}} \times \dot{\boldsymbol{\gamma}} + \mathbf{b} \times \ddot{\boldsymbol{\gamma}} \\ &= (-\tau \mathbf{n} \times \dot{\boldsymbol{\gamma}}) + \mathbf{b} \times \kappa \mathbf{n} \\ &= \tau(\dot{\boldsymbol{\gamma}} \times \mathbf{n}) - \kappa(\mathbf{n} \times \mathbf{b}) \\ &= \tau \mathbf{b} - \kappa \dot{\boldsymbol{\gamma}} \,, \end{aligned}$$

where in the last equality we used the first and third equations in (2.17). The above is exactly the second Frenet-Serret equation.

Remark 2.48

We can write the Frenet-Serret ODE sysyem in vectorial form. To this end, introduce the matrix

$$F := \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & \tau & 0 \end{pmatrix} .$$

It is immediate to check that the Frenet-Serret equations are equivalent to

$$\begin{pmatrix} \ddot{\boldsymbol{\gamma}} \\ \dot{\mathbf{n}} \\ \dot{\mathbf{b}} \end{pmatrix} = F \begin{pmatrix} \dot{\boldsymbol{\gamma}} \\ \mathbf{n} \\ \mathbf{b} \end{pmatrix} .$$

Important: Summary

Recall that:

1. Curvature κ is defined only for regular curves.
2. Torsion τ is defined only for regular curves with non-vanishing κ.

The two strategies for computing κ and τ are discussed in the diagram in Figure 2.10 below.

Let us conclude the section with an example. We compute the Frenet frame of the helix. As a consequence we obtain curvature and torsion.

Example 2.49: Frenet frame of helix

Consider the helix of radius 1 and rise 1 given by

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

for $t \in \mathbb{R}$. We now proceed following the diagram at Figure 2.10. We ask the first question:

Is $\boldsymbol{\gamma}$ unit speed?

We have that

$$\dot{\boldsymbol{\gamma}}(t) = (-\sin(t), \cos(t), 1) ,$$

and therefore

$$\|\dot{\boldsymbol{\gamma}}\| = \sqrt{2} .$$

This shows that $\boldsymbol{\gamma}$ is regular but not unit speed. We ask the second question in the diagram:

Can we find a unit speed reparametrization of $\boldsymbol{\gamma}$?

Let us try. We compute the arc length of $\boldsymbol{\gamma}$ starting at $t_0 = 0$

$$s(t) := \int_0^t \|\dot{\boldsymbol{\gamma}}(u)\| \, du = \sqrt{2}\, t .$$

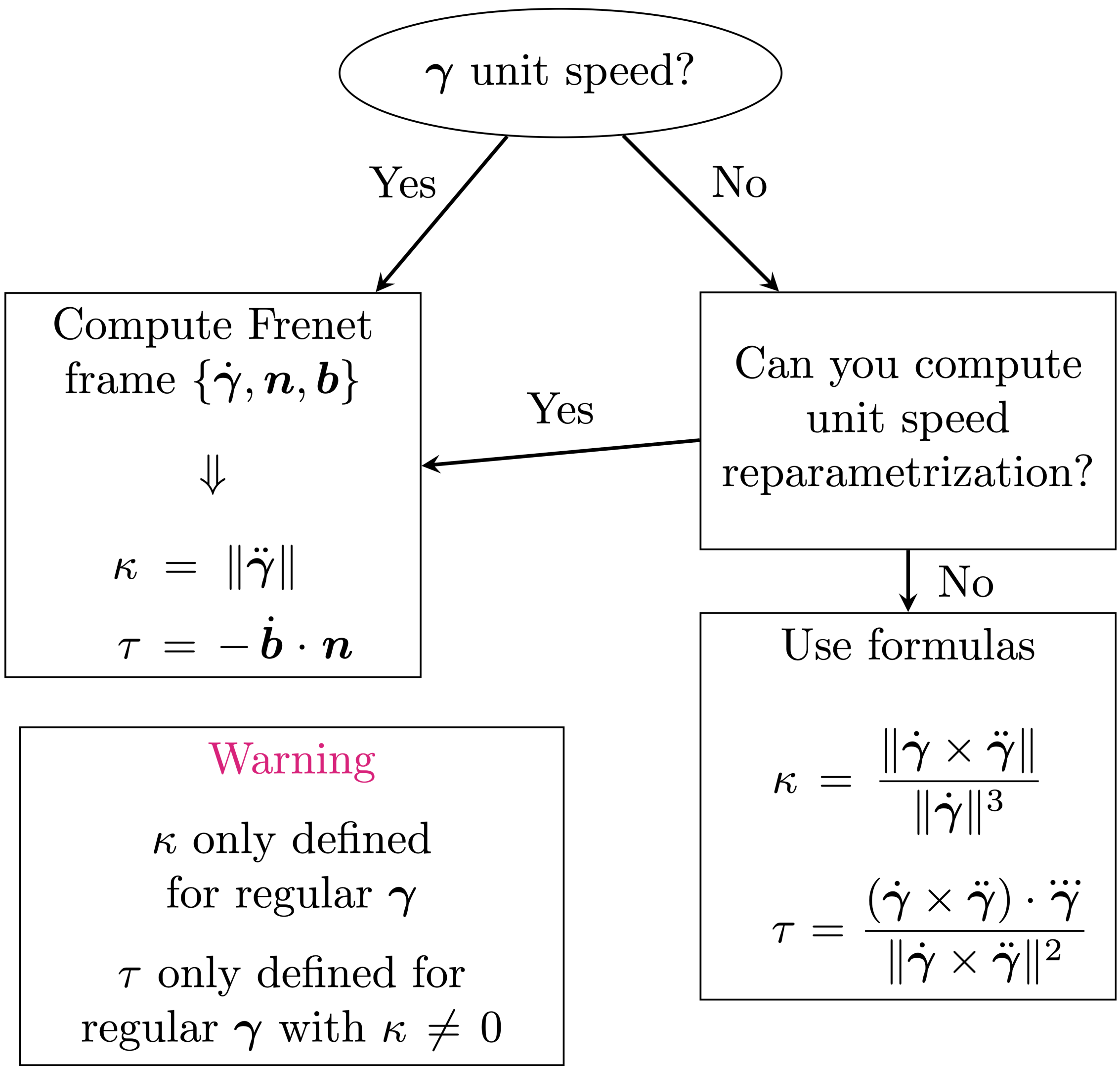


Figure 2.10: Summary for computing κ and τ for regular curve $\boldsymbol{\gamma}$.

The arc length is invertible with

$$\psi(t) := s^{-1}(t) = \frac{t}{\sqrt{2}}.$$

Therefore a unit speed reparametrization of $\boldsymbol{\gamma}$ is given by

$$\tilde{\boldsymbol{\gamma}}(t) := \boldsymbol{\gamma}(\psi(t)) = \left(\cos\left(\frac{t}{\sqrt{2}}\right), \sin\left(\frac{t}{\sqrt{2}}\right), \frac{t}{\sqrt{2}}\right).$$

The next step in the diagram is

Compute Frenet frame $\{\dot{\boldsymbol{\gamma}}, \mathbf{n}, \mathbf{b}\}$ and curvature κ, torsion τ

We compute

$$\begin{aligned}
\dot{\tilde{\boldsymbol{\gamma}}}(t) &= \frac{1}{\sqrt{2}}\left(-\sin\left(\frac{t}{\sqrt{2}}\right), \cos\left(\frac{t}{\sqrt{2}}\right), 1\right) \\
\ddot{\tilde{\boldsymbol{\gamma}}}(t) &= \frac{1}{2}\left(-\cos\left(\frac{t}{\sqrt{2}}\right), -\sin\left(\frac{t}{\sqrt{2}}\right), 0\right)
\end{aligned}$$

Therefore the curvature is

$$\kappa(t) = \|\ddot{\tilde{\boldsymbol{\gamma}}}(t)\| = \frac{1}{2}.$$

From the curvature we obtain the principal normal vector

$$\mathbf{n}(t) = \frac{1}{\kappa(t)}\ddot{\tilde{\boldsymbol{\gamma}}}(t) = \left(-\cos\left(\frac{t}{\sqrt{2}}\right), -\sin\left(\frac{t}{\sqrt{2}}\right), 0\right).$$

We can now compute the binormal

$$\begin{aligned}
\mathbf{b}(t) &= \dot{\tilde{\boldsymbol{\gamma}}} \times \mathbf{n} \\
&= \frac{1}{\sqrt{2}}\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\sin\left(\frac{t}{\sqrt{2}}\right) & \cos\left(\frac{t}{\sqrt{2}}\right) & 1 \\ -\cos\left(\frac{t}{\sqrt{2}}\right) & -\sin\left(\frac{t}{\sqrt{2}}\right) & 0 \end{vmatrix} \\
&= \frac{1}{\sqrt{2}}\left(\sin\left(\frac{t}{\sqrt{2}}\right), -\cos\left(\frac{t}{\sqrt{2}}\right), 1\right).
\end{aligned}$$

We have therefore computed the Frenet frame of $\boldsymbol{\gamma}$. This is given by

$$\begin{aligned}
\dot{\tilde{\boldsymbol{\gamma}}}(t) &= \frac{1}{\sqrt{2}}\left(-\sin\left(\frac{t}{\sqrt{2}}\right), \cos\left(\frac{t}{\sqrt{2}}\right), 1\right) \\
\mathbf{n}(t) &= \left(-\cos\left(\frac{t}{\sqrt{2}}\right), -\sin\left(\frac{t}{\sqrt{2}}\right), 0\right) \\
\mathbf{b}(t) &= \frac{1}{\sqrt{2}}\left(\sin\left(\frac{t}{\sqrt{2}}\right), -\cos\left(\frac{t}{\sqrt{2}}\right), 1\right).
\end{aligned}$$

See below for a picture of the Frenet frame of the helix. Given the Frenet frame, we can compute the torsion via the formula

$$\tau(t) = -\dot{\mathbf{b}} \cdot \mathbf{n}\,.$$

Indeed, we have

$$\dot{\mathbf{b}} = \frac{1}{2}\left(\cos\left(\frac{t}{\sqrt{2}}\right), -\sin\left(\frac{t}{\sqrt{2}}\right), 0\right)$$

and therefore

$$\dot{\mathbf{b}} \cdot \mathbf{n} = \frac{1}{2}\left(-\cos^2\left(\frac{t}{\sqrt{2}}\right) - \sin^2\left(\frac{t}{\sqrt{2}}\right)\right) = -\frac{1}{2}\,.$$

The torsion is then

$$\tau(t) = -\dot{\mathbf{b}} \cdot \mathbf{n} = \frac{1}{2}\,.$$

The Frenet-Frame of the unit-speed Helix is plotted in Figure 2.11.

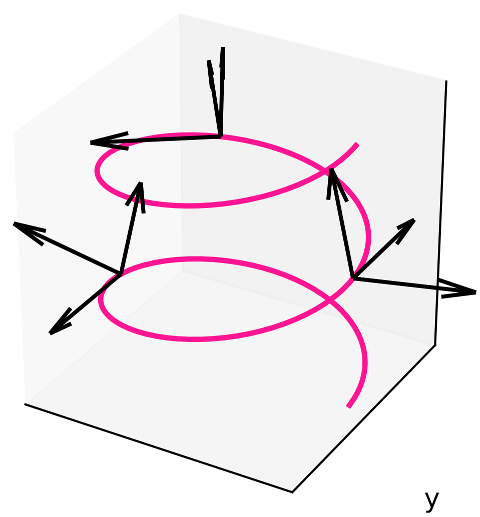


Figure 2.11: Frenet frame of the helix of radius 1 and rise 1.

2.7 Consequences of Frenet-Serret

The most important consequence of the Frenet-Serret equations is that they allow to fully characterize space curves in terms of curvature and torsion. Precisely, the following theorem holds.

Theorem 2.50: Characterization of space curves

Let $\kappa, \tau : \mathbb{R} \to \mathbb{R}$ be smooth functions, with $\kappa > 0$. Then:

1. There exists aunit speed curve $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^3$ such that its curvature $\kappa^{\boldsymbol{\gamma}}$ and torsion $\tau^{\boldsymbol{\gamma}}$ satisfy
$$\kappa^{\boldsymbol{\gamma}}(t) = \kappa(t), \quad \tau^{\boldsymbol{\gamma}}(t) = \tau(t), \ \forall t \in \mathbb{R}.$$
2. Suppose that $\tilde{\boldsymbol{\gamma}} : \mathbb{R} \to \mathbb{R}^3$ is a unit speed curve such that its curvature $\tilde{\kappa}^{\tilde{\boldsymbol{\gamma}}}$ and torsion $\tau^{\tilde{\boldsymbol{\gamma}}}$ satisfy
$$\kappa^{\tilde{\boldsymbol{\gamma}}}(t) = \kappa(t), \quad \tau^{\tilde{\boldsymbol{\gamma}}}(t) = \tau(t), \ \forall t \in \mathbb{R}.$$
Then
$$\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$$
up to rotations and translations.

The proof of Theorem 2.50 is omitted, and it can be found in Theorem 2.3.6 in [6].

Theorem 2.50 is a very strong result. It is saying two things:

1. If we prescribe curvature and torsion, then there exists a unit speed curve which has such curvature and torsion.
2. If two unit speed curves have same curvature and torsion, then they must be the same curve, up to translations and rotations.

In other words, curvature and torsion fully characterize space curves. This result is the 3D counterpart of Theorem 2.27, which said that signed curvature characterizes 2D curves.

Example 2.51

In Example 2.43 we have considered the unit speed curve
$$\boldsymbol{\gamma}(t) := \left(\frac{4}{5}\cos(t), 1 - \sin(t), -\frac{3}{5}\cos(t) \right),$$
for $t \in [0, 2\pi]$. We have computed that
$$\kappa^{\boldsymbol{\gamma}} = 1, \quad \tau^{\boldsymbol{\gamma}} = 0.$$
If we plot $\boldsymbol{\gamma}$, we clearly see that $\boldsymbol{\gamma}$ is just obtained by translating and rotating a unit circle, see plot below. Theorem 2.50 enables us to rigorously prove this claim. Indeed, consider the unit speed circle
$$\tilde{\boldsymbol{\gamma}}(t) := (\cos(t), \sin(t), 0),$$
for $t \in [0, 2\pi]$. In Example 2.42 we have proven that curvature and torsion are
$$\kappa^{\tilde{\boldsymbol{\gamma}}} = 1, \quad \tau^{\tilde{\boldsymbol{\gamma}}} = 1.$$

Therefore

$$\kappa^{\boldsymbol{\gamma}} = \kappa^{\tilde{\boldsymbol{\gamma}}}, \quad \tau^{\boldsymbol{\gamma}} = \tau^{\tilde{\boldsymbol{\gamma}}},$$

and by Theorem 2.50 we conclude that $\boldsymbol{\gamma}$ is equal to $\tilde{\boldsymbol{\gamma}}$ up to rotations and translations.

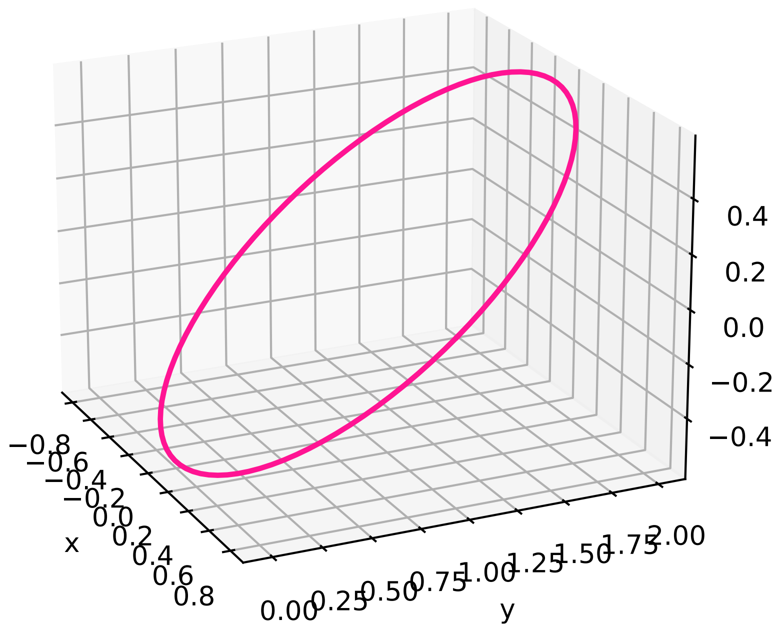


Figure 2.12: Plot of the curve in example above

Another consequence of the Frenet-Serret equations is that they allow us to finally prove the curvature and torsion formulas given in Proposition 2.19 and Proposition 2.40. For reader's convenience we recall these two results.

Proposition 2.52: Curvature and torsion formulas

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be a regular curve. The curvature $\kappa(t)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is given by

$$\kappa(t) = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3}.$$

Suppose in addition that $\boldsymbol{\gamma}$ has non-vanishing curvature. The torsion $\tau(t)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is given by

$$\tau(t) = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2}.$$

Before proceeding with the proof, we need to establish some notation.

Notation: Compact notation for arc length reparametrization

Suppose $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^n$ is regular and denote by

$$s : (a, b) \to (\tilde{a}, \tilde{b}) , \quad t \mapsto s(t)$$

its arc length. We already know that in this case s invertible, with inverse s^{-1} giving a unit speed reparametrization $\tilde{\boldsymbol{\gamma}} : (\tilde{a}, \tilde{b}) \to \mathbb{R}^n$ of $\boldsymbol{\gamma}$, defined by

$$\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} \circ \psi , \quad \psi := s^{-1} : (\tilde{a}, \tilde{b}) \to (a, b)$$

Sometimes it is more convenient to adopt more compact notation. In the new notation the unit speed reparametrization $\tilde{\boldsymbol{\gamma}}$ is denoted by $\boldsymbol{\gamma}(s)$:

$$t \mapsto \tilde{\boldsymbol{\gamma}}(t) \qquad \rightsquigarrow \qquad s \mapsto \boldsymbol{\gamma}(s) .$$

Thus, the reparametrization is denoted with the same symbol $\boldsymbol{\gamma}$, but this time $\boldsymbol{\gamma}$ is considered as a function of the **arc length parameter**

$$s \in (\tilde{a}, \tilde{b}) .$$

We will denote:

- The derivative of s by
$$\frac{ds}{dt}$$
- The derivative of $\psi = s^{-1}$ by
$$\frac{dt}{ds} .$$

Moreover:

- The derivative of $\boldsymbol{\gamma}(t)$ is denoted by
$$\frac{d\boldsymbol{\gamma}}{dt}(t) = \dot{\boldsymbol{\gamma}}(t) , \quad t \in (a, b)$$
- The derivative of $\boldsymbol{\gamma}(s)$ is denoted by
$$\frac{d\boldsymbol{\gamma}}{ds}(s) = \dot{\boldsymbol{\gamma}}(s) , \quad s \in (\tilde{a}, \tilde{b}) .$$

We also have new notations for the **chain rule**:

- The chain rule for $\boldsymbol{\gamma}$ is the old notations is:

$$\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\gamma}}(s(t)) \quad \Longrightarrow \quad \dot{\boldsymbol{\gamma}}(t) = \dot{\tilde{\boldsymbol{\gamma}}}(s(t))\, \dot{s}(t)\,, \quad t \in (a,b)\,.$$

  In the new notations the above chain rule is written

$$\frac{d\boldsymbol{\gamma}}{dt}(t) = \frac{d\boldsymbol{\gamma}}{ds}(s(t))\, \frac{ds}{dt}(t)\,, \quad t \in (a,b)\,.$$

  We will often omit the dependence on the point t by writing

$$\frac{d\boldsymbol{\gamma}}{ds} = \frac{d\boldsymbol{\gamma}}{dt}\,\frac{dt}{ds}\,.$$

- The chain rule for the reparametrization $\tilde{\boldsymbol{\gamma}}$ in the old notation is:

$$\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(\psi(t)) \quad \Longrightarrow \quad \dot{\tilde{\boldsymbol{\gamma}}}(t) = \dot{\boldsymbol{\gamma}}(\psi(t))\, \dot{\psi}(t)\,, \quad t \in (\tilde{a},\tilde{b})\,.$$

  In the new notations the above chain rule is written

$$\frac{d\boldsymbol{\gamma}}{ds}(s) = \frac{d\boldsymbol{\gamma}}{dt}(\psi(s))\, \frac{dt}{ds}(s)\,, \quad s \in (\tilde{a},\tilde{b})\,,$$

  since $\dot{\psi}$ is written dt/ds in the new notations. Without dependence on the point s, the above reads

$$\frac{d\boldsymbol{\gamma}}{ds} = \frac{d\boldsymbol{\gamma}}{dt}\,\frac{dt}{ds}\,.$$

Example 2.53: How to use the new notations

Let $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ be as above. We know that $\tilde{\boldsymbol{\gamma}}$ is unit speed. Thus $\boldsymbol{\gamma}(s)$ is unit speed with respect to s, that is,

$$\|\dot{\boldsymbol{\gamma}}(s)\| = 1\,, \quad \forall\, s \in (\tilde{a},\tilde{b})\,. \tag{2.18}$$

As an exercise, let us check that (2.18) holds, using the new notations. By chain rule we have

$$\begin{aligned}
\|\dot{\boldsymbol{\gamma}}(s)\| & = \left\| \frac{d\boldsymbol{\gamma}}{ds}(s) \right\| \\
& = \left\| \frac{d\boldsymbol{\gamma}}{dt}(\psi(s)) \right\| \left| \frac{dt}{ds}(s) \right| \\
& = \|\dot{\boldsymbol{\gamma}}(\psi(s))\| \left| \frac{dt}{ds}(s) \right|\,.
\end{aligned}$$

Now, recall that

$$\frac{ds}{dt}(t) = \dot{s}(t) = \|\dot{\boldsymbol{\gamma}}(t)\|\,, \quad \forall\, t \in (a,b)\,. \tag{2.19}$$

According to the new notations and the inverse function theorem,

$$\frac{dt}{ds}(s) = \frac{1}{\left(\frac{ds}{dt}(\psi(s))\right)} = \frac{1}{\|\dot{\boldsymbol{\gamma}}(\psi(s))\|}, \quad \forall\, s \in (\tilde{a}, \tilde{b}),$$

where we used (2.19) evaluated at $t = \psi(s)$. Thus

$$\begin{aligned} \|\dot{\boldsymbol{\gamma}}(s)\| &= \|\dot{\boldsymbol{\gamma}}(\psi(s))\| \left| \frac{dt}{ds}(s) \right| \\ &= \|\dot{\boldsymbol{\gamma}}(\psi(s))\| \, \frac{1}{\|\dot{\boldsymbol{\gamma}}(\psi(s))\|} \\ &= 1, \end{aligned}$$

concluding (2.18).

Let us highlight the main feature of the above notation.

Important: New Notation!

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a regular curve:

1. We denote by
$$t \mapsto \boldsymbol{\gamma}(t), \quad t \in (a,b)$$
the **given** curve $\boldsymbol{\gamma}$.
2. We denote by
$$s \mapsto \boldsymbol{\gamma}(s), \quad s \in (\tilde{a}, \tilde{b})$$
the **arc length reparametrization** of the curve $\boldsymbol{\gamma}$. The parameter s is the **arc length parameter**. In particular $\boldsymbol{\gamma}(s)$ is unit speed with respect to s.

We will heavily rely on the new notations for proving Proposition 2.52.

