Differential Geometry

Lecture Notes

Dr. Silvio Fanzon

23 Oct 2024

Academic Year 2024/25

Department of Mathematics

University of Hull

Table of contents

Welcome

These are the Lecture Notes of **Differential Geometry 661955** for 2024/25 at the University of Hull. I will use this material during lectures. 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/73612

Digital Notes

Digital version of these notes available at

silviofanzon.com/2024-Differential-Geometry-Notes

Readings

We will study curves and surfaces in $\mathbb{R}^3$, as well as some general topology. The main textbooks are:

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

Other good readings are the books:

- do Carmo [2], a classic and really nice textbook
- Abate, Tovena [1], for a more in depth analysis

I will assume some knowledge from Analysis and Linear Algebra. A good place to revise these topics are the books by Zorich [8, 9].

Visualization

It is important to visualize the geometrical objects and concepts we are going to talk about in this module. Chapter 5 contains a basic Python tutorial to plot curves and surfaces. This part of the notes is **`not examinable`**.

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 module using

- Desmos
- CalcPlot3D

> **!** 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 Homework and Final Exam.

1 Curves

Curves are 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

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

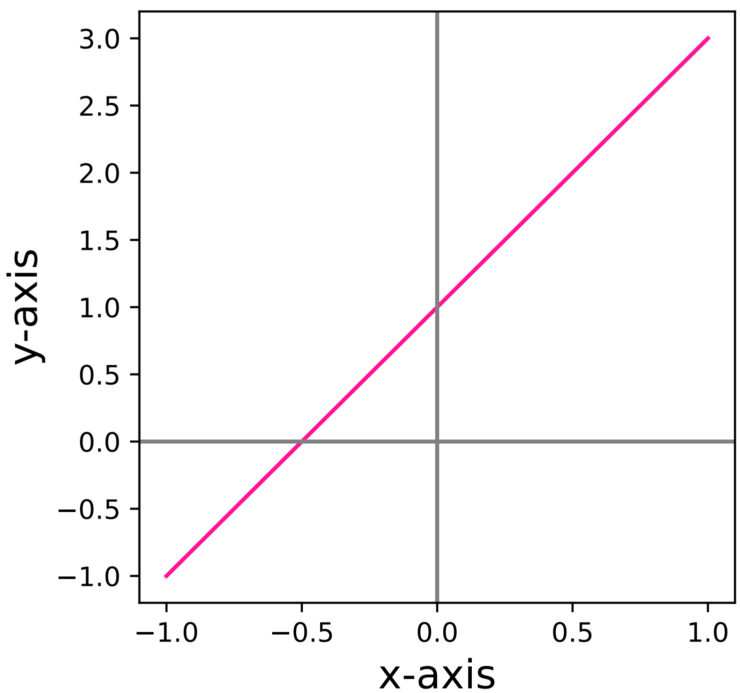


Figure 1.1: Plotting straight line $y = 2x + 1$

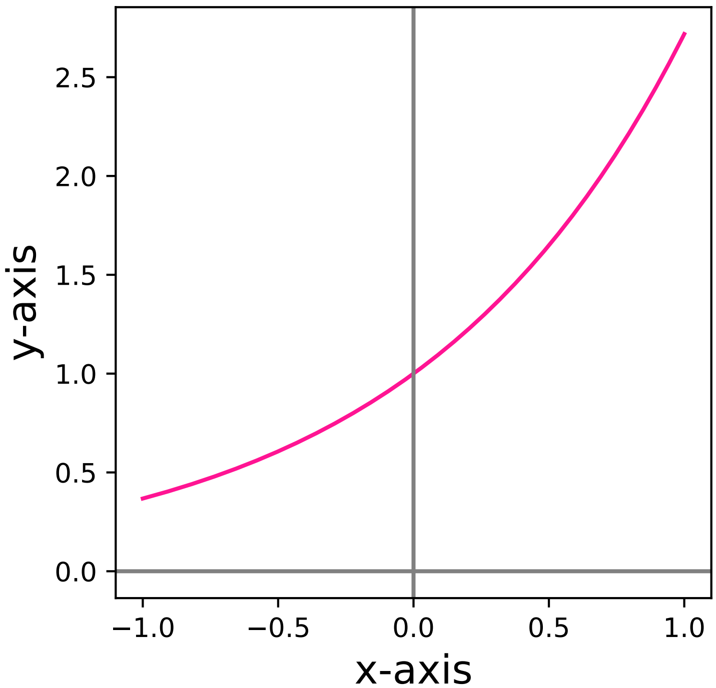


Figure 1.2: Plot of hyperbole $y = e^x$

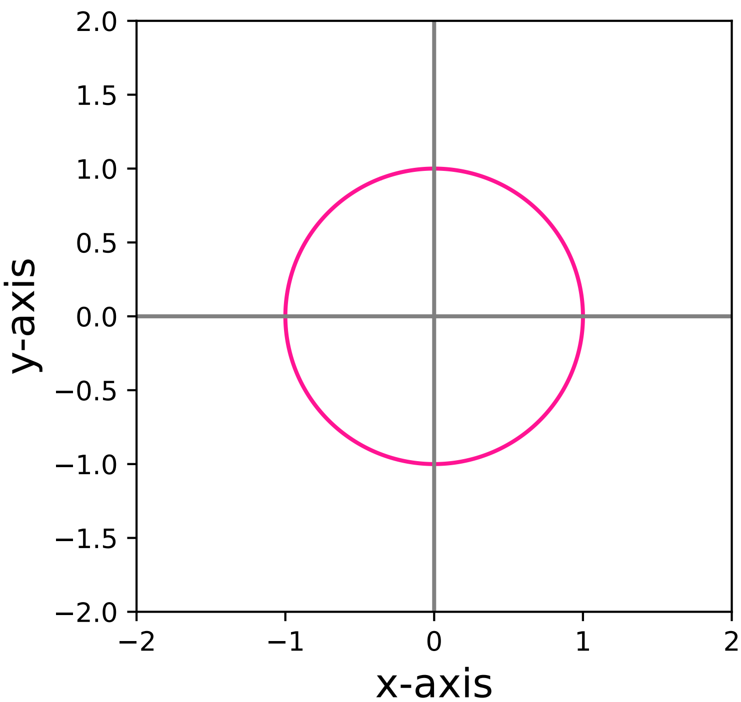


Figure 1.3: Plot of unit circle of equation $x^2 + y^2 = 1$

Goal

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

Question

In what sense do we differentiate the above curves?

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.

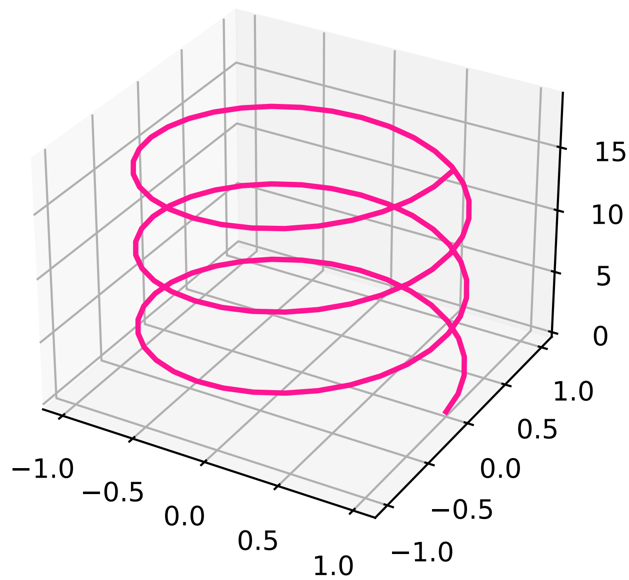


Figure 1.4: Plot of a 3D Helix

Problem

We need a unified and convenient way to describe curves.

This can be done via parametrization.

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.

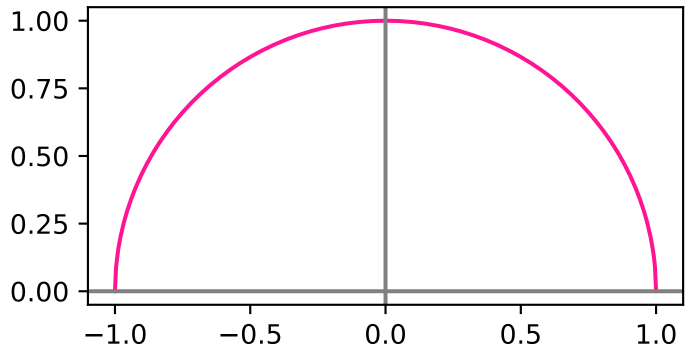


Figure 1.5: Upper semi-circle

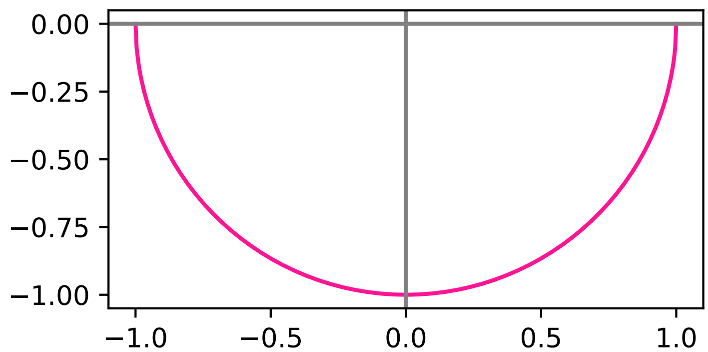


Figure 1.6: Lower semi-circle

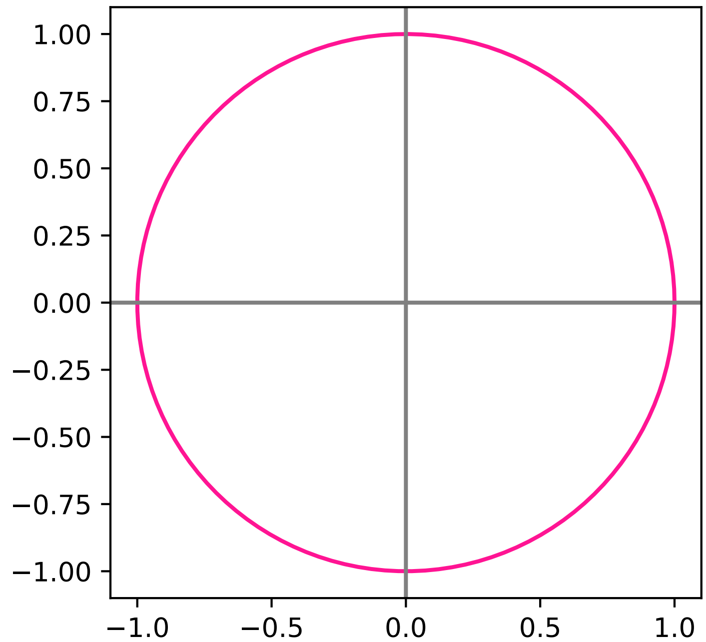


Figure 1.7: Lower semi-circle

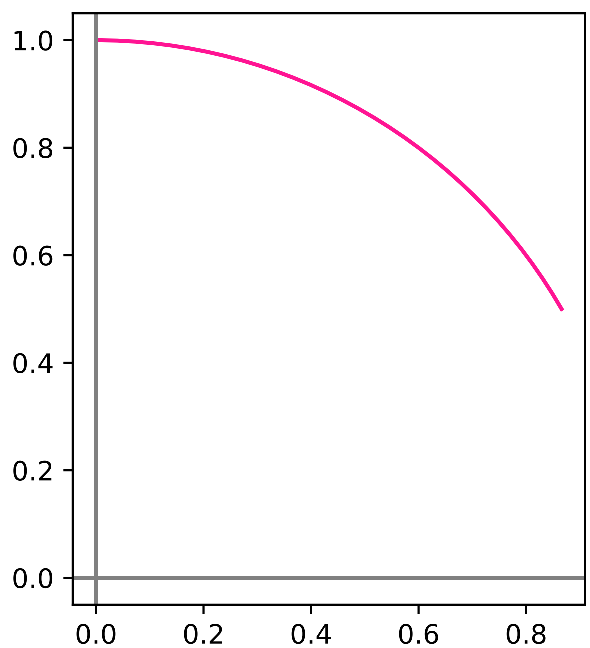


Figure 1.8: Plotting a portion of C

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

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, second and third derivatives of f as follows:

$$\dot{f} := \frac{df}{dt}, \quad \ddot{f} := \frac{d^2 f}{dt^2}, \quad \dddot{f} := \frac{d^3 f}{dt^3}.$$

Example 1.10

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

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

Other examples smooth functions are polynomials, as well as

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

Definition 1.11

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

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

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 $i = 1, \ldots, n$. The derivatives of $\boldsymbol{\gamma}$ are

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

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

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

and the second by

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

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.

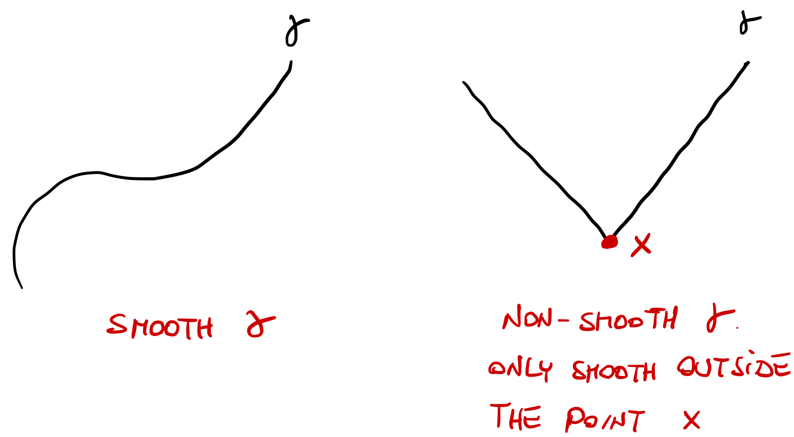


Figure 1.9: Example of smooth and non-smooth curves

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

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

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

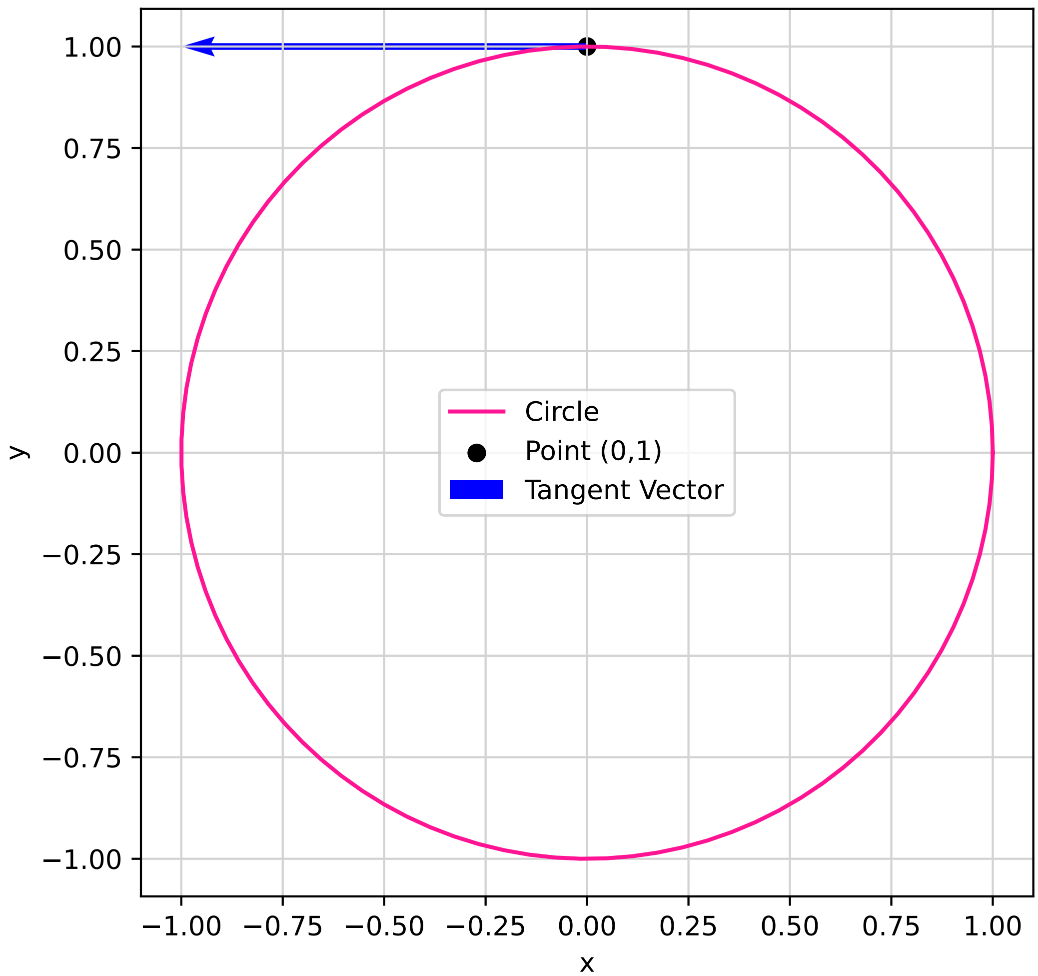


Figure 1.10: Plot of Circle and Tangent Vector at $(0, 1)$

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** 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.

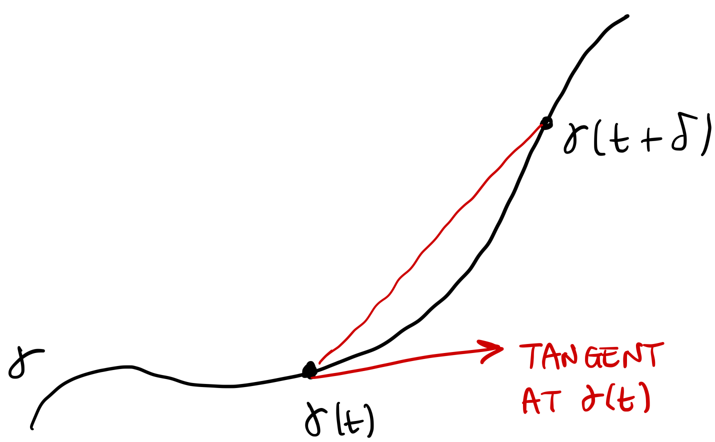


Figure 1.11: Approximating the tangent vector

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

$$\dot{\boldsymbol{\gamma}}(t) \in \mathbb{R}^n \,.$$

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,

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

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.

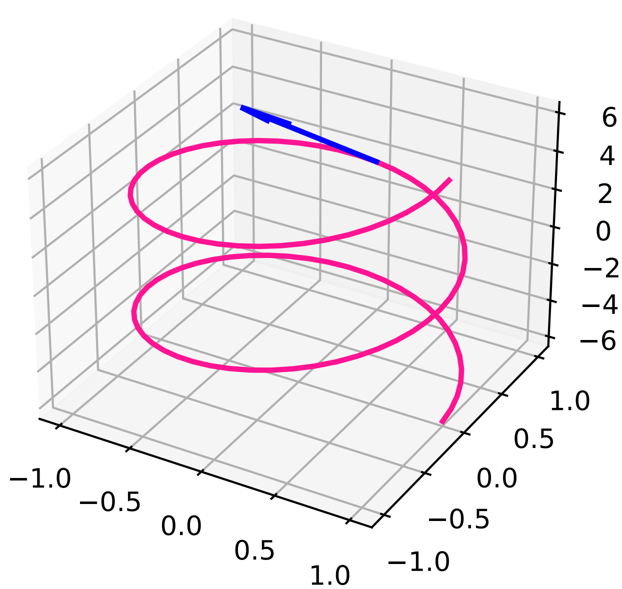


Figure 1.12: Plot of Helix with tangent vector

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,

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

1.5 Length of curves

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

$$\mathbf{v} = (v_1, \dots, v_n),$$

its **length** is defined by

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

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

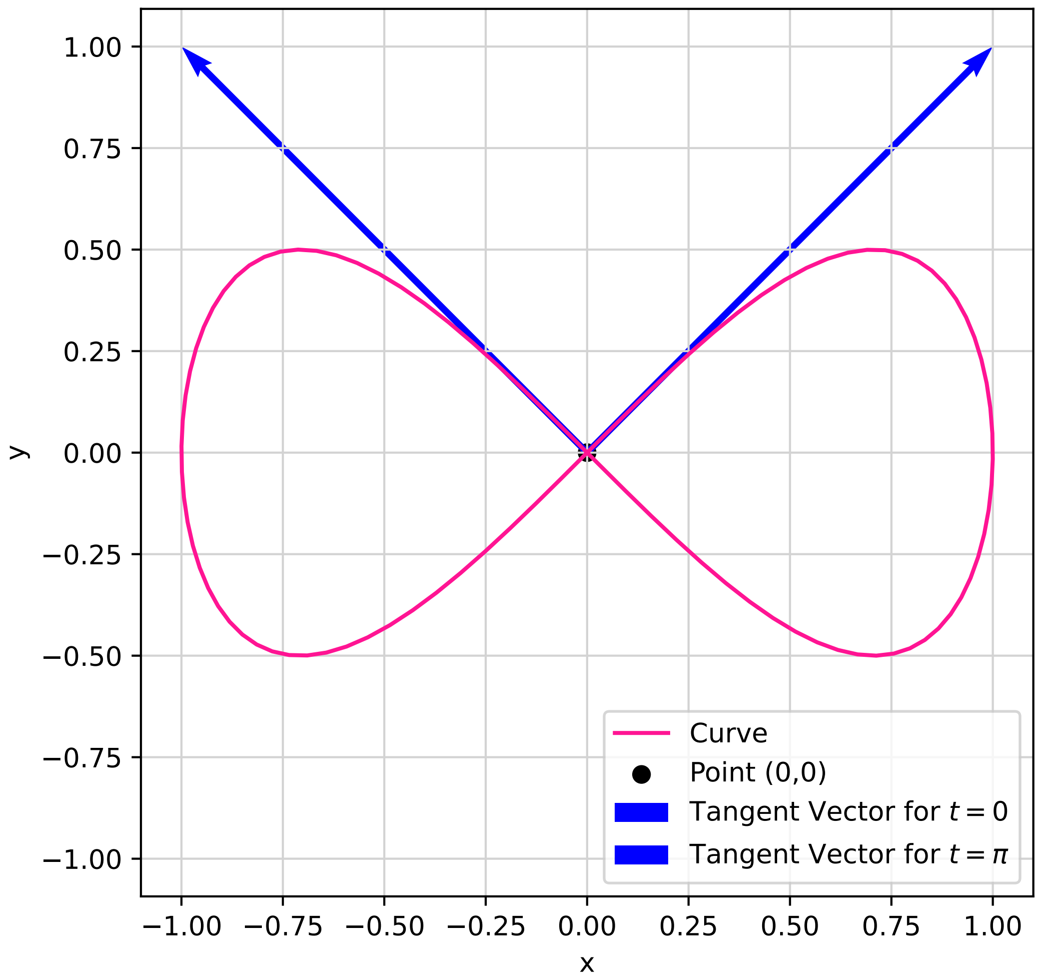


Figure 1.13: The Lemniscate curve

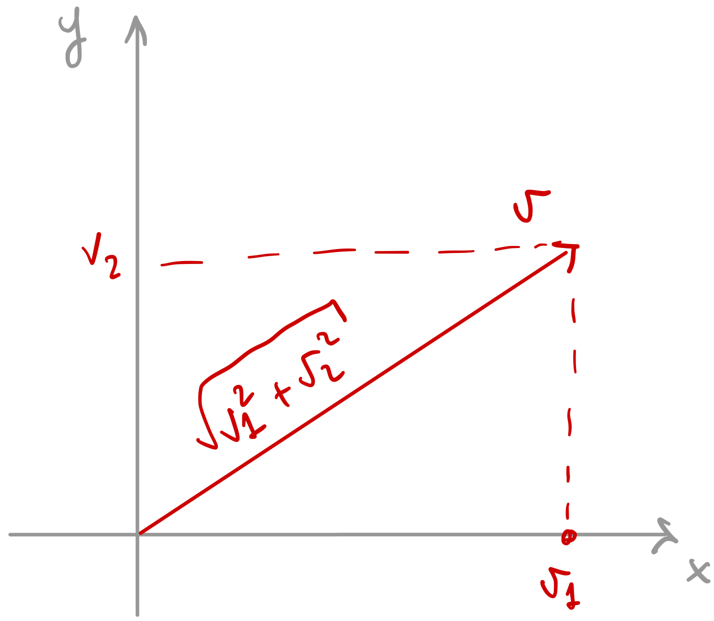


Figure 1.14: Interpretation of $\|v\|$ in $\mathbb{R}^2$

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

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

measures the length of the difference between u and v.

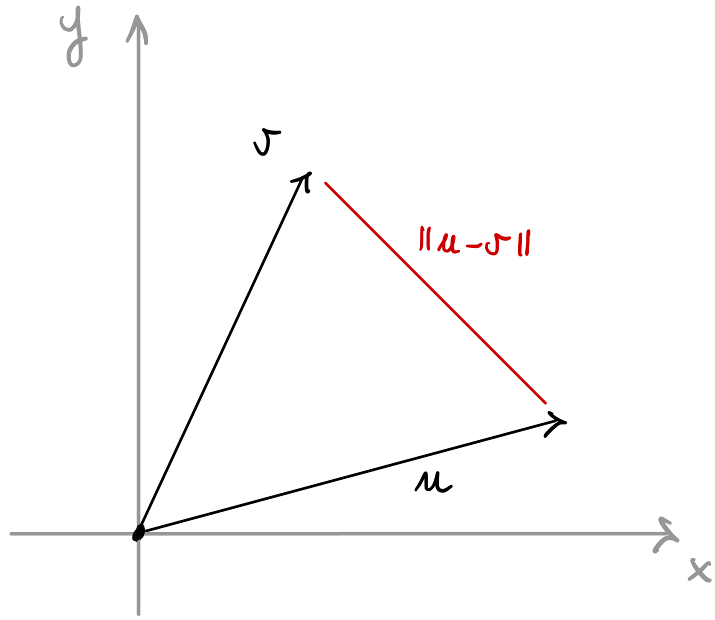


Figure 1.15: Interpretation of $\|\mathbf{u} - \mathbf{v}\|$ in $\mathbb{R}^2$

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

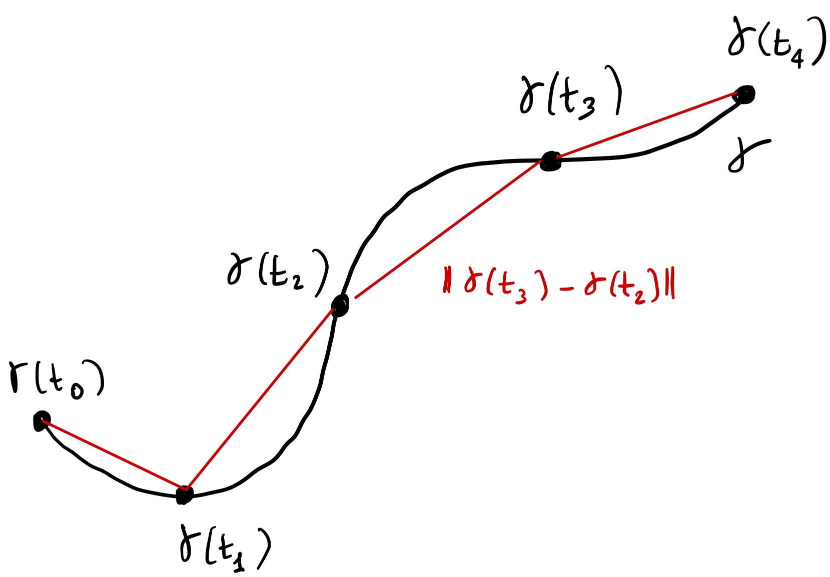


Figure 1.16: Approximating the length of $\boldsymbol{\gamma}$

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

A partition $\mathscr{P}$ of the interval $[a, b]$ is a vector of time instants

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

with

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

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

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

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.

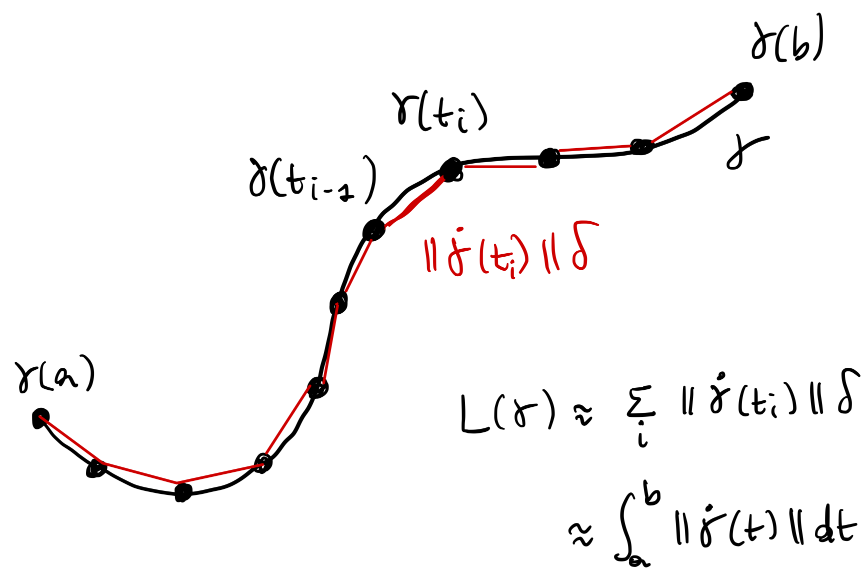


Figure 1.17: Approximating $L(\boldsymbol{\gamma})$ via $\dot{\boldsymbol{\gamma}}$

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

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

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

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

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

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

The idea now is that the integral on the RHS can be made arbitrarily small by choosing a sufficiently fine partition, thanks to the uniform continuity of $\dot{\boldsymbol{\gamma}}$ on the compact interval $[a,b]$. In details, taking the absolute value of the above equation yields

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

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 $i = 1, \ldots, m$ 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

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

Note that if $H > 0$ then

$$2\pi\sqrt{R^2 + H^2} > 2\pi R ,$$

showing that the length of one full turn of the Helix is larger than the length of a disk. This might seem counterintuitive as it looks like one turn of the Helix can be superimposed to the circle by *squashing* the Helix on the plane. However this *squashing* action clearly causes a bit of shrinkage, as shown in the above estimate.

1.6 Arc-length

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

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

Using this formula, we introduce the notion of length of a portion of γ.

Definition 1.24: Arc-length

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

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

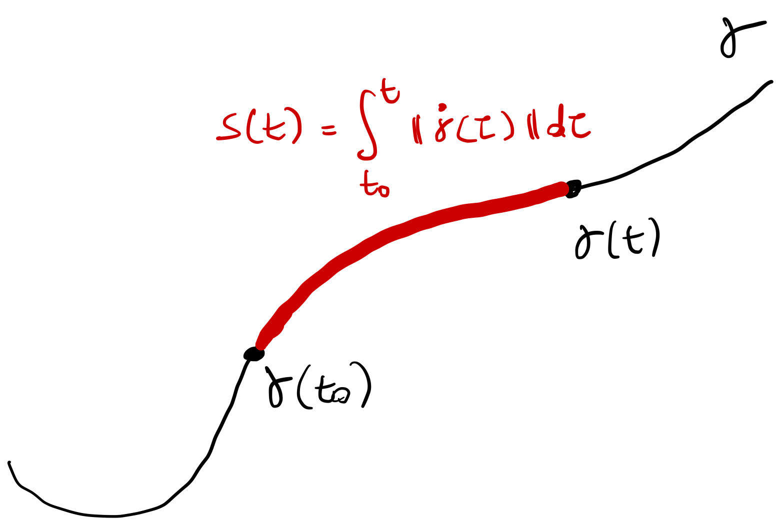


Figure 1.18: Arc-length of γ starting at $\gamma(t_0)$

Remark 1.25

A few remarks:

- Arc-length is well-defined

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 $\boldsymbol{\gamma} : [0, 2\pi] \to \mathbb{R}^2$ 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,

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

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 $\mathbb{R}^2$.