Proof: Proof of Proposition 2.52

We only prove the formula for κ, as the one for τ can be obtained similarly, just with more calculations. For a proof see Proposition 2.3.1 in [6].
Since $\boldsymbol{\gamma}$ is regular, we can reparametrize $\boldsymbol{\gamma}$ by arc length $s(t)$. We denote the arc lenght reparametrization by $\boldsymbol{\gamma}(s)$. We know that $\boldsymbol{\gamma}(s)$ is unit speed, that is,

$$\left\| \frac{d\boldsymbol{\gamma}}{ds} \right\| = 1.$$

Therefore is well define the Frenet frame

$$\{\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)\}, \quad \mathbf{t}(s) := \dot{\boldsymbol{\gamma}}(s) = \frac{d\boldsymbol{\gamma}}{ds}(s).$$

The Frenet-Serret equations are

$$\begin{aligned}
\dot{\mathbf{t}}(s) &= \kappa(s)\mathbf{n}(s) \\
\dot{\mathbf{n}}(s) &= -\kappa(s)\mathbf{t}(s) + \tau(s)\mathbf{b}(s) \\
\dot{\mathbf{b}}(s) &= -\tau(s)\mathbf{n}(s)
\end{aligned}$$

By chain rule

$$\frac{d\boldsymbol{\gamma}}{dt} = \frac{d\boldsymbol{\gamma}}{ds}\frac{ds}{dt} = \left(\frac{ds}{dt}\right)\mathbf{t}.$$

Differentiating the above we infer

$$\begin{aligned}
\frac{d^2\boldsymbol{\gamma}}{dt^2} &= \frac{d}{dt}\left[\left(\frac{ds}{dt}\right)\mathbf{t}\right] \\
&= \frac{d^2 s}{dt^2}\mathbf{t} + \left(\frac{ds}{dt}\right)\frac{d\mathbf{t}}{dt}.
\end{aligned}$$

By chain rule we have

$$\frac{d\mathbf{t}}{dt} = \frac{d\mathbf{t}}{ds}\frac{dt}{ds},$$

and therefore

$$\begin{aligned}
\frac{d^2\boldsymbol{\gamma}}{dt^2} &= \frac{d^2 s}{dt^2}\mathbf{t} + \left(\frac{ds}{dt}\right)\frac{d\mathbf{t}}{dt} \\
&= \frac{d^2 s}{dt^2}\mathbf{t} + \left(\frac{ds}{dt}\right)^2\frac{d\mathbf{t}}{ds}.
\end{aligned}$$

Hence

$$\begin{aligned}
\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t) &= \frac{d\boldsymbol{\gamma}}{dt} \times \frac{d^2\boldsymbol{\gamma}}{dt^2} \\
&= \left(\frac{ds}{dt}\right)\mathbf{t} \times \left[\frac{d^2 s}{dt^2}\mathbf{t} + \left(\frac{ds}{dt}\right)^2\frac{d\mathbf{t}}{ds}\right] \\
&= \left[\left(\frac{ds}{dt}\right)\left(\frac{d^2 s}{dt^2}\right)\mathbf{t} \times \mathbf{t}\right] + \left[\left(\frac{ds}{dt}\right)^3 \mathbf{t} \times \frac{d\mathbf{t}}{ds}\right] \\
&= \left(\frac{ds}{dt}\right)^3 \mathbf{t} \times \frac{d\mathbf{t}}{ds},
\end{aligned}$$

since $\mathbf{t} \times \mathbf{t} = 0$ by the properties of the cross product. Now we recall that

$$\frac{d\mathbf{t}}{ds} = \kappa(s)\,\mathbf{n}(s)$$

by the first Frenet-Serret equation. Moreover

$$\frac{ds}{dt}(t) = \left\lVert \dot{\boldsymbol{\gamma}}(t) \right\rVert^2 .$$

Therefore

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t) &= \left(\frac{ds}{dt} \right)^3 \ \mathbf{t} \times \frac{d\mathbf{t}}{ds} \\ &= \left\lVert \dot{\boldsymbol{\gamma}}(t) \right\rVert^3 \ \kappa(s(t)) \ \mathbf{t} \times \mathbf{n} \\ &= \left\lVert \dot{\boldsymbol{\gamma}}(t) \right\rVert^3 \ \kappa(s(t)) \ \mathbf{b} \, , \end{aligned}$$

where in the last line we used the definition of $\mathbf{b}$

$$\mathbf{b}(s) = \dot{\boldsymbol{\gamma}}(s) \times \mathbf{n}(s) = \mathbf{t}(s) \times \mathbf{n}(s) \, .$$

We can now take the norms and obtain

$$\begin{aligned} \left\lVert \dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t) \right\rVert &= \left\lVert \dot{\boldsymbol{\gamma}}(t) \right\rVert^3 \ \kappa(s(t)) \ \left\lVert \mathbf{b} \right\rVert \\ &= \left\lVert \dot{\boldsymbol{\gamma}}(t) \right\rVert^3 \ \kappa(s(t)) \end{aligned}$$

using that $\left\lVert \mathbf{b} \right\rVert = 1$. As $\boldsymbol{\gamma}$ is regular, we can divide by $\left\lVert \dot{\boldsymbol{\gamma}}(t) \right\rVert^3$ and obtain

$$\kappa(s(t)) = \frac{\left\lVert \dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t) \right\rVert}{\left\lVert \dot{\boldsymbol{\gamma}}(t) \right\rVert^3} \, .$$

Recalling that the curvature of $\boldsymbol{\gamma}$ at t is defined as the curvature of $\boldsymbol{\gamma}(s)$ at $s(t)$, we conclude that the above is the desired formula.

We now state and prove two more results which directly follow from the Frenet-Serret equations. They state, respectivley:

1. A curve has torsion $\tau = 0$ if and only if it is contained in a plane.
2. A curve has constant curvature and zero torsion if and only if it is part of a circle.

Before proceeding, we recall the following.

Remark 2.54: Equation of a plane

The general equation of a plane $\boldsymbol{\pi}_d$ in $\mathbb{R}^3$ is given by

$$\boldsymbol{\pi}_d = \{\mathbf{x} \in \mathbb{R}^3 \ : \ \mathbf{x} \cdot \mathbf{P} = d\},$$

for some vector $\mathbf{P} \in \mathbb{R}^3$ and scalar $d \in \mathbb{R}$. Note that:

- If $d = 0$, the condition
$$\mathbf{x} \cdot \mathbf{P} = 0$$
is saying that the plane $\boldsymbol{\pi}_0$ contains all the points $\mathbf{x}$ in $\mathbb{R}^3$ which are orthogonal to $\mathbf{P}$. In particular $\boldsymbol{\pi}_0$ contains the origin $\mathbf{0}$.
- If $d \neq 0$, then $\boldsymbol{\pi}_d$ is the translation of $\boldsymbol{\pi}_0$ by the quantity d in direction $\mathbf{P}$.

In both cases, $\mathbf{P}$ is the normal vector to the plane, as shown in Figure 2.13 below.

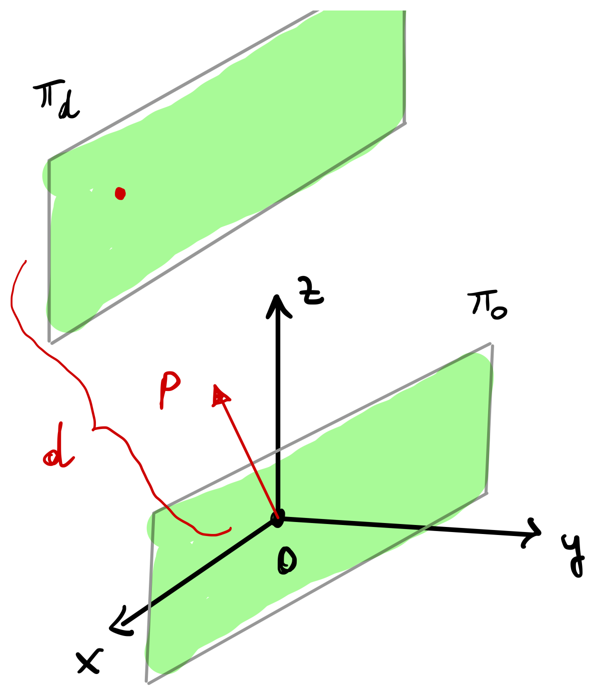


Figure 2.13: The plane $\boldsymbol{\pi}_0$ is the set of points of $\mathbb{R}^3$ orthogonal to $\mathbf{P}$. The plane $\boldsymbol{\pi}_d$ is obtained by translating $\boldsymbol{\pi}_0$ by a quantity d in direction $\mathbf{P}$.

Proposition 2.55

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be regular and such that $\kappa \neq 0$. They are equivalent:

1. The torsion of $\boldsymbol{\gamma}$ satisfies $\tau(t) = 0$ for all $t \in (a,b)$.
2. The image of $\boldsymbol{\gamma}$ is contained in a plane, that is, there exists a vector $\mathbf{P} \in \mathbb{R}^3$ and a scalar $d \in \mathbb{R}$ such

that

$$\boldsymbol{\gamma}(t) \cdot \mathbf{P} = d \,, \quad \forall t \in (a,b) \,.$$

Proof

Without loss of generality we can assume that $\boldsymbol{\gamma}$ is unit speed. Indeed, if we were to consider $\tilde{\boldsymbol{\gamma}}$ a unit speed reparametrization of $\boldsymbol{\gamma}$, then

- $\tilde{\boldsymbol{\gamma}}$ would still be contained in the same plane in which $\boldsymbol{\gamma}$ is contained.
- The torsion of $\tilde{\boldsymbol{\gamma}}$ would not change, i.e., it would still be identically zero.

Thefore the Frenet frame of $\boldsymbol{\gamma}$ exists. We denote it by

$$\{\dot{\boldsymbol{\gamma}}(t), \mathbf{n}(t), \mathbf{b}(t)\} \,.$$

Step 1. Suppose that $\tau = 0$ for all t. By the Frenet-Serret equations we have

$$\dot{\mathbf{b}} = -\tau(t)\mathbf{n} = \mathbf{0} \,,$$

so that $\mathbf{b}(t)$ is constant. As by definition

$$\mathbf{b} = \dot{\boldsymbol{\gamma}} \times \mathbf{n} \,,$$

we conclude that the vectors $\dot{\boldsymbol{\gamma}}(t)$ and $\mathbf{n}(t)$ always span the same plane, which has constant normal vector $\mathbf{b}$. Intuition suggests that $\boldsymbol{\gamma}$ should be contained in such plane, see Figure Figure 2.14 below. Indeed, recall that the Frenet frame is orthonormal. Hence

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{b} = 0 \,, \quad \forall \, t \in (a,b) \,.$$

Then

$$\frac{d}{dt}(\boldsymbol{\gamma} \cdot \mathbf{b}) = \dot{\boldsymbol{\gamma}} \cdot \mathbf{b} + \boldsymbol{\gamma} \cdot \dot{\mathbf{b}} = 0 \,, \quad \forall \, t \in (a,b) \,,$$

since $\dot{\mathbf{b}} = 0$. Thus $\boldsymbol{\gamma} \cdot \mathbf{b}$ is a constant scalar function, meaning that there exists costant $d \in \mathbb{R}$ such that

$$\boldsymbol{\gamma}(t) \cdot \mathbf{b} = d \,, \ \forall \, t \in (a,b) \,.$$

The says that $\boldsymbol{\gamma}$ is contained in a plane.
Step 2. Suppose that $\boldsymbol{\gamma}$ is contained in a plane. Hence there exists $\mathbf{P} \in \mathbb{R}^3$ and $d \in \mathbb{R}$ such that

$$\boldsymbol{\gamma}(t) \cdot \mathbf{P} = d \,, \quad \forall \, t \in (a,b) \,.$$

We can differentiate the above equation twice to obtain

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{P} = 0 \,, \quad \ddot{\boldsymbol{\gamma}} \cdot \mathbf{P} = 0 \,,$$

where we used that $\mathbf{P}$ and d are constant. By Frenet-Serret we have

$$\ddot{\boldsymbol{\gamma}}(t) = \kappa(t)\mathbf{n}(t)\,.$$

Therefore the already proven relation $\ddot{\boldsymbol{\gamma}} \cdot \mathbf{P} = 0$ implies

$$\kappa(t)\mathbf{n}(t) \cdot \mathbf{P} = 0\,.$$

As we are assuming $\kappa \neq 0$, we deduce that

$$\mathbf{n}(t) \cdot \mathbf{P} = 0\,, \quad \forall\, t \in (a,b)\,.$$

We have shown that $\dot{\boldsymbol{\gamma}}(t)$ and $\mathbf{n}(t)$ are both orthogonal to $\mathbf{P}$. Since $\mathbf{b}(t)$ is orthogonal to $\dot{\boldsymbol{\gamma}}(t)$ and $\mathbf{n}(t)$, we conclude that $\mathbf{b}(t)$ is parallel to $\mathbf{P}$. Hence, there exists $\lambda(t) \in \mathbb{R}$ such that

$$\mathbf{b}(t) = \lambda(t)\mathbf{P}\,\forall\, t \in (a,b)\,. \tag{2.20}$$

Since $\|\mathbf{b}\| = 1$ and $\mathbf{P}$ is constant, from (2.20) we conclude that $\lambda(t)$ is constant. Differentiating (2.20) we obtain

$$\dot{\mathbf{b}}(t) = 0\,, \quad \forall\, t \in (a,b)\,.$$

By definition of torsion we thus have

$$\tau(t) = -\dot{\mathbf{b}} \cdot \mathbf{n}(t) = 0\,, \quad \forall\, t \in (a,b)\,.$$

Proposition 2.56

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a unit speed curve. They are equivalent:

1. The image of $\boldsymbol{\gamma}$ is contained in a circle of radius $1/c$.
2. The curvature and torsion of $\boldsymbol{\gamma}$ satisfy

$$\kappa(t) = c\,, \quad \tau(t) = 0\,, \quad \forall\, t \in (a,b)\,,$$

for some constant $c \in \mathbb{R}$.

Proposition 2.56 is actually a consequence of Theorem 2.50, and of the fact that we have computed that for a circle of radius R one has

$$\kappa = \frac{1}{R}\,, \quad \tau = 0\,.$$

Therefore, by Theorem 2.50, every unit speed curve $\boldsymbol{\gamma}$ with constant curvature and torsion must be equal to a circle, up to rigid motions.

Nevertheless, we still give a proof of Proposition 2.56, to show yet another application of the Frenet-Serret equations.

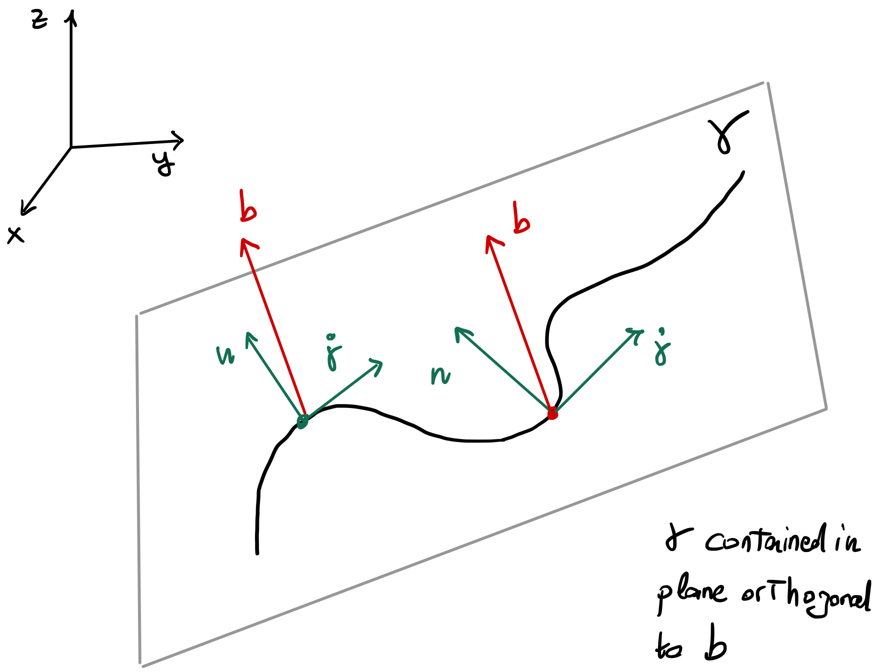


Figure 2.14: If $\mathbf{b}$ is constant, then $\boldsymbol{\gamma}$ lies in the plane spanned by $\dot{\boldsymbol{\gamma}}$ and $\mathbf{n}$.

Proof

Step 1. Suppose the image of $\boldsymbol{\gamma}$ is contained in a circle of radius $1/c$. Then, up to a translation, $\boldsymbol{\gamma}$ is parametrized by

$$\boldsymbol{\gamma}(t) = \left(\frac{1}{c}\cos(t), \frac{1}{c}\sin(t), 0 \right)$$

for t in some interval $(\tilde{a}, \tilde{b})$. We have already seen that in this case

$$\kappa = c\,, \quad \tau = 0\,,$$

concluding the proof.
Step 2. Suppose that

$$\kappa(t) = c\,, \quad \tau(t) = 0\,, \quad \forall\, t \in (a,b)\,,$$

for some constant $c \in \mathbb{R}$. Since $\boldsymbol{\gamma}$ is unit speed, its Frenet-Serret equations are:

$$\begin{aligned} \ddot{\boldsymbol{\gamma}} &= \kappa \mathbf{n} = c\mathbf{n} \\ \dot{\mathbf{n}} &= -\kappa \dot{\boldsymbol{\gamma}} + \tau \mathbf{b} = -c\dot{\boldsymbol{\gamma}} \\ \dot{\mathbf{b}} &= -\tau \mathbf{n} = 0 \end{aligned}$$

In particular $\dot{\mathbf{b}} = 0$ and so $\mathbf{b}$ is a constant vector. As seen in the proof Proposition 2.55, this implies that $\boldsymbol{\gamma}$ is contained in a plane $\boldsymbol{\pi}$ orthogonal to $\mathbf{b}$, see Figure 2.14. As c is constant we get

$$\frac{d}{dt}\left(\boldsymbol{\gamma} + \frac{1}{c}\mathbf{n} \right) = \dot{\boldsymbol{\gamma}} + \frac{1}{c}\dot{\mathbf{n}} = \dot{\boldsymbol{\gamma}} - \frac{1}{c}c\dot{\boldsymbol{\gamma}} = 0\,,$$

where we used the second Frenet-Serret equation. Therefore

$$\boldsymbol{\gamma}(t) + \frac{1}{c}\mathbf{n}(t) = \mathbf{p}\,, \quad t \in (a,b)\,,$$

for some constant point $\mathbf{p} \in \mathbb{R}^3$. In particular

$$\|\boldsymbol{\gamma}(t) - \mathbf{p}\| = \left\|-\frac{1}{c}\mathbf{n}(t)\right\| = \frac{1}{c}\,,$$

since $\mathbf{n}$ is a unit vector. The above shows that $\boldsymbol{\gamma}$ is contained in a sphere of radius $1/c$ and center $\mathbf{p}$. In formulas:

$$\boldsymbol{\gamma}((a,b)) \subset \mathscr{S} := \{\mathbf{x} \in \mathbb{R}^3 \ : \ \|\mathbf{x} - \mathbf{p}\| = 1/c\}\,.$$

The intersection of $\mathscr{S}$ with the plane $\boldsymbol{\pi}$ is a circle $\mathscr{C}$ with some radius R. Since

$$\boldsymbol{\gamma}((a,b)) \subset \boldsymbol{\pi}\,, \quad \boldsymbol{\gamma}((a,b)) \subset \mathscr{S}\,,$$

this implies

$$\boldsymbol{\gamma}((a,b)) \subset \boldsymbol{\pi} \cap \mathscr{S} = \mathscr{C}\,. \tag{2.21}$$

Thus $\boldsymbol{\gamma}$ parametrizes part of $\mathscr{C}$. From Step 1 it follows that the curvature and torsion of $\boldsymbol{\gamma}$ must satisfy

$$\kappa = \frac{1}{R}\,, \quad \tau = 0\,.$$

Since we already know that $\kappa = c$, we conclude that $R = 1/c$. Therefore the circle $\mathscr{C}$ has radius $1/c$ and the thesis follows by (2.21).

3 Topology

So far we have worked in $\mathbb{R}^n$, where for example we have the notions of open set, continuous function and compact set. Topology is what allows us to extend these notions to arbitrary sets.

Definition 3.1: Topological space

Let X be a set and $\mathscr{T}$ a collection of subsets of X. We say that $\mathscr{T}$ is a **topology** on X if the following 3 properties hold:

- (A1) We have $\emptyset, X \in \mathscr{T}$,
- (A2) If $\{A_i\}_{i\in I}$ is an arbitrary family of elements of $\mathscr{T}$, then

$$\bigcup_{i\in I} A_i \in \mathscr{T} .$$

- (A3) If $A, B \in \mathscr{T}$ then

$$A \cap B \in \mathscr{T} .$$

Further, we say:

- The pair $(X, \mathscr{T})$ is a **topological space**.
- The elements of X are called **points**.
- The sets in the topology $\mathscr{T}$ are called **open sets**.

Remark 3.2

The intersection property of $\mathscr{T}$, Property (A3) in Definition 3.1, is equivalent to the following:

- (A3') If $A_1, \dots, A_M \in \mathscr{T}$ for some $M \in \mathbb{N}$, then

$$\bigcap_{n=1}^{M} A_n \in \mathscr{T} .$$

The equivalence between (A3) and (A3') can be immediately obtained by induction.

Warning

Notice:

- The union property (A2) of $\mathcal{T}$ holds for an **arbitrary** number of sets, even uncountable!
- The intersection property (A3') of $\mathcal{T}$ holds only for a **finite** number of sets.

There are two main examples of topologies that one should always keep in mind. These are:

- **Trivial topology**: The topology with the smallest possible number of sets.
- **Discrete topology**: The topology with the highest possible number of sets.

Definition 3.3: Trivial topology

Let X be a set. The trivial topology on X is the topology $\mathcal{T}$ defined by

$$\mathcal{T} := \{\emptyset, X\}.$$

Let us check that $\mathcal{T}$ is indeed a topology. We need to verify the 3 properties of a topology:

- (A1) We clearly have $\emptyset, X \in \mathcal{T}$.
- (A2) The only non-trivial union to check is the one between $\emptyset$ and X. We have

$$\emptyset \cup X = X \in \mathcal{T}.$$

- (A3) The only non-trivial intersection to check is the one between $\emptyset$ and X. We have

$$\emptyset \cap X = \emptyset \in \mathcal{T}.$$

Therefore $\mathcal{T}$ is a topology on X.

Definition 3.4: Discrete topology

Let X be a set. The discrete topology on X is the topology $\mathcal{T}$ defined by

$$\mathcal{T} := \{A : A \subseteq X\},$$

that is, every subset of X is open.

Let us check that $\mathcal{T}$ is a topology:

- (A1) We have $\emptyset, X \in \mathcal{T}$, since $\emptyset$ and X are subsets of X.
- (A2) The arbitrary union of subsets of X is still a subset of X. Therefore

$$\bigcup_{i \in I} A_i \in \mathcal{T},$$

whenever $A_i \in \mathcal{T}$ for all $i \in I$.

- (A3) The intersection of two subsets of X is still a subset of X. Therefore

$$A \cap B \in \mathcal{T} ,$$

whenever $A, B \in \mathcal{T}$.

Therefore $\mathcal{T}$ is a topology on X.

We anticipated that topology is the extension of familiar concepts of open set, continuity, etc. that we have in $\mathbb{R}^n$. Let us see how the usual definition of open set of $\mathbb{R}^n$ can fit in our new abstract framework of topology.

Definition 3.5: Open set of $\mathbb{R}^n$

Let $A \subseteq \mathbb{R}^n$. We say that the set A is **open** if it holds:

$$\forall \, \mathbf{x} \in A \,, \; \exists \, r > 0 \; \text{ s.t. } \; B_r(\mathbf{x}) \subseteq A \,, \tag{3.1}$$

where $B_r(\mathbf{x})$ is the ball of radius $r > 0$ centered at $\mathbf{x}$

$$B_r(\mathbf{x}) := \{ \mathbf{y} \in \mathbb{R}^n \; : \; \|\mathbf{y} - \mathbf{x}\| < r \} ,$$

and the **Euclidean norm** of $\mathbf{x} \in \mathbb{R}^n$ is defined by

$$\|\mathbf{x}\| := \sqrt{\sum_{i=1}^n x_i^2} \, .$$

See Figure 3.1 for a schematic picture of an open set.

Definition 3.6: Euclidean topology of $\mathbb{R}^n$

The Euclidean topology on $\mathbb{R}^n$ is the topology $\mathcal{T}$ defined by

$$\mathcal{T} := \{ A \; : \; A \subseteq \mathbb{R}^n \,, \; A \text{ is open} \} .$$

We need to check that the above definition is well-posed, in the sense that we have to prove that $\mathcal{T}$ is a topology on $\mathbb{R}^n$.

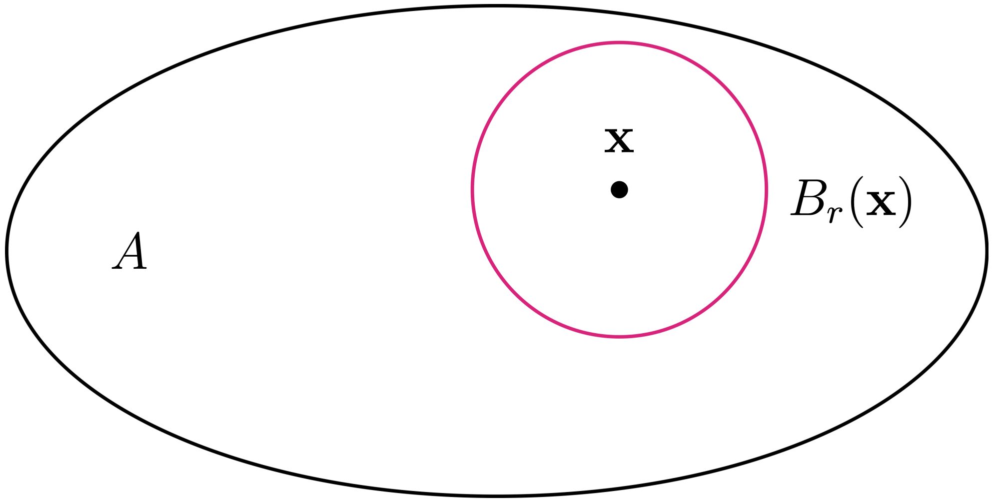


Figure 3.1: The set $A \subseteq \mathbb{R}^n$ is open if for every $\mathbf{x} \in A$ there exists $r > 0$ such that $B_r(\mathbf{x}) \subseteq A$.

Proof: Well-posedness of Definition 4.6

Let us check that $\mathscr{T}$ is a topology on $\mathbb{R}^n$:

- (A1) We have $\emptyset, \mathbb{R}^n \in \mathscr{T}$: Indeed $\emptyset$ is open because there is no point $\mathbf{x}$ for which (3.1) needs to be checked. Moreover $\mathbb{R}^n$ is open because (3.1) holds with any radius $r > 0$.

- (A2) Let $A_i \in \mathscr{T}$ for all $i \in I$ and define the union set

$$A := \bigcup_{i \in I} A_i \,.$$

  We need to check that A is open. Let $\mathbf{x} \in A$. By definition of union, there exists an index $i_0 \in I$ such that $\mathbf{x} \in A_{i_0}$. Since A_{i_0} is open, by (3.1) there exists $r > 0$ such that $B_r(\mathbf{x}) \subseteq A_{i_0}$. As $A_{i_0} \subseteq A$, we conclude that $B_r(\mathbf{x}) \subseteq A$. Thus A is open and $A \in \mathscr{T}$.

- (A3) Let $A, B \in \mathscr{T}$. We need to check that $A \cap B$ is open. Let $\mathbf{x} \in A \cap B$. Therefore $\mathbf{x} \in A$ and $\mathbf{x} \in B$. Since A and B are open, by (3.1) there exist $r_1, r_2 > 0$ such that $B_{r_1}(\mathbf{x}) \subseteq A$ and $B_{r_2}(\mathbf{x}) \subseteq B$. Set $r := \min\{r_1, r_2\}$. Then

$$B_r(\mathbf{x}) \subseteq B_{r_1}(\mathbf{x}) \subseteq A \,, \quad B_r(\mathbf{x}) \subseteq B_{r_2}(\mathbf{x}) \subseteq B \,,$$

  Hence $B_r(\mathbf{x}) \subseteq A \cap B$, showing that $A \cap B$ open, so that $A \cap B \in \mathscr{T}$.

This proves that $\mathscr{T}$ is a topology on $\mathbb{R}^n$.