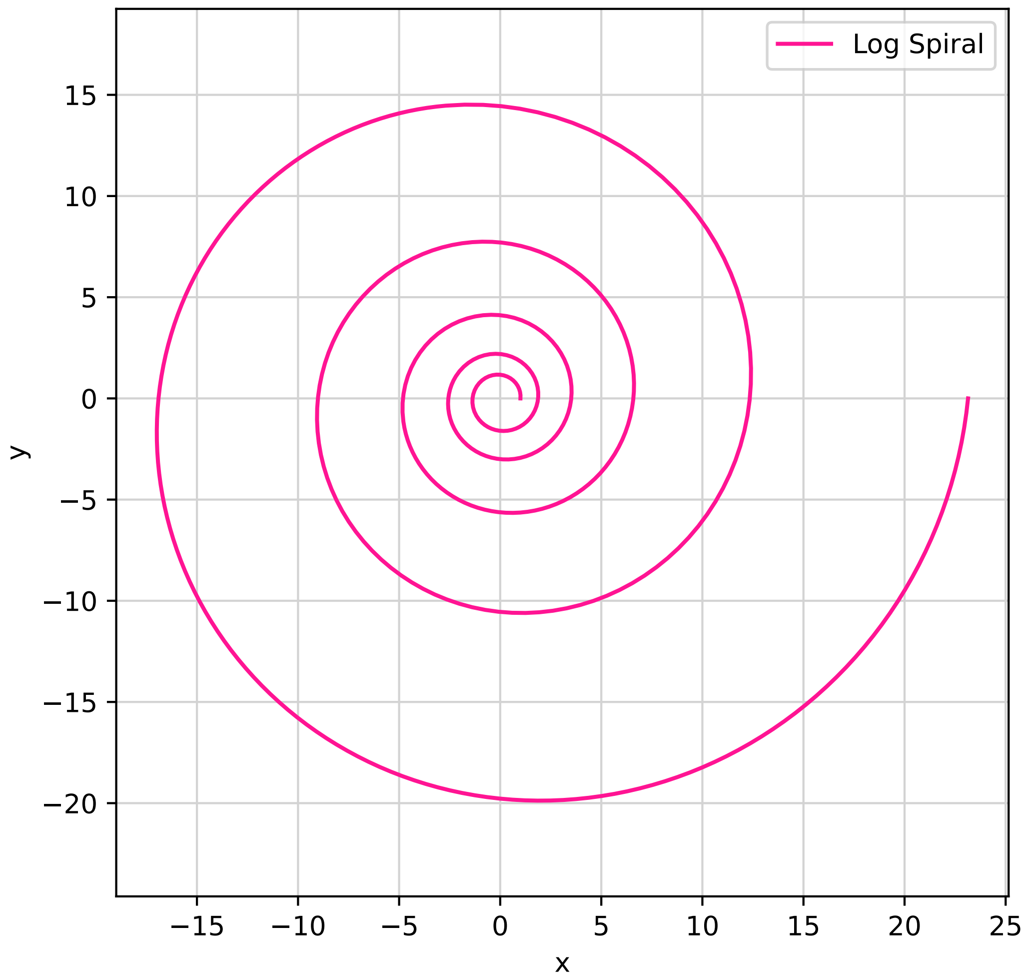


Figure 1.19: Plot of Logarithmic Spiral with $k = 0.1$

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

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

$$\mathbf{u} \cdot \mathbf{v} \ := \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta) \,.$$

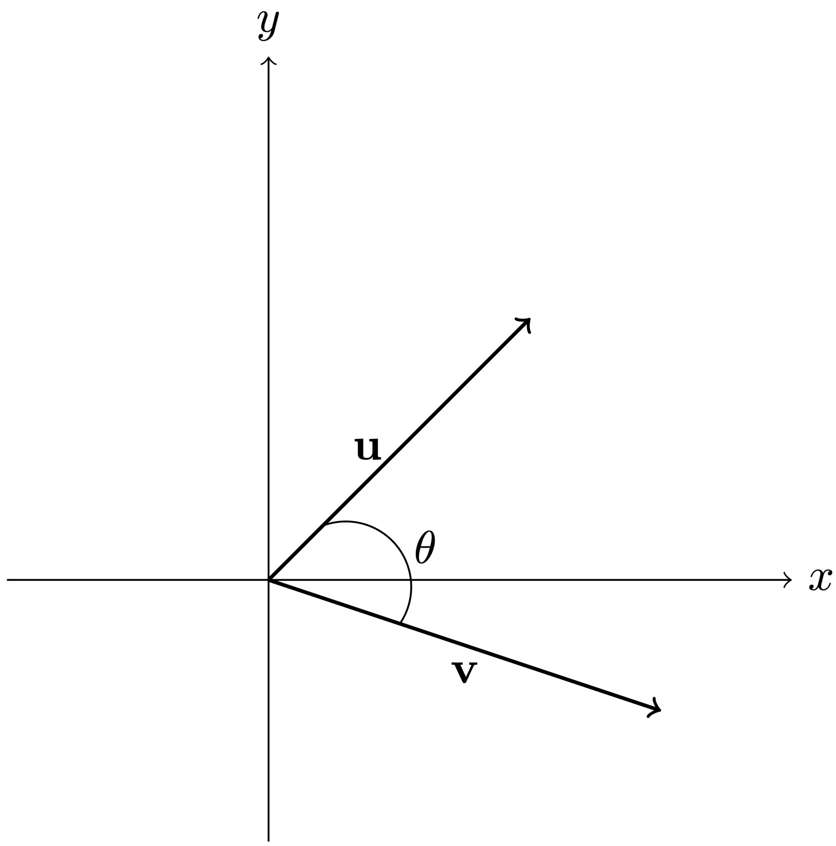


Figure 1.20: Vectors u and v in $\mathbb{R}^2$ forming angle θ

Remark 1.29

1. Two vectors in the plane form two complementary angles. To avoid ambiguity, we choose the smallest of the two angles. This is enforced in Definition 1.28 by requiring that $\theta \in [0, \pi]$.
2. The scalar product is maximized for $\theta = 0$, for which we have
$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta) = \|\mathbf{u}\| \|\mathbf{v}\| \,.$$
3. It is instead minimized for $\theta = \pi$, for which
$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta) = -\|\mathbf{u}\| \|\mathbf{v}\| \,.$$

4. For each $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ the Cauchy-Schwarz inequality holds:

$$-\|\mathbf{u}\|\|\mathbf{v}\| \leq \mathbf{u} \cdot \mathbf{v} \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

   The above is immediate from the observation that $|\cos(\theta)| \leq 1$.

5. The observations in points 2-3 imply that the Cauchy-Schwarz inequality is sharp, in the sense that both inequalities are attained.

6. Usually the Cauchy-Schwarz inequality is written in the equivalent form

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

7. By the above observations it follows that equality holds if and only if $\mathbf{u}$ and $\mathbf{v}$ are parallel.

Definition 1.30: Orthogonal vectors

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$. We say that $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if

$$\mathbf{u} \cdot \mathbf{v} = 0$$

Proposition 1.31: Bilinearity and symmetry of scalar product

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

- **Symmetry**: It holds
$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$$
- **Bilinearity**: They hold
$$\lambda(\mathbf{u} \cdot \mathbf{v}) = (\lambda \mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (\lambda \mathbf{v}),$$
$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}.$$

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

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}.$$

is non-trivial, due to the presence of 3 vectors.

Proposition 1.32: Scalar products written wrt euclidean coordinates

Denote by

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

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

$$\mathbf{u} = (u_1, u_2) = u_1 \mathbf{e}_1 + u_2 \mathbf{e}_2$$

$$\mathbf{v} = (v_1, v_2) = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2$$

their coordinates with respect to $\mathbf{e}_1, \mathbf{e}_2$. Then

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

Proof

Note that

$$\mathbf{e}_1 \cdot \mathbf{e}_1 = 1 \, , \quad \mathbf{e}_2 \cdot \mathbf{e}_2 = 1 \, , \quad \mathbf{e}_1 \cdot \mathbf{e}_2 = \mathbf{e}_2 \cdot \mathbf{e}_1 = 0 \, .$$

Using the bilinearity of scalar product we have

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= (u_1 \mathbf{e}_1 + u_2 \mathbf{e}_2) \cdot (v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2) \\ &= u_1 v_1 \mathbf{e}_1 \cdot \mathbf{e}_1 + u_1 v_2 \mathbf{e}_1 \cdot \mathbf{e}_2 + u_2 v_1 \mathbf{e}_2 \cdot \mathbf{e}_1 + u_2 v_2 \mathbf{e}_2 \cdot \mathbf{e}_2 \\ &= u_1 v_1 + u_2 v_2 \, . \end{aligned}$$

The above proposition provides a natural way to define a scalar product in $\mathbb{R}^n$.

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

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

$$\mathbf{u} = (u_1, \ldots, u_n) \, , \quad \mathbf{v} = (v_1, \ldots, v_n) \, .$$

We define the scalar product between $\mathbf{u}$ and $\mathbf{v}$ by

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

The scalar product in $\mathbb{R}^n$ is still bilinear and symmetric, as detailed in the following proposition:

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

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

- **Symmetry**: It holds
$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$$
- **Bilinearity**: They hold
$$\lambda(\mathbf{u} \cdot \mathbf{v}) = (\lambda \mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (\lambda \mathbf{v}) \, ,$$

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

The proof of the above proposition is an easy check, and is left to the reader for exercise. We can now define orthogonal vectors in $\mathbb{R}^n$.

Definition 1.35

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

$$\mathbf{u} \cdot \mathbf{v} = 0 \,.$$

Proposition 1.36: Differentiating the scalar product

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

$$\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n) \,, \quad \boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$$

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

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

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

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

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 a curve. Therefore the magnitude of $\dot{\boldsymbol{\gamma}}$ can be interpreted as the rate of change, i.e. **speed**, of the curve.

Definition 1.37: Speed of a curve

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

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

Remark 1.38

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

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

Definition 1.39: unit-speed curve

Let $\boldsymbol{\gamma} \,:\, (a,b) \to \mathbb{R}^n$ be a curve. We say that $\boldsymbol{\gamma}$ is a **unit-speed** curve if

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

The reason why we introduce unit-speed curves is because they make calculations easy. A crucial identity which allows to simplify calculations for unit-speed curves is given in the next proposition.

Proposition 1.40

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

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

Proof

Let us consider the identity

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

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

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

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

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

which gives the thesis.

Remark 1.41

Proposition 1.40 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 information will be used in the next Chapter to define the Frenet Frame: an orthonormal basis of vectors which moves smoothly along the curve. The Frenet frame will be crucial for studing local behavior of curves.

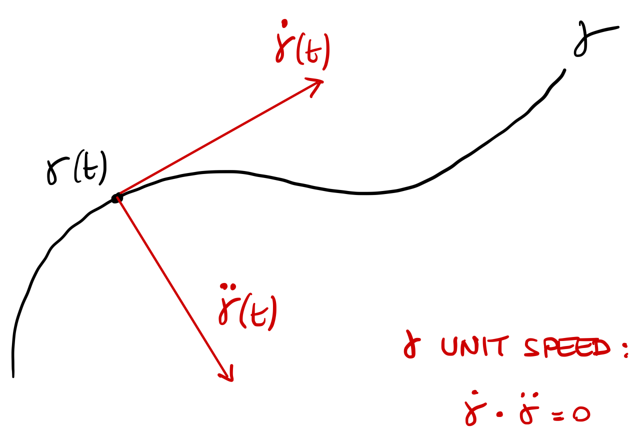


Figure 1.21: If $\boldsymbol{\gamma}$ is unit-speed then $\dot{\boldsymbol{\gamma}}$ and $\ddot{\boldsymbol{\gamma}}$ are orthogonal

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. First we need some notation:

Notation

1. Let X, Y, Z be sets and
$$f : X \to Y, \quad g : Y \to Z$$
two maps. The **composition** of f and g is the map
$$g \circ f : X \to Z, \quad (g \circ f)(x) := g(f(x))$$

2. The **identity map** on X is denoted by
$$\mathrm{Id}_X : X \to X, \quad \mathrm{Id}_X(x) := x, \quad \forall\, x \in X.$$
The identity in $\mathbb{R}$ will just be denoted by Id.

3. The function $f : X \to Y$ is **invertible** if there exists a function $g : Y \to X$ such that
$$g \circ f = \mathrm{Id}_X, \quad f \circ g = \mathrm{Id}_Y,$$
The map g, if it exists, is called the **inverse** of f and is denoted by
$$g := f^{-1}.$$

4. It is elementary to check that the inverse is unique if it exists.

Definition 1.42: 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: There exists a map
$$\phi^{-1} : (\tilde{a}, \tilde{b}) \to (a, b)$$
such that
$$\phi^{-1} \circ \phi = \phi \circ \phi^{-1} = \mathrm{Id},$$
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 1.43: 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

$$\tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\gamma}(\phi(t)) \quad \forall\, t \in (\tilde{a},\tilde{b})\,, \tag{1.21}$$

for a suitable diffeomorphism

$$\phi : (\tilde{a},\tilde{b}) \to (a,b)\,.$$

We call both ϕ and ϕ^{-1} **reparametrization maps**.

Remark 1.44

Since ϕ is invertible with smooth inverse, $\tilde{\boldsymbol{\gamma}}$ is a reparametrization of $\boldsymbol{\gamma}$

$$\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(\phi(\phi^{-1}(t))) = \tilde{\boldsymbol{\gamma}}(\phi^{-1}(t))\,, \quad \forall\, t \in (a,b)\,.$$

Remark 1.45

1. Given a parametrized curve $\boldsymbol{\gamma}$, this identifies a 1D shape $\Gamma \subset \mathbb{R}^n$ defined by

   $$\Gamma := \{\boldsymbol{\gamma}(t) \,:\, t \in (a,b)\}\,.$$

   Γ is called the support of γ.

2. A reparametrization $\tilde{\boldsymbol{\gamma}}$ is just an equivalent way to describe Γ.

3. For $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ to be reparametrizations of each other, there must exist a smooth rule ϕ (the diffeomorphism) to switch from one to another, according to formula (1.21). This concept is sketched in Figure 1.22.

Example 1.46: 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.

Example 1.47: Reversing orientation of circle

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

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

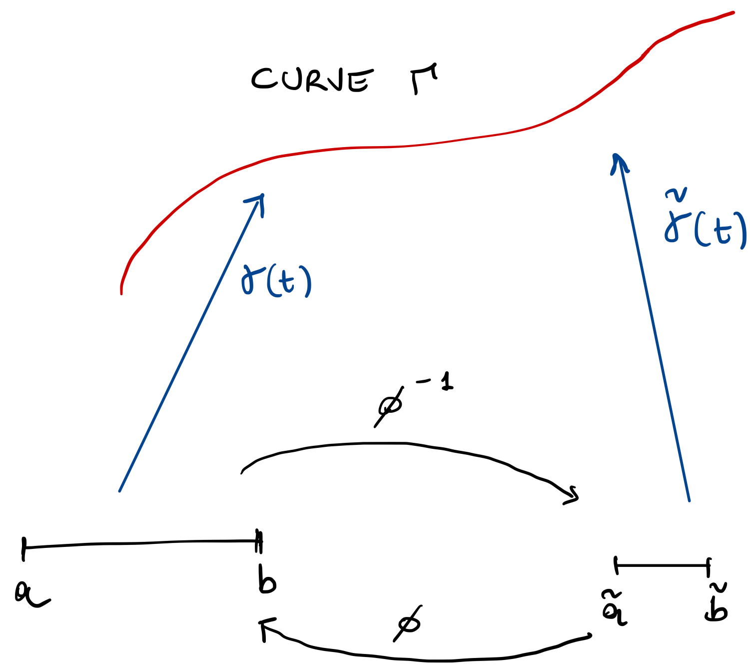


Figure 1.22: Sketch of 1D shape Γ parametrized by $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$. We can switch parametrization by means of the diffeomorphism ϕ.

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,

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

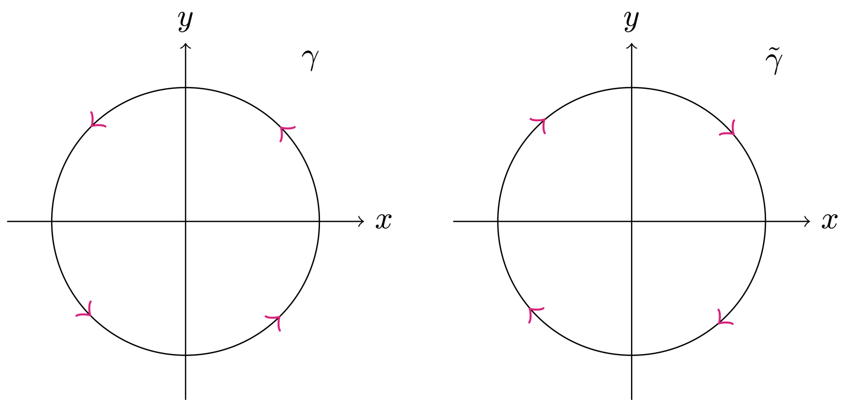


Figure 1.23: Unit circle with usual parametrization $\boldsymbol{\gamma}$, and with reversed orientation $\tilde{\boldsymbol{\gamma}}$

Example 1.48: 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 1.49: 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

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

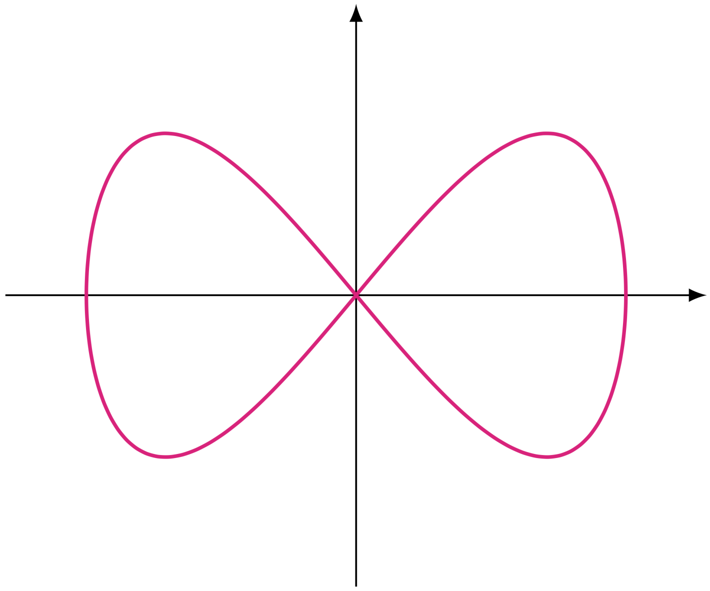


Figure 1.24: Lemniscate curve

1.10 Unit-speed reparametrization

For a curve $\boldsymbol{\gamma}$ we wish to find a reparametrization $\tilde{\boldsymbol{\gamma}}$ which is unit-speed:

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

We will see that this is possible if and only if the curve $\boldsymbol{\gamma}$ is regular, in the sense of Definition 1.50 below.

Definition 1.50: Regular points

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

1. $\boldsymbol{\gamma}(t_0)$ is a **regular point** if
$$\dot{\boldsymbol{\gamma}}(t_0) \neq 0\,.$$
2. A point $\boldsymbol{\gamma}(t_0)$ is **singular** if it is not regular.
3. 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 1.51

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

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

We can differentiate the LHS by chain rule

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

From (1.22) we then get

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

As the RHS is non-zero, we must have that both the elements in the LHS product are non-zero. In particular we conclude

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

Example 1.52: A curve with one singular point

Consider the parabola

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

Both curves $\boldsymbol{\gamma}, \boldsymbol{\eta} : (-1, 1) \to \mathbb{R}^2$

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

are parametrizations of Γ. However $\boldsymbol{\eta}$ is not a reparametrization of $\boldsymbol{\gamma}$.