Let us make a basic bus useful observation: balls in $\mathbb{R}^n$ are open for the Euclidean topology.

Proposition 3.7

Let $\mathbb{R}^n$ be equipped with $\mathcal{T}$ the Euclidean topology. Let $r > 0$ and $\mathbf{x} \in \mathbb{R}^n$. Then

$$B_r(\mathbf{x}) \in \mathcal{T} .$$

Proof

We need to shown that $B_r(\mathbf{x})$ satisfies (3.1). Therefore, let $\mathbf{y} \in B_r(\mathbf{x})$. In particular

$$\|\mathbf{x} - \mathbf{y}\| < r . \tag{3.2}$$

Define

$$\varepsilon := r - \|\mathbf{x} - \mathbf{y}\| .$$

Note that $\varepsilon > 0$ by (3.2). We claim that

$$B_\varepsilon(\mathbf{y}) \subseteq B_r(\mathbf{x}) , \tag{3.3}$$

see Figure 3.2. Indeed, let $\mathbf{z} \in B_\varepsilon(\mathbf{y})$. By triangle inequality we have

$$\|\mathbf{z} - \mathbf{x}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y} - \mathbf{z}\| < \|\mathbf{x} - \mathbf{y}\| + \varepsilon = r ,$$

where we used that $\|\mathbf{y} - \mathbf{z}\| < \varepsilon$ and the definition of ε. Hence $\mathbf{z} \in B_r(\mathbf{x})$, proving (3.3). This proves that $B_r(\mathbf{x})$ satisfies (3.1), and is therefore open.

3.1 Closed sets

The opposite of open sets are closed sets.

Definition 3.8: Closed set

Let $(X, \mathcal{T})$ be a topological space. A set $C \subseteq X$ is **closed** if

$$C^c \in \mathcal{T} ,$$

where $C^c := X \setminus C$ is the complement of C in X.

In words, a set is closed if its complement is open.

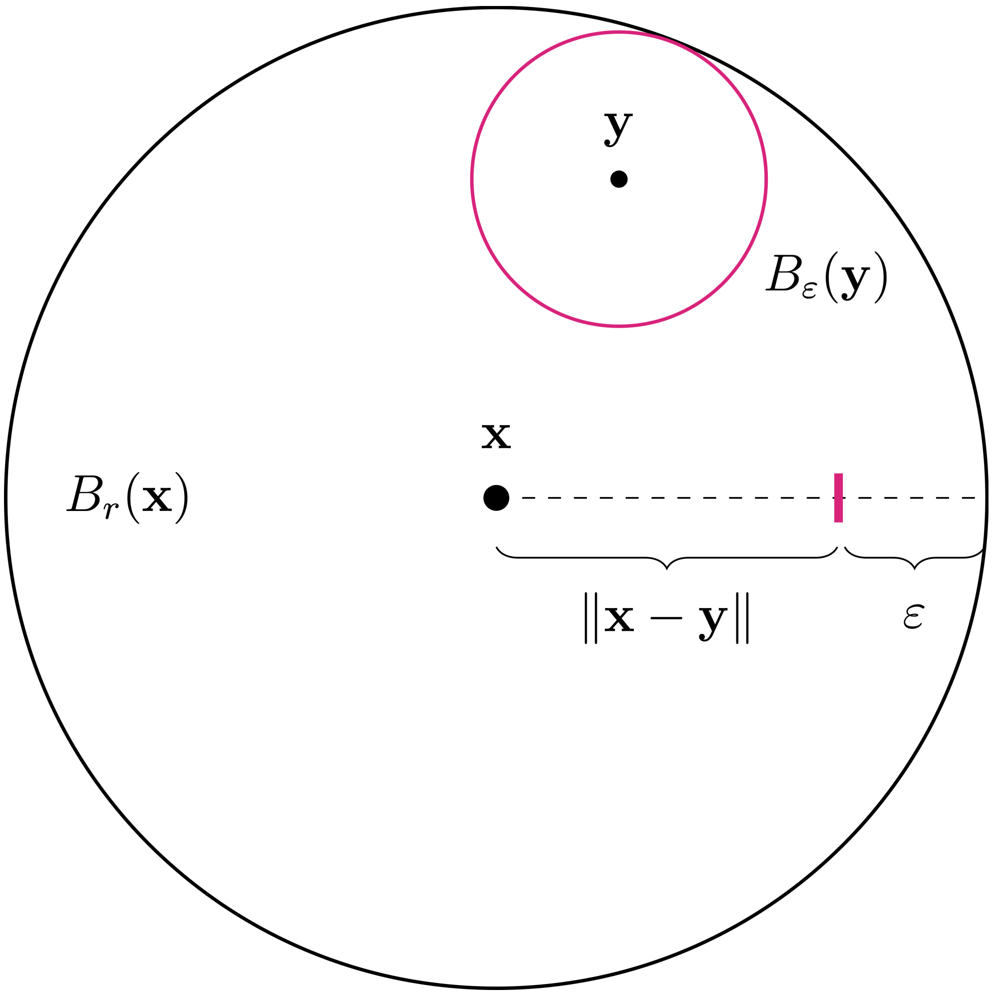


Figure 3.2: The ball $B_\varepsilon(\mathbf{y})$ is contained in $B_r(\mathbf{x})$ if $\varepsilon := r - \|\mathbf{x}-\mathbf{y}\|$.

Warning

There are sets which are neither open nor closed. For example consider $\mathbb{R}$ equipped with Euclidean topology. Then the interval

$$A := [0, 1)$$

is neither open nor closed.

For the moment we do not have the tools to prove this. We will have them shortly.

We could have defined a topology starting from closed sets. We would have had to replace the properties (A1)-(A2)-(A3) with suitable properties for closed sets. Such properties are detailed in the following proposition.

Proposition 3.9

Let $(X, \mathcal{T})$ be a topological space. Properties (A1)-(A2)-(A3) of $\mathcal{T}$ are equivalent to (C1)-(C2)-(C3), where

- (C1) $\emptyset, X$ are closed.
- (C2) If C_i is closed for all $i \in I$, then
$$\bigcap_{i \in I} C_i$$
is closed.
- (C3) If C_1, C_2 are closed then
$$C_1 \cup C_2$$
is closed.

Proof

We have 3 points to check:

- The equivalence between (A1) and (C1) is clear, since
$$\emptyset^c = X \,, \quad X^c = \emptyset \,.$$
- Suppose C_i are closed for all $i \in I$. Therefore C_i^c are open for all $i \in I$. By De Morgan's laws we have that
$$\left(\bigcap_{i \in I} C_i \right)^c = \bigcup_{i \in I} C_i^c$$
showing that
$$\bigcap_{i \in I} C_i \text{ is closed} \quad \iff \quad \bigcup_{i \in I} C_i^c \text{ is open} \,.$$
Therefore (A2) and (C2) are equivalent.
- Suppose C_1, C_2 are closed. Therefore C_1^c, C_2^c are open. By De Morgan's laws we have that
$$(C_1 \cup C_2)^c = C_1^c \cap C_2^c$$

showing that

$$C_1 \cup C_2 \ \text{ is closed} \qquad \Longleftrightarrow \qquad C_1^c \cap C_2^c \ \text{ is open}\,.$$

Therefore (A3) and (C3) are equivalent.

As a consequence of the above proposition, we can define a topology by declaring what the closed sets are. We then need to verify that (C1)-(C2)-(C3) are satisfied by such topology. Let us make an example.

Example 3.10: The Zariski topology

Let $(\mathbb{K}, +, \cdot)$ be a field. Define

$$X := \mathbb{K}^n := \{(a_1, \dots, a_n) \, : \, a_i \in \mathbb{K}\}\,.$$

Consider the set of polynomials with coefficients in the field

$$\mathbb{K}[x_1, \dots, x_n]\,.$$

Therefore $f \in \mathbb{K}[x_1, \dots, x_n]$ has the form

$$f(x_1, \dots, x_n) = \lambda_1 x_1 + \dots + \lambda_n x_n\,,$$

where $\lambda_1, \dots, \lambda_n$ are given elements of $\mathbb{K}$. For $I \subset \mathbb{K}[x_1, \dots, x_n]$ define

$$V(I) := \{(a_1, \dots, a_n) \in \mathbb{K}^n \, : \, f(a_1, \dots, a_n) = 0\,, \ \forall \, f \in I\}\,.$$

Define

$$\mathscr{C} := \{V(I) \, : \, I \subset \mathbb{K}[x_1, \dots, x_n]\}\,.$$

Then $\mathscr{C}$ satisfies (C1), (C2) and (C3). This is an easy check, and is left as exercise. $\mathscr{C}$ is called the **Zariski Topology** on the field $\mathbb{K}^n$. This is used in algebraic geometry to study Affine Varieties, an algebraic version of surfaces, see Wikipedia page.

3.2 Comparing topologies

Consider the situation where you have two topologies $\mathscr{T}_1$ and $\mathscr{T}_2$ on the same set X. We would like to have some notions of comparison between $\mathscr{T}_1$ and $\mathscr{T}_2$.

Definition 3.11: Finer and coarser topology

Let X be a set and let $\mathcal{T}_1, \mathcal{T}_2$ be topologies on X. Suppose that

$$\mathcal{T}_2 \subseteq \mathcal{T}_1 .$$

We say that:

- $\mathcal{T}_1$ is **finer** than $\mathcal{T}_2$.
- $\mathcal{T}_2$ is **coarser** than $\mathcal{T}_1$.

If it holds

$$\mathcal{T}_2 \subsetneq \mathcal{T}_1 ,$$

we say that:

- $\mathcal{T}_1$ is **strictly finer** than $\mathcal{T}_2$.
- $\mathcal{T}_2$ is **strictly coarser** than $\mathcal{T}_1$.

We say that $\mathcal{T}_1$ and $\mathcal{T}_2$ are the **same** topology if

$$\mathcal{T}_1 = \mathcal{T}_2 .$$

Example 3.12

Let X be a set and consider the trivial and discrete topologies

$$\mathcal{T}_{\text{trivial}} = \{\emptyset, X\} , \quad \mathcal{T}_{\text{discrete}} = \{A \, : \, A \subseteq X\} .$$

Then

$$\mathcal{T}_{\text{trivial}} \subsetneq \mathcal{T}_{\text{discrete}} ,$$

so that $\mathcal{T}_{\text{discrete}}$ is strictly finer than $\mathcal{T}_{\text{trivial}}$.

Another interesting example is given by the **cofinite topology** on $\mathbb{R}$. The sets in this topology are open if they are either empty, or coincide with $\mathbb{R}$ with a finite number of points removed.

Example 3.13: Cofinite topology on $\mathbb{R}$

Consider the following family $\mathcal{T}_{\text{cofinite}}$ of subsets of $\mathbb{R}$

$$\mathcal{T}_{\text{cofinite}} := \{U \subseteq \mathbb{R} \, : \, U^c \text{ is finite, or } U^c = \mathbb{R}\} .$$

Then $(\mathbb{R}, \mathcal{T}_{\text{cofinite}})$ is a topological space, and $\mathcal{T}_{\text{cofinite}}$ is called the **cofinite topology**. We have that

$$\mathcal{T}_{\text{cofinite}} \subsetneq \mathcal{T}_{\text{euclidean}} .$$

Exercise: Show that $\mathcal{T}_{\text{cofinite}}$ is a topology on $\mathbb{R}$ and that $\mathcal{T}_{\text{cofinite}} \subsetneq \mathcal{T}_{\text{euclidean}}$.

3.3 Convergence

We have generalized the notion of open set to arbitrary sets. Next we generalize the notion of convergence of sequences.

Definition 3.14: Convergent sequence

Let $(X, \mathscr{T})$ be a topological. Consider a sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq X$ and a point $x \in X$. We say that x_n converges to x_0 if the following property holds:

$$\forall U \in \mathscr{T} \text{ s.t. } x_0 \in U\,, \ \exists N = N(U) \in \mathbb{N} \text{ s.t. } x_n \in U\,, \forall n \geq N\,. \tag{3.4}$$

Notation

The convergence of x_n to x_0 is denoted by

$$x_n \to x_0 \quad \text{ or } \quad \lim_{n\to\infty} x_n = x_0\,.$$

Let us analyze the definition of convergence in the topologies we have encountered so far. We will have that:

- **Trivial topology**: Every sequence converges to every point.
- **Discrete topology**: A sequence converges if and only if it is eventually constant.
- **Euclidean topology**: Topological convergence coincides with classical notion of convergence.

We now precisely state and prove the above claims.

Proposition 3.15: Convergence for trivial topology

Let $(X, \mathscr{T})$ be topological space, with $\mathscr{T}$ the trivial topology, that is,

$$\mathscr{T} = \{\emptyset, X\}\,.$$

Let $\{x_n\} \subseteq X$ be a sequence and $x_0 \in X$ a point. Then

$$x_n \to x_0\,.$$

Proof

To show that $x_n \to x_0$ we need to check that (3.4) holds. Therefore, let $U \in \mathcal{T}$ with $x_0 \in U$. We have two cases:

- $U = \emptyset$: This case is not possible, since x_0 cannot be in U.
- $U = X$: Take $N = 1$. Since U is the whole space, then $x_n \in U$ for all $n \geq 1$.

As these are all the open sets, we conclude that $x_n \to x_0$.

Warning

This example is saying that in general the topological limit of a sequence is **not unique**!

Proposition 3.16: Convergence for discrete topology

Let $(X, \mathcal{T})$ be topological space, with $\mathcal{T}$ the discrete topology, that is,

$$\mathcal{T} = \{A \, : \, A \subseteq X\} .$$

Let $\{x_n\} \subseteq X$ be a sequence and $x_0 \in X$ a point. They are equivalent:

1. $x_n \to x_0$.
2. $\{x_n\}$ is eventually constant, that is, there exists $N \in \mathbb{N}$ such that

$$x_n = x_0 \, , \quad \forall \, n \geq N \, .$$

Proof

Part 1. Assume that $x_n \to x_0$.
We have to prove that $\{x_n\}$ is eventually constant. To this end, let

$$U = \{x_0\} .$$

Then $U \in \mathcal{T}$. Since $x_n \to x_0$, by (3.4) there exists $N \in \mathbb{N}$ such that

$$x_n \in U \, , \quad \forall \, n \geq N \, .$$

As $U = \{x_0\}$, the above is saying that $x_n = x_0$ for all $n \geq N$. Hence x_n is eventually constant.
Part 2. Assume that x_n is eventually equal to x_0.
By assumption there exists $N \in \mathbb{N}$ such that

$$x_n = x_0 \, , \quad \forall \, n \geq N \, . \tag{3.5}$$

Let $U \in \mathcal{T}$ be an open set such that $x_0 \in U$. By (3.5) we have that

$$x_n \in U\,, \quad \forall\, n \geq N\,.$$

Since U was arbitrary, we conclude that $x_n \to x_0$.

Before proceeding to examining convergence in the Euclidean topology, let us recall the classical definition of convergence in $\mathbb{R}^n$.

Definition 3.17: Classical convergence in $\mathbb{R}^n$

Let $\{\mathbf{x}_n\} \subseteq \mathbb{R}^n$ and $\mathbf{x}_0 \in \mathbb{R}^n$. We say that $\mathbf{x}_n$ converges $\mathbf{x}_0$ in the classical sense if

$$\lim_{n \to \infty} \|\mathbf{x}_n - \mathbf{x}_0\| = 0\,.$$

The above is equivalent to: For all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$\|\mathbf{x}_n - \mathbf{x}_0\| < \varepsilon\,, \quad \forall\, n \geq N\,.$$

Proposition 3.18: Convergence for Euclidean topology

Let $\mathbb{R}^n$ be equipped with $\mathcal{T}$ the Euclidean topology. Let $\{\mathbf{x}_n\} \subseteq \mathbb{R}^n$ be a sequence and $\mathbf{x}_0 \in \mathbb{R}^n$ a point. They are equivalent:

1. $\mathbf{x}_n \to \mathbf{x}_0$ with respect to $\mathcal{T}$.
2. $\mathbf{x}_n \to \mathbf{x}_0$ in the classical sense.

Proof

Part 1. Assume $\mathbf{x}_n \to \mathbf{x}_0$ with respect to $\mathcal{T}$.
Fix $\varepsilon > 0$ and consider the set

$$U := B_\varepsilon(\mathbf{x}_0)\,.$$

By Proposition 3.7 we know that $U \in \mathcal{T}$. Moreover $\mathbf{x}_0 \in U$. By the convergence $\mathbf{x}_n \to \mathbf{x}_0$ with respect to $\mathcal{T}$, there exists $N \in \mathbb{N}$ such that

$$\mathbf{x}_n \in U\,, \quad \forall\, n \geq N\,.$$

As $U = B_\varepsilon(\mathbf{x}_0)$, the above reads

$$\|\mathbf{x}_n - \mathbf{x}_0\| < \varepsilon\,, \quad \forall\, n \geq N\,,$$

showing that $\mathbf{x}_n \to \mathbf{x}_0$ in the classical sense.
Part 2. Assume $\mathbf{x}_n \to \mathbf{x}_0$ in the classical sense.

Let $U \in \mathcal{T}$ be such that $\mathbf{x}_0 \in U$. By definition of Euclidean topology, this means that there exists $r > 0$ such that

$$B_r(\mathbf{x}_0) \subseteq U \,.$$

As $\mathbf{x}_n \to \mathbf{x}_0$ in the classical sense, there exists $N \in \mathbb{N}$ such that

$$\|\mathbf{x}_n - \mathbf{x}_0\| < r \,, \quad \forall \, n \geq N \,.$$

The above is equivalent to

$$\mathbf{x}_n \in B_r(\mathbf{x}_0) \,, \quad \forall \, n \geq N \,.$$

Since $B_r(\mathbf{x}_0) \subseteq U$, we have proven that

$$\mathbf{x}_n \in U \,, \quad \forall \, n \geq N \,.$$

Since U is arbitrary, we conclude that $\mathbf{x}_n \to \mathbf{x}_0$ with respect to $\mathcal{T}$.

Notation

Since classical convergence in $\mathbb{R}^n$ agrees with topological convergence with respect to $\mathcal{T}$, we will just say that $\mathbf{x}_n \to \mathbf{x}_0$ in $\mathbb{R}^n$ without ambiguity.

We conclude with a useful proposition which relates convergences when multiple topologies are present.

Proposition 3.19

Let X be a set and $\mathcal{T}_1, \mathcal{T}_2$ be topologies on X. Suppose that

$$\mathcal{T}_2 \subseteq \mathcal{T}_1 \,.$$

Let $\{x_n\} \subset X$ and $x_0 \in X$. We have

$$x_n \to x_0 \;\text{ in }\; \mathcal{T}_1 \quad \implies \quad x_n \to x_0 \;\text{ in }\; \mathcal{T}_2 \,.$$

Proof

Assume $x_n \to x_0$ in $\mathcal{T}_1$. We need to prove that $x_n \to x_0$ in $\mathcal{T}_2$. Therefore, let $U \in \mathcal{T}_2$ be such that $x_0 \in U$. Since $\mathcal{T}_2 \subseteq \mathcal{T}_1$, we have that $U \in \mathcal{T}_1$. As $x_n \to x_0$ in $\mathcal{T}_1$, there exists $N \in \mathbb{N}$ such that

$$x_n \in U \,, \quad \forall \, n \geq N \,.$$

Since $U \in \mathcal{T}_2$, the above proves $x_n \to x_0$ in $\mathcal{T}_2$.

3.4 Metric spaces

We will now define a class of topological spaces known as metric spaces.

Definition 3.20: Distance

Let X be a set. A **distance** on X is a function

$$d : X \times X \to \mathbb{R}$$

such that, for all $x, y, z \in X$ they hold:

- (M1) Positivity: The distance is non-negative

$$d(x, y) \geq 0 .$$

  Moreover

$$d(x, y) = 0 \quad \iff \quad x = y .$$

- (M2) Symmetry: The distance is symmetric

$$d(x, y) = d(y, x) .$$

- (M3) Triangle Inequality: It holds

$$d(x, z) \leq d(x, y) + d(y, z) .$$

Definition 3.21: Metric space

Let X be a set and $d : X \times X \to \mathbb{R}$ be a distance on X. We say that the pair (X, d) is a **metric space**.

Example 3.22: $\mathbb{R}^n$ as metric space

The Euclidean norm naturally induces a distance over $\mathbb{R}^n$ by setting

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| .$$

Then $(\mathbb{R}^n, d)$ is a metric space.

It is trivial to check that the Euclidean distance satisfies (M1) and (M2). To show (M3), recalling the triangle inequality in $\mathbb{R}^n$:

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| ,$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Using the above we obtain

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\| \\ &= \|(\mathbf{x} - \mathbf{z}) + (\mathbf{z} - \mathbf{y})\| \\ &\leq \|\mathbf{x} - \mathbf{z}\| + \|\mathbf{z} - \mathbf{y}\| \\ &= d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}) \,, \end{aligned}$$

proving that d satisfies (M3). This prove that $(\mathbb{R}^n, d)$ is a metric space.

Example 3.23: p-distance on $\mathbb{R}^n$

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $p \in [1, \infty)$ define

$$d_p(\mathbf{x}, \mathbf{y}) := \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}} .$$

Note that d_2 coincides with the Euclidean distance. For $p = \infty$ we set

$$d_\infty(\mathbf{x}, \mathbf{y}) := \max_{i=1\dots,n} |x_i - y_i| \,.$$

We have that $(\mathbb{R}^n, d_p)$ is a metric space.

Indeed properties (M1)-(M2) hold trivially. The triangle inequality is also trivially satisfied by d_∞. We are left with checking the triangle inequality for d_p with $p \geq 1$. To this end, define

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} .$$

Minkowski's inequality, see Wikipedia page, states that

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p \,,$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Therefore

$$\begin{aligned} d_p(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\|_p \\ &= \|(\mathbf{x} - \mathbf{z}) + (\mathbf{z} - \mathbf{y})\|_p \\ &\leq \|\mathbf{x} - \mathbf{z}\|_p + \|\mathbf{z} - \mathbf{y}\|_p \\ &= d_p(\mathbf{x}, \mathbf{z}) + d_p(\mathbf{z}, \mathbf{y}) \,, \end{aligned}$$

proving that d_p satisfies (M3). Hence $(\mathbb{R}^n, d_p)$ is a metric space.

A metric d on a set X naturally induces a topology which is **compatible** with the metric.

Definition 3.24: Topology induced by the metric

Let (X,d) be a metric space. We define the topology $\mathcal{T}_d$ **induced by the metric** d as the collection of sets $U \subseteq X$ that satisfy the following property:

$$\forall\, x \in U\,,\ \exists\, r \in \mathbb{R},\ r > 0 \ \text{ s.t. } \ B_r(x) \subseteq U\,,$$

where $B_r(x)$ is the ball centered at x of radius r. This is defined by

$$B_r(x) := \{y \in X \ : \ d(x,y) < r\}\,.$$

We need to check that the above definition is well-posed, that is, we need to show that $\mathcal{T}_d$ is actually a topology on X. The proof follows, line by line, the proof that the Euclidean topology is indeed a topology, see proof immediately below Definition 4.6. This is left as an exercise.

Example 3.25: Topology induced by Euclidean distance

Consider the metric space $(\mathbb{R}^n, d)$ with d the Euclidean distance. Then

$$\mathcal{T}_d = \mathcal{T}_{\text{euclidean}}\,,$$

where $\mathcal{T}_{\text{euclidean}}$ is the Euclidean topology on $\mathbb{R}^n$.

Exercise: Prove the above statement. It is an immediate consequence of definitions.

Example 3.26: Discrete distance

Let X be a set. Define the function $d : \ X \times X \to \mathbb{R}$ by

$$d(x,y) := \begin{cases} 0 & \text{if } \ x = y \\ 1 & \text{if } \ x \neq y \end{cases}$$

Then (X,d) is a metric space, and d is called the **discrete distance**. Moreover

$$\mathcal{T}_d = \mathcal{T}_{\text{discrete}}$$

where $\mathcal{T}_{\text{discrete}}$ is the **discrete topology** on X.

Exercise: Prove that (X,d) is a metric space and $\mathcal{T}_d = \mathcal{T}_{\text{discrete}}$.

The following proposition tells us that balls in a metric space X are open sets. Moreover balls are the building blocks of all open sets in X. The proof is left as an exercise.

Proposition 3.27

Let (X,d) be a metric space and $\mathcal{T}_d$ the topology induced by d. Then:

- For all $x \in X$, $r > 0$ we have $B_r(x) \subseteq \mathcal{T}_d$.
- $U \in \mathcal{T}_d$ if and only if
$$U = \bigcup_{i \in I} B_{r_i}(x_i) \,,$$
with I family of indices and $x_i \in X$, $r_i > 0$.

We now define the concept of equivalent metrics.

Definition 3.28: Equivalent metrics

Let X be a set and d_1, d_2 be metrics on X. We say that d_1 and d_2 are equivalent if

$$\mathcal{T}_{d_1} = \mathcal{T}_{d_2} \,.$$

The following proposition gives a practical way to check if two metrics are equivalent.

Proposition 3.29

Let X be a set and d_1, d_2 be metrics on X. They are equivalent:

1. d_1 and d_2 are equivalent metrics.
2. There exists a constant $\alpha > 0$ such that
$$\frac{1}{\alpha}\, d_2(x,y) \le d_1(x,y) \le \alpha\, d_2(x,y) \,, \quad \forall\, x,y \in X \,.$$

The proof of Proposition 3.29 is trivial, and is left as an exercise.

Example 3.30

Let $p > 1$. The metrics d_p and d_∞ on $\mathbb{R}^n$ are equivalent.

This follows from Proposition 3.29 and the estimate

$$d_\infty(\mathbf{x},\mathbf{y}) \le d_p(\mathbf{x},\mathbf{y}) \le n\, d_\infty(\mathbf{x},\mathbf{y}) \,, \quad \forall \mathbf{x} \,,\ \mathbf{y} \in \mathbb{R}^n \,.$$

Warning

If two metrics are equivalent, that does not mean they have the same balls. For example the balls of the metrics d_1, d_2 and d_∞ on $\mathbb{R}^n$ look very different, see Figure 3.3.