Indeed, suppose by contradiction there exist a diffeomorphism

$$\phi : (-1, 1) \to (-1, 1)$$

such that

$$\boldsymbol{\eta}(t) = \boldsymbol{\gamma}(\phi(t)) , \quad \forall \, t \in (-1, 1) .$$

Substituting the definitions of $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ we obtain

$$(t^3, t^6) = (\phi(t), \phi(t)^2) , \quad \forall \, t \in (-1, 1) ,$$

which forces

$$\phi(t) = t^3 , \quad \forall \, t \in (-1, 1) .$$

Note that f is invertible in $(-1,1)$ with inverse

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

However ϕ^{-1} is not smooth at $t=0$, and therefore 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 1.51.

To understand what is going on with the two parametrizations, 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, as the first component of $\dot{\boldsymbol{\gamma}}$ is non-zero and so $\dot{\boldsymbol{\gamma}} \neq 0$.
- $\boldsymbol{\eta}$ is regular if and only if $t \neq 0$.

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 $t>0$.

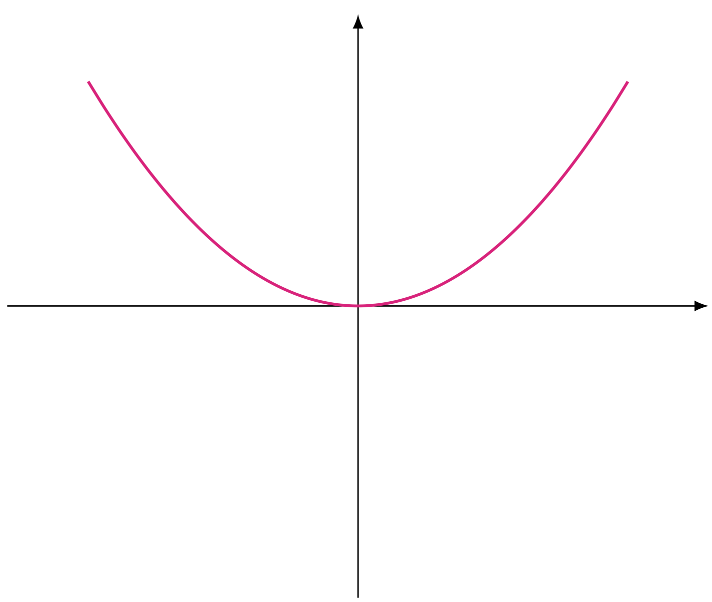


Figure 1.25: Parabola Γ

The previous example shows that, altough $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ describe the same parabola Γ, they are not reparametrizations of each other. We have seen that this is due to the fact that $\boldsymbol{\gamma}$ is regular, while $\boldsymbol{\eta}$ is not. Indeed we can prove that regularity is invariant by reparametrization.

Proposition 1.53: 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

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

As ϕ is a diffeomorphism, by Lemma 1.51 it holds

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

Therefore

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

Since $\boldsymbol{\gamma}$ is regular we infer

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

From (1.23) we conclude that $\tilde{\boldsymbol{\gamma}}$ is regular.

Example 1.54

Let us go back to the parabola

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

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 1.53 implies that $\boldsymbol{\eta}$ is **not** a reparametrization of $\boldsymbol{\gamma}$.

We now define unit-speed reparametrizations:

Definition 1.55: unit-speed reparametrization

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$ be a parametrized curve. A **unit-speed reparametrization** of $\boldsymbol{\gamma}$ is a reparametrization $\tilde{\boldsymbol{\gamma}} : (\tilde{a}, \tilde{b}) \to \mathbb{R}^n$ which is unit-speed, i.e.,

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

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. Notice that

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

Therefore

$$\boldsymbol{\gamma} \text{ regular} \quad \iff \quad \dot{s}(t) \neq 0 \,.$$

In this case the arc-length s is a diffeomorphism by the Inverse Function Theorem. As it turns out, all the unit-speed reparametrizations of $\boldsymbol{\gamma}$ are of the form

$$\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} \circ \psi \,, \qquad \psi := \pm s^{-1} + c$$

The above statements will be proved in Theorem 1.56 and Proposition 1.57 below.

Theorem 1.56: Existence of unit-speed reparametrization

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

1. $\boldsymbol{\gamma}$ is regular,
2. $\boldsymbol{\gamma}$ 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

$$\dot{s}(t) = \|\dot{\boldsymbol{\gamma}}(t)\| \tag{1.24}$$

so that

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

The above condition and the Inverse Function Theorem guarantee the existence of a smooth inverse

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

for some $\tilde{a} < \tilde{b}$. Define the reparametrization map ϕ as

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

and the corresponding reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ as

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

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

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

or in other words

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

By chain rule

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

where in the last equality we used (1.24). Taking the absolute value of the above yields

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \left\|\dot{\tilde{\boldsymbol{\gamma}}}(s(t))\right\| \,\|\dot{\boldsymbol{\gamma}}(t)\|\,. \tag{1.25}$$

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

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

Therefore we can divide (1.25) 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

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

By chain rule

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

Taking the norm

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

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

$$\|\dot{\boldsymbol{\gamma}}(\phi(t))\|\;|\dot{\phi}(t)| = 1\,, \quad \forall\, t \in (\tilde{a}, \tilde{b})\,. \tag{1.26}$$

Hence none of terms on the LHS can be zero, meaning that

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

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 1.56 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 1.57: 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 there exists $c \in \mathbb{R}$ such that

$$\phi(t) = \pm s(t) + c\,, \quad \forall\, t \in (a,b)\,. \tag{1.27}$$

2. If ϕ is given by (1.27) for some $c \in \mathbb{R}$, then $\tilde{\boldsymbol{\gamma}}$ is unit-speed.

Proof

Step 1. First Point. Assume $\tilde{\boldsymbol{\gamma}}$ is a unit-speed reparametrization of $\boldsymbol{\gamma}$: such $\tilde{\boldsymbol{\gamma}}$ exists by Theorem 1.56, since $\boldsymbol{\gamma}$ is assumed to be regular. This means there exists a reparametrization map ϕ such that

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

Differentiating the above we get

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

Taking the norms we then have

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

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

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

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

$$\dot{s}(t) = |\dot{\phi}(t)| \quad \forall\, t \in (a,b) \tag{1.28}$$

As ϕ is a diffeomorphism, by Lemma 1.51 we have

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

By continuity of $\dot{\phi}$ and the Mean Value Theorem we conclude that either

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

or

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

Therefore (1.28) reads either

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

or

$$\dot{s}(t) = -\dot{\phi}(t) \quad \forall\, t \in (a,b)$$

Integrating the last two equations we get

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

for some $c \in \mathbb{R}$, concluding the proof.
Step 2. Second Point. Let $\tilde{\boldsymbol{\gamma}}$ be reparametrization of $\boldsymbol{\gamma}$

$$\boldsymbol{\gamma}(t) = \tilde{\boldsymbol{\gamma}}(\phi(t)), \quad \forall\, t \in (a, b) \tag{1.29}$$

with

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

for some $c \in \mathbb{R}$. Differentiating (1.29) we get

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

where in the last equality we used the Fundamental Theorem of Calculus and definition of s. Taking the absolute values

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

Since $\boldsymbol{\gamma}$ is regular we have $\|\dot{\boldsymbol{\gamma}}(t)\| \neq 0$. Hence we can divide by $\|\dot{\boldsymbol{\gamma}}(t)\|$ and obtain that

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

As 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.

Definition 1.58: Arc-length reparametrization

Let $\boldsymbol{\gamma}$ be regular. Proposition 1.57 tells us that a unit speed reparametrization of $\boldsymbol{\gamma}$ is given by

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

with s^{-1} inverse of the arc-length of $\boldsymbol{\gamma}$. We call $\tilde{\boldsymbol{\gamma}}$ the **arc-length reparametrization** of $\boldsymbol{\gamma}$.

Notation: Arc-length parameter

In the following we will use the letters

1. s to denote the *arc-length parameter*
2. t to denote an *arbitrary parameter*

Accordingly the parameter of the arc-length function is t

$$s = s(t)$$

and the parameter of the inverse $\psi := s^{-1}$ of the arc-length is s

$$\psi = \psi(s)$$

As the arc-length function allows to transition from an arbitrary parameter t to the arc-length parameter s, the inverse of s will also be denoted by

$$t = t(s)$$

The above notation might seem confusing, but it actually makes a lot of sense in the long run, as calculations get heavier. Accordingly we have the following notation for the arc-length reparametrization.

Notation: Arc-length reparametrization

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^n$ be a regular curve and $s \colon (a,b) \to (\tilde{a}, \tilde{b})$ its arc-length function. Denote by

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

the arc-length reparametrization of $\boldsymbol{\gamma}$. According to the above notations, we will write

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

and also

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

Example 1.59: Arc-length reparametrization of the Circle

Let $\boldsymbol{\gamma} \colon [0, 2\pi] \to \mathbb{R}^2$ be the parametrization of the circle with center (x_0, y_0) and radius $R > 0$

$$\boldsymbol{\gamma}(t) = (x_0 + R\cos(t), t_0 + \sin(t))$$

The arc-length of $\boldsymbol{\gamma}$ starting from $t_0 = 0$ is

$$s(t) = \int_0^t \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau = Rt$$

Rearranging we obtain

$$s = Rt \quad \Longrightarrow \quad t = \frac{s}{R}$$

The inverse $\psi = s^{-1}$ is therefore

$$\psi(s) = \frac{s}{R}$$

According to our notation ψ can also be denoted as

$$t(s) = \frac{s}{R}$$

By Proposition 1.57

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

is a unit-speed reparametrization of $\boldsymbol{\gamma}$. Plugging the formula for s^{-1} into the expression for $\boldsymbol{\gamma}$ yields the explicit formula for $\tilde{\boldsymbol{\gamma}}$

$$\tilde{\boldsymbol{\gamma}}(s) = \boldsymbol{\gamma}(t(s)) = \left(x_0 + R\cos\left(\frac{s}{R}\right), y_0 + \sin\left(\frac{s}{R}\right)\right)$$

Warning

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

Example 1.60: 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

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

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

$$s(t) = \int_{t_0}^{t} \sqrt{1 + 4\tau^2 + 9\tau^4}\, d\tau . \tag{1.30}$$

Since $\boldsymbol{\gamma}$ is regular, by Proposition 1.57 we know that $\boldsymbol{\gamma}$ admits the unit-speed reparametrization

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

with s^{-1} the inverse of the arc-length function. It can be shown that the integral at (1.30) cannot be written in terms of elementary functions. Therefore there are not explicit formulas for s and s^{-1}. As a consequence the unit-speed parametrization $\tilde{\boldsymbol{\gamma}}$ cannot be computed explicitly in this case.

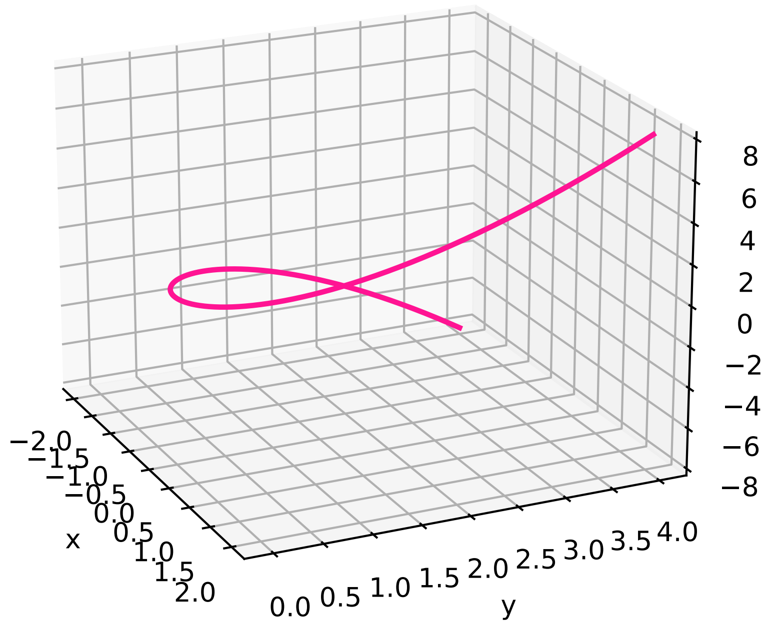


Figure 1.26: Plot of Twisted Cubic for t between -2 and 2

1.11 Closed curves

So far we have seen examples of:

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

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

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

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

3. 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 1.61

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:

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

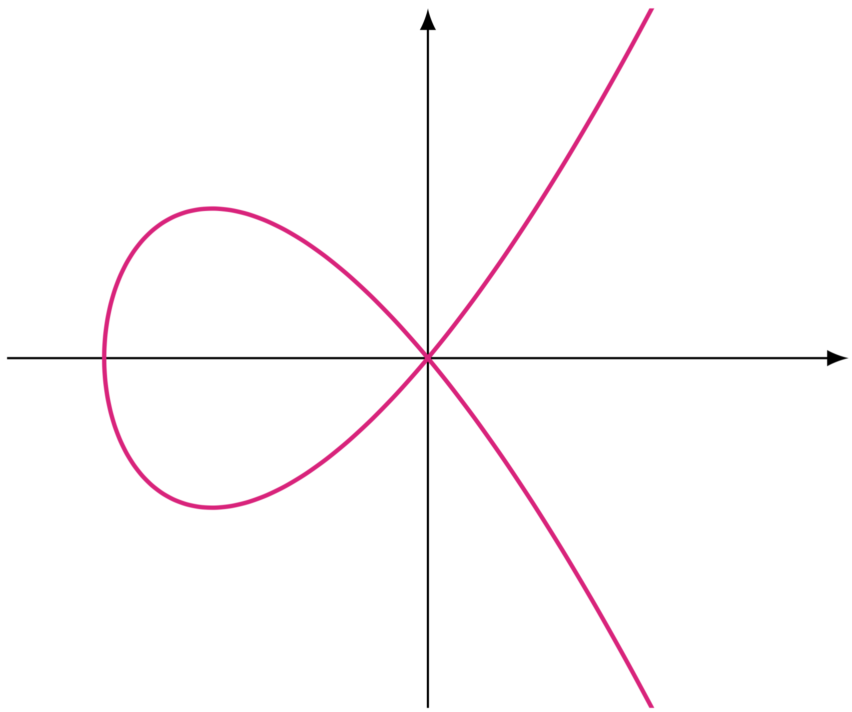


Figure 1.27: Plot of curve $\boldsymbol{\gamma}(t) = (t^2 - 1, t^3 - 1)$ for $t \in [-2, 2]$

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

Definition 1.62: 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 1.63: Closed curve

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

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

Remark 1.64

The following basic observations hold:

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: 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
   $$\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b)\,, \quad \frac{d^k\boldsymbol{\gamma}}{dt^k}(a) = \frac{d^k\boldsymbol{\gamma}}{dt^k}(b) \tag{1.31}$$
   for all $k \in \mathbb{N}$. Set
   $$T := b - a\,.$$
   Then $\boldsymbol{\gamma}$ can be extended to a smooth 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 and smooth.

   Note that assumption (1.31) must hold for all $k \in \mathbb{N}$ for the extension $\tilde{\boldsymbol{\gamma}}$ to be smooth. As a counterexample consider $f(x) := x^2$ for $x \in [-1,1]$. As seen by plotting f, it is clear that f cannot be extended to a smooth periodic function. And indeed in this case (1.31) is violated, because
   $$f(-1) = f(1) = 1\,, \quad f'(-1) = -2 \neq 2 = f'(1)$$
   showing that the periodic extension is continuous but not differentiable.

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 then $\boldsymbol{\gamma}$ 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.

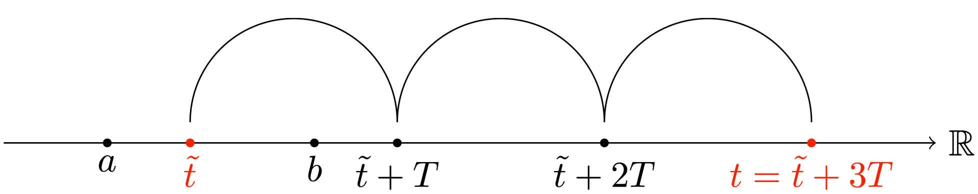


Figure 1.28: The points $t \in \mathbb{R}$ and $\tilde{t} \in [a,b]$ from Point 3 in Remark 1.64. In this skecth $t = \tilde{t}+3T$, with $T = b-a$.

Definition 1.65

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 1.66

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

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

admits positive minumum

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

Proof

We make 2 observations about the set S:

- Since $\boldsymbol{\gamma}$ is closed, we have that $\boldsymbol{\gamma}$ is T-periodic for some $T \neq 0$. By Remark 1.64 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:

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

Since $P = \inf S$, the above is equivalent to showing that
Claim: $P \in S$
Proof of claim. To prove that $P \in S$, by definition of S we need to show that

1. $\boldsymbol{\gamma}$ is P-periodic
2. $P > 0$

Since $P = \inf S$, 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 $T_n \in S$, by definition $\boldsymbol{\gamma}$ is T_n-periodic. Then

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

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. We have shown that $\boldsymbol{\gamma}$ is P-periodic. Therefore

$$P \in S \qquad \Longleftrightarrow \qquad P > 0\,.$$

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 1.64 Point 6 implies that $\boldsymbol{\gamma}$ is also $k_n T_n$-periodic, since $k_n \in \mathbb{Z}$. Thus

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

Therefore

$$\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(t_n)\,, \quad \forall\, n \in \mathbb{N}\,. \tag{1.32}$$

Also notice that

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

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 (1.32) 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.

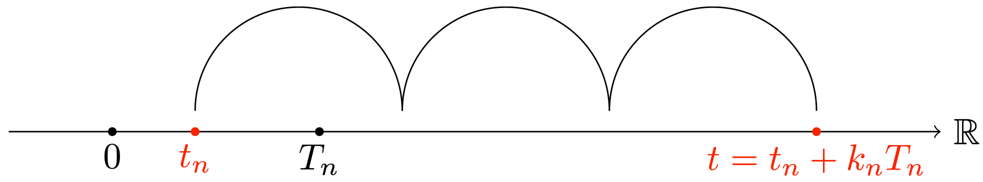


Figure 1.29: For each $t \in \mathbb{R}$ there exist unique $k_n \in \mathbb{Z}$ and $\tilde{t}_n \in [0, T_n]$ such that $t = \tilde{t} + k_n T_n$. In this skecth $k_n = 3$.

Example 1.67: Examples of closed curves

1. The circumference
$$\boldsymbol{\gamma}(t) = (\cos(t), \sin(t)) , \quad t \in \mathbb{R}$$
is not costant and is 2π-periodic. Thus $\boldsymbol{\gamma}$ is closed. The period of $\boldsymbol{\gamma}$ is 2π.

2. The Lemniscate
$$\boldsymbol{\gamma}(t) = (\sin(t), \sin(t)\cos(t)) , \quad t \in \mathbb{R}$$
is not costant and is 2π-periodic. Thus $\boldsymbol{\gamma}$ is closed. The period of $\boldsymbol{\gamma}$ is 2π.

3. Consider again the curve from Example 1.61
$$\boldsymbol{\gamma}(t) := (t^2 - 1, t^3 - t) , \quad t \in \mathbb{R} .$$
According to our definition, $\boldsymbol{\gamma}$ is not periodic. Therefore $\boldsymbol{\gamma}$ is not closed. However there is a point of **self-intersection** on $\boldsymbol{\gamma}$, namely
$$p := (0, 0) ,$$
for which we have
$$p = \boldsymbol{\gamma}(-1) = \boldsymbol{\gamma}(1) .$$

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

Definition 1.68: Self-intersecting curve

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

1. There exist two times $a \neq b$ such that
$$p = \boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b) ,$$

2. If $\boldsymbol{\gamma}$ is closed with period T, then $b - a$ is not an integer multiple of T.

Remark 1.69

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 1.64 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 1.70

Let us go back to the curve of Example 1.61, 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 1.71: 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 $(0,0)$.

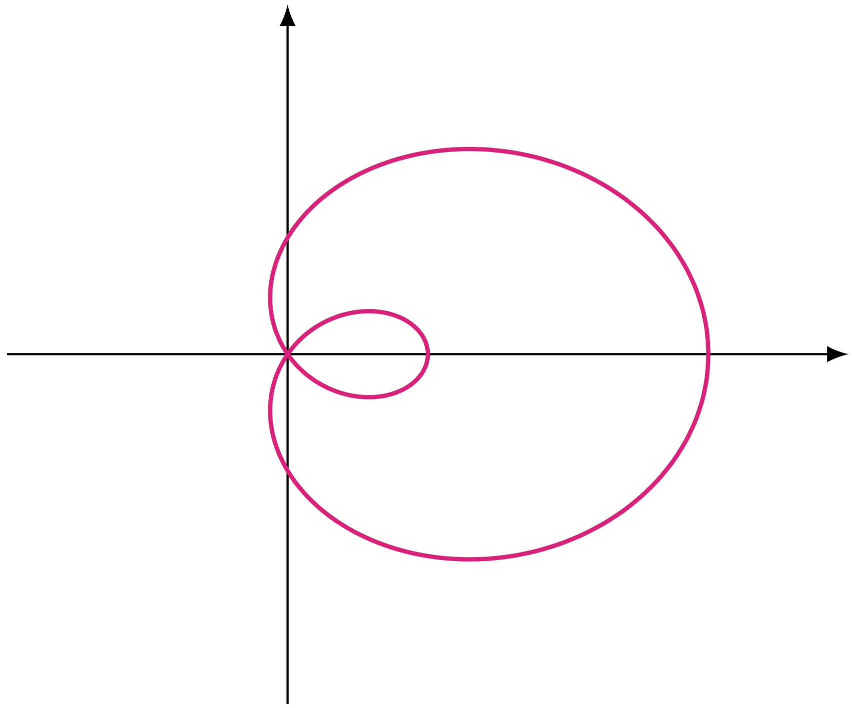


Figure 1.30: Limaçon curve

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 $\boldsymbol{\gamma}$ deviates from 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 $\mathbf{a}, \mathbf{v} \in \mathbb{R}^3$. Whichever the definition of curvature will be, we expect the curvature of a straight line to be zero. The tangent vector to $\boldsymbol{\gamma}$ is constant

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

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 $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^2$ is a **planar curve** 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

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

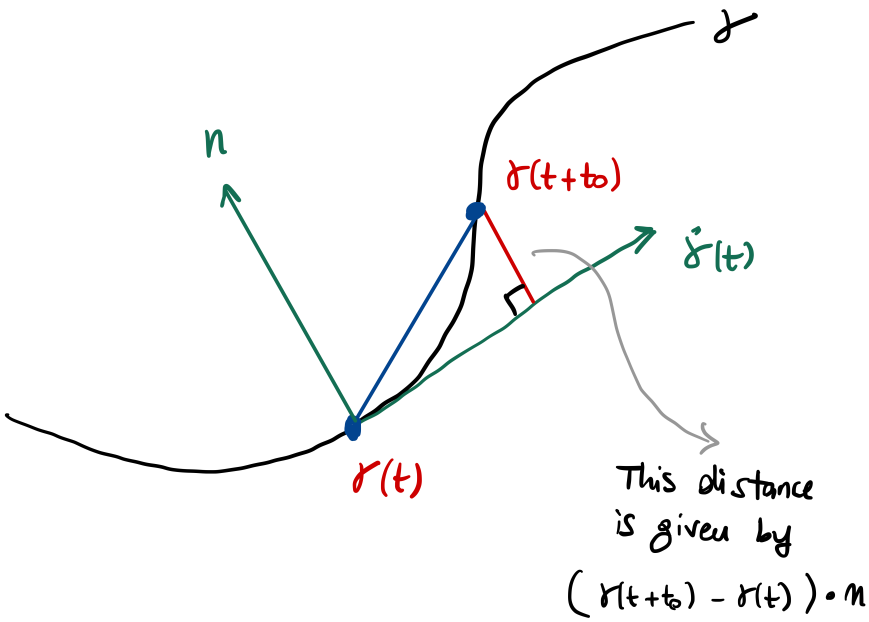


Figure 2.1: Amount that $\boldsymbol{\gamma}$ deviates from tangent is $[\boldsymbol{\gamma}(t+t_0) - \boldsymbol{\gamma}(t)] \cdot \mathbf{n}(t)$

as seen in Figure Figure 2.1.

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

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

Taking the norms of the above and recalling that $\|\mathbf{n}\| = 1$ gives

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

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

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

where the remainder $o(t_0^2)$ is such that

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

Therefore, discarding 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: Curvature of unit-speed curve

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

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

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

Example 2.2: Curvature of the circle

We have seen that the arc-length parametrization of the circle of center (x_0, y_0) and radius $R > 0$ is

$$\boldsymbol{\gamma}(t) = \left(x_0 + R\cos\left(\frac{t}{R}\right), y_0 + \sin\left(\frac{t}{R}\right) \right)$$

As $\boldsymbol{\gamma}$ is parametrized by arc-length, we have that $\boldsymbol{\gamma}$ is unit-speed. To compute the curvature we need the second derivative $\ddot{\boldsymbol{\gamma}}$

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

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

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

In this case κ is constant. The above formula also tells us that the curvature is inversely proportional to the radius of the circle: A large circle will look (locally) almost like a straight line.

Question: How do we define curvature for arbitrary curves?

Answer: When $\boldsymbol{\gamma}$ is regular we can use unit-speed reparametrizations to define κ.

Definition 2.3: Curvature of regular curve

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a regular curve and $\tilde{\boldsymbol{\gamma}}$ be a unit-speed reparametrization of $\boldsymbol{\gamma}$, with

$$\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi \,, \quad \phi \colon (a,b) \to (\tilde{a}, \tilde{b}) \,.$$

Denote by

$$\tilde{\kappa} \colon (\tilde{a}, \tilde{b}) \to \mathbb{R} \,, \qquad \tilde{\kappa}(t) := \left\| \ddot{\tilde{\boldsymbol{\gamma}}}(t) \right\|$$

the curvature of $\tilde{\boldsymbol{\gamma}}$. The **curvature** of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is defined by

$$\kappa(t) := \tilde{\kappa}(\phi(t))$$

In order for the above definition to make sense, we need to check that the curvature κ does not change if we choose a different unit-speed reparametrization. This is shown in the next Proposition.

Proposition 2.4: κ is invariant for unit-speed reparametrization

Consider the setting of Definition 2.3. If $\hat{\boldsymbol{\gamma}}$ is another unit-speed reparametrization of $\boldsymbol{\gamma}$, with $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}} \circ \psi$, then

$$\kappa(t) = \tilde{\kappa}(\phi(t)) = \hat{\kappa}(\psi(t)) \,, \quad \forall \, t \in (a,b)$$

where

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

is the curvature of $\hat{\boldsymbol{\gamma}}$.

Proof

Since $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ are both reparametrizations of $\boldsymbol{\gamma}$

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

Using that ϕ is invertible we obtain

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

and ξ is a diffeomorphism, being composition of diffeomorphisms. Differentiating (2.2)

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

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

$$|\dot{\xi}(t)| = 1 \,, \quad \forall \, t \,.$$

Since $\dot{\xi}$ is continuous we infer

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

In both cases
$$\ddot{\xi} \equiv 0 \,.$$
Differentiating (2.3) we then obtain
$$\begin{aligned}\ddot{\tilde{\boldsymbol{\gamma}}}(t) &= \ddot{\hat{\boldsymbol{\gamma}}}(\xi(t))\dot{\xi}^2(t) + \dot{\hat{\boldsymbol{\gamma}}}(\xi(t))\ddot{\xi}(t) \\ &= \ddot{\hat{\boldsymbol{\gamma}}}(\xi(t))\dot{\xi}^2(t)\,,\end{aligned}$$
where we used that $\ddot{\xi} = 0$. Taking the norms and using again that $|\dot{\xi}| \equiv 1$
$$\left\|\ddot{\tilde{\boldsymbol{\gamma}}}(t)\right\| = \left\|\ddot{\hat{\boldsymbol{\gamma}}}(\xi(t))\right\| \,.$$
Recalling that $\xi = \psi \circ \phi^{-1}$ and the definitions of $\tilde{\kappa}$ and $\hat{\kappa}$ we conclude
$$\tilde{\kappa}(\phi(t)) = \left\|\ddot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\right\| = \left\|\ddot{\hat{\boldsymbol{\gamma}}}(\psi(t))\right\| = \hat{\kappa}(\psi(t))\,.$$

Remark 2.5: Method 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:

1. Find a unit-speed reparametrization $\tilde{\boldsymbol{\gamma}}$ of the regular curve $\boldsymbol{\gamma}$
2. This can be done by computing s the arc-length of $\boldsymbol{\gamma}$, and then defining the arc-length reparametrization
$$\tilde{\boldsymbol{\gamma}} \;:= \boldsymbol{\gamma} \circ s^{-1}$$
3. Compute curvature of $\tilde{\boldsymbol{\gamma}}$
$$\tilde{\kappa}(s) = \|\ddot{\tilde{\boldsymbol{\gamma}}}(s)\|$$
4. The curvature of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is
$$\kappa(t) = \tilde{\kappa}(s(t))$$

Important

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

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 2.2 below. The properties of the hyperbolic functions which are of interest to us are:

1. Identities:
$$\begin{gathered}
\cosh(t) + \sinh(t) = e^t \\
\cosh(t) - \sinh(t) = e^{-t} \\
\cosh^2(t) - \sinh^2(t) = 1 \\
\operatorname{sech}^2(t) - \tanh^2(t) = 1
\end{gathered}$$

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:

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

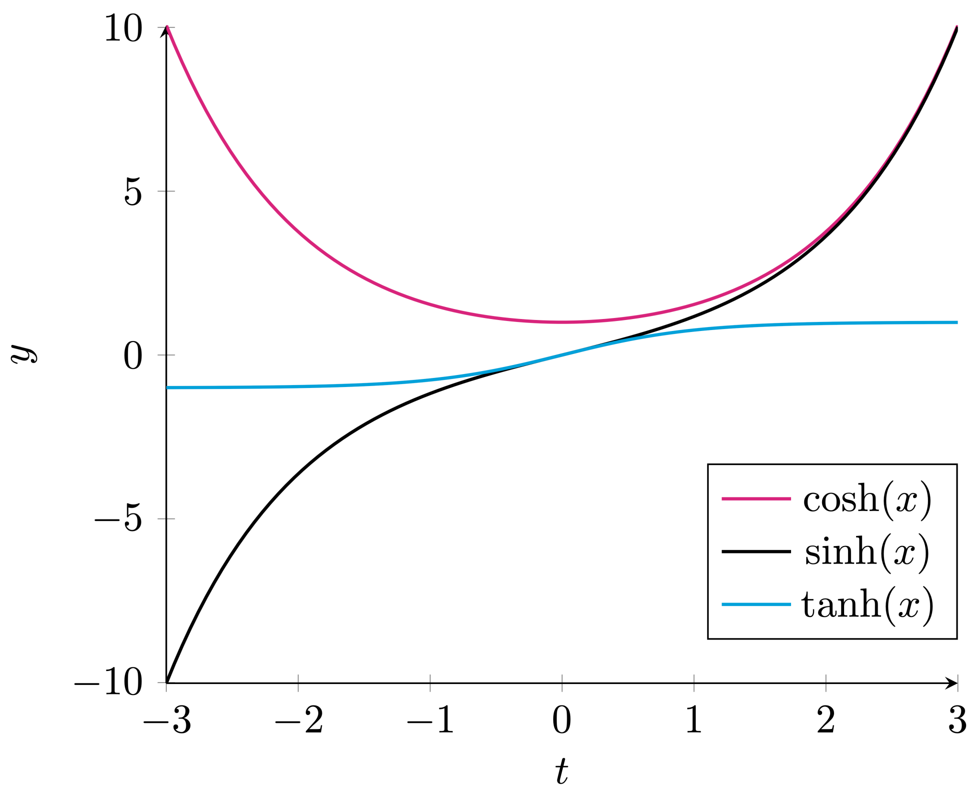


Figure 2.2: Plot of cosh, sinh, tanh.

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 2.3. 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 2.4 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 $t_0 = 0$

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

where we used that $\sinh(0) = 0$. We need to invert s. We have

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

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

$$y = e^t$$

and obtain

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

Notice that

$$y_+ = s + \sqrt{1 + s^2} \geq s + \sqrt{s^2} = s + |s| \geq 0$$

by definition of absolute value. As we were looking for y in the form $y = e^t$, we only consider the positive solution y_+. Then

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

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

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

Therefore the arc-length reparametrization of $\boldsymbol{\gamma}$ is

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

We can now compute the curvature. We have:

$$\dot{\tilde{\boldsymbol{\gamma}}}(s) = \left(\frac{1}{\sqrt{1+s^2}}, \frac{s}{\sqrt{1+s^2}}\right)$$

$$\ddot{\tilde{\boldsymbol{\gamma}}}(s) = \left(-\frac{s}{(1+s^2)^{3/2}}, \frac{1}{(1+s^2)^{3/2}}\right)$$

Hence

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

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

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

Recalling that $s(t) = \sinh(t)$, we conclude that the curvature of $\boldsymbol{\gamma}$ is

$$\kappa(t) = \tilde{\kappa}(s(t)) = \frac{1}{1+\sinh(t)^2} = \frac{1}{\cosh(t)^2}$$

where we used the identity $1 + \sinh^2 = \cosh^2$.