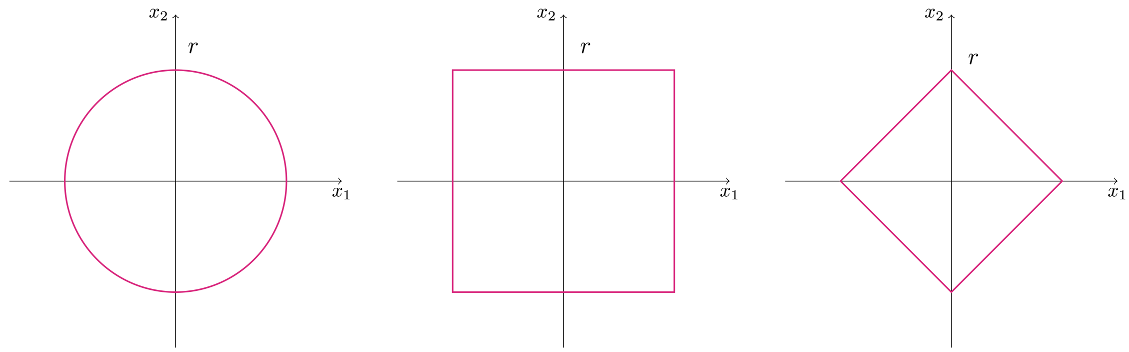


Figure 3.3: Balls $B_r(0)$ for the metrics d_2, d_∞, d_1 in $\mathbb{R}^2$.

We can characterize the convergence of sequences in metric spaces.

Proposition 3.31: Convergence in metric space

Suppose (X, d) is a metric space and denote by $\mathscr{T}_d$ the topology induce by d. Let $\{x_n\} \subseteq X$ and $x_0 \in X$. They are equivalent:

1. $x_n \to x_0$ with respect to the topology $\mathscr{T}_d$.
2. $d(x_n, x_0) \to 0$ in $\mathbb{R}$.
3. For all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$x_n \in B_r(x_0)\,, \ \ \forall\, n \geq \mathbb{N}\,.$$

The proof is similar to the one of Proposition 3.18, and it is left as an exercise.

3.5 Interior, closure and boundary

We now define interior, closure and boundary of a set A contained in a topological space.

Definition 3.32: Interior of a set

Let $(X, \mathscr{T})$ be a topological space and $A \subseteq X$ a set. The **interior** of A is the set

$$\operatorname{Int} A := \bigcup_{\substack{U \subseteq A \\ U \in \mathscr{T}}} U\,.$$

Remark 3.33

The definition of $\operatorname{Int} A$ is well-posed, since $\emptyset \subseteq A$ and $\emptyset \in \mathcal{T}$. Therefore the union is taken over a non-empty family.

Proposition 3.34

Let $(X, \mathcal{T})$ be a topological space and $A \subseteq X$ a set. Then $\operatorname{Int} A$ is the largest open set contained in A, that is:

1. $\operatorname{Int} A$ is open.
2. $\operatorname{Int} A \subseteq A$.
3. If $V \in \mathcal{T}$ and $V \subseteq A$, then $V \subseteq \operatorname{Int} A$.
4. A is open if and only if
$$A = \operatorname{Int} A\,.$$

Proof

We have:

1. $\operatorname{Int} A$ is open, since it is union of open sets, see property (A2).
2. $\operatorname{Int} A \subseteq A$, since $\operatorname{Int} A$ is union of sets contained in A.
3. Suppose $V \in \mathcal{T}$ and $V \subseteq A$. Therefore
$$V \subseteq \bigcup_{\substack{U \subseteq A \\ U \in \mathcal{T}}} U = \operatorname{Int} A\,.$$
4. Suppose that A is open. Then
$$A \subseteq \bigcup_{\substack{U \subseteq A \\ U \in \mathcal{T}}} U = \operatorname{Int} A\,.$$
As we already know that $\operatorname{Int} A \subseteq A$, we conclude that $A = \operatorname{Int} A$.
Conversely, suppose that $A = \operatorname{Int} A$. Since $\operatorname{Int} A$ is open, then also A is open.

Definition 3.35: Closure of a set

Let $(X, \mathcal{T})$ be a topological space and $A \subseteq X$ a set. The **closure** of A is the set
$$\overline{A} := \bigcap_{\substack{A \subseteq C \\ C \text{ closed}}} C\,,$$

that is, $\overline{A}$ is the intersection of all closed sets containing A.

Remark 3.36

The definition of $\overline{A}$ is well-posed, since $A \subseteq X$, and X is closed. Therefore the intersection is taken over a non-empty family.

Proposition 3.37

Let $(X, \mathscr{T})$ be a topological space and $A \subseteq X$ a set. Then $\overline{A}$ is the smallest closed set containing A, that is:

1. $\overline{A}$ is closed.
2. $A \subseteq \overline{A}$.
3. If V is closed $A \subseteq V$, then $\overline{A} \subseteq V$.
4. A is closed if and only if
$$A = \overline{A} .$$

Proof

We have:

1. $\overline{A}$ is closed, since it is intersection of closed sets, see property (C2).
2. $A \subseteq \overline{A}$, since $\overline{A}$ is intersection of sets which contain A.
3. Suppose V is closed and $A \subseteq V$. Therefore
$$\overline{A} = \bigcap_{\substack{A \subseteq C \\ C \text{ closed}}} C \subseteq V .$$
4. Suppose that A is closed. Then
$$\overline{A} = \bigcap_{\substack{A \subseteq C \\ C \text{ closed}}} C \subseteq A ,$$
showing that $\overline{A} \subseteq A$. As we already know that $A \subseteq \overline{A}$, we conclude that $A = \overline{A}$. Conversely, suppose that $A = \overline{A}$. Since $\overline{A}$ is closed, then also A is closed.

Lemma 3.38

Let $(X, \mathscr{T})$ be a topological space and $A \subseteq X$ a set. They are equivalent:

1. $x_0 \in \overline{A}$.
2. For every $U \in \mathscr{T}$ such that $x_0 \in U$, it holds

$$U \cap A \neq \emptyset .$$

Proof

We prove the contronominal statement:

$$x_0 \notin \overline{A} \qquad \Longleftrightarrow \qquad \exists\, U \in \mathscr{T} \text{ s.t. } x_0 \in U \,, \; U \cap A = \emptyset \,.$$

Let us check the two implications hold:

- Suppose $x_0 \notin \overline{A}$. Then $x_0 \in U := (\overline{A})^c$. Note that U is open, since $U^c = \overline{A}$ is closed. We have

$$A \cap U = A \cap (\overline{A})^c = \emptyset \,,$$

  since $A \subseteq \overline{A}$.
- Assume there exists $U \in \mathscr{T}$ such that $x_0 \in U$ and $U \cap A = \emptyset$. Therefore $A \subseteq U^c$. Since U is open, U^c is closed. Then

$$\overline{A} = \bigcap_{\substack{A \subseteq C \\ C \text{ closed}}} C \subseteq U^c \,.$$

  Since $x_0 \notin U^c$, we conclude that $x_0 \notin \overline{A}$.

Definition 3.39: Boundary of a set

Let $(X, \mathscr{T})$ be a topological space and $A \subseteq X$ a set. The **boundary** of A is the set

$$\partial A := \overline{A} \setminus \operatorname{Int} A \,.$$

Proposition 3.40

Let $(X, \mathscr{T})$ be a topological space and $A \subseteq X$ a set. Then ∂A is closed.

Proof

We can write

$$\partial A = \overline{A} \setminus \operatorname{Int} A = \overline{A} \cap (\operatorname{Int} A)^c\,.$$

Note that $\overline{A}$ is closed and $(\operatorname{Int} A)^c$ is closed, since $\operatorname{Int} A$ is open. Then ∂A is intersection of two closed sets, and in hence closed by (C2).

We can characterize $\overline{A}$ as the set of limit points of sequences in A.

Definition 3.41

Let $(X, \mathcal{T})$ be a topological space and $A \subseteq X$. The set of limit points of A is defined as

$$L(A) := \{x \in X \,:\, \exists \{x_n\} \subseteq A \text{ s.t. } x_n \to x\}\,.$$

Proposition 3.42

Let $(X, \mathcal{T})$ be a topological space and $A \subseteq X$ a set. Let $\{x_n\} \subseteq A$ and $x_0 \in X$ be such that $x_n \to x_0$. Then $x_0 \in \overline{A}$. Therefore

$$L(A) \subseteq \overline{A}\,.$$

Proof

Suppose by contradiction $x_0 \notin \overline{A}$, so that

$$x_0 \in (\overline{A})^c\,.$$

Since $(\overline{A})^c$ is open and $x_n \to x_0$, there exists $N \in \mathbb{N}$ such that

$$x_n \in (\overline{A})^c\,, \quad \forall\, n \geq N\,.$$

This is a contradiction, since we were assuming that $\{x_n\} \subseteq A$. This shows $x_0 \in \overline{A}$ and therefore $L(A) \subseteq \overline{A}$.

Warning

The converse of Proposition 3.42 is false in general, that is,

$$\overline{A} \not\subseteq L(A)\,.$$

We show a counterexample of the above in Example 3.43. The above relation holds in the so-called first countable topological spaces, such as metric spaces, see Proposition 3.44 below.

Example 3.43: Co-countable topology

Let $X = \mathbb{R}$ with the co-countable topology

$$\mathcal{T} := \{A \subseteq \mathbb{R} \, : \, A^c = \mathbb{R} \, \text{ or } \, A^c \, \text{ countable} \} \, .$$

The set

$$A = (-\infty, 0]$$

is not closed and $\overline{A} = \mathbb{R}$. Moreover, convergent sequences in $(X, \mathcal{T})$ are eventually constant. Therefore $L(A) = A$, showing that $\overline{A} \not\subset L(A)$.

Exercise: Prove all the above statements.

In metric spaces we can characterize the interior of a set and the closure in the following way.

Proposition 3.44

Let (X, d) be a metric space. Denote by $\mathcal{T}_d$ the topology induced by d. Let $A \subseteq X$. We have

$$\operatorname{Int} A = \{x \in A \, : \, \exists \, r > 0 \, \text{ s.t. } \, B_r(x) \subseteq A\} \, . \tag{3.6}$$

and

$$\overline{A} = L(A) := \{x \in X \, \text{ s.t. } \, \exists \, \{x_n\} \subseteq A \, \text{ s.t. } \, x_n \to x\} \, . \tag{3.7}$$

Proof

The proof of (3.6) is left as an exercise. Let us prove (3.7). The inclusion $L(A) \subseteq \overline{A}$ holds by Proposition 3.42. We are left to show that

$$\overline{A} \subseteq L(A) \, .$$

To this end, let $x_0 \in \overline{A}$. For $n \in \mathbb{N}$, consider the ball $B_{1/n}(x_0)$. Since $B_{1/n}(x_0) \in \mathcal{T}_d$ and $x_0 \in B_\varepsilon(x_0)$, we can apply Lemma 3.38 and deduce that

$$B_{1/n}(x_0) \cap A \neq \emptyset \, .$$

Let $x_n \in B_{1/n}(x_0) \cap A$. Since n was arbitrary, we have constructed a sequence $\{x_n\} \subseteq A$ such that

$$x_n \in B_{1/n}(x_0) \, , \quad \forall \, n \in \mathbb{N} \, .$$

In particular, we have that

$$d(x_n, x_0) < \frac{1}{n} \to 0$$

as $n \to \infty$. Thus $x_n \to x_0$, showning that $x_0 \in L(A)$.

Example 3.45

Consider $\mathbb{R}$ with the Euclidean topology and $A := [0,1)$. We have that

$$\operatorname{Int} A = (0,1)\,, \quad \overline{A} = [0,1]\,, \quad \partial A = \{0,1\}\,.$$

In particular

$$\operatorname{Int} A \neq A\,, \quad \overline{A} \neq A\,,$$

showing that A is neither open, nor closed.

The proof of the above statements is left as an exercise.

3.6 Density

Definition 3.46: Density

Let $(X, \mathcal{T})$ be a topological space and $A \subseteq X$ a set. We say that A is **dense** in X if

$$A \cap U \neq \emptyset\,, \quad \forall\, U \in \mathcal{T}\,,\; U \neq \emptyset\,.$$

Density can be characterized in terms of closure.

Proposition 3.47

Let $(X, \mathcal{T})$ be a topological space and $A \subseteq X$ a set. They are equivalent:

1. A is **dense** in X.
2. It holds
$$\overline{A} = X\,.$$

Proof

Part 1. Let A be dense in X. Suppose by contradiction that

$$\overline{A} \neq X\,.$$

This means $(\overline{A})^c \neq \emptyset$. Note that $(\overline{A})^c$ is open, being $\overline{A}$ closed. By density of A in X we have

$$A \cap (\overline{A})^c \neq \emptyset\,.$$

Since $A \subseteq \overline{A}$, the above is a contradiction.

Part 2. Suppose that $\overline{A} = X$. Let $U \in \mathcal{T}$ with $U \neq \emptyset$. By contradiction, assume that

$$A \cap U = \emptyset\,.$$

Therefore $A \subseteq U^c$. As U^c is closed, we have

$$\overline{A} \subseteq U^c\,,$$

because $\overline{A}$ is the smallest closed set containing A. Recalling that $\overline{A} = X$, we conclude that $U^c = X$. Therefore $U = \emptyset$, which is a contradiction.

Example 3.48

Consider $\mathbb{R}$ with the Euclidean topology.

1. We have that the set of integers $\mathbb{Z}$ is closed in $\mathbb{R}$. Indeed,

$$\mathbb{Z}^c = \bigcup_{z \in \mathbb{Z}} (z, z+1)\,.$$

Since $(z, z+1)$ is open in $\mathbb{R}$, by (A2) we conclude that $\mathbb{Z}^c$ is open, so that $\mathbb{Z}$ is closed. Therefore

$$\overline{\mathbb{Z}} = \mathbb{Z}\,,$$

showing that $\mathbb{Z}$ is not dense in $\mathbb{R}$.

2. The rational numbers $\mathbb{Q}$ are instead dense in $\mathbb{R}$, as proven in the Analysis module. Therefore

$$\overline{\mathbb{Q}} = \mathbb{R}\,.$$

It is also easy to check that

$$\operatorname{Int} \mathbb{Q} = \emptyset\,.$$

Therefore

$$\operatorname{Int} \mathbb{Q} \neq \mathbb{Q}\,, \quad \overline{\mathbb{Q}} \neq \mathbb{Q}\,,$$

showing that $\mathbb{Q}$ is neither open, nor closed.

Example 3.49

Consider $\mathbb{R}$ with the cofinite topology

$$\mathcal{T}_{\text{cofinite}} := \{U \subset \mathbb{R} \, : \, U^c \, \text{ is finite, or } U^c = \mathbb{R}\}\,.$$

We have that

$$\overline{\mathbb{Z}} = \mathbb{R}\,,$$

showing that $\mathbb{Z}$ is dense in $\mathbb{R}$.

Proof. Suppose C is a closed set such that $\mathbb{Z} \subseteq C$. By definition of $\mathcal{T}_{\text{cofinite}}$ we have $C = \mathbb{R}$ or C finite. Since $\mathbb{Z} \subseteq C$ and $\mathbb{Z}$ is not finite, we conclude $C = \mathbb{R}$. This proves that $\mathbb{R}$ is the only closed set containing $\mathbb{Z}$, and so $\overline{\mathbb{Z}} = \mathbb{R}$.

3.7 Hausdorff spaces

Hausdorff space are topological spaces in which points can be separated by means of disjoint open sets.

Definition 3.50

Let $(X, \mathcal{T})$ be a topological space. We say that X is a Hausdorff space if for every two points $x, y \in X$ with $x \neq y$ there exist $U, V \in \mathcal{T}$ such that

$$x \in U\,, \quad y \in V\,, \quad U \cap V = \emptyset\,.$$

The main example of Hausdorff spaces are metrizable spaces.

Proposition 3.51

Let (X, d) be a metric space with $\mathcal{T}_d$ the topology induced by d. Then $(X, \mathcal{T}_d)$ is a Hausdorff space.

Proof

Let $x, y \in X$ with $x \neq y$. Set

$$\varepsilon := \frac{1}{2}\, d(x, y)\,,$$

and define

$$U := B_\varepsilon(x)\,, \quad V := B_\varepsilon(y)\,.$$

By Proposition 3.27 we know that $U, V \in \mathcal{T}_d$. Moreover $x \in U$, $y \in V$. We are left to show that

$$U \cap V = \emptyset\,.$$

Suppose by contradiction that $U \cap V \neq \emptyset$ and let $z \in U \cap V$. Therefore

$$d(x, z) < \varepsilon\,, \quad d(y, z) < \varepsilon\,.$$

By triangle inequality we have

$$d(x,y) \leq d(x,z) + d(y,z) < \varepsilon + \varepsilon = d(x,y)\,,$$

where in the last inequality we used the definition of ε. This is a contradiction. Therefore $U \cap V = \emptyset$ and $(X, \mathscr{T}_d)$ is Hausdorff.

In general, every metrizable space is Hausdorff.

Definition 3.52: Metrizable space

Let $(X, \mathscr{T})$ be a topological space. We say that the topology $\mathscr{T}$ is metrizable if there exists a metric d on X such that

$$\mathscr{T} = \mathscr{T}_d\,,$$

with $\mathscr{T}_d$ the topology induced by d.

Corollary 3.53

Let $(X, \mathscr{T})$ be a metrizable space. Then X is Hausforff.

Proof

Since $(X, \mathscr{T})$ is metrizable, there exists a metric d on X such that

$$\mathscr{T} = \mathscr{T}_d\,.$$

By Proposition 3.51 we know that $(X, \mathscr{T}_d)$ is Hausdorff. Hence $(X, \mathscr{T})$ is Hausdorff.

As a conseuqence of Corollary 3.53 we have that spaces which are not metrizable are not Hausdorff. Let us make a few examples.

Example 3.54: Trivial topology is not Hausdorff

Let $(X, \mathscr{T})$ be a topological space with $\mathscr{T}$ trivial topology. Assume that X has more than one element. Then X is not Hausdorff.

Indeed, let $x, y \in X$ with $x \neq y$. Suppose by contradiction that X is Hausdorff. Then there exist $U, V \in \mathscr{T}$ such that

$$x \in U\,, \quad y \in V\,, \quad U \cap V = \emptyset\,.$$

Recall that

$$\mathscr{T} = \{\emptyset, X\}.$$

Since $x \in U$ and $y \in V$, we deduce that U and V are non-empty. Since U and V are open, the only possibility is that

$$U = V = X.$$

In this case we have

$$U \cap V = X \cap X = X \neq \emptyset,$$

leading to a contradiciton. Hence X is not Hausdorff.

Example 3.55: Cofinite topology on $\mathbb{R}$

Consider the following family $\mathscr{T}$ of subsets of $\mathbb{R}$

$$\mathscr{T} := \{U \subseteq \mathbb{R} \, : \, U^c \text{ is finite, or } U^c = \mathbb{R}\}.$$

Then $(\mathbb{R}, \mathscr{T})$ is a topological space which is not Hausdorff. The topology $\mathscr{T}$ is called the **cofinite topology**.

Exercise: Show that $(\mathbb{R}, \mathscr{T})$ is not Hausdorff.

Example 3.56

Consider the following family $\mathscr{T}$ of subsets of $\mathbb{R}$

$$\mathscr{T} := \{U = (-\infty, a) \, : \, -\infty \leq a \leq \infty\}.$$

Then $(\mathbb{R}, \mathscr{T})$ is a topological space which is not Hausdorff.

We start by showing that $(\mathbb{R}, \mathscr{T})$ is a topological space. We need to check the properties of topologies:

- (A1) We have that
$$(\infty, \infty) = \emptyset \in \mathscr{T}, \quad (-\infty, \infty) = \mathbb{R} \in \mathscr{T}.$$
- (A2) Suppose that $A_i \in \mathscr{T}$ for all $i \in I$. By definition
$$A_i = (-\infty, a_i), \quad -\infty \leq a_i \leq \infty.$$
Set
$$a := \sup_{i \in I} a_i, \quad A := (-\infty, a).$$
Note that a always exists, and possibly $a = \infty$. Moreover $A \in \mathscr{T}$. We claim
$$A = \bigcup_{i \in I} A_i. \tag{3.8}$$

To prove (3.8) first suppose that $x \in A$. Then $x < a$. Set $\varepsilon := a - x$, so that $\varepsilon > 0$. By definition of supremum there exists $i_0 \in I$ such that

$$a - \varepsilon < a_{i_0} .$$

From the above, and from the definition of ε, we deduce

$$a_{i_0} > a - \varepsilon = a - a + x = x ,$$

showing that $x \in (-\infty, a_{i_0}) = A_{i_0}$. Therefore

$$A \subseteq \bigcup_{i \in I} A_i .$$

Conversely, assume that $x \in \cup_{i \in I} A_i$. Therefore there exists $i_0 \in I$ such that $x \in A_{i_0} = (-\infty, a_{i_0})$. In particular

$$x < a_{i_0} \leq \sup_{i \in I} a_i = a ,$$

showing that $x \in (-\infty, a) = A$. Therefore

$$\bigcup_{i \in I} A_i \subseteq A ,$$

and (3.8) is proven.

- (A3) Let $A, B \in \mathscr{T}$. Therefore

  $$A = (-\infty, a) , \quad B = (-\infty, b) ,$$

  for some $a, b \in [-\infty, \infty]$. Set

  $$U := A \cap B , \quad z := \min\{a, b\} .$$

  It is immediate to check that

  $$U = (-\infty, z) ,$$

  showing that $U \in \mathscr{T}$.

Therefore $(\mathbb{R}, \mathscr{T})$ is a topological space. We now show that $(\mathbb{R}, \mathscr{T})$ is not Hausdorff. Suppose by contradiction that $(\mathbb{R}, \mathscr{T})$ is Hausdorff. Let $x, y \in \mathbb{R}$ with $x \neq y$. By assumption there exist $U, V \in \mathscr{T}$ such that

$$x \in U , \quad y \in V , \quad U \cap V = \emptyset .$$

By definition of $\mathscr{T}$ there exist $a, b \in [-\infty, \infty]$ such that

$$U = (-\infty, a) , \quad V = (-\infty, b) .$$

Since $x \in U$ and $y \in V$, in particular U and V are non-empty. Therefore $a, b > -\infty$. Set

$$z := \min\{a, b\} , \quad Z := U \cap V = (-\infty, z) .$$

As $a, b > -\infty$, we have $z > -\infty$. Therefore $Z \neq \emptyset$. This is a contradiction, since $U \cap V = \emptyset$. Therefore $(\mathbb{R}, \mathscr{T})$ is not Hausdorff.

In Hausdorff spaces the limit of a sequence is unique.

Proposition 3.57: Uniqueness of limit in Hausdorff spaces

Let $(X, \mathscr{T})$ be a Hausdorff space. If a sequence $\{x_n\} \subseteq X$ converges, then the limit is unique.

Proof

Let $\{x_n\} \subseteq X$ be a convergent sequence. Suppose by contradiction that

$$x_n \to x_0 \,, \quad x_n \to y_0$$

in X, for some $x_0, y_0 \in X$ with $x_0 \neq y_0$. Since X is Hausdorff, there exist $U, V \in \mathscr{T}$ such that

$$x_0 \in U \,, \quad y_0 \in V \,, \quad U \cap V = \emptyset \,.$$

As $x_n \to x_0$ and $U \in \mathscr{T}$ with $x_0 \in U$, there exists $N_1 \in \mathbb{N}$ such that

$$x_n \in U \,, \quad \forall \, n \geq N_1 \,.$$

Similarly, since $x_n \to y_0$ and $V \in \mathscr{T}$ with $y_0 \in U$, there exists $N_2 \in \mathbb{N}$ such that

$$x_n \in V \,, \quad \forall \, n \geq N_2 \,.$$

Take $N := \max\{N_1, N_2\}$. Then

$$x_n \in U \cap V \,, \quad \forall \, n \geq N \,.$$

Since $U \cap V = \emptyset$, the above is a contradiction. Therefore the limit of x_n is unique.

3.8 Continuity

We extend the notion of continuity to topological spaces. To this end, we need the concept of pre-image of a set under a function.

Definition 3.58: Images and Pre-images

Let X, Y be sets and $f \colon X \to Y$ be a function.

- Let $U \subseteq X$. The image of U under f is the subset of Y defined by

$$f(U) := \{ y \in Y \, : \, \exists \, x \in X \text{ s.t. } y = f(x) \} = \{ f(x) \, : \, x \in X \} \,.$$

- Let $V \subseteq Y$. The pre-image of V under f is the subset of X defined by

$$f^{-1}(V) := \{x \in X : \ f(x) \in V\} .$$

Warning

The notation $f^{-1}(V)$ does not mean that we are inverting f. In fact, the pre-image is defined for all functions.

Let us gather useful properties of images and pre-images.

Proposition 3.59

Let X, Y be sets and $f : X \to Y$. We denote with the letter A sets in X and with the letter B sets in Y. We have

- $A \subseteq f^{-1}(f(A))$
- $A = f^{-1}(f(A))$ if f is injective
- $f(f^{-1}(B)) \subseteq B$
- $f(f^{-1}(B)) = B$ if f is surjective
- If $A_1 \subseteq A_2$ then $f(A_1) \subseteq f(A_2)$
- If $B_1 \subseteq B_2$ then $f^{-1}(B_1) \subseteq f^{-1}(B_2)$
- If $A_i \subseteq X$ for $i \in I$ we have

$$f\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f(A_i)$$
$$f\left(\bigcap_{i\in I} A_i\right) \subseteq \bigcap_{i\in I} f(A_i)$$

- If $B_i \subseteq Y$ for $i \in I$ we have

$$f^{-1}\left(\bigcup_{i\in I} B_i\right) = \bigcup_{i\in I} f^{-1}(B_i)$$
$$f^{-1}\left(\bigcap_{i\in I} B_i\right) = \bigcap_{i\in I} f^{-1}(B_i)$$

Suppose Z is another set and $g : Y \to Z$. Let $C \subseteq Z$. Then

$$(g \circ f)(A) = g(f(A))$$
$$(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$$

It is a good exercise to try and prove a few of the above properties. We omit the proof. We can now define continuous functions between topological spaces.

Definition 3.60: Continuous function

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Let $f : X \to Y$ be a function.

- Let $x_0 \in X$. We say that f is continuous at x_0 if it holds:

$$\forall \, V \in \mathcal{T}_Y \;\text{ s.t. }\; f(x_0) \in V \,, \;\exists U \in \mathcal{T}_X \;\text{ s.t. }\; x_0 \in U \,, \; f(U) \subseteq V \,.$$

- We say that f is continuous from $(X, \mathcal{T}_X)$ to $(Y, \mathcal{T}_Y)$ if f is continuous at each point $x_0 \in X$.

The following proposition presents a useful characterization of continuous functions in terms of pre-images.

Proposition 3.61

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Let $f : X \to Y$ be a function. They are equivalent:

1. f is continuous from $(X, \mathcal{T}_X)$ to $(Y, \mathcal{T}_Y)$.
2. It holds:
$$f^{-1}(V) \in \mathcal{T}_X \,, \quad \forall \, V \in \mathcal{T}_Y \,.$$

Important

In other words, a function $f : X \to Y$ is continuous if and only if the pre-image of open sets in Y are open sets in X.

The proof of Proposition 3.61 is simple, but very tedious. We choose to skip it.

Example 3.62

Let X be a set and $\mathcal{T}_1$, $\mathcal{T}_2$ be topologies on X. Define the identity map

$$\mathrm{Id}_X : (X, \mathcal{T}_1) \to (X, \mathcal{T}_2) \,, \quad \mathrm{Id}_X(x) := x \,.$$

They are equivalent:

1. Id_X is continuous from $(X, \mathcal{T}_1)$ to $(X, \mathcal{T}_2)$.
2. $\mathcal{T}_1$ is finer than $\mathcal{T}_2$
$$\mathcal{T}_2 \subseteq \mathcal{T}_1 \,.$$

   Indeed, Id_X is continuous if and only if

$$\mathrm{Id}_X^{-1}(V) \in \mathcal{T}_1 \,, \quad \forall \, V \in \mathcal{T}_2 \,.$$

But $\mathrm{Id}_X^{-1}(V) = V$, so that the above reads

$$V \in \mathcal{T}_1 \,, \quad \forall \, V \in \mathcal{T}_2 \,,$$

which is equivalent to $\mathcal{T}_2 \subseteq \mathcal{T}_1$.

Let us compare our new definition of contiuity with the classical notion of continuity in $\mathbb{R}^n$. Let us recall the definition of continuous function in $\mathbb{R}^n$.

Definition 3.63: Continuity in the classical sense

Let $f : \subseteq \mathbb{R}^n \to \mathbb{R}^m$. We say that f is continuous at $\mathbf{x}_0$ if it holds:

$$\forall \, \varepsilon > 0 \,, \, \exists \, \delta > 0 \;\text{ s.t. }\; \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \varepsilon \;\text{ if }\; \|\mathbf{x} - \mathbf{x}_0\| < \delta \,.$$

Proposition 3.64

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and suppose $\mathbb{R}^n, \mathbb{R}^m$ are equipped with the Euclidean topology. Let $\mathbf{x}_0 \in \mathbb{R}^n$. They are equivalent:

1. f is continuous at $\mathbf{x}_0$ in the topological sense.
2. f is continuous at $\mathbf{x}_0$ in the classical sense.

Proof

Part 1. Suppose that f is continuous at $\mathbf{x}_0$ in the topological sense. Let $\varepsilon > 0$ and consider the set

$$V := B_\varepsilon(f(\mathbf{x}_0)) \,.$$

We have that $V \subset \mathbb{R}^m$ is open and $f(\mathbf{x}_0) \in V$. As f is continuous in the topological sense, there exists $U \subset \mathbb{R}^n$ open with $\mathbf{x}_0 \in U$ and such that

$$f(U) \subset V = B_\varepsilon(f(\mathbf{x}_0)) \,. \tag{3.9}$$

Since U is open and $\mathbf{x}_0 \in U$, there exists $\delta > 0$ such that

$$B_\delta(\mathbf{x}_0) \subset U \,.$$

By the above inclusion and (3.9) we conclude that

$$f(B_\delta(\mathbf{x}_0)) \subset f(U) \subset V = B_\varepsilon(f(\mathbf{x}_0)) \,.$$

This is equivalent to

$$\mathbf{x} \in B_\delta(\mathbf{x}_0) \quad \Longrightarrow \quad f(\mathbf{x}) \in B_\varepsilon(f(\mathbf{x}_0)),$$

which reads

$$\|\mathbf{x} - \mathbf{x}_0\| < \delta \quad \Longrightarrow \quad \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \varepsilon .$$

Therefore f is continuous at $\mathbf{x}_0$ in the classical sense.
Part 2. Suppose f is continuous at x_0 in the classical sense. Let $V \subset \mathbb{R}^m$ be open and such that $f(\mathbf{x}_0) \in V$. Since V is open, there exists $\varepsilon > 0$ such that

$$B_\varepsilon(f(\mathbf{x}_0)) \subset V . \tag{3.10}$$

Since f is continous in the classical sense, there exists $\delta > 0$ such that

$$\|\mathbf{x} - \mathbf{x}_0\| < \delta \quad \Longrightarrow \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \varepsilon .$$

The above is equivalent to

$$\mathbf{x} \in B_\delta(\mathbf{x}_0) \quad \Longrightarrow \quad f(\mathbf{x}) \in B_\varepsilon(f(\mathbf{x}_0)) . \tag{3.11}$$

Set

$$U := B_\delta(\mathbf{x}_0)$$

and note that U is open in $\mathbb{R}^n$ and $\mathbf{x}_0 \in U$. By definition of image of a set, (3.11) reads

$$f(U) = f(B_\delta(\mathbf{x}_0)) \subseteq B_\varepsilon(f(\mathbf{x}_0)) .$$

Recalling (3.10) we conclude that

$$f(U) \subset V .$$

In summary, we have shown that given $V \subset \mathbb{R}^m$ open and such that $f(\mathbf{x}_0) \in V$, there exists U open in $\mathbb{R}^n$ such that $\mathbf{x}_0 \in U$ and $f(U) \subset V$. Therefore f is continuous at $\mathbf{x}_0$ in the topological sense.

A similar proof yields the characterization of continuity in metric spaces. The proof is left as an exercise.

Proposition 3.65

Let (X, d_X) and (Y, d_Y) be metric spaces. Denote by $\mathcal{T}_X$ and $\mathcal{T}_Y$ the topologies induced by the metrics. Let $f \colon X \to Y$ and $x_0 \in X$. They are equivalent:

1. f is continuous at x_0 in the topological sense.
2. It holds:

$$\forall\, \varepsilon > 0 \,, \, \exists\, \delta > 0 \ \text{ s.t. } \ d_Y(f(x), f(x_0)) < \varepsilon \ \text{ if } \ d_X(x, x_0) < \delta .$$

Let us examine continuity in the cases of the trivial and discrete topologies.

Example 3.66

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be a topological space. Suppose that $\mathcal{T}_Y$ is the trivial topology, that is,

$$\mathcal{T}_Y = \{\emptyset, Y\}.$$

Then every function $f : X \to Y$ is continuous.

Indeed, we know that f is continuous if and only if it holds:

$$f^{-1}(V) \in \mathcal{T}_X, \quad \forall\, V \in \mathcal{T}_Y.$$

We have two cases:

- $V = \emptyset$: Then
$$f^{-1}(V) = f^{-1}(\emptyset) = \emptyset \in \mathcal{T}_X.$$
- $V = Y$: Then
$$f^{-1}(V) = f^{-1}(Y) = X \in \mathcal{T}_X.$$

Therefore f is continuous.

Example 3.67

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Suppose that $\mathcal{T}_Y$ is the discrete topology, that is,

$$\mathcal{T}_Y = \{V \text{ s.t. } V \subseteq Y\}.$$

Let $f : X \to Y$. They are equivalent:

1. f is continuous from X to Y.
2. $f^{-1}(\{y\}) \in \mathcal{T}_X$ for all $y \in Y$.

Indeed, suppose that f is continuous. Then

$$f^{-1}(V) \in \mathcal{T}_X, \quad \forall\, V \in \mathcal{T}_Y.$$

As $V = \{y\} \in \mathcal{T}_Y$, we conclude that $f^{-1}(\{y\}) \in \mathcal{T}_X$.
Conversely, assume that $f^{-1}(\{y\}) \in \mathcal{T}_X$ for all $y \in Y$. Let $V \in \mathcal{T}_Y$. Trivially, we have

$$V = \bigcup_{y \in V} \{y\}.$$

Therefore

$$f^{-1}(V) = f^{-1}\left(\bigcup_{y \in V} \{y\} \right) = \bigcup_{y \in V} f^{-1}(\{y\}).$$

As $f^{-1}(\{y\}) \in \mathcal{T}_X$ for all $y \in Y$, by property (A2) we conclude that $f^{-1}(V) \in \mathcal{T}_X$. Therefore f is continuous.

In a topological space, continuity preserves limits of sequences.

Proposition 3.68

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Let $f : X \to Y$ be continuous. Let $\{x_n\} \subset X$ and $x_0 \in X$. We have

$$x_n \to x_0 \text{ in } X \quad \implies \quad f(x_n) \to f(x_0) \text{ in } Y .$$

Proof

Let $V \in \mathcal{T}_Y$ be such that $f(x_0) \in V$. Since f is continuous there exists $U \in \mathcal{T}_X$ with $x_0 \in U$ such that

$$f(U) \subset V .$$

Since $U \in \mathcal{T}_X$ and $x_n \to x_0$ in X, there exists $N \in \mathbb{N}$ such that

$$x_n \in U , \quad \forall \, n \geq N .$$

Therefore

$$f(x_n) \in f(U) , \quad \forall \, n \geq N .$$

Seeing that $f(U) \subset V$, we conclude

$$f(x_n) \in V , \quad \forall \, n \geq N ,$$

showing that $f(x_n) \to f(x_0)$ in Y.

Warning

The converse implication of Proposition 3.68 is false. That is, even if it holds

$$x_n \to x_0 \text{ in } X \quad \implies \quad f(x_n) \to f(x_0) \text{ in } Y .$$

for all sequences $\{x_n\} \subset X$, the function f might **not** be continuous. For this to hold, it is necessary for the topologies on X and Y to be metrizable.

Let us make an observation on continuity of compositions.

Proposition 3.69

Let $(X, \mathcal{T}_X), (Y, \mathcal{T}_Y), (Z, \mathcal{T}_Z)$ be topological spaces. Let

$$f \colon X \to Y , \quad g \colon Y \to Z ,$$

be given functions. If f and g are continuous, then

$$(g \circ f) \colon X \to Z$$

is continuous.

Proof

Let $C \in \mathcal{T}_Z$. As g is continuous, we have that

$$g^{-1}(C) \in \mathcal{T}_Y .$$

Since f is continuous, we also have

$$f^{-1}(g^{-1}(C)) \in \mathcal{T}_X .$$

Therefore

$$(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C)) \in \mathcal{T}_X ,$$

so that $g \circ f$ is continuous.

We conclude the section by introducing homeomorphisms.

Definition 3.70: Homeomoprhim

Let $(X, \mathcal{T}_X)$, $(Y, \mathcal{T}_Y)$ be topological space. A function $f \colon X \to Y$ is called an **homeomorphism** if they hold:

1. f is continuous.
2. There exists $g \colon Y \to X$ continuous such that

$$g \circ f = \mathrm{Id}_X , \quad f \circ g = \mathrm{Id}_Y .$$

The above is saying that f is a homeomorphism if it is continuous and has continuous inverse. Homeomorphisms are the way we say that two topological spaces look the same.

3.9 Subspace topology

Any subset Y in a topological space X inherits naturally a topological structure. Such structure is called **subspace topology**.

Definition 3.71: Subspace topology

Let $(X, \mathscr{T})$ be a topological space and $Y \subseteq X$ a subset. Define the family of sets

$$\mathscr{S} := \{A \subset Y \, : \, \exists \, U \in \mathscr{T} \text{ s.t. } A = U \cap Y\} .$$

The family $\mathscr{S}$ is called subspace topology on Y induced by the inclusion $Y \subset X$.

Proof: Well-posedness of Definition 3.71

We have to show that $(Y, \mathscr{S})$ is a topological space:

- (A1) $\emptyset \in \mathscr{S}$ since
$$\emptyset = \emptyset \cap Y$$
and $\emptyset \in \mathscr{T}$. Similarly we have $Y \in \mathscr{S}$, since
$$Y = X \cap Y ,$$
and $X \in \mathscr{T}$.
- (A2) Let $A_i \in \mathscr{S}$ for $i \in I$. By definition there exist $U_i \in \mathscr{T}$ such that
$$A_i = U_i \cap Y , \quad \forall \, i \in I .$$
Therefore
$$\bigcup_{i \in I} A_i = \bigcup_{i \in I} (U_i \cap Y) = \left(\bigcup_{i \in I} U_i \right) \cap Y .$$
The above proves that $\cup_{i \in I} A_i \in \mathscr{S}$, since $\cup_{i \in I} U_i \in \mathscr{T}$.
- (A3) Let $A_1, A_2 \in \mathscr{S}$. By definition there exist $U_1, U_2 \in \mathscr{T}$ such that
$$A_1 = U_1 \cap Y , \quad A_2 = U_2 \cap Y$$
Therefore
$$A_1 \cap A_2 = (U_1 \cap Y) \cap (U_2 \cap Y) = (U_1 \cap U_2) \cap Y$$
The above proves that $A_1 \cap A_2 \in \mathscr{S}$, since $U_1 \cap U_2 \in \mathscr{T}$.

If the set Y is open, then sets are open in the subspace topology if and only if they are open in X.

Proposition 3.72

Let $(X, \mathcal{T})$ be a topological space and $Y \in \mathcal{T}$ a subset. Let $A \subset Y$. Then

$$A \in \mathcal{S} \quad \Longleftrightarrow \quad A \in \mathcal{T} \,.$$

Proof

Suppose $A \in \mathcal{S}$. Then there exists $U \in \mathcal{T}$ such that

$$A = U \cap Y \,.$$

Since $U, Y \in \mathcal{T}$, by property (A3) of topologies it follows that

$$A = U \cap Y \in \mathcal{T} \,.$$

Conversely, assume that $A \in \mathcal{T}$. Then

$$A = A \cap Y \,,$$

showing that $A \in \mathcal{S}$.

Warning

Let $(X, \mathcal{T})$ be a topological space, $A \subset Y \subset X$. In general we could have

$$A \in \mathcal{S} \quad \text{and} \quad A \notin \mathcal{T}$$

For example consider $X = \mathbb{R}$ with $\mathcal{T}$ the euclidean topology. Consider the subset $Y = [0, 2)$ and equip Y with the subspace topology $\mathcal{S}$. Let $A = [0, 1)$. Then $A \notin \mathcal{T}$ but $A \in \mathcal{S}$, since

$$A = (-1, 1) \cap Y$$

and $(-1, 1) \in \mathcal{T}$.

Example 3.73

Let $X = \mathbb{R}$ be equipped with $\mathcal{T}$ the euclidean topology. Let $\mathcal{S}$ be the subspace topology on $\mathbb{Z}$. Then $\mathcal{S}$ coincides with the discrete topology.

Proof. The set $\{z\}$ is open in $\mathcal{S}$ for all $z \in \mathbb{Z}$. Indeed,

$$\{z\} = (z-1, z+1) \cap \mathbb{Z}$$

and $(z-1, z+1) \in \mathcal{T}$. Thus $\{z\} \in \mathcal{S}$. Let now $A \subseteq \mathbb{Z}$. Then

$$A = \bigcup_{z \in A} \{z\} \,,$$

and therefore $A \in \mathcal{S}$ by (A2). This proves that

$$\mathcal{S} = \{A \text{ s.t. } A \subseteq \mathbb{Z}\},$$

that is, $\mathcal{S}$ is the discrete topology on $\mathbb{Z}$.

3.10 Topological basis

We have seen that in metric spaces every open set is union of open balls, see Propostion 3.27. We can then regard the open balls as building blocks for the whole topology. In this context, we call the open balls a basis for the topology.

We can generalize the concept of basis to arbitrary topological spaces.

Definition 3.74: Topological basis

Let $(X, \mathcal{T})$ be a topological space and let $\mathcal{B} \subseteq \mathcal{T}$. We say that $\mathcal{B}$ is a **topological basis** for the topology $\mathcal{T}$ if for all $U \in \mathcal{T}$ there exist open sets $\{B_i\} \subseteq \mathcal{B}$, with I family of indices, such that

$$U = \bigcup_{i \in I} B_i \,. \tag{3.12}$$

Example 3.75

1. Let $(X, \mathcal{T})$ be a topological space. Then $\mathcal{B} := \mathcal{T}$ is a basis for $\mathcal{T}$.

   This is true because one can just take $B = U$ in (3.12).

2. (X, d) metric space with topology $\mathcal{T}_d$ induced by the metric. Then

   $$\mathcal{B} := \{B_r(x) \, : \, x \in X \, , \, r > 0\}$$

   is a basis for $\mathcal{T}_d$.

   This is true by Propostion 3.27.

3. Let $(X, \mathcal{T})$ with X the discrete topology. Then

   $$\mathcal{B} := \{\{x\} \, : \, x \in X\}$$

   is a basis for $\mathcal{T}$.

   This is true because for any $U \in \mathcal{T}$ we have

   $$U = \bigcup_{x \in U} \{x\} \,.$$

Proposition 3.76

Let $(X, \mathscr{T})$ be a topological space and $\mathscr{B}$ a basis for $\mathscr{T}$. They hold:

- (B1) We have
$$\bigcup_{B \in \mathscr{B}} B = X\,.$$
- (B2) If $U_1, U_2 \in \mathscr{B}$ then there exist $\{B_i\} \subseteq \mathscr{B}$ such that
$$U_1 \cap U_2 = \bigcup_{i \in I} B_i\,.$$

Proof

- (B1) This holds because $X \in \mathscr{T}$. Therefore by definition of basis there exist $B_i \in \mathscr{B}$ such that
$$X = \bigcup_{i \in I} B_i\,.$$
Therefore taking the union over all $B \in \mathscr{B}$ yields X, and (B1) follows.
- (B2) Let $U_1, U_2 \in \mathscr{B}$. Then $U_1, U_2 \in \mathscr{T}$, since $\mathscr{B} \subseteq \mathscr{T}$. By property (A3) we get that $U_1 \cap U_2 \in \mathscr{T}$. Since $\mathscr{B}$ is a basis we conclude (B2).

Properties (B1) and (B2) from Proposition 3.76 are sufficient for generating a topology.

Proposition 3.77

Let X be a set and $\mathscr{B}$ a collection of subsets of X such that (B1)-(B2) hold. Define
$$\mathscr{T} := \left\{ U \subseteq X \,:\, U = \bigcup_{i \in I} B_i\,,\ B_i \in \mathscr{B} \right\}.$$

Then:

1. $\mathscr{T}$ is a topology on X.
2. $\mathscr{B}$ is a basis for $\mathscr{T}$.

Proof

1. We need to verify that $\mathscr{T}$ is a topology:

- (A1) We have that $X \in \mathscr{T}$ by (B1). Moreover $\emptyset \in \mathscr{T}$, since $\emptyset$ can be obtained as empty union. Therefore (A1) holds.

- (A2) Let $U_i \in \mathcal{T}$ for all $i \in I$. By definition of $\mathcal{T}$ we have

$$U_i = \bigcup_{k \in K_i} B_k^i$$

for some family of indices K_i and $B_k^i \in \mathcal{B}$. Therefore

$$U := \bigcup_{i \in I} U_i = \bigcup_{i \in I,\, k \in K_i} B_k^i \,,$$

showing that $U \in \mathcal{T}$.

- (A3) Suppose that $U_1, U_2 \in \mathcal{T}$. Then

$$U_1 = \bigcup_{i \in I_1} B_i^1 \,, \quad U_2 = \bigcup_{i \in I_2} B_i^2$$

for $B_i^1, B_i^2 \in \mathcal{B}$. From the above we have

$$U_1 \cap U_2 = \bigcup_{i \in I_1,\, k \in I_2} B_i^1 \cap B_k^2 \,.$$

From property (B2) we have that for each pair of indices (i, k) the set $B_i^1 \cap B_k^2$ is the union of sets in $\mathcal{B}$. Therefore $U_1 \cap U_2$ is union of sets in $\mathcal{B}$, showing that $U_1 \cap U_2 \in \mathcal{T}$.

2. This trivially follows from defintion of $\mathcal{T}$ and definition of basis.

3.11 Product topology

Given two topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ we would like to equip the cartesian product

$$X \times Y = \{(x, y) \;:\; x \in X \,,\; y \in Y\}$$

with a topology. We proceed as follows.

Proposition 3.78

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Define the family $\mathcal{B}$ of subsets of $X \times Y$ as

$$\mathcal{B} := \{U \times V \;:\; U \in \mathcal{T}_X \,,\; V \in \mathcal{T}_Y\} \subset X \times Y \,.$$

Then $\mathcal{B}$ satisfies properties (B1) and (B2) from Proposition 3.76.

The proof is an easy check, and is left as an exercise. As $\mathscr{B}$ satisfies (B1)-(B2), by Proposition 3.77 we know that

$$\mathscr{T}_{X\times Y} := \left\{ U \times V \, : \, U \times V = \bigcup_{i \in I} B_i \, , \, , \; B_i \in \mathscr{B} \right\} \tag{3.13}$$

is a topology on $X \times Y$.

Definition 3.79: Product topology

Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be topological spaces. We call $\mathscr{T}_{X\times Y}$ at (3.13) the **product topology** on $X \times Y$.

Example 3.80

Let $\mathbb{R}$ be equipped with the Euclidean topology. The product topology on $\mathbb{R}\times\mathbb{R}$ coincides with the topology on $\mathbb{R}^2$ equipped with the Euclidean topology.

Consider the projection maps

$$\pi_X \, : \, X \times Y \to X \, , \quad \pi_X(x, y) := x$$

and

$$\pi_Y \, : \, X \times Y \to Y \, , \quad \pi_Y(x, y) := y$$

Proposition 3.81

Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be topological spaces and equip $X \times Y$ with the product topology $\mathscr{T}_{X\times Y}$. Then π_X and π_Y are continuous.

Proof

Let $U \in \mathscr{T}_X$. Then

$$\pi_X^{-1}(U) = U \times Y \, .$$

We have that $U \times Y \in \mathscr{T}_{X\times Y}$ since $U \in \mathscr{T}_X$ and $Y \in \mathscr{T}_Y$. Therefore π_X is continuous. The proof that π_Y is continuous is similar, and is left as an exercise.

The following proposition gives a useful criterion to check whether a map into $X \times Y$ is continuous.

Proposition 3.82

Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be topological spaces and equip $X \times Y$ with the product topology $\mathscr{T}_{X\times Y}$. Let

$(Z, \mathcal{T}_Z)$ be a topological space and

$$f : Z \to X \times Y$$

a function. They are equivalent:

1. f is continuous.
2. The compositions
$$\pi_X \circ f : Z \to X , \quad \pi_Y \circ f : Z \to Y$$
are continuous.

The proof is left as an exercise.

3.12 Connectedness

Suppose that $(X, \mathcal{T})$ is a topological space. By property (A1) we have that

$$\emptyset , \ X \in \mathcal{T}$$

Therefore

$$\emptyset^c = X , \quad X^c = \emptyset$$

are closed. It follows that $\emptyset$ and X are both open and closed.

Definition 3.83: Connected space

Let $(X, \mathcal{T})$ be a topological space. We say that:

- X is **connected** if the only subsets of X which are both open and closed are $\emptyset$ and X.
- X is **disconnected** if it is not connected.

The following proposition gives two extremely useful equivalent definitions of connectedness. Before stating it, we define the concept of proper set.

Definition 3.84: Proper subset

Let X be a set. A subset $A \subseteq X$ is **proper** if

$$A \neq \emptyset , \quad A \neq X .$$

Proposition 3.85: Equivalent definition for connectedness

Let $(X, \mathcal{T})$ be a topological space. They are equivalent:

1. X is disconnected.
2. X is the disjoint union of two proper open subsets.
3. X is the disjoint union of two proper closed subsets.

Proof

Part 1. Point 1 implies Points 2 and 3.
Suppose X is disconnected. Then there exists $U \subseteq X$ which is open, closed, and such that

$$U \neq \emptyset \,, \quad U \neq X \,. \tag{3.14}$$

Define

$$A := U \,, \quad B := U^c \,.$$

By definition of complement we have

$$X = A \cup B \,, \quad A \cap B = \emptyset \,.$$

Moreover:

- A and B are both open and closed, since U is both open and closed.
- A and B are proper, since (3.14) holds.

Therefore we conclude Points 2, 3.
Part 2. Point 2 implies Point 1. Suppose A, B are open, proper, and such that

$$X = A \cup B \,, \quad A \cap B = \emptyset \,.$$

This implies

$$A^c = X \smallsetminus A = B \,,$$

showing that A^c is open, and hence A is closed. Therefore A is proper, open and closed, showing that X is disconnected.
Part 3. Point 3 implies Point 1. Suppose A, B are closed, proper, and such that

$$X = A \cup B \,, \quad A \cap B = \emptyset \,.$$

This implies

$$A^c = X \smallsetminus A = B \,,$$

showing that A^c is closed, and hence A is open. Therefore A is proper, open and closed, showing that X is disconnected.

In the following we will use Point 2 and Point 3 in Proposition 3.85 as equivalent definitions of disconnected topological space.

Example 3.86

Consider the set $X = \{0, 1\}$ with the subspace topology induced by the inclusion $X \subset \mathbb{R}$, where $\mathbb{R}$ is equipped with the Euclidean topology $\mathcal{T}_{\text{euclidean}}$. Then X is disconnected.