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, that is, if

$$\det P > 0$$

Figure 2.3: The catenary is the shape of a heavy chain suspended at its ends. Image from Wikipedia.

where $P \in \mathbb{R}^{3\times 3}$ is such that

$$\tilde{B} = P^{-1} B P \,.$$

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

$$B \sim \widetilde{B} \,.$$

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

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

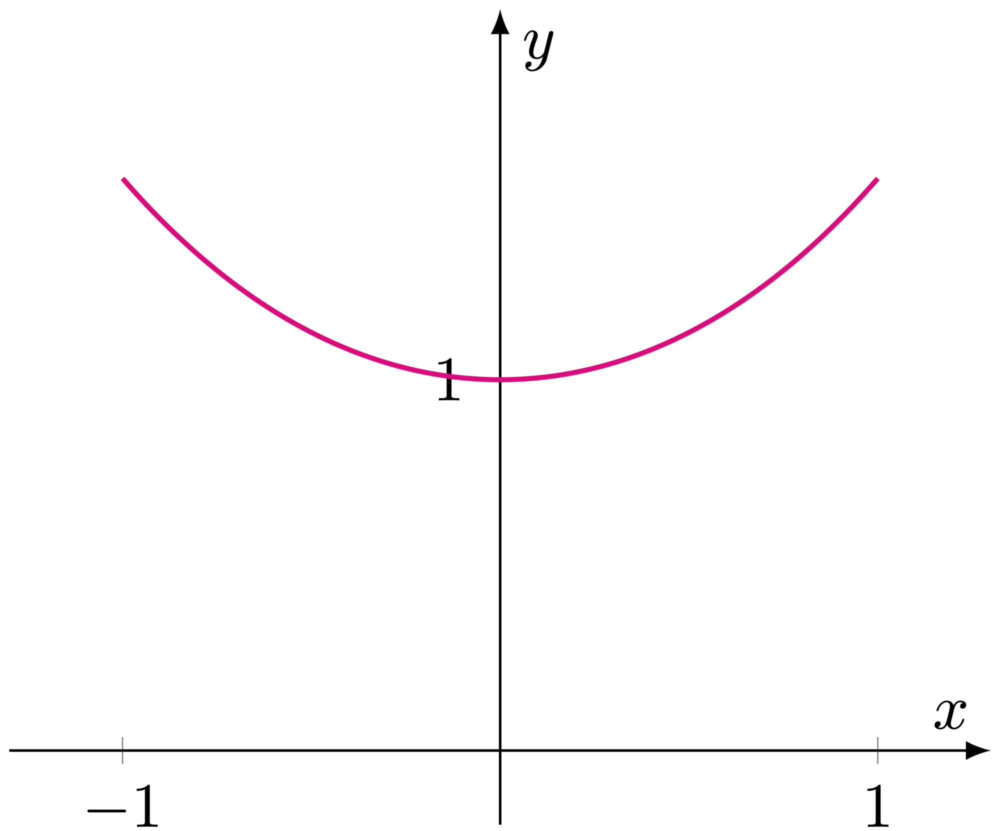


Figure 2.4: Plot of the catenary curve $\boldsymbol{\gamma}(t) = (t, \cosh(t))$.

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

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

and clearly

$$\det P = -1 < 0\,.$$

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:

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}\,, \quad \forall\, \mathbf{w} \in \mathbb{R}^3\,. \tag{2.4}$$

Here $|A|$ denotes the determinant of the matrix $A = (a_{ij})_{ij}$, and u_i, v_i, w_i are the components of $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, i.e.

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

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

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{e}_3 \,. \tag{2.5}$$

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 (2.4) 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.

Notation

In some cases we will denote formula (2.5) 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 $\mathbf{w} \in \mathbb{R}^3$, $a, b \in \mathbb{R}$
$$(a\mathbf{u} + b\mathbf{w}) \times \mathbf{v} = a\mathbf{u} \times \mathbf{v} + b\mathbf{w} \times \mathbf{w}$$
5. For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ it holds
$$(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{x} \times \mathbf{y}) = \begin{vmatrix} \mathbf{u} \cdot \mathbf{x} & \mathbf{v} \cdot \mathbf{x} \\ \mathbf{u} \cdot \mathbf{y} & \mathbf{v} \cdot \mathbf{y} \end{vmatrix} \tag{2.6}$$
6. Denote by θ the angle between $\mathbf{v}$ and $\mathbf{w}$
$$\theta := \cos^{-1}\left(\frac{\mathbf{u} \cdot \mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|} \right) .$$
The following identity holds
$$\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2(\theta) \tag{2.7}$$

Proof

- For point (1) we have

$$
\begin{aligned}
(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} &= \begin{vmatrix} u_1 & u_2 & u_2 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \\
&= - \begin{vmatrix} v_1 & v_2 & v_3 \\ u_1 & u_2 & u_2 \\ w_1 & w_2 & w_3 \end{vmatrix} \\
&= -(\mathbf{v} \times \mathbf{u}) \cdot \mathbf{w}
\end{aligned}
$$

  where we used that swapping two rows in a matrix changes the sign of the determinant. Since $\mathbf{w}$ is arbitrary, we conclude point (1).

- Let $A \in \mathbb{R}^{3 \times 3}$ be a matrix. Points (2)-(3) follow from the fact that

$$
\det(A) = 0
$$

  if and only if at least 2 rows or columns of A are linearly dependent.

- Point (4) can be easily verified by direct calculation.

- Point (5) can be obtained as follows: Check by hand that formula (2.6) holds for the vectors of the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$; Then write the vectors $\mathbf{v}, \mathbf{w}, \mathbf{x}, \mathbf{y}$ in coordinates with respect to the standard basis; Using the linearity of the vector product obtained in point (4), conclude that (2.6) holds for $\mathbf{v}, \mathbf{w}, \mathbf{x}, \mathbf{y}$.

- By definition of θ we have

$$
\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos(\theta)
$$

  Applying (2.6) with $\mathbf{x} = \mathbf{v}$ and $\mathbf{y} = \mathbf{w}$ we conclude

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

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. Let θ be the angle between $\mathbf{v}$ and $\mathbf{w}$ and A the area of the parallelogram with sides $\mathbf{u}$ and $\mathbf{v}$, see Figure 2.5. Basic trigonometry gives that

$$A = \|u\|\|v\| \sin(\theta)\,.$$

   Using (2.6) we have

$$\|\mathbf{u} \times \mathbf{v}\| = \|\mathbf{u}\|\|\mathbf{v}\| \sin(\theta) = A$$

We have therefore proven the following theorem.

Theorem 2.16: Geometric Properties of vector product

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 $\mathbf{u}, \mathbf{v}$

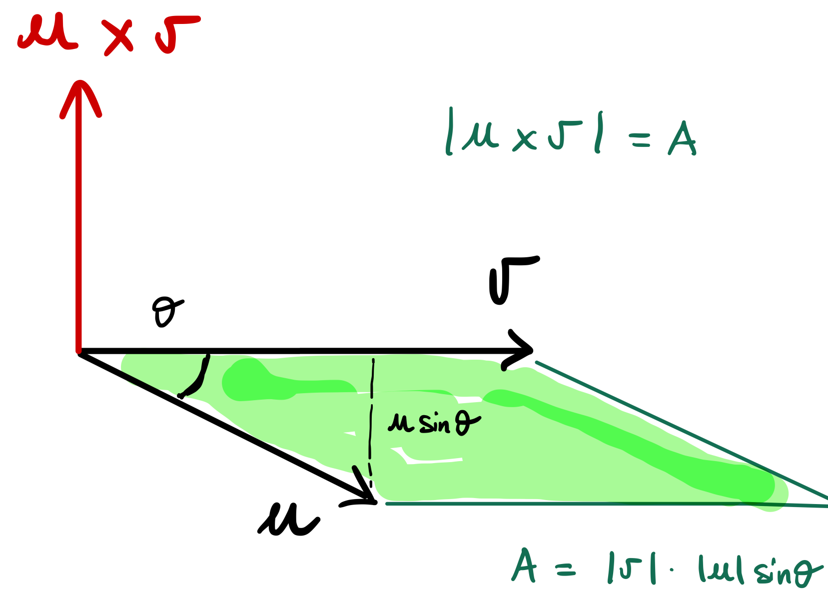


Figure 2.5: For $\mathbf{u}, \mathbf{v}$ linearly independent, $\mathbf{u} \times \mathbf{v}$ is orthogonal to the plane generated by $\mathbf{u}, \mathbf{v}$. Moreover $\|\mathbf{u} \times \mathbf{v}\|$ is the area of the parallelogram with sides $\mathbf{u}, \mathbf{v}$, and $(\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v})$ is a positive basis of $\mathbb{R}^3$

- $\mathbf{u} \times \mathbf{v}$ 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:

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u}\,. \tag{2.8}$$

Proof

Observe 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 can be checked by direct calculation.

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 (2.5).

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

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

When $\boldsymbol{\gamma}$ is regular we defined the curvature as

$$\kappa(t) := \tilde{\kappa}(s(t))$$

where

$$\tilde{\kappa}(s) := \|\ddot{\tilde{\boldsymbol{\gamma}}}(s)\|$$

is the curvature of the arc-length reparametrization $\tilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ s^{-1}$ of $\boldsymbol{\gamma}$.

If $\boldsymbol{\gamma}$ is a regular curve in $\mathbb{R}^3$ the following formula can be used to compute κ without passing through $\tilde{\boldsymbol{\gamma}}$.

Proposition 2.19: Curvature formula

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

$$\kappa(t) = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} \,. \tag{2.10}$$

We delay the proof of the above Proposition, as this will get easier when the **Frenet-Serret equations** are introduced. For a proof which does not make use of the Frenet-Serret equations see Proposition 2.1.2 in [6].

For now we use (2.10) to compute the curvature of specific curves.

Example 2.20: Curvature of the straight line

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

$$\kappa(t) = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} = 0 \,,$$

as expected.

Example 2.21: Curvature of Helix

Consider the Helix of radius $R > 0$ and rise $H > 0$

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

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:

1. 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.

2. 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

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

We deduce that 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 determines 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

We will now deal with **space curves**, that is, curves with values in $\mathbb{R}^3$. There are several issues compared to the plane case:

- A 3D counterpart of the signed curvature is not available, 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
$$\kappa^{\boldsymbol{\gamma}} = \kappa^{\boldsymbol{\eta}} \quad \text{but} \quad \boldsymbol{\gamma} \neq \boldsymbol{\eta} \,,$$
that is, $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ have same curvature but are different curves.

Example 2.28: Different curves with same curvature

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
$$\frac{1}{2} = \frac{S}{S^2 + H^2} \qquad \Longrightarrow \qquad H^2 = 2S - S^2$$
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.

We start by introducing the *principal normal*, which is just the unit vector obtained by renormalizing $\ddot{\boldsymbol{\gamma}}$.

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

The principal normal is a unit vector orthogonal to $\dot{\boldsymbol{\gamma}}$, that is,

$$\|\mathbf{n}(t)\| = 1 , \qquad \dot{\boldsymbol{\gamma}} \cdot \mathbf{n} = 0 .$$

Proof. For $\boldsymbol{\gamma}$ unit-speed we defined the curvature as

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

Therefore

$$\|\mathbf{n}\| = \frac{1}{\|\ddot{\boldsymbol{\gamma}}(t)\|} \|\ddot{\boldsymbol{\gamma}}(t)\| = 1$$

In addition for $\boldsymbol{\gamma}$ unit-speed it holds that $\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0$. Therefore

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{n} = \frac{1}{\kappa} \dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = 0 .$$

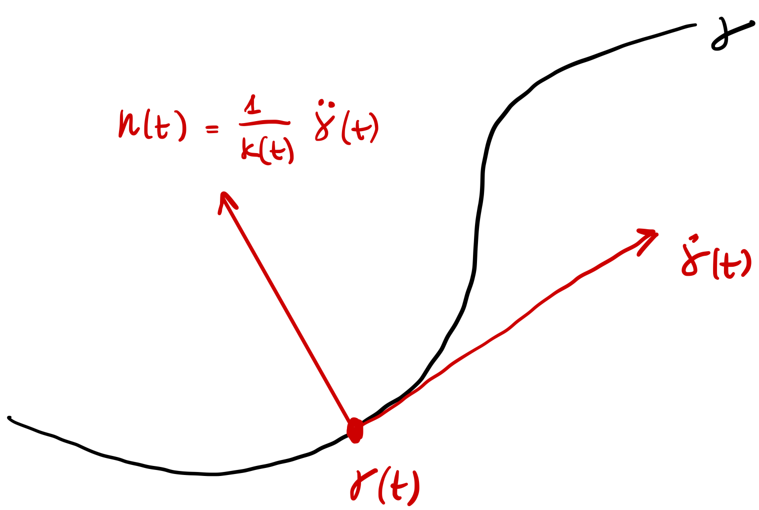


Figure 2.6: Principal normal vector $\mathbf{n}(t)$ to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$.

Question 2.31

Why is the principal normal interesting? Because it can tell the difference beween a plane curve and a space curve, see Figure 2.7.

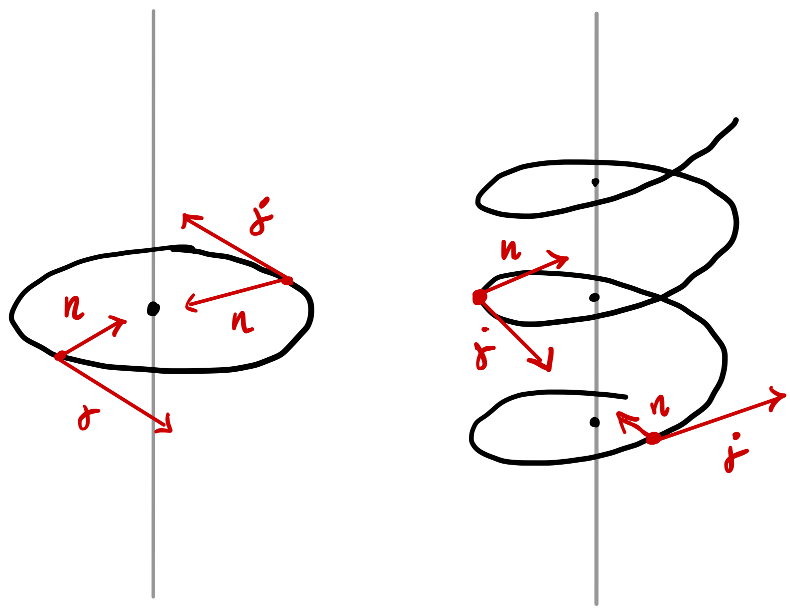


Figure 2.7: Left: Principal normal to a circle. Note that $\mathbf{n}$ always points towards the origin $\mathbf{0}$. Right: Principal normal to a helix. Note that $\mathbf{n}$ points towards the z-axis, but never towards the same point.

We now introduce the binormal vector $\mathbf{b}$ as the vector product of $\dot{\boldsymbol{\gamma}}$ and $\mathbf{n}$. By the properties of vector product we will see that the triple

$$(\dot{\boldsymbol{\gamma}}, \mathbf{n}, \mathbf{b})$$

forms a positive orthonormal basis of $\mathbb{R}^3$.

Definition 2.32: Binormal vector

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be a unit-speed curve with

$$\kappa(t) \neq 0\,, \quad \forall\, t \in (a, b)\,.$$

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 $\mathbb{R}^3$.

Therefore to each unit-speed curve $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ with non-vanishing curvature we can associate a positive orthonormal basis of $\mathbb{R}^3$. Such basis is known as Frenet frame.

Definition 2.35: Frenet frame

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be unit-speed with $\kappa \neq 0$. The positive orthonormal basis of $\mathbb{R}^3$

$$(\dot{\boldsymbol{\gamma}}, \mathbf{n}, \mathbf{b})$$

is called **Frenet frame** of $\boldsymbol{\gamma}$.

Notation

For $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ unit-speed the **tangent vector** is often denoted by

$$\mathbf{t} := \dot{\boldsymbol{\gamma}}$$

Therefore the Frenet frame of $\boldsymbol{\gamma}$ can be equivalently written as

$$(\mathbf{t}, \mathbf{n}, \mathbf{b}) .$$

By using unit-speed reparametrizations we can also compute the Frenet frame for regular curves with non-vanishing curvature. In doing so, we need to be aware of the following:

Warning

The Frenet frame depends on the **orientation** of the curve, see next Definition and Proposition.