Proof. Note that
$$X = \{0\} \cup \{1\}, \quad \{0\} \cap \{1\} = \emptyset .$$
The set $\{0\}$ is open for the subspace topology, since
$$\{0\} = X \cap (-1, 1), \quad (-1, 1) \in \mathcal{T}_{\text{euclidean}} .$$
Similarly, also $\{1\}$ is open for the subspace topology, since
$$\{1\} = X \cap (0, 2), \quad (0, 2) \in \mathcal{T}_{\text{euclidean}} .$$
Clearly
$$\{0\} \neq \emptyset, \quad \{1\} \neq \emptyset ,$$
showing that X is disconnected.

Example 3.87

Let $p \in \mathbb{R}$. The set $X = \mathbb{R} \setminus \{p\}$ is disconnected.

Proof. Define the sets
$$A = (-\infty, p), \quad B = (p, \infty) .$$
Then A, B are proper subsets of X, since $p \notin X$. Moreover
$$X = A \cup B, \quad A \cap B = \emptyset .$$
Finally we have that A, B are open for the subspace topology, since they are open in $\mathbb{R}$. Therefore X is disconnected.

Example 3.88

Let $n \geq 2$ and $A \subseteq \mathbb{R}^n$ be open and connected. Let $p \in A$. Then $X = A \setminus \{p\}$ is connected.

Exercise: Prove that X is connected.

The next theorem shows that connectedness is preserved by continuous maps.

Theorem 3.89

Let $(X, \mathcal{T}_X)$, $(Y, \mathcal{T}_Y)$ be topological spaces. Suppose that $f : X \to Y$ is continuous and let $f(X) \subseteq Y$ be equipped with the subspace topology. If X is connected, then $f(X)$ is connected.

Proof

Suppose that A, B are open in $f(X)$ and such that

$$f(X) = A \cup B\,, \quad A \cap B = \emptyset\,.$$

if we show that

$$A = \emptyset \ \text{ or } \ B = \emptyset \tag{3.15}$$

the proof is concluded. Since A, B are open for the subspace topology, there exist $\widetilde{A}, \widetilde{B} \in \mathcal{T}_Y$ such that

$$A = \widetilde{A} \cap f(X)\,, \quad B = \widetilde{B} \cap f(X)\,. \tag{3.16}$$

Since $f(X) = A \cup B$ we have

$$\begin{aligned} X &= f^{-1}(A \cup B) \\ &= f^{-1}(A) \cup f^{-1}(B) \\ &= f^{-1}(\widetilde{A}) \cup f^{-1}(\widetilde{B}) \end{aligned}$$

where in the last equality we used (3.16). Since $A \cap B = \emptyset$, we also have that

$$\begin{aligned} f^{-1}(\widetilde{A}) \cap f^{-1}(\widetilde{B}) &= f^{-1}(A) \cap f^{-1}(B) \\ &= f^{-1}(A \cap B) \\ &= f^{-1}(\emptyset) \\ &= \emptyset \end{aligned}$$

where in the first equality we used (3.16). By continuity of f we have that

$$f^{-1}(\widetilde{A})\,, \ f^{-1}(\widetilde{B}) \in \mathcal{T}_X\,.$$

Therefore, using that X is connected, we deduce that

$$f^{-1}(\widetilde{A}) = \emptyset \ \text{ or } \ f^{-1}(\widetilde{B}) = \emptyset\,.$$

The above implies

$$\widetilde{A} \cap f(X) = \emptyset \ \text{ or } \ \widetilde{B} \cap f(X) = \emptyset\,.$$

Recalling (3.16), we obtain (3.15), ending the proof.

An immediate corollary of Theorem 3.89 is that connectedness is a topological invariant, e.g., connectedness is preserved by homeomorphisms.

Corollary 3.90

Let $(X, \mathscr{T}_X)$, $(Y, \mathscr{T}_Y)$ be homeomorhic topological spaces. Then

$$X \text{ is connected} \quad \Longleftrightarrow \quad Y \text{ is connected}$$

The proof follows immediately by Theorem 3.89, and is left to the reader as an exercise.

Example 3.91

Let $n \geq 2$. $\mathbb{R}^n$ not homeomorphic to $\mathbb{R}$.

Proof. Suppose by contradiction that there exists an omeomorphism

$$f : \mathbb{R}^n \to \mathbb{R}\,.$$

Define $p = f(0)$ and the restriction

$$g : \mathbb{R}^n \setminus \{0\} \to \mathbb{R} \setminus \{p\}\,, \quad g(x) = f(x)\,.$$

Since g is a restriction of an omeomorphism, then g is an omeomorphism. We have that $\mathbb{R}^n \setminus \{0\}$ is connected, as a consequence of
Example 3.88. Hence, by Corollary 3.90, we infer that $\mathbb{R} \setminus \{p\}$ is connected. This is a contradiction, since $\mathbb{R} \setminus \{p\}$ is disconnected, as shown in Example 3.87.

Example 3.92

Define the 1D unit circle

$$\mathbb{S}^1 := \{(x, y) \in \mathbb{R}^2 \;:\; x^2 + y^2 = 1\}\,.$$

Then $\mathbb{S}^1$ and $[0, 1]$ are not homeomorphic.

Proof. Suppose by contradiction that there exists and omeomorphism

$$f : [0, 1] \to \mathbb{S}^1\,.$$

The restriction of f to $[0, 1] \setminus \{\frac{1}{2}\}$ defines an omeomorphism

$$g \;:\; \left([0,1] \setminus \left\{\frac{1}{2}\right\}\right) \to \left(\mathbb{S}^1 \setminus \{\mathbf{p}\}\right)\,, \quad \mathbf{p} := f\left(\frac{1}{2}\right)\,.$$

The set $[0, 1] \setminus \left\{\frac{1}{2}\right\}$ is disconnected, since

$$[0, 1] \setminus \{1/2\} = [0, 1/2) \cup (1/2, 1]$$

with $[0, 1/2)$ and $(1/2, 1]$ open for the subset topology, non-empty and disjoint. Therefore, using that g is an omeomorphism, we conclude that also $\mathbb{S}^1 \smallsetminus \{\mathbf{p}\}$ is disconnected. Let $\theta_0 \in [0, 2\pi)$ be the unique angle such that

$$\mathbf{p} = (\cos(\theta_0), \sin(\theta_0)) .$$

Thus $\mathbb{S}^1 \smallsetminus \{\mathbf{p}\}$ is parametrized by

$$\boldsymbol{\gamma}(t) := (\cos(t), \sin(t)) , \quad t \in (\theta_0, \theta_0 + 2\pi) .$$

Since $\boldsymbol{\gamma}$ is continuous and $(\theta_0, \theta_0 + 2\pi)$ is connected, by Theorem 3.89, we conclude that $\mathbb{S}^1 \smallsetminus \{\mathbf{p}\}$ is connected. Contradiction.

3.13 Intermediate Value Theorem

Another consequence of Theorem 3.89 is a generalization of the Intermediate Value Theorem to arbitrary topological spaces. Before providing statement and proof of such Theorem, we need to characterize the connected subsets of $\mathbb{R}$.

Definition 3.93: Interval

A subset $I \subset \mathbb{R}$ is an interval if it holds:

$$\forall \, a, b \in I \, , \; x \in \mathbb{R} \;\; \text{s.t.} \;\; a < x < b \quad \Longrightarrow \quad x \in I \, .$$

Theorem 3.94

Let $\mathbb{R}$ be equipped with the Euclidean topology and let $I \subseteq \mathbb{R}$. They are equivalent:

1. I is connected.
2. I is an interval.

Proof

Part 1. Suppose I is connected. If $I = \{p\}$ for some $p \in \mathbb{R}$ then I is an interval and the thesis is achieved. Otherwise there exist $a, b \in I$ with $a < b$. Assume that $x \in \mathbb{R}$ is such that

$$a < x < b \, .$$

We need to show that $x \in I$. Suppose by contradiction that $x \notin I$ and define the open sets

$$A = (-\infty, x)\,, \quad B = (x, \infty)\,.$$

Then

$$\widetilde{A} = (-\infty, x) \cap I\,, \quad \widetilde{B} = (x, \infty) \cap I$$

are open in I for the subspace topology. Clearly

$$\widetilde{A} \cap \widetilde{B} = \emptyset\,.$$

Moreover

$$I = \widetilde{A} \cup \widetilde{B}$$

since $x \notin I$. We have:

- Since $a < x$ and $a \in I$, we have that $a \in \widetilde{A}$. Therefore $\widetilde{A} \neq \emptyset$.
- Similarly, $b > x$ and $b \in I$, therefore $b \in \widetilde{B}$. Hence $\widetilde{B} \neq \emptyset$.

Therefore I is disconnected, which is a contradiction.
Part 2. Suppose I is an interval. Suppose by contradiction that I is disconnected. Then there exist A, B proper and closed, such that

$$I = A \cup B\,, \quad A \cap B = \emptyset\,.$$

Since A and B are proper, there exist points $a \in A$, $b \in B$. WLOG we can assume $a < b$. Define

$$\alpha = \sup\ S\,, \quad S := \{x \in \mathbb{R} : \ [a, x) \cap I \subseteq A\}\,.$$

Note that α exists finite since b is an upper bound for the set S.

> Suppose by contradiction b is not an upper bound for S. Hence there exists $x \in \mathbb{R}$ such that $[a, x) \cap I \subseteq A$ and that $x > b$. As $b > a$, we conclude that $b \in [a, x) \cap I \subseteq A$. Thus $b \in A$, which is a contradiction, since $b \in B$ and $A \cap B = \emptyset$.

Moreover we have that $\alpha \in A$.

> This is because the supremum α is the limit of a sequence in S, and hence of a sequence in A. Therefore α belongs to $\overline{A}$. Since A is closed, we infer $\alpha \in A$.

Note that $A^c = B$, which is closed. Therefore A^c is closed, showing that A is open. As $\alpha \in A$ and A is open in I, there exists $\varepsilon > 0$ such that

$$(\alpha - \varepsilon, \alpha + \varepsilon) \cap I \subseteq A\,.$$

In particular

$$[a, \alpha + \varepsilon) \cap I \subseteq A\,,$$

showing that $\alpha + \varepsilon \in S$. This is a contradiction, since α is the supremum of S.

We are finally ready to prove the Intermediate Value Theorem.

Theorem 3.95: Intermediate Value Theorem

Let $(X, \mathcal{T})$ be a connected topological space. Suppose that $f : X \to \mathbb{R}$ is continuous. Suppose that $a, b \in X$ are such that $f(a) < f(b)$. It holds:

$$\forall \, c \in \mathbb{R} \ \text{ s.t. } \ f(a) < c < f(b) \,, \ \exists \, \xi \in X \ \text{ s.t. } \ f(\xi) = c \,.$$

Proof

As f is continuous and X is connected, by Theorem 3.89 we know that $f(X)$ is connected in $\mathbb{R}$. By Theorem 3.94 we have that $f(X)$ is an interval. Since $a, b \in X$ it follows $f(a), f(b) \in f(X)$. Therefore, if $c \in \mathbb{R}$ is such that

$$f(a) < c < f(b)$$

we conclude that $c \in f(X)$, since $f(X)$ is an interval. Hence there exists $\xi \in X$ such that $f(\xi) = c$.

3.14 Path connectedness

Definition 3.96: Path connectedness

Let $(X, \mathcal{T})$ be a topological space. We say that X is **path connected** if for every $x, y \in X$ there exist $a, b \in \mathbb{R}$ with $a < b$, and a continuous function

$$\alpha : \, [a, b] \to X$$

such that

$$\alpha(a) = x \,, \quad \alpha(b) = y \,.$$

Example 3.97

Let $A \subset \mathbb{R}^n$ be convex. Then A is path connected.

A is convex if for all $x, y \in A$ the segment connecting x to y is contained in A, namely,

$$[x, y] \, := \{(1 - t)x + ty \, : \, t \in [0, 1]\} \subseteq A \,.$$

Therefore we can define

$$\alpha : \, [0, 1] \to A \,, \quad \alpha(t) \, := (1 - t)x + ty \,.$$

Clearly α is continuous, and $\alpha(0) = x, \alpha(1) = y$.

It turns out that path-connectedness implies connectedness.

Theorem 3.98

Let $(X, \mathcal{T})$ be a path connected topological space. Then X is connected.

Proof

Suppose that $X = A \cup B$ with $A, B \in \mathcal{T}$ and non-empty. In order to conclude that X is connected, we need to show that

$$A \cap B \neq \emptyset \,.$$

Since A and B are non-empty, we can find two points $x \in A$ and $b \in B$. As X is path connected, there exists $\alpha \colon [0, 1] \to X$ continuous such that

$$\alpha(0) = x \,, \quad \alpha(1) = y \,.$$

In particular

$$\alpha^{-1}(A) \neq \emptyset \,, \quad \alpha^{-1}(B) \neq \emptyset \,.$$

Moreover

$$\begin{aligned} [0, 1] & = \alpha^{-1}(X) \\ & = \alpha^{-1}(A \cup B) \\ & = \alpha^{-1}(A) \cup \alpha^{-1}(B) \,. \end{aligned}$$

As α is continuous, $\alpha^{-1}(A)$ and $\alpha^{-1}(B)$ are open in $[0, 1]$. Suppose by contradiction that $A \cap B = \emptyset$. Then

$$\alpha^{-1}(A) \cap \alpha^{-1}(B) = \alpha^{-1}(A \cap B) = \alpha^{-1}(\emptyset) = \emptyset \,.$$

Hence $[0, 1]$ is disconnected, which is a contradiction. Therefore $A \cap B \neq \emptyset$ and X is connected.

3.15 Compactness

4 Surfaces

Definition 4.1

A set $\mathcal{S} \subset \mathbb{R}^3$ is a **surface** if for every point $\mathbf{p} \in \mathcal{S}$ there exist open sets $U \subset \mathbb{R}^2$, $V \subset \mathbb{R}^3$ such that

- $\mathbf{p} \in V$,
- U is diffeomorphic to $V \cap \mathcal{S}$.

Further:

- A diffeomorphism of U into $V \cap \mathcal{S}$, denoted by
$$\boldsymbol{\sigma} : U \to V \cap \mathcal{S}$$
is called a **surface chart**.
- For each $i \in I$ suppose to have a surface chart
$$\boldsymbol{\sigma}_i : U_i \to V_i \cap \mathcal{S} \,.$$
We say that the family $\{\boldsymbol{\sigma}_i\}_{i \in I}$ is an **atlas** of $\mathcal{S}$ if
$$\bigcup_{i \in I} (V_i \cap \mathcal{S}) = \mathcal{S} \,.$$

Note that a surface chart $\boldsymbol{\sigma}$ is a map from $\mathbb{R}^2$ into $\mathbb{R}^3$. Points in U will be denoted by the pair (u, v), while points on $\mathcal{S}$ by $\mathbf{p}$ and points in $\mathbb{R}^3$ by $\mathbf{x}$.

Definition 4.2

Let $U \subset \mathbb{R}^2$ be open A surface chart
$$\boldsymbol{\sigma} = \boldsymbol{\sigma}(u, v) : U \to \mathbb{R}^3$$
is called **regular** if the partial derivatives
$$\boldsymbol{\sigma}_u(u, v) = \frac{d\boldsymbol{\sigma}}{du}(u, v) \,, \quad \boldsymbol{\sigma}_v(u, v) = \frac{d\boldsymbol{\sigma}}{dv}(u, v)$$
are linearly independent vectors of $\mathbb{R}^3$ for all $(u, v) \in U$.

Definition 4.3

Let $U \subset \mathbb{R}^2$ be open. A surface chart

$$\boldsymbol{\sigma} \colon U \to \mathbb{R}^3$$

is called a **conformal parametrization** if the first fundamental form satisfies

$$E\,du^2 + F\,du\,dv + G\,dv^2 = \lambda(u,v)(du^2 + dv^2)$$

for some smooth function $\lambda \colon U \to \mathbb{R}$.

Definition 4.4

Let $\mathcal{S}$ be a surface and $\boldsymbol{\sigma} \colon U \to V \cap \mathcal{S}$ a surface chart. The principal unit normal to $\mathcal{S}$ is

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|}\,.$$

Denote by

$$\mathbb{S}^2 = \{\mathbf{x} \in \mathbb{R}^3 \ : \ \ \|\mathbf{x}\| = 1\}$$

the unit sphere in $\mathbb{R}^3$.

Definition 4.5

Let $\mathcal{S}$ be a surface and denote by $\mathbf{N}$ the principal unit normal. The **Gauss map** of $\mathcal{S}$ is the map $G \colon \ \mathcal{S} \to \mathbb{S}^2$ defined by

$$G(\mathbf{p}) \ := \mathbf{N}(\mathbf{p})\,.$$

The **Weingarten map** is the derivative of the Gauss map, denoted by $\mathscr{W} \ : \ T_{\mathbf{p}}\mathcal{S} \to T_{G(\mathbf{p})}\mathbb{S}^2$.

5 Plots with Python

5.1 Curves in Python

5.1.1 Curves in 2D

Suppose we want to plot the parabola $y = t^2$ for t in the interval $[-3, 3]$. In our language, this is the two-dimensional curve

$$\boldsymbol{\gamma}(t) = (t, t^2), \quad t \in [-3, 3].$$

The two Python libraries we use to plot $\boldsymbol{\gamma}$ are **numpy** and **matplotlib**. In short, **numpy** handles multi-dimensional arrays and matrices, and can perform high-level mathematical functions on them. For any question you may have about numpy, answers can be found in the searchable documentation available here. Instead **matplotlib** is a plotting library, with documentation here. Python libraries need to be imported every time you want to use them. In our case we will import:

```
import numpy as np
import matplotlib.pyplot as plt
```

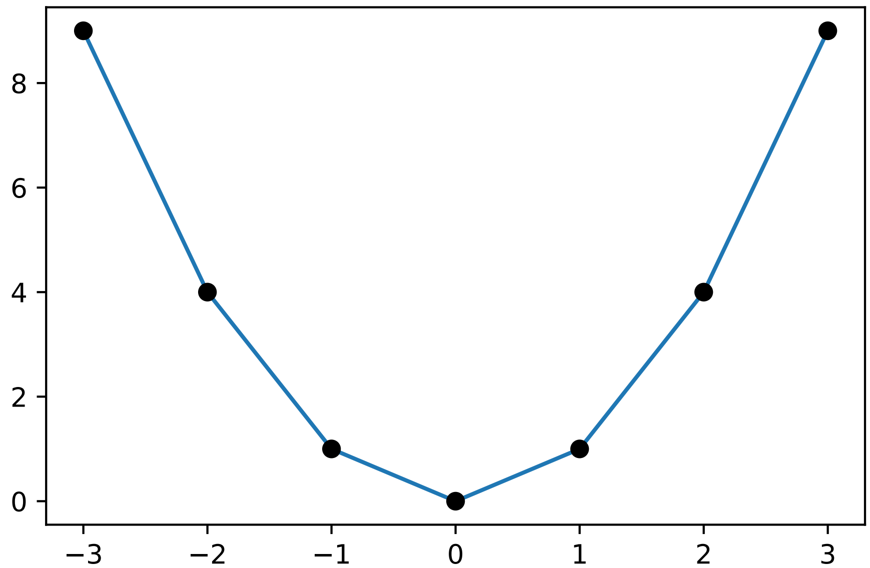
The above imports **numpy** and the module **pyplot** from **matplotlib**, and renames them to `np` and `plt`, respectively. These shorthands are standard in the literature, and they make code much more readable.
The function for plotting 2D graphs is called `plot(x,y)` and is contained in `plt`. As the syntax suggests, `plot` takes as arguments two arrays

$$x = [x_1, \dots, x_n], \quad y = [y_1, \dots, y_n].$$

As output it produces a graph which is the linear interpolation of the points (x_i, y_i) in $\mathbb{R}^2$, that is, consecutive points (x_i, y_i) and (x_{i+1}, y_{i+1}) are connected by a segment. Using `plot`, we can graph the curve $\boldsymbol{\gamma}(t) = (t, t^2)$ like so:

```
# Code for plotting gamma

import numpy as np
import matplotlib.pyplot as plt

# Generating array t
t = np.array([-3,-2,-1,0,1,2,3])

# Computing array f
```

```
f = t**2

# Plotting the curve
plt.plot(t,f)

# Plotting dots
plt.plot(t,f,"ko")

# Showing the plot
plt.show()
```



Let us comment the above code. The variable `t` is a numpy array containing the ordered values

$$t = [-3, -2, -1, 0, 1, 2, 3]\,. \tag{5.1}$$

This array is then squared entry-by-entry via the operation t `**2` and saved in the new numpy array `f`, that is,

$$f = [9, 4, 1, 0, 1, 4, 9]\,.$$

The arrays `t` and `f` are then passed to `plot(t,f)`, which produces the above linear interpolation, with `t` on the *x-axis* and `f` on the *y-axis*. The command `plot(t,f,'ko')` instead plots a black dot at each point (t_i, f_i). The latter is clearly not needed to obtain a plot, and it was only included to highlight the fact that `plot` is actually producing a linear interpolation between points. Finally `plt.show()` displays the figure in the user window[1].

Of course one can refine the plot so that it resembles the continuous curve $\boldsymbol{\gamma}(t) = (t, t^2)$ that we all have in mind. This is achieved by generating a numpy array `t` with a finer stepsize, invoking the function

[1]The command `plt.show()` can be omitted if working in Jupyter Notebook, as it is called by default.

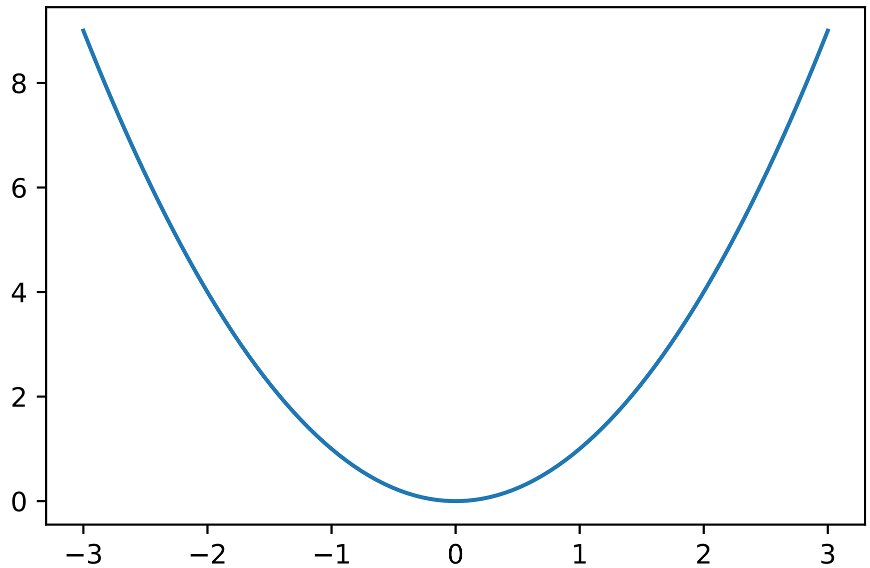
`np.linspace(a,b,n)`. Such call will return a numpy array which contains `n` evenly spaced points, starts at `a`, and ends in `b`. For example `np.linspace(-3,3,7)` returns our original array `t` at 5.1, as shown below

```
# Displaying output of np.linspace

import numpy as np

# Generates array t by dividing interval
# (-3,3) in 7 parts
t = np.linspace(-3,3, 7)

# Prints array t
print("t =", t)
```

```
t = [-3. -2. -1.  0.  1.  2.  3.]
```

In order to have a more refined plot of $\boldsymbol{\gamma}$, we just need to increase n.

```
# Plotting gamma with finer step-size

import numpy as np
import matplotlib.pyplot as plt

# Generates array t by dividing interval
# (-3,3) in 100 parts
t = np.linspace(-3,3, 100)

# Computes f
f = t**2

# Plotting
plt.plot(t,f)
plt.show()
```



We now want to plot a parametric curve $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^2$ with

$$\boldsymbol{\gamma}(t) = (x(t), y(t)) \,.$$

Clearly we need to modify the above code. The variable `t` will still be a numpy array produced by `linspace`. We then need to introduce the arrays x and y which ecode the first and second components of $\boldsymbol{\gamma}$, respectively.

```
import numpy as np
import matplotlib.pyplot as plt

# Divides time interval (a,b) in n parts
# and saves output to numpy array t
t = np.linspace(a, b, n)

# Computes gamma from given functions x(y) and y(t)
x = x(t)
y = y(t)

# Plots the curve
plt.plot(x,y)

# Shows the plot
plt.show()
```

We use the above code to plot the 2D curve known as the Fermat's spiral

$$\boldsymbol{\gamma}(t) = (\sqrt{t}\cos(t), \sqrt{t}\sin(t)) \quad \text{for} \quad t \in [0, 50] \,. \tag{5.2}$$

```
# Plotting Fermat's spiral

import numpy as np
import matplotlib.pyplot as plt

# Divides time interval (0,50) in 500 parts
t = np.linspace(0, 50, 500)

# Computes Fermat's Spiral
x = np.sqrt(t) * np.cos(t)
y = np.sqrt(t) * np.sin(t)

# Plots the Spiral
plt.plot(x,y)
plt.show()
```

Before displaying the output of the above code, a few comments are in order. The array `t` has size 500, due to the behavior of `linspace`. You can also fact check this information by printing `np.size(t)`, which is the numpy function that returns the size of an array. We then use the numpy function `np.sqrt` to compute the square root of the array `t`. The outcome is still an array with the same size of `t`, that is,

$$t = [t_1, \dots, t_n] \quad \implies \quad \sqrt{t} = [\sqrt{t_1}, \dots, \sqrt{t_n}] \,.$$

Similary, the call `np.cos(t)` returns the array

$$\cos(t) = [\cos(t_1), \dots, \cos(t_n)] \,.$$

The two arrays `np.sqrt(t)` and `np.cos(t)` are then multiplied, term-by-term, and saved in the array `x`. The array `y` is computed similarly. The command `plt.plot(x,y)` then yields the graph of the Fermat's spiral:

The above plots can be styled a bit. For example we can give a title to the plot, label the axes, plot the spiral by means of green dots, and add a plot legend, as coded below:

```
# Adding some style

import numpy as np
import matplotlib.pyplot as plt

# Computing Spiral
t = np.linspace(0, 50, 500)
x = np.sqrt(t) * np.cos(t)
y = np.sqrt(t) * np.sin(t)

# Generating figure
plt.figure(1, figsize = (4,4))
```

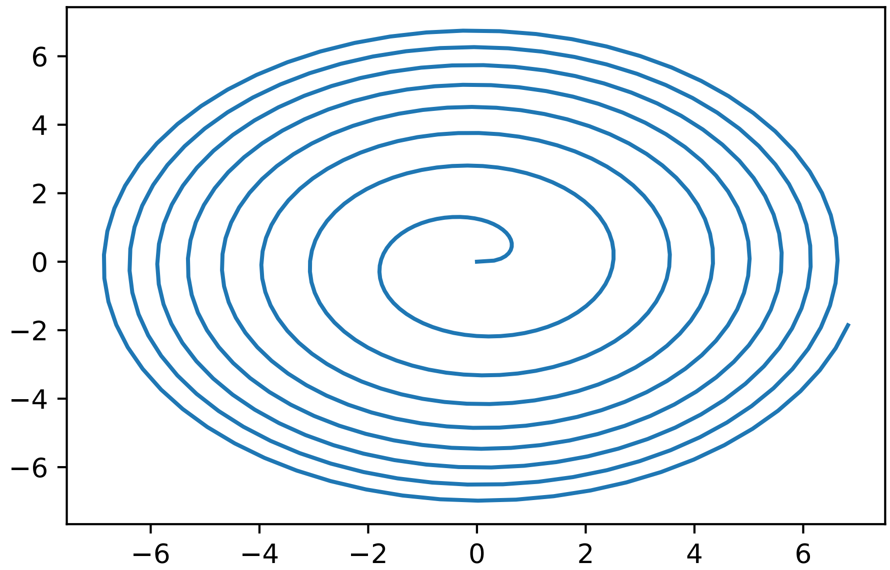


Figure 5.1: Fermat's spiral

```
# Plotting the Spiral with some options
plt.plot(x, y, '--', color = 'deeppink', linewidth = 1.5, label = 'Spiral')

# Adding grid
plt.grid(True, color = 'lightgray')

# Adding title
plt.title("Fermat's spiral for t between 0 and 50")

# Adding axes labels
plt.xlabel("x-axis", fontsize = 15)
plt.ylabel("y-axis", fontsize = 15)

# Showing plot legend
plt.legend()

# Show the plot
plt.show()
```

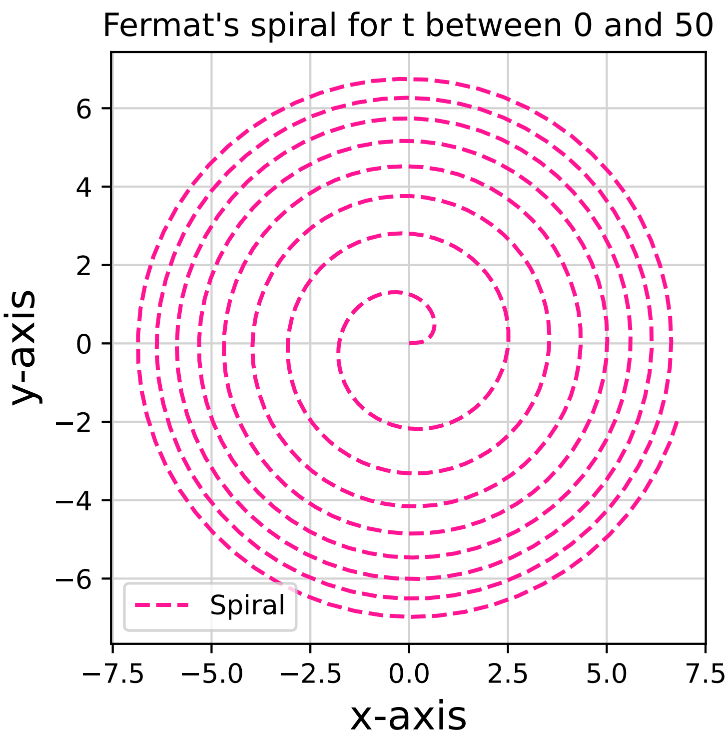


Figure 5.2: Adding a bit of style

Let us go over the novel part of the above code:

- `plt.figure()`: This command generates a figure object. If you are planning on plotting just one figure at a time, then this command is optional: a figure object is generated implicitly when calling `plt.plot`. Otherwise, if working with `n` figures, you need to generate a figure object with `plt.figure(i)` for each `i` between `1` and `n`. The number `i` uniquely identifies the i-th figure: whenever you call `plt.figure(i)`, Python knows that the next commands will refer to the i-th figure. In our case we only have one figure, so we have used the identifier `1`. The second argument `figsize = (a,b)` in `plt.figure()` specifies the size of `figure 1` in inches. In this case we generated a figure 4 x 4 inches.
- `plt.plot`: This is plotting the arrays `x` and `y`, as usual. However we are adding a few aestethic touches: the curve is plotted in *dashed* style with `--`, in *deep pink* color and with a line width of 1.5. Finally this plot is labelled *Spiral*.
- `plt.grid`: This enables a grid in *light gray* color.
- `plt.title`: This gives a title to the figure, displayed on top.
- `plt.xlabel` and `plt.ylabel`: These assign labels to the axes, with font size 15 points.
- `plt.legend()`: This plots the legend, with all the labels assigned in the `plt.plot` call. In this case the only label is *Spiral*.

Matplotlib styles

There are countless plot types and options you can specify in **matplotlib**, see for example the Matplotlib Gallery. Of course there is no need to remember every single command: a quick Google search can do wonders.

Generating arrays

There are several ways of generating evenly spaced arrays in Python. For example the function `np.arange(a,b,s)` returns an array with values within the half-open interval $[a, b)$, with spacing between values given by `s`. For example

```
import numpy as np

t = np.arange(0,1, 0.2)
print("t =",t)
```

```
t = [0.  0.2 0.4 0.6 0.8]
```

5.1.2 Implicit curves 2D

A curve $\boldsymbol{\gamma}$ in $\mathbb{R}^2$ can also be defined as the set of points $(x, y) \in \mathbb{R}^2$ satisfying

$$f(x, y) = 0$$

for some given $f : \mathbb{R}^2 \to \mathbb{R}$. For example let us plot the curve $\boldsymbol{\gamma}$ implicitly defined by

$$f(x, y) = (3x^2 - y^2)^2\, y^2 - (x^2 + y^2)^4$$

for $-1 \leq x, y \leq 1$. First, we need a way to generate a grid in $\mathbb{R}^2$ so that we can evaluate f on such grid. To illustrate how to do this, let us generate a grid of spacing 1 in the 2D square $[0, 4]^2$. The goal is to obtain the 5 x 5 matrix of coordinates

$$A = \begin{pmatrix} (0,0) & (1,0) & (2,0) & (3,0) & (4,0) \\ (0,1) & (1,1) & (2,1) & (3,1) & (4,1) \\ (0,2) & (1,2) & (2,2) & (2,3) & (2,4) \\ (0,3) & (1,3) & (2,3) & (3,3) & (3,4) \\ (0,4) & (1,4) & (2,4) & (3,4) & (4,4) \end{pmatrix}$$

which corresponds to the grid of points

To achieve this, first generate x and y coordinates using

```
x = np.linspace(0, 4, 5)
y = np.linspace(0, 4, 5)
```

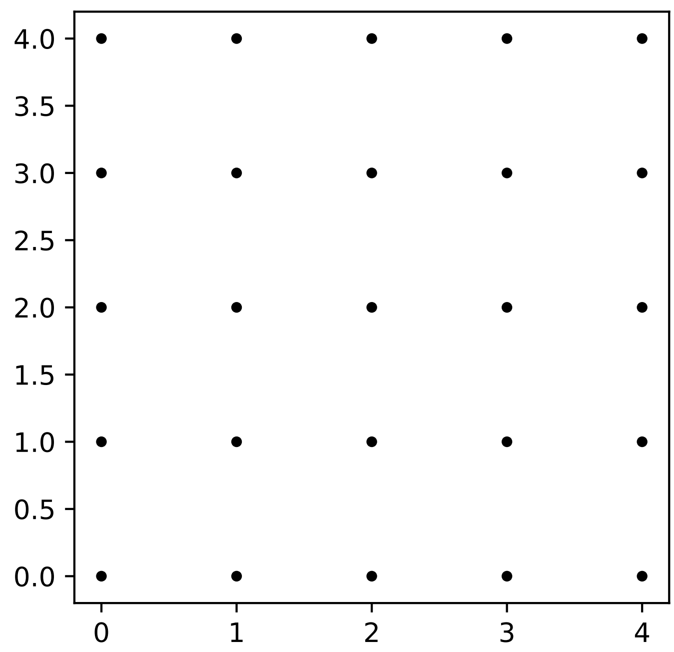


Figure 5.3: The 5 x 5 grid corresponding to the matrix A

This generates coordinates

$$x = [0, 1, 2, 3, 4] \,, \quad y = [0, 1, 2, 3, 4] \,.$$

We then need to obtain two matrices X and Y: one for the x coordinates in A, and one for the y coordinates in A. This can be achieved with the code

```
X[0,0] = 0
X[0,1] = 1
X[0,2] = 2
X[0,3] = 3
X[0,4] = 4
X[1,0] = 0
X[1,1] = 1
...
x[4,3] = 3
x[4,4] = 4
```

and similarly for Y. The output would be the two matrices X and Y

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix} \,, \quad Y = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 \end{pmatrix}$$

If now we plot X against Y via the command

```
plt.plot(X, Y, 'k.')
```

we obtain Figure 5.3. In the above command the style `'k.'` represents black dots. This procedure would be impossible with large vectors. Thankfully there is a function in numpy doing exactly what we need: `np.meshgrid`.

```
# Demonstrating np.meshgrid

import numpy as np

# Generating x and y coordinates
xlist = np.linspace(0, 4, 5)
ylist = np.linspace(0, 4, 5)

# Generating grid X, Y
X, Y = np.meshgrid(xlist, ylist)

# Printing the matrices X and Y
# np.array2string is only needed to align outputs
print('X =', np.array2string(X, prefix='X= '))
print('\n')
print('Y =', np.array2string(Y, prefix='Y= '))
```

```
X = [[0. 1. 2. 3. 4.]
    [0. 1. 2. 3. 4.]
    [0. 1. 2. 3. 4.]
    [0. 1. 2. 3. 4.]
    [0. 1. 2. 3. 4.]]


Y = [[0. 0. 0. 0. 0.]
    [1. 1. 1. 1. 1.]
    [2. 2. 2. 2. 2.]
    [3. 3. 3. 3. 3.]
    [4. 4. 4. 4. 4.]]
```

Now that we have our grid, we can evaluate the function f on it. This is simply done with the command

```
Z =((3*(X**2) - Y**2)**2)*(Y**2) - (X**2 + Y**2)**4
```

This will return the matrix Z containing the values $f(x_i, y_i)$ for all (x_i, y_i) in the grid $[X, Y]$. We are now interested in plotting the points in the grid $[X, Y]$ for which Z is zero. This is achieved with the command

```
plt.contour(X, Y, Z, [0])
```

Putting the above observations together, we have the code for plotting the curve $f = 0$ for $-1 \leq x, y \leq 1$.

```
# Plotting f=0

import numpy as np
import matplotlib.pyplot as plt

# Generates coordinates and grid
xlist = np.linspace(-1, 1, 5000)
ylist = np.linspace(-1, 1, 5000)
X, Y = np.meshgrid(xlist, ylist)

# Computes f
Z =((3*(X**2) - Y**2)**2)*(Y**2) - (X**2 + Y**2)**4

# Creates figure object
plt.figure(figsize = (4,4))

# Plots level set Z = 0
plt.contour(X, Y, Z, [0])

# Set axes labels
plt.xlabel("x-axis", fontsize = 15)
plt.ylabel("y-axis", fontsize = 15)

# Shows plot
plt.show()
```

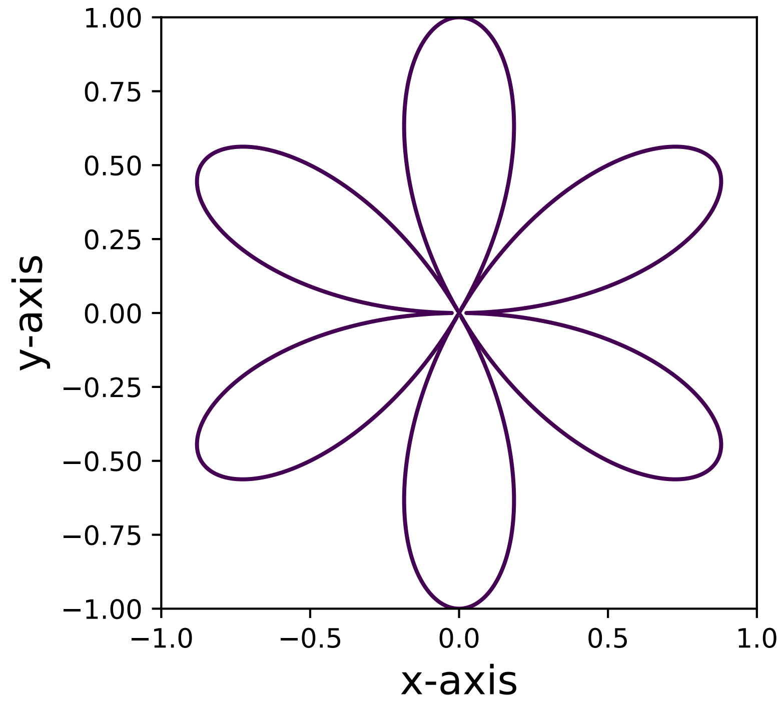


Figure 5.4: Plot of the curve defined by f=0

5.1.3 Curves in 3D

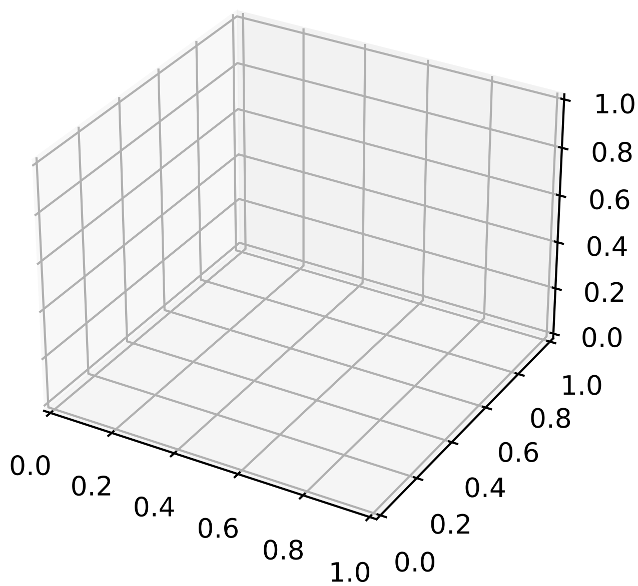
Plotting in 3D with matplotlib requires the `mplot3d` toolkit, see here for documentation. Therefore our first lines will always be

```
# Packages for 3D plots

import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d
```

We can now generate empty 3D axes

```
# Generates and plots empty 3D axes

import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

# Creates figure object
fig = plt.figure(figsize = (4,4))
```

```
# Creates 3D axes object
ax = plt.axes(projection = '3d')

# Shows the plot
plt.show()
```



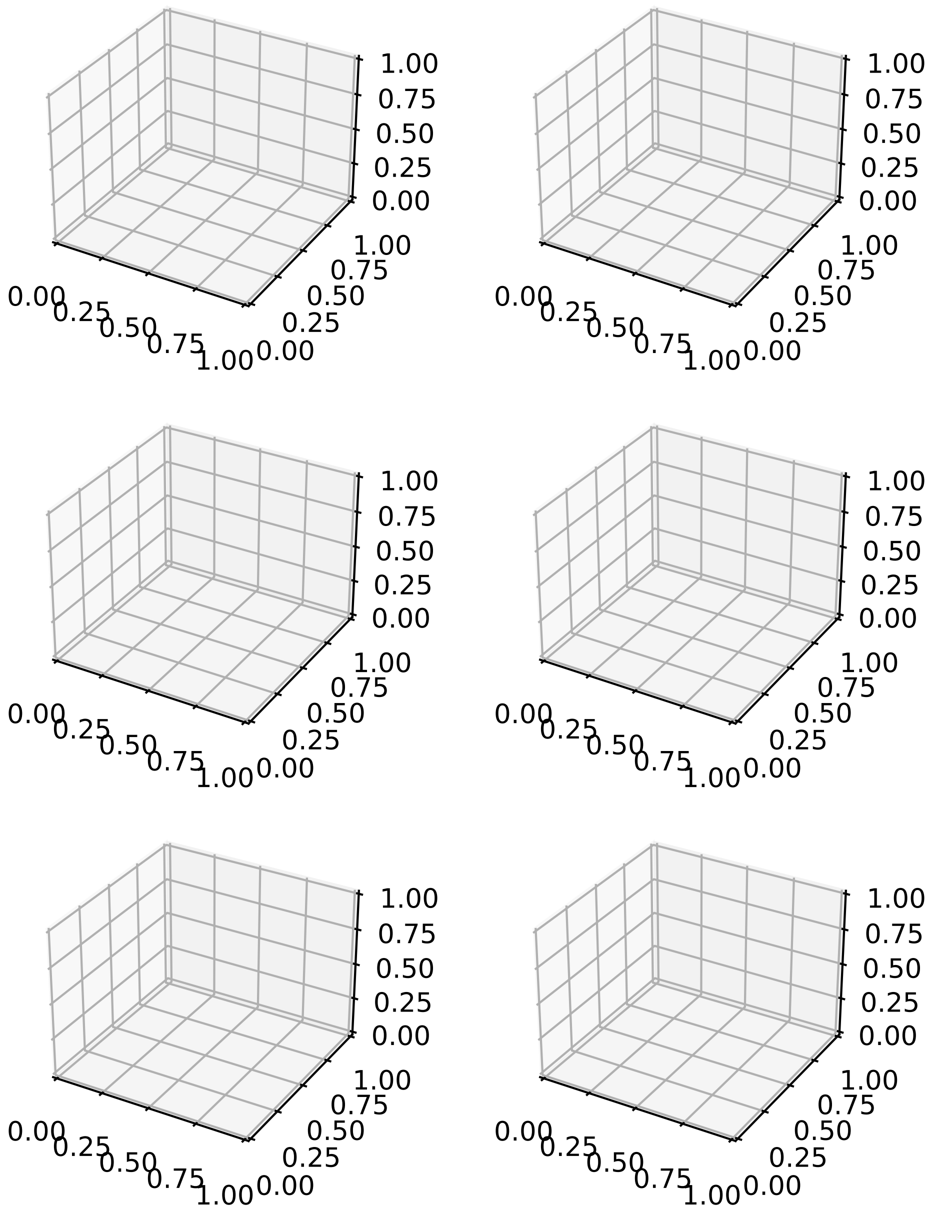
In the above code `fig` is a figure object, while `ax` is an axes object. In practice, the figure object contains the axes objects, and the actual plot information will be contained in axes. If you want multiple plots in the figure container, you should use the command

```
ax = fig.add_subplot(nrows = m, ncols = n, pos = k)
```

This generates an axes object `ax` in position `k` with respect to a `m x n` grid of plots in the container figure. For example we can create a 3 x 2 grid of empty 3D axes as follows

```
# Generates 3 x 2 empty 3D axes

import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

# Creates container figure object
fig = plt.figure(figsize = (6,8))
```

```
# Creates 6 empty 3D axes objects
ax1 = fig.add_subplot(3, 2, 1, projection = '3d')
ax2 = fig.add_subplot(3, 2, 2, projection = '3d')
ax3 = fig.add_subplot(3, 2, 3, projection = '3d')
ax4 = fig.add_subplot(3, 2, 4, projection = '3d')
ax5 = fig.add_subplot(3, 2, 5, projection = '3d')
ax6 = fig.add_subplot(3, 2, 6, projection = '3d')

# Shows the plot
plt.show()
```



We are now ready to plot a 3D parametric curve $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ of the form

$$\boldsymbol{\gamma}(t) = (x(t), y(t), z(t))$$

with the code

```
# Code to plot 3D curve

import numpy as np
import matplotlib.pyplot as plt
```

```
from mpl_toolkits import mplot3d

# Generates figure and 3D axes
fig = plt.figure(figsize = (size1,size2))
ax = plt.axes(projection = '3d')

# Plots grid
ax.grid(True)

# Divides time interval (a,b)
# into n parts and saves them in array t
t = np.linspace(a, b, n)

# Computes the curve gamma on array t
# for given functions x(t), y(t), z(t)
x = x(t)
y = y(t)
z = z(t)

# Plots gamma
ax.plot3D(x, y, z)

# Setting title for plot
ax.set_title('3D Plot of gamma')

# Setting axes labels
ax.set_xlabel('x', labelpad = 'p')
ax.set_ylabel('y', labelpad = 'p')
ax.set_zlabel('z', labelpad = 'p')

# Shows the plot
plt.show()
```

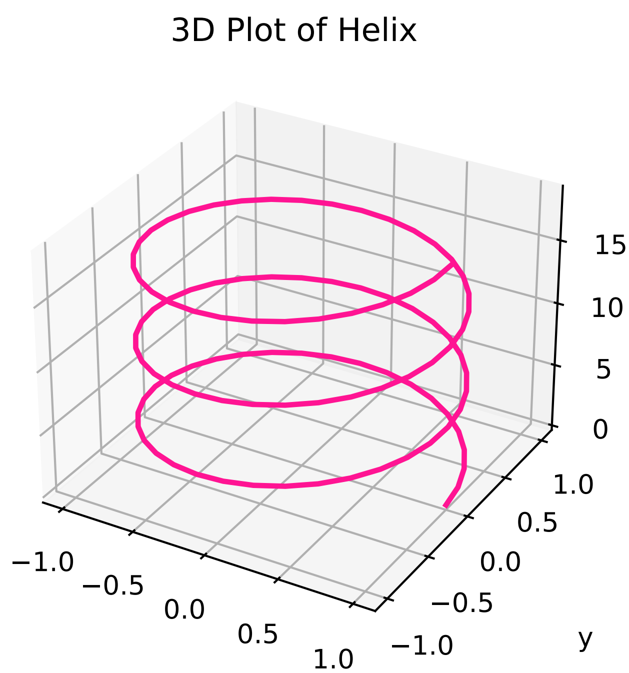
For example we can use the above code to plot the Helix

$$x(t) = \cos(t)\,, \quad y(t) = \sin(t)\,, \quad z(t) = t \tag{5.3}$$

for $t \in [0, 6\pi]$.

```
# Plotting 3D Helix

import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d
```

```
# Generates figure and 3D axes
fig = plt.figure(figsize = (4,4))
ax = plt.axes(projection = '3d')

# Plots grid
ax.grid(True)

# Divides time interval (0,6pi) in 100 parts
t = np.linspace(0, 6*np.pi, 100)

# Computes Helix
x = np.cos(t)
y = np.sin(t)
z = t

# Plots Helix - We added some styling
ax.plot3D(x, y, z, color = "deeppink", linewidth = 2)

# Setting title for plot
ax.set_title('3D Plot of Helix')

# Setting axes labels
ax.set_xlabel('x', labelpad = 20)
ax.set_ylabel('y', labelpad = 20)
ax.set_zlabel('z', labelpad = 20)

# Shows the plot
plt.show()
```



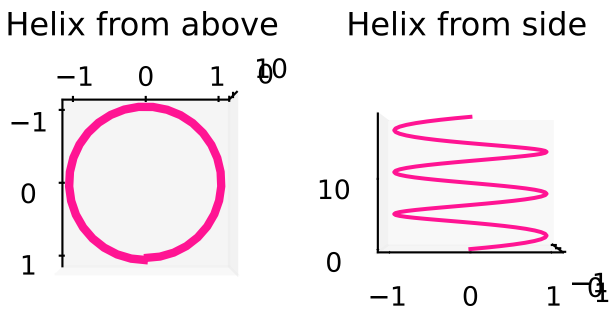
We can also change the viewing angle for a 3D plot store in ax. This is done via

```
ax.view_init(elev = e, azim = a)
```

which displays the 3D axes with an elevation angle `elev` of `e` degrees and an azimuthal angle `azim` of a degrees. In other words, the 3D plot will be rotated by e degrees above the xy-plane and by a degrees around the z-axis. For example, let us plot the helix with 2 viewing angles. Note that we generate 2 sets of axes with the `add_subplot` command discussed above.

```
# Plotting 3D Helix

import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

# Generates figure object
fig = plt.figure(figsize = (4,4))

# Generates 2 sets of 3D axes
ax1 = fig.add_subplot(1, 2, 1, projection = '3d')
ax2 = fig.add_subplot(1, 2, 2, projection = '3d')

# We will not show a grid this time
ax1.grid(False)
```

```
ax2.grid(False)

# Divides time interval (0,6pi) in 100 parts
t = np.linspace(0, 6*np.pi, 100)

# Computes Helix
x = np.cos(t)
y = np.sin(t)
z = t

# Plots Helix on both axes
ax1.plot3D(x, y, z, color = "deeppink", linewidth = 1.5)
ax2.plot3D(x, y, z, color = "deeppink", linewidth = 1.5)

# Setting title for plots
ax1.set_title('Helix from above')
ax2.set_title('Helix from side')

# Changing viewing angle of ax1
# View from above has elev = 90 and azim = 0
ax1.view_init(elev = 90, azim = 0)

# Changing viewing angle of ax2
# View from side has elev = 0 and azim = 0
ax2.view_init(elev = 0, azim = 0)

# Shows the plot
plt.show()
```



5.1.4 Interactive plots

`Matplotlib` produces beautiful static plots; however it lacks built in interactivity. For this reason I would also like to show you how to plot curves with `Plotly`, a very popular Python graphic library which has built in interactivity. Documentation for `Plotly` and lots of examples can be found here.