Definition 2.36

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be regular and $\tilde{\boldsymbol{\gamma}}$ be a reparametrization with

$$\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi , \quad \phi \colon (a,b) \to (\tilde{a}, \tilde{b}) .$$

We say that

1. $\tilde{\boldsymbol{\gamma}}$ is *orientation preserving* if $\dot{\phi} > 0$
2. $\tilde{\boldsymbol{\gamma}}$ is *orientation reversing* if $\dot{\phi} < 0$

Proposition 2.37: Frenet frame of reparametrization

Let $\boldsymbol{\gamma}$ be a unit-speed curve with $\kappa \neq 0$. Let $\tilde{\boldsymbol{\gamma}}$ be a unit-speed reparametrization with

$$\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi , \quad \phi \colon (a,b) \to (\tilde{a}, \tilde{b}) .$$

Then $\dot{\phi}$ is constant, with either

$$\dot{\phi} \equiv 1 \quad \text{or} \quad \dot{\phi} \equiv -1$$

Denote by

$$(\mathbf{t}, \mathbf{n}, \mathbf{b}) , \qquad (\tilde{\mathbf{t}}, \tilde{\mathbf{n}}, \tilde{\mathbf{b}})$$

the Frenet frames of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$, respectively. We have:

1. If $\tilde{\boldsymbol{\gamma}}$ is *orientation preserving* then $\dot{\phi} \equiv 1$ and
$$\mathbf{t} = \tilde{\mathbf{t}} \circ \phi, \quad \mathbf{n} = \tilde{\mathbf{n}} \circ \phi, \quad \mathbf{b} = \tilde{\mathbf{b}} \circ \phi$$
2. If $\tilde{\boldsymbol{\gamma}}$ is *orientation reversing* then $\dot{\phi} \equiv -1$ and
$$\mathbf{t} = -\tilde{\mathbf{t}} \circ \phi, \quad \mathbf{n} = \tilde{\mathbf{n}} \circ \phi, \quad \mathbf{b} = -\tilde{\mathbf{b}} \circ \phi$$

Proof

Differentiating $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi$ gives
$$\dot{\boldsymbol{\gamma}}(t) = \dot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\, \dot{\phi}(t) \tag{2.11}$$
Taking the norms in (2.11) and recalling that $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ are unit speed yields $|\dot{\phi}| = 1$. By continuity of $\dot{\phi}$ either
$$\dot{\phi} \equiv 1 \quad \text{or} \quad \dot{\phi} \equiv -1 \tag{2.12}$$
Differentiating (2.11) one more time
$$\begin{aligned} \ddot{\boldsymbol{\gamma}}(t) &= \ddot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\, \dot{\phi}^2(t) + \dot{\tilde{\boldsymbol{\gamma}}}(\phi(t))\, \ddot{\phi}(t) \\ &= \ddot{\tilde{\boldsymbol{\gamma}}}(\phi(t)) \end{aligned} \tag{2.13}$$
where we used (2.12). By definition
$$\mathbf{t} := \dot{\boldsymbol{\gamma}}\,, \qquad \tilde{\mathbf{t}} := \dot{\tilde{\boldsymbol{\gamma}}}$$
Therefore (2.11) reads
$$\mathbf{t}(t) = \tilde{\mathbf{t}}(\phi(t))\, \dot{\phi}(t) \tag{2.14}$$
By Proposition 2.4 the curvatures $\kappa, \tilde{\kappa}$ of $\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}$ are related by
$$\kappa(t) = \tilde{\kappa}(\phi(t))\,. \tag{2.15}$$
Dividing both sides of (2.13) by $\kappa(t)$ and using (2.15) gives
$$\begin{aligned} \frac{1}{\kappa(t)}\, \ddot{\boldsymbol{\gamma}}(t) &= \frac{1}{\kappa(t)}\, \ddot{\tilde{\boldsymbol{\gamma}}}(\phi(t)) \\ &= \frac{1}{\tilde{\kappa}(\phi(t))}\, \ddot{\tilde{\boldsymbol{\gamma}}}(\phi(t)) \end{aligned} \tag{2.16}$$
By definition the principal normals are
$$\mathbf{n} := \frac{1}{\kappa}\, \ddot{\boldsymbol{\gamma}}\,, \qquad \tilde{\mathbf{n}} := \frac{1}{\tilde{\kappa}}\, \ddot{\tilde{\boldsymbol{\gamma}}}$$
and therefore (2.16) reads
$$\mathbf{n}(t) = \tilde{\mathbf{n}}(\phi(t)) \tag{2.17}$$

Recall the definition of binormal

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}\,, \qquad \tilde{\mathbf{b}} = \tilde{\mathbf{t}} \times \tilde{\mathbf{n}}$$

Using (2.14) and (2.17) then gives

$$\begin{aligned} \mathbf{b}(t) &= \mathbf{t}(t) \times \mathbf{n}(t) \\ &= \tilde{\mathbf{t}}(\phi(t)) \times \tilde{\mathbf{n}}(\phi(t))\, \dot{\phi}(t) \\ &= \tilde{\mathbf{b}}(\phi(t))\, \dot{\phi}(t) \end{aligned}$$

To summarize, we have shown the following relations between the Frenet frames of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$

$$\begin{aligned} \mathbf{t}(t) &= \tilde{\mathbf{t}}(\phi(t))\, \dot{\phi}(t) \\ \mathbf{n}(t) &= \tilde{\mathbf{n}}(\phi(t)) \\ \mathbf{b}(t) &= \tilde{\mathbf{b}}(\phi(t))\, \dot{\phi}(t) \end{aligned} \tag{2.18}$$

We can finally conclude:

1. If $\tilde{\boldsymbol{\gamma}}$ is orientation preserving then $\dot{\phi} > 0$. By (2.12) we infer $\dot{\phi} \equiv 1$, so that the equations at (2.18) read
$$\mathbf{t} = \tilde{\mathbf{t}} \circ \phi, \quad \mathbf{n} = \tilde{\mathbf{n}} \circ \phi, \quad \mathbf{b} = \tilde{\mathbf{b}} \circ \phi$$
2. If $\tilde{\boldsymbol{\gamma}}$ is orientation reversing then $\dot{\phi} < 0$. By (2.12) we infer $\dot{\phi} \equiv -1$, so that the equations at (2.18) read
$$\mathbf{t} = -\tilde{\mathbf{t}} \circ \phi, \quad \mathbf{n} = \tilde{\mathbf{n}} \circ \phi, \quad \mathbf{b} = -\tilde{\mathbf{b}} \circ \phi$$

In conclusion, the Frenet frame is not invariant under reparametrization. However the Frenet vectors stay the same, up to changing the sign of tangent and binormal.

$$\mathbf{t} = \pm\tilde{\mathbf{t}} \circ \phi, \quad \mathbf{n} = \tilde{\mathbf{n}} \circ \phi, \quad \mathbf{b} = \pm\tilde{\mathbf{b}} \circ \phi$$

Let us conclude the section with an example, where we compute the Frenet frame of the Helix.

Example 2.38: Frenet frame of Helix

Consider the helix of radius $R > 0$ and rise H given by

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

for $t \in \mathbb{R}$. We have that

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

which implies

$$\|\dot{\boldsymbol{\gamma}}\| = \rho, \qquad \rho := \sqrt{R^2 + H^2}\,.$$

This shows that $\boldsymbol{\gamma}$ is regular, since

$$\|\dot{\boldsymbol{\gamma}}\| = \rho = \sqrt{R^2 + H^2} \geq \sqrt{R^2} = R > 0$$

However $\boldsymbol{\gamma}$ is not unit-speed in general. We know that all unit-speed reparametrizations are of the form

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

where s is the arc-length and $c \in \mathbb{R}$. Therefore we compute the arc-length of $\boldsymbol{\gamma}$ starting at $t_0 = 0$

$$s(t) := \int_0^t \|\dot{\boldsymbol{\gamma}}(u)\| \, du = \rho t\,.$$

The arc-length function is invertible with inverse

$$\psi(s) := t(s) = \frac{s}{\rho}\,.$$

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

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

We compute the tangent vector to $\tilde{\boldsymbol{\gamma}}$ and its derivative

$$\begin{aligned} \tilde{\mathbf{t}} &= \dot{\tilde{\boldsymbol{\gamma}}} = \frac{1}{\rho}\left(-R\sin\left(\frac{s}{\rho}\right), R\cos\left(\frac{s}{\rho}\right), H \right) \\ \dot{\tilde{\mathbf{t}}} &= \frac{R}{\rho^2}\left(-\cos\left(\frac{s}{\rho}\right), -\sin\left(\frac{s}{\rho}\right), 0 \right) \end{aligned}$$

We have already computed the curvature of the Helix in Example 2.21 by using the general formula. We now compute it again, but this time by using the unit-speed reparametrization just obtained:

$$\tilde{\kappa}(s) = \left\|\dot{\tilde{\mathbf{t}}}\right\| = \frac{R}{\rho^2} = \frac{R}{R^2 + H^2}\,.$$

The principal normal vector is then

$$\tilde{\mathbf{n}} = \frac{1}{\tilde{\kappa}(t)}\,\dot{\tilde{\mathbf{t}}} = \left(-\cos\left(\frac{s}{\rho}\right), -\sin\left(\frac{s}{\rho}\right), 0 \right)\,.$$

We can now compute the binormal

$$\begin{aligned} \tilde{\mathbf{b}} &= \tilde{\mathbf{t}} \times \tilde{\mathbf{n}} \\ &= \frac{1}{\rho} \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -R\sin\left(\frac{s}{\rho}\right) & R\cos\left(\frac{s}{\rho}\right) & H \\ -\cos\left(\frac{s}{\rho}\right) & -\sin\left(\frac{s}{\rho}\right) & 0 \end{vmatrix} \\ &= \frac{1}{\rho}\left(H\sin\left(\frac{s}{\rho}\right), -H\cos\left(\frac{s}{\rho}\right), R \right)\,. \end{aligned}$$

This concludes the calculation of the Frenet frame of $\tilde{\boldsymbol{\gamma}}$. For the choice of $R = 1$ and $H = 1$ the Frenet frame is plotted in Figure 2.8.

Note: If we had reparametrized $\boldsymbol{\gamma}$ by $-s$ instead of s, we would have obtained the Frenet frame

$$(-\tilde{\mathbf{t}}, \tilde{\mathbf{n}}, -\tilde{\mathbf{b}})$$

in accordance with Proposition 2.37.

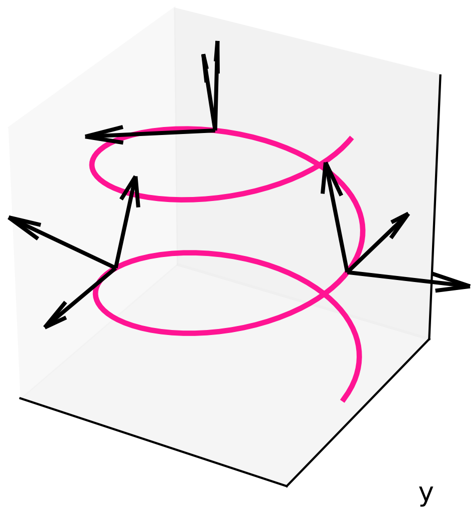


Figure 2.8: Frenet frame of the Helix with R = 1 and H = 1.

2.6 Torsion

For space curves with non-vanishing curvature we can define another scalar quantity, known as *torsion*. Such quantity allows to measure by how much a curve fails to be planar.

The torsion can be defined by computing the derivative of the binormal vector $\mathbf{b}$.

Proposition 2.39

Let $\boldsymbol{\gamma} \colon (a, b) \to \mathbb{R}^3$ be a unit-speed curve with $\kappa \neq 0$. Then

$$\dot{\mathbf{b}}(t) = -\tau(t)\mathbf{n}(t) \,, \tag{2.19}$$

for some $\tau(t) \in \mathbb{R}$.

Proof

By definition $\mathbf{b} = \mathbf{t} \times \mathbf{n}$. Using the formula of derivation of the cross product (2.9) we have

$$\begin{aligned} \dot{\mathbf{b}} &= \frac{d}{dt}(\mathbf{t} \times \mathbf{n}) \\ &= \dot{\mathbf{t}} \times \mathbf{n} + \mathbf{t} \times \dot{\mathbf{n}} \\ &= \mathbf{t} \times \dot{\mathbf{n}}\,, \end{aligned}$$

where we used that, by definition of $\mathbf{n}$,

$$\dot{\mathbf{t}} \times \mathbf{n} = \frac{1}{\kappa}\dot{\mathbf{t}} \times \dot{\mathbf{t}} = \mathbf{0}\,.$$

This shows

$$\dot{\mathbf{b}} = \dot{\boldsymbol{\gamma}} \times \dot{\mathbf{n}}\,. \tag{2.20}$$

By the properties of the cross product we have that $\mathbf{t} \times \dot{\mathbf{n}}$ is orthogonal to both $\mathbf{t}$ and $\dot{\mathbf{n}}$. Thus (2.20) implies that

$$\dot{\mathbf{b}} \cdot \mathbf{t} = 0\,.$$

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

$$\dot{\mathbf{b}} \cdot \mathbf{b} = 0\,.$$

showing that $\dot{\mathbf{b}}$ is orthogonal to $\mathbf{b}$. We also shown that $\dot{\mathbf{b}}$ is orthogonal to $\mathbf{t}$. Since the Frenet frame

$$(\mathbf{t}, \mathbf{n}, \mathbf{b})$$

is an orthonormal basis of $\mathbb{R}^3$, and $\dot{\mathbf{b}}$ is orthogonal to both $\mathbf{t}$ and $\mathbf{b}$, 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 (2.19) is called the torsion of $\boldsymbol{\gamma}$.

Definition 2.40: Torsion of unit-speed curve with $\kappa \neq 0$

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 2.41: Formula for torsion

In particular the torsion satisfies:

$$\tau(t) = -\dot{\mathbf{b}}(t) \cdot \mathbf{n}(t)\,.$$

Proof. Multiply (2.19) by $\mathbf{n}$ and note that $\|\mathbf{n}\| = 1$.

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

$$\|\dot{\boldsymbol{\gamma}}(t)\| = 1\,, \quad \kappa(t) = \|\ddot{\boldsymbol{\gamma}}(t)\| \neq 0\,, \qquad \forall\, t \in (a,b)\,.$$

As we did for curvature, we can extend the definition of torsion to regular curves $\boldsymbol{\gamma}$ with non-vanishing curvature.

Definition 2.42: Torsion of regular curve with $\kappa \neq 0$

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a regular curve, with $\kappa \neq 0$. Let $\tilde{\boldsymbol{\gamma}}$ be a unit-speed reparametrization of $\boldsymbol{\gamma}$, with

$$\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi\,, \quad \phi \colon (a,b) \to (\tilde{a},\tilde{b})\,.$$

Denote by $\tilde{\tau} \colon (\tilde{a},\tilde{b}) \to \mathbb{R}$ the torsion of $\tilde{\boldsymbol{\gamma}}$. The **torsion** of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is defined by

$$\tau(t) := \tilde{\tau}(\phi(t))$$

As usual, we need to check that the above definition of torsion does not depend on the choice of unit-speed reparametrization $\tilde{\boldsymbol{\gamma}}$.

Proposition 2.43: τ is invariant for unit-speed reparametrization

Consider the setting of Definition 2.42. Let $\hat{\boldsymbol{\gamma}}$ is another unit-speed reparametrization of $\boldsymbol{\gamma}$, with $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}} \circ \psi$. Then

$$\tau(t) = \tilde{\tau}(\phi(t)) = \hat{\tau}(\psi(t))$$

where $\hat{\tau}$ is the torsion of $\hat{\boldsymbol{\gamma}}$.

Proof

The curves $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ are unit-speed, therefore they are defined their Frenet frames

$$(\tilde{\mathbf{t}}, \tilde{\mathbf{n}}, \tilde{\mathbf{b}}) \,, \qquad (\hat{\mathbf{t}}, \hat{\mathbf{n}}, \hat{\mathbf{b}})$$

Since $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ are both reparametrization of $\boldsymbol{\gamma}$

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

Using that ϕ is invertible we obtain

$$\tilde{\boldsymbol{\gamma}}(t) = \hat{\boldsymbol{\gamma}}(\xi(t)) \,, \qquad \xi := \psi \circ \phi^{-1}$$

with ξ diffeomorphims. The above formula is saying that $\hat{\boldsymbol{\gamma}}$ is a reparametrization of $\tilde{\boldsymbol{\gamma}}$. As both $\tilde{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}$ are unit-speed, we can apply Proposition 2.37 and infer that the Frenet frames are linked by the formulas

$$\tilde{\mathbf{t}} = \pm \hat{\mathbf{t}} \circ \xi \,, \qquad \tilde{\mathbf{n}} = \hat{\mathbf{n}} \circ \xi \,, \qquad \tilde{\mathbf{b}} = \pm \hat{\mathbf{b}} \circ \xi \tag{2.21}$$

and ξ satisfies

$$\dot{\xi} \equiv \pm 1 \,.$$

Differentiating the third equation in (2.21) gives

$$\dot{\tilde{\mathbf{b}}}(t) = \pm \frac{d}{dt} \hat{\mathbf{b}}(\xi(t)) = \pm \dot{\hat{\mathbf{b}}}(\xi(t)) \dot{\xi}(t) = \dot{\hat{\mathbf{b}}}(\xi(t)) \tag{2.22}$$

where we used that $\dot{\xi} \equiv \pm 1$. By Remark 2.41 the torsions of $\tilde{\boldsymbol{\gamma}}$ and $\dot{\hat{\boldsymbol{\gamma}}}$ are computed by

$$\tilde{\tau} = -\dot{\tilde{\mathbf{b}}} \cdot \tilde{\mathbf{n}} \,, \qquad \hat{\tau} = -\dot{\hat{\mathbf{b}}} \cdot \hat{\mathbf{n}}$$

Using the second equation in (2.21) and (2.22) allows to infer

$$\tilde{\tau}(t) = -\dot{\tilde{\mathbf{b}}}(t) \cdot \tilde{\mathbf{n}}(t) = -\dot{\hat{\mathbf{b}}}(\xi(t)) \cdot \hat{\mathbf{n}}(\xi(t)) = \hat{\tau}(\xi(t))$$

Recalling that $\xi = \psi \circ \phi^{-1}$ we conclude

$$\tilde{\tau}(\phi(t)) = \hat{\tau}(\psi(t))$$

as required.

As with the curvature, there is a general formula to compute the torsion without having to reparametrize.

Proposition 2.44: Torsion formula

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be a regular curve with non-vanishing curvature. The torsion $\tau(t)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is given by

$$\tau(t) = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} \,. \tag{2.23}$$

We delay the proof of the above Proposition, as this will get easier when the **Frenet-Serret equations** are introduced. For a proof which does not make use of the Frenet-Serret equations, see the proof of Proposition 2.3.1 in [6].