5.1.4.1 2D Plots

Say we want to plot the 2D curve $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^2$ parametrized by

$$\boldsymbol{\gamma}(t) = (x(t), y(t)) .$$

The `Plotly` module needed is called `graph_objects`, usually imported as `go`. The function for line plots is called `Scatter`. For documentation and examples see link. The code for plotting $\boldsymbol{\gamma}$ is as follows.

```
# Plotting gamma 2D

# Import libraries
import numpy as np
import plotly.graph_objects as go

# Compute times grid by dividing (a,b) in
# n equal parts
t = np.linspace(a, b, n)

# Compute the parametric curve gamma
# for given functions x(t) and y(t)
x = x(t)
y = y(t)

# Create empty figure object and saves
# it in the variable "fig"
fig = go.Figure()

# Create the line plot object
data = go.Scatter(x = x, y = y, mode = 'lines', name = 'gamma')

# Add "data" plot to the figure "fig"
fig.add_trace(data)

# Display the figure
fig.show()
```

Some comments about the functions called above:

- `go.Figure`: generates an empty Plotly figure
- `go.Scatter`: generates the actual plot. By default a scatter plot is produced. To obtain linear interpolation of the points, set `mode = 'lines'`. You can also label the plot with `name = "string"`
- `add_trace`: adds a plot to a figure
- `show`: displays a figure

As an example, let us plot the Fermat's Spiral defined at 5.2. Compared to the above code, we also add a bit of styling.

```
# Plotting Fermat's Spiral

# Import libraries
import numpy as np
import plotly.graph_objects as go

# Compute times grid by dividing (0,50) in
# 500 equal parts
t = np.linspace(0, 50, 500)

# Computes Fermat's Spiral
x = np.sqrt(t) * np.cos(t)
y = np.sqrt(t) * np.sin(t)

# Create empty figure object and saves
# it in the variable "fig"
fig = go.Figure()

# Create the line plot object
data = go.Scatter(x = x, y = y, mode = 'lines', name = 'gamma')

# Add "data" plot to the figure "fig"
fig.add_trace(data)

# Here we start with the styling options
# First we set a figure title
fig.update_layout(title_text = "Plotting Fermat's Spiral with Plotly")

# Adjust figure size
fig.update_layout(autosize = False, width = 600, height = 600)

# Change background canvas color
fig.update_layout(paper_bgcolor = "snow")

# Axes styling: adding title and ticks positions
```

```
fig.update_layout(
xaxis=dict(
        title_text="X-axis Title",
        titlefont=dict(size=20),
        tickvals=[-6,-4,-2,0,2,4,6],
        ),

yaxis=dict(
        title_text="Y-axis Title",
        titlefont=dict(size=20),
        tickvals=[-6,-4,-2,0,2,4,6],
        )
)

# Display the figure
fig.show()
```

```
Unable to display output for mime type(s): text/html
```

```
Unable to display output for mime type(s): text/html
```

The above code generates an image that cannot be rendered in pdf. To see the output, please click here for the digital version of these notes. Note that the style customizations could be listed in a single call of the function `update_layout`. There are also pretty buit-in themes available, see here. The layout can be specified with the command

```
fig.update_layout(template = template_name)
```

where `template_name` can be `"plotly"`, `"plotly_white"`, `"plotly_dark"`, `"ggplot2"`, `"seaborn"`, `"simple_white`“.

5.1.4.2 3D Plots

We now want to plot a 3D curve $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ parametrized by

$$\boldsymbol{\gamma}(t) = (x(t), y(t), z(t)) .$$

Again we use the `Plotly` module `graph_objects`, imported as `go`. The function for 3D line plots is called `Scatter3d`, and documentation and examples can be found at link. The code for plotting $\boldsymbol{\gamma}$ is as follows.

```
# Plotting gamma 3D
```

```
# Import libraries
import numpy as np
import plotly.graph_objects as go

# Compute times grid by dividing (a,b) in
# n equal parts
t = np.linspace(a, b, n)

# Compute the parametric curve gamma
# for given functions x(t), y(t), z(t)
x = x(t)
y = y(t)
z = z(t)

# Create empty figure object and saves
# it in the variable "fig"
fig = go.Figure()

# Create the line plot object
data = go.Scatter3d(x = x, y = y, z = z, mode = 'lines', name = 'gamma')

# Add "data" plot to the figure "fig"
fig.add_trace(data)

# Display the figure
fig.show()
```

The functions `go.Figure`, `add_trace` and `show` appearing above are described in the previous Section. The new addition is `go.Scatter3d`, which generates a 3D scatter plot of the points stored in the array `[x,y,z]`. Setting `mode = 'lines'` results in a linear interpolation of such points. As before, the curve can be labeled by setting `name = "string"`.

As an example, we plot the 3D Helix defined at 5.3. We also add some styling. We can also use the same predefined templates described for `go.Scatter` in the previous section, see here for official documentation.

```
# Plotting 3D Helix

# Import libraries
import numpy as np
import plotly.graph_objects as go

# Divides time interval (0,6pi) in 100 parts
t = np.linspace(0, 6*np.pi, 100)
```

```
# Computes Helix
x = np.cos(t)
y = np.sin(t)
z = t

# Create empty figure object and saves
# it in the variable "fig"
fig = go.Figure()

# Create the line plot object
# We add options for the line width and color
data = go.Scatter3d(
    x = x, y = y, z = z,
    mode = 'lines', name = 'gamma',
    line = dict(width = 10, color = "darkblue")
    )

# Add "data" plot to the figure "fig"
fig.add_trace(data)

# Here we start with the styling options
# First we set a figure title
fig.update_layout(title_text = "Plotting 3D Helix with Plotly")

# Adjust figure size
fig.update_layout(
    autosize = False,
    width = 600,
    height = 600
    )

# Set pre-defined template
fig.update_layout(template = "seaborn")

# Options for curve line style


# Display the figure
fig.show()
```

```
Unable to display output for mime type(s): text/html
```

The above code generates an image that cannot be rendered in pdf. To see the output, please click here for the digital version of these notes. Once again, the style customizations could be listed in a single call of the function `update_layout`.

5.2 Surfaces in Python

5.2.1 Plots with Matplotlib

I will take for granted all the commands explained in Section 5.1. Suppose we want to plot a surface S which is defined by the parametric equations

$$x = x(u,v)\,, \quad y = y(u,v)\,, \quad z = z(u,v)$$

for $u \in (a,b)$ and $v \in (c,d)$. This can be done via the function called `plot_surface` contained in the mplot3d Toolkit. This function works as follows: first we generate a mesh-grid $[U,V]$ from the coordinates (u,v) via the command

```
[U, V] = np.meshgrid(u, v)
```

Then we compute the parametric surface on the mesh

```
x = x (U, V)
y = y (U, V)
z = z (U, V)
```

Finally we can plot the surface with the command

```
plt.plot_surface(x, y, z)
```

The complete code looks as follows.

```
# Plotting surface S

# Importing numpy, matplotlib and mplot3d
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

# Generates figure object of size m x n
fig = plt.figure(figsize = (m,n))

# Generates 3D axes
ax = plt.axes(projection = '3d')
```

```
# Shows axes grid
ax.grid(True)

# Generates coordinates u and v
# by dividing the interval (a,b) in n parts
# and the interval (c,d) in m parts
u = np.linspace(a, b, m)
v = np.linspace(c, d, n)

# Generates grid [U,V] from the coordinates u, v
U, V = np.meshgrid(u, v)

# Computes S given the functions x, y, z
# on the grid [U,V]
x = x(U,V)
y = y(U,V)
z = z(U,V)

# Plots the surface S
ax.plot_surface(x, y, z)

# Setting plot title
ax.set_title('The surface S')

# Setting axes labels
ax.set_xlabel('x', labelpad=10)
ax.set_ylabel('y', labelpad=10)
ax.set_zlabel('z', labelpad=10)

# Setting viewing angle
ax.view_init(elev = e, azim = a)

# Showing the plot
plt.show()
```

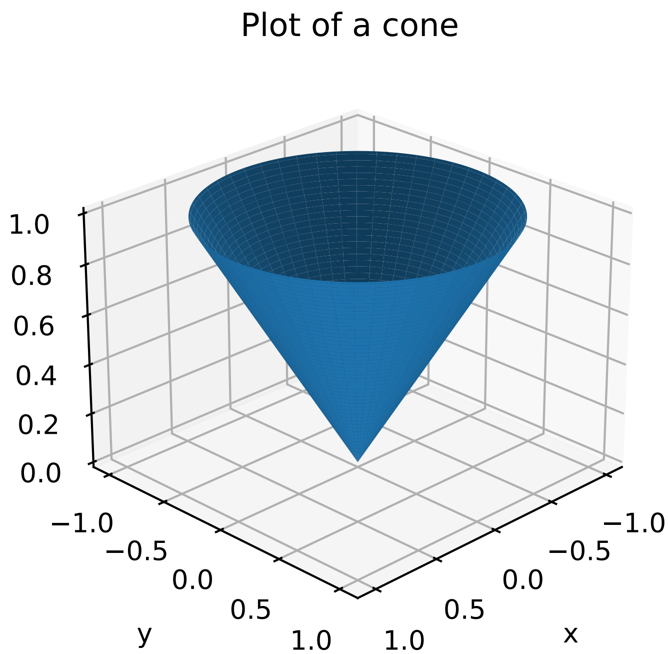
For example let us plot a cone described parametrically by:

$$x = u\cos(v), \quad y = u\sin(v), \quad z = u$$

for $u \in (0,1)$ and $v \in (0, 2\pi)$. We adapt the above code:

```
# Plotting a cone
```

```
# Importing numpy, matplotlib and mplot3d
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

# Generates figure object of size 4 x 4
fig = plt.figure(figsize = (4,4))

# Generates 3D axes
ax = plt.axes(projection = '3d')

# Shows axes grid
ax.grid(True)

# Generates coordinates u and v by dividing
# the intervals (0,1) and (0,2pi) in 100 parts
u = np.linspace(0, 1, 100)
v = np.linspace(0, 2*np.pi, 100)

# Generates grid [U,V] from the coordinates u, v
U, V = np.meshgrid(u, v)

# Computes the surface on grid [U,V]
x = U * np.cos(V)
y = U * np.sin(V)
z = U

# Plots the cone
ax.plot_surface(x, y, z)

# Setting plot title
ax.set_title('Plot of a cone')

# Setting axes labels
ax.set_xlabel('x', labelpad=10)
ax.set_ylabel('y', labelpad=10)
ax.set_zlabel('z', labelpad=10)

# Setting viewing angle
ax.view_init(elev = 25, azim = 45)

# Showing the plot
plt.show()
```



As discussed in Section 5.1, we can have multiple plots in the same figure. For example let us plot the torus viewed from 2 angles. The parametric equations are:

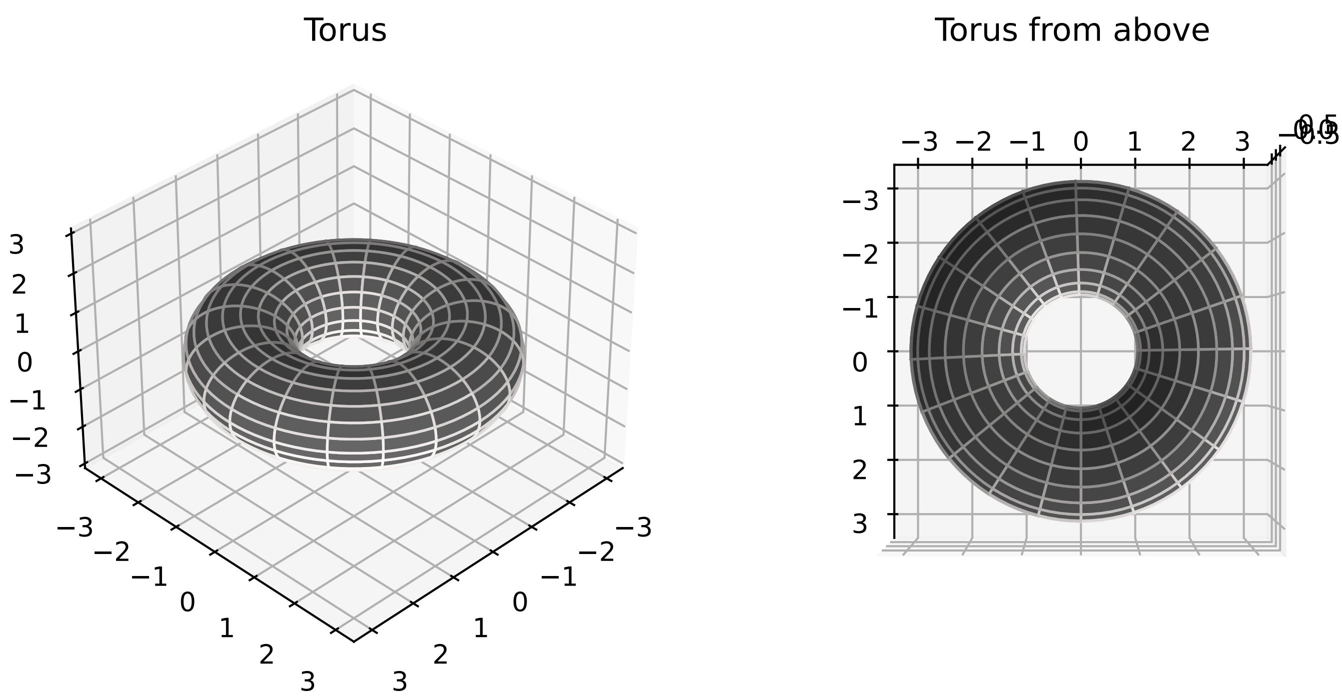
$$
\begin{aligned}
x &= (R + r\cos(u))\cos(v) \\
y &= (R + r\cos(u))\sin(v) \\
z &= r\sin(u)
\end{aligned}
$$

for $u, v \in (0, 2\pi)$ and with

- R distance from the center of the tube to the center of the torus
- r radius of the tube

```
# Plotting torus seen from 2 angles

# Importing numpy, matplotlib and mplot3d
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

# Generates figure object of size 9 x 5
fig = plt.figure(figsize = (9,5))

# Generates 2 sets of 3D axes
ax1 = fig.add_subplot(1, 2, 1, projection = '3d')
ax2 = fig.add_subplot(1, 2, 2, projection = '3d')
```

```
# Shows axes grid
ax1.grid(True)
ax2.grid(True)

# Generates coordinates u and v by dividing
# the interval (0,2pi) in 100 parts
u = np.linspace(0, 2*np.pi, 100)
v = np.linspace(0, 2*np.pi, 100)

# Generates grid [U,V] from the coordinates u, v
U, V = np.meshgrid(u, v)

# Computes the torus on grid [U,V]
# with radii r = 1 and R = 2
R = 2
r = 1

x = (R + r * np.cos(U)) * np.cos(V)
y = (R + r * np.cos(U)) * np.sin(V)
z = r * np.sin(U)

# Plots the torus on both axes
ax1.plot_surface(x, y, z, rstride = 5, cstride = 5, color = 'dimgray', edgecolors =
 ↪  'snow')

ax2.plot_surface(x, y, z, rstride = 5, cstride = 5, color = 'dimgray', edgecolors =
 ↪  'snow')

# Setting plot titles
ax1.set_title('Torus')
ax2.set_title('Torus from above')

# Setting range for z axis in ax1
ax1.set_zlim(-3,3)

# Setting viewing angles
ax1.view_init(elev = 35, azim = 45)
ax2.view_init(elev = 90, azim = 0)

# Showing the plot
plt.show()
```



Notice that we have added some customization to the `plot_surface` command. Namely, we have set the color of the figure with `color = 'dimgray'` and of the edges with `edgecolors = 'snow'`. Moreover the commands `rstride` and `cstride` set the number of *wires* you see in the plot. More precisely, they set by how much the data in the mesh $[U, V]$ is downsampled in each direction, where rstride sets the row direction, and cstride sets the column direction. On the torus this is a bit difficult to visualize, due to the fact that $[U, V]$ represents angular coordinates. To appreciate the effect, we can plot for example the paraboiloid

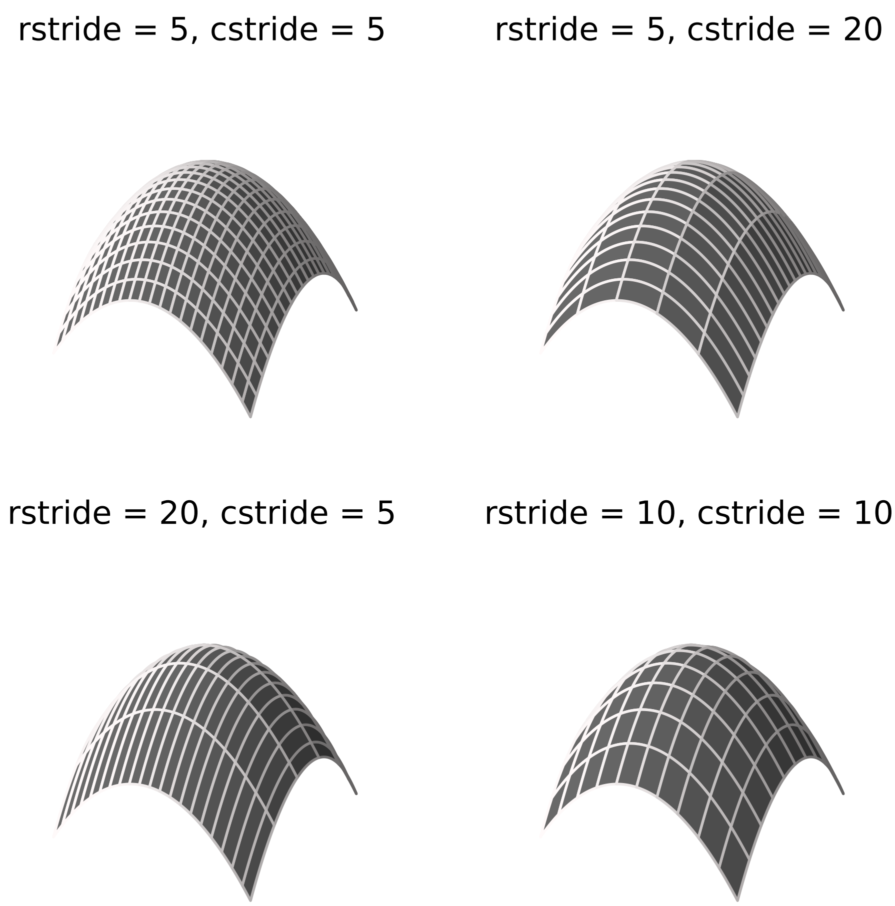
$$
\begin{aligned}
x &= u \\
y &= v \\
z &= -u^2 - v^2
\end{aligned}
$$

for $u, v \in [-1, 1]$.

```
# Showing the effect of rstride and cstride

# Importing numpy, matplotlib and mplot3d
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

# Generates figure object of size 6 x 6
fig = plt.figure(figsize = (6,6))

# Generates 2 sets of 3D axes
```

```
ax1 = fig.add_subplot(2, 2, 1, projection = '3d')
ax2 = fig.add_subplot(2, 2, 2, projection = '3d')
ax3 = fig.add_subplot(2, 2, 3, projection = '3d')
ax4 = fig.add_subplot(2, 2, 4, projection = '3d')

# Generates coordinates u and v by dividing
# the interval (-1,1) in 100 parts
u = np.linspace(-1, 1, 100)
v = np.linspace(-1, 1, 100)

# Generates grid [U,V] from the coordinates u, v
U, V = np.meshgrid(u, v)

# Computes the paraboloid on grid [U,V]
x = U
y = V
z = - U**2 - V**2

# Plots the paraboloid on the 4 axes
# but with different stride settings
ax1.plot_surface(x, y, z, rstride = 5, cstride = 5, color = 'dimgray', edgecolors = 'snow')

ax2.plot_surface(x, y, z, rstride = 5, cstride = 20, color = 'dimgray', edgecolors = 'snow')

ax3.plot_surface(x, y, z, rstride = 20, cstride = 5, color = 'dimgray', edgecolors = 'snow')

ax4.plot_surface(x, y, z, rstride = 10, cstride = 10, color = 'dimgray', edgecolors = 'snow')

# Setting plot titles
ax1.set_title('rstride = 5, cstride = 5')
ax2.set_title('rstride = 5, cstride = 20')
ax3.set_title('rstride = 20, cstride = 5')
ax4.set_title('rstride = 10, cstride = 10')

# We do not plot axes, to get cleaner pictures
ax1.axis('off')
ax2.axis('off')
ax3.axis('off')
ax4.axis('off')
```

```
# Showing the plot
plt.show()
```



In this case our mesh is `100 x 100`, since u and v both have 100 components. Therefore setting `rstride` and `cstride` to 5 implies that each row and column of the mesh is sampled one time every 5 elements, for a total of

$$100/5 = 20$$

samples in each direction. This is why in the first picture you see a `20 x 20` grid. If instead one sets `rstride` and `cstride` to 10, then each row and column of the mesh is sampled one time every 10 elements, for a total of

$$100/10 = 10$$

samples in each direction. This is why in the fourth figure you see a `10x10` grid.

5.2.2 Plots with Plotly

As done in Section 5.1.4, we now see how to use `Plotly` to generate an interactive 3D plot of a surface. This can be done by means of functions contained in the `Plotly` module `graph_objects`, usually imported as `go`. Specifically, we will use the function `go.Surface`. The code will look similar to the one used to plot surfaces with `matplotlib`:

- generate meshgrid on which to compute the parametric surface,
- store such surface in the numpy array `[x,y,z]`,
- pass the array `[x,y,z]` to `go.Surface` to produce the plot.

The full code is below.

```
# Plotting a Torus with Plotly

# Import "numpy" and the "graph_objects" module from Plotly
import numpy as np
import plotly.graph_objects as go

# Generates coordinates u and v by dividing
# the interval (0,2pi) in 100 parts
u = np.linspace(0, 2*np.pi, 100)
v = np.linspace(0, 2*np.pi, 100)

# Generates grid [U,V] from the coordinates u, v
U, V = np.meshgrid(u, v)

# Computes the torus on grid [U,V]
# with radii r = 1 and R = 2
R = 2
r = 1

x = (R + r * np.cos(U)) * np.cos(V)
y = (R + r * np.cos(U)) * np.sin(V)
z = r * np.sin(U)

# Generate and empty figure object with Plotly
# and saves it to the variable called "fig"
fig = go.Figure()

# Plot the torus with go.Surface and store it
# in the variable "data". We also do now show the
# plot scale, and set the color map to "teal"
data = go.Surface(
```

```
    x = x , y = y, z = z,
    showscale = False,
    colorscale='teal'
    )

# Add the plot stored in "data" to the figure "fig"
# This is done with the command add_trace
fig.add_trace(data)

# Set the title of the figure in "fig"
fig.update_layout(title_text="Plotting a Torus with Plotly")

# Show the figure
fig.show()
```

```
Unable to display output for mime type(s): text/html
```

The above code generates an image that cannot be rendered in pdf. To see the output, see the link to the digital version of these notes. To further customize your plots, you can check out the documentation of `go.Surface` at this link. For example, note that we have set the colormap to `teal`: for all the pretty colorscales available in Plotly, see this page.

One could go even fancier and use the tri-surf plots in `Plotly`. This is done with the function `create_trisurf` contained in the module `figure_factory` of `Plotly`, usually imported as `ff`. The documentation can be found here. We also need to import the Python library `scipy`, which we use to generate a *Delaunay triangulation* for our plot. Let us for example plot the torus.

```
# Plotting Torus with tri-surf

# Importing libraries
import numpy as np
import plotly.figure_factory as ff
from scipy.spatial import Delaunay

# Generates coordinates u and v by dividing
# the interval (0,2pi) in 100 parts
u = np.linspace(0, 2*np.pi, 20)
v = np.linspace(0, 2*np.pi, 20)

# Generates grid [U,V] from the coordinates u, v
U, V = np.meshgrid(u, v)

# Collapse meshes to 1D array
```

```
# This is needed for create_trisurf
U = U.flatten()
V = V.flatten()

# Computes the torus on grid [U,V]
# with radii r = 1 and R = 2
R = 2
r = 1

x = (R + r * np.cos(U)) * np.cos(V)
y = (R + r * np.cos(U)) * np.sin(V)
z = r * np.sin(U)

# Generate Delaunay triangulation
points2D = np.vstack([U,V]).T
tri = Delaunay(points2D)
simplices = tri.simplices

# Plot the Torus
fig = ff.create_trisurf(
    x=x, y=y, z=z,
    colormap = "Portland",
    simplices=simplices,
    title="Torus with tri-surf",
    aspectratio=dict(x=1, y=1, z=0.3),
    show_colorbar = False
    )

# Adjust figure size
fig.update_layout(autosize = False, width = 700, height = 700)

# Show the figure
fig.show()
```

```
Unable to display output for mime type(s): text/html
```

Again, the above code generates an image that cannot be rendered in pdf. To see the output, see the link to the digital version of these notes.

License

Reuse

Citation

For attribution, please cite this work as:

> Fanzon, Silvio. (2023). *Lecture Notes on Differential Geometry*.
> https://www.silviofanzon.com/2023-Differential-Geometry-Notes/

BibTex citation:

```
@electronic{Danzon-Diff-Geom-2023,
    author = {Fanzon, Silvio},
    title = {Lecture Notes on Differential Geometry},
    url = {https://www.silviofanzon.com/2023-Differential-Geometry-Notes/},
    year = {2023}}
```

References

[1] Abate, Marco and Tovena, Francesca. *Curves and Surfaces*. Springer, 2011.

[2] M. P. do Carmo. *Differential Geometry of Curves and Surfaces*. Second Edition. Dover Books on Mathematics, 2017.

[3] R. Johansson. *Numerical Python. Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib*. Second Edition. Apress, 2019.

[4] Kong, Qingkai, Siauw, Timmy, and Bayen, Alexandre. *Python Programming and Numerical Methods*. Academic Press, 2020.

[5] M. Manetti. *Topology*. Second Edition. Springer, 2023.

[6] A. Pressley. *Elementary Differential Geometry*. Second Edition. Springer, 2010.

[7] V. A. Zorich. *Mathematical Analysis I*. Second Edition. Springer, 2015.

[8] V. A. Zorich. *Mathematical Analysis II*. Second Edition. Springer, 2016.