For now we use (2.23) to compute the curvature of specific curves.

Example 2.45: 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:

$$\tau = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} = \frac{H}{R^2 + H^2}\,.$$

As a consequence of the above example, we can immediately infer curvature and torsion formulas for the circle.

Example 2.46: 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 2.47

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 in Figure 2.9, $\boldsymbol{\gamma}$ 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. Therefore we can compute curvature and torsion by using the Frenet frame. We have

$$\ddot{\boldsymbol{\gamma}} = \left(-\frac{4}{5}\cos(t), \sin(t), \frac{3}{5}\cos(t)\right) .$$

As $\boldsymbol{\gamma}$ is unit-speed, the curvature and principal normal are

$$\kappa = \|\ddot{\boldsymbol{\gamma}}\| = \frac{16}{25}\cos^2(t) + \sin^2(t) + \frac{9}{25}\cos^2(t) = 1$$

$$\mathbf{n} = \frac{1}{\kappa}\ddot{\boldsymbol{\gamma}} = \left(-\frac{4}{5}\cos(t), \sin(t), \frac{3}{5}\cos(t)\right) .$$

We can then compute the binormal

$$
\begin{aligned}
\mathbf{b} &= \dot{\boldsymbol{\gamma}} \times \mathbf{n} \\
&= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\frac{4}{5}\sin(t) & -\cos(t) & \frac{3}{5}\sin(t) \\ -\frac{4}{5}\cos(t) & \sin(t) & \frac{3}{5}\cos(t) \end{vmatrix} \\
&= \left(-\frac{3}{5}\cos^2(t) - \frac{3}{5}\sin^2(t), -\frac{12}{25}\cos(t)\sin(t) + \frac{12}{25}\cos(t)\sin(t), -\frac{4}{5}\sin^2(t) - \frac{4}{5}\cos^2(t)\right) \\
&= \left(-\frac{3}{5}, 0, -\frac{4}{5}\right) .
\end{aligned}
$$

In particular

$$\dot{\mathbf{b}} = 0 \,,$$

and we obtain that the torsion is

$$\tau = -\dot{\mathbf{b}} \cdot \mathbf{n} = 0 \,.$$

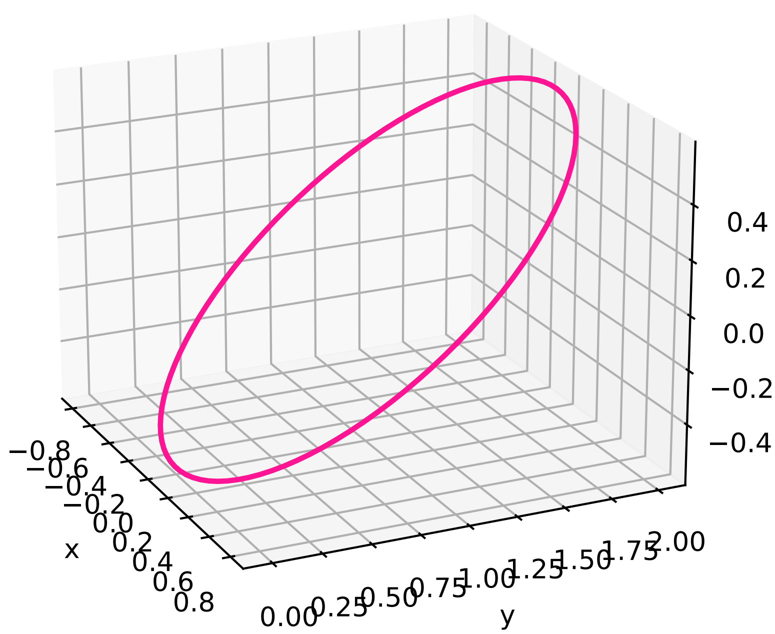


Figure 2.9: Plot of the curve in Example above

Let us summarize our findings about curvature and torsion.

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.
3. Both κ and τ are invariant under unit-speed reparametrizations

The two strategies for computing κ and τ are summarized in the diagram in Figure 2.10.

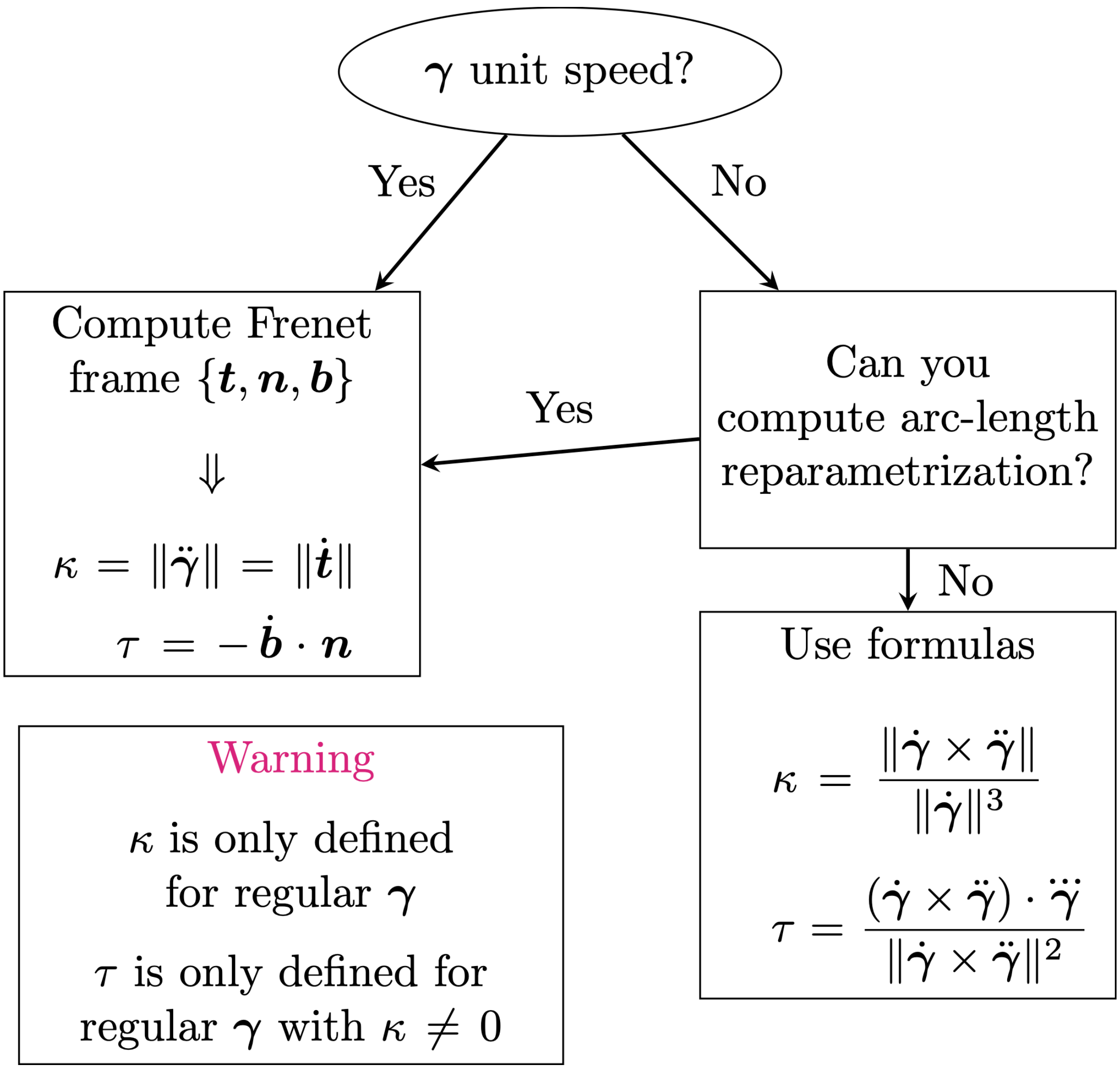


Figure 2.10: How to compute κ and τ for regular curve $\boldsymbol{\gamma}$.

We have already made an example in which we compute curvature and torsion of the Helix using the general formulas

$$\kappa = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^2}, \qquad \tau = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2}.$$

We conclude the section by providing an example where we compute curvature and torsion by making use of the Frenet frame.

Example 2.48: Curvature and torsion of Helix with Frenet frame

Consider the helix of radius $R > 0$ and rise H given by

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

for $t \in \mathbb{R}$. We want to compute curvature and torsion by following the diagram at Figure 2.10. We ask the first question:

Is $\boldsymbol{\gamma}$ unit-speed?

We have already computed in Example 2.38 that

$$\|\dot{\boldsymbol{\gamma}}\| = \rho, \qquad \rho := \sqrt{R^2 + H^2}.$$

This shows that $\boldsymbol{\gamma}$ is regular but not unit-speed. We ask the second question in the diagram:

Can we find the arc-length reparametrization of $\boldsymbol{\gamma}$?

We have already computed the arc-length reparametrization of $\boldsymbol{\gamma}$ in Example 2.38. This is given by

$$\tilde{\boldsymbol{\gamma}}(s) = \left(R\cos\left(\frac{s}{\rho}\right), R\sin\left(\frac{s}{\rho}\right), \frac{Hs}{\rho} \right).$$

The next step in the diagram is

Compute Frenet frame $\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}$ and curvature κ, torsion τ

From Example 2.38, the Frenet frame and curvature of $\tilde{\boldsymbol{\gamma}}$ are

$$\begin{aligned}
\tilde{\mathbf{t}} &= \frac{1}{\rho}\left(-R\sin\left(\frac{s}{\rho}\right), R\cos\left(\frac{s}{\rho}\right), H \right) \\
\tilde{\mathbf{n}} &= \left(-\cos\left(\frac{s}{\rho}\right), -\sin\left(\frac{s}{\rho}\right), 0 \right) \\
\tilde{\mathbf{b}} &= \frac{1}{\rho}\left(H\sin\left(\frac{s}{\rho}\right), -H\cos\left(\frac{s}{\rho}\right), R \right) \\
\tilde{\kappa} &= \left\|\dot{\tilde{\mathbf{t}}}\right\| = \frac{R}{\rho^2} = \frac{R}{R^2 + H^2}
\end{aligned}$$

we are left to compute the torsion using formula

$$\tilde{\tau} = -\dot{\tilde{\mathbf{b}}} \cdot \tilde{\mathbf{n}}$$

Indeed, we have

$$\dot{\tilde{\mathbf{b}}} = \frac{H}{\rho^2}\left(\cos\left(\frac{s}{\rho}\right), \sin\left(\frac{s}{\rho}\right), 0 \right)$$

and therefore

$$\dot{\tilde{\mathbf{b}}} \cdot \mathbf{n} = \frac{H}{\rho^2}\left(-\cos^2\left(\frac{s}{\rho}\right) - \sin^2\left(\frac{s}{\rho}\right)\right) = -\frac{H}{\rho^2}\,.$$

The torsion is then

$$\tilde{\tau} = -\dot{\tilde{\mathbf{b}}} \cdot \tilde{\mathbf{n}} = \frac{H}{\rho^2} = \frac{H}{R^2 + H^2}\,,$$

which of course agrees with the calculation for τ in Example 2.45.

2.7 Frenet-Serret equations

For unit-speed curves $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ with non-vanishing curvature we introduced the Frenet frame

$$\{\mathbf{t}, \mathbf{n}, \mathbf{b}\}\,.$$

We proved that the Frenet frame is a positive orthonormal basis of $\mathbb{R}^3$. We also used such basis to compute curvature κ and torsion τ of $\boldsymbol{\gamma}$:

$$\kappa := \|\dot{\mathbf{t}}\|\,, \quad \tau := -\dot{\mathbf{b}} \cdot \mathbf{n}\,.$$

In this section we show that the Frenet frame satisfies a linear system of ODEs known as the Frenet-Serret equations. In order to do this, we first need prove that the Frenet frame

$$(\mathbf{t}, \mathbf{n}, \mathbf{b})$$

is *right-handed*. Such property holds in general for any positive basis of $\mathbb{R}^3$ of the form

$$(\mathbf{u}, \mathbf{v}, \mathbf{w})\,, \qquad \mathbf{w} := \mathbf{u} \times \mathbf{v}\,.$$

Proposition 2.49: Frenet frame is right-handed

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be a unit-speed curve with $\kappa \neq 0$. The Frenet frame is *right-handed*, in the sense that

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}\,, \quad \mathbf{n} = \mathbf{b} \times \mathbf{t}\,, \quad \mathbf{t} = \mathbf{n} \times \mathbf{b}\,. \tag{2.24}$$

Proof

The first equation in (2.24) is true by definition of $\mathbf{b}$. For the remaining 2 equations, recall formula (2.8):

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u}\,, \tag{2.25}$$

which holds for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. Applying (2.25) to

$$\mathbf{u} = \mathbf{t}\,, \quad \mathbf{v} = \mathbf{n}\,, \quad \mathbf{w} = \mathbf{t}\,,$$

yields

$$\begin{aligned}(\mathbf{t} \times \mathbf{n}) \times \mathbf{t} &= (\mathbf{t} \cdot \mathbf{t})\mathbf{n} - (\mathbf{n} \cdot \mathbf{t})\mathbf{t} \\ &= \|\mathbf{t}\|^2 \, \mathbf{n} - \mathbf{0} \\ &= \mathbf{n} \,,\end{aligned}$$

where we used that $\mathbf{t}$ is a unit vector and $\mathbf{n} \cdot \mathbf{t} = 0$. Therefore, by definition of $\mathbf{b}$, we have

$$\mathbf{b} \times \mathbf{t} = (\mathbf{t} \times \mathbf{n}) \times \mathbf{t} = \mathbf{n}$$

obtaining the second equation in (2.24). Now we apply (2.25) to

$$\mathbf{u} = \mathbf{t} \,, \quad \mathbf{v} = \mathbf{n} \,, \quad \mathbf{w} = \mathbf{n} \,,$$

to get

$$\begin{aligned}(\mathbf{t} \times \mathbf{n}) \times \mathbf{n} &= (\mathbf{t} \cdot \mathbf{n})\mathbf{n} - (\mathbf{n} \cdot \mathbf{n})\mathbf{t} \\ &= \mathbf{0} - \|\mathbf{n}\|^2 \mathbf{t} \\ &= -\mathbf{t}\end{aligned}$$

where we used that $\mathbf{n}$ is a unit vector and $\mathbf{t} \cdot \mathbf{n} = 0$. Therefore, by definition of $\mathbf{b}$ and anti-commutativity of the vector product, we have

$$\mathbf{n} \times \mathbf{b} = -\mathbf{b} \times \mathbf{n} = -(\mathbf{t} \times \mathbf{n}) \times \mathbf{n} = \mathbf{t} \,,$$

obtaining the last equation in (2.24).

Theorem 2.50: Frenet-Serret equations

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be unit-speed with $\kappa \neq 0$. The **Frenet-Serret** equations are

$$\begin{aligned}\dot{\mathbf{t}} &= \kappa \mathbf{n} \\ \dot{\mathbf{n}} &= -\kappa \mathbf{t} + \tau \mathbf{b} \\ \dot{\mathbf{b}} &= -\tau \mathbf{n}\end{aligned}$$

Proof

The first Frenet-Serret equation

$$\dot{\mathbf{t}} = \kappa \mathbf{n} \tag{2.26}$$

is just the definition of $\mathbf{n}$. The third Frenet-Serret equation

$$\dot{\mathbf{b}} = -\tau \mathbf{n} \tag{2.27}$$

holds by Proposition 2.39. Now, recall that in Proposition 2.49 we have proven

$$\mathbf{b} = \mathbf{t} \times \mathbf{n}\,, \quad \mathbf{n} = \mathbf{b} \times \mathbf{t}\,, \quad \mathbf{t} = \mathbf{n} \times \mathbf{b}\,. \tag{2.28}$$

Differentiating the second equation in (2.28) and using (2.26)-(2.27) we get

$$\begin{aligned} \dot{\mathbf{n}} &= \dot{\mathbf{b}} \times \mathbf{t} + \mathbf{b} \times \dot{\mathbf{t}} \\ &= (-\tau \mathbf{n} \times \mathbf{t}) + \mathbf{b} \times \kappa \mathbf{n} \\ &= \tau(\mathbf{t} \times \mathbf{n}) - \kappa(\mathbf{n} \times \mathbf{b}) \\ &= \tau \mathbf{b} - \kappa \mathbf{t}\,, \end{aligned}$$

where in the last equality we used the first and third equations in (2.28). The above is exactly the second Frenet-Serret equation.

Remark 2.51: Vectorial form of Frenet-Serret equations

We can write the Frenet-Serret ODE sysyem in vectorial form. Introduce the vector of the Frenet frame

$$\mathbf{\Gamma} = (\mathbf{t}, \mathbf{n}, \mathbf{b})$$

This way $\mathbf{\Gamma}$ is a 9 dimensional time-dependent vector

$$\mathbf{\Gamma} : \ (a, b) \to \mathbb{R}^9$$

Also define the block matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \kappa I & \mathbf{0} \\ -\kappa I & \mathbf{0} & \tau I \\ \mathbf{0} & -\tau I & \mathbf{0} \end{pmatrix},$$

where we denoted

$$\mathbf{0} := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad I := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This way $\mathbf{A}$ is a 9×9 time-dependent matrix

$$\mathbf{A} : \ (a, b) \to \mathbb{R}^{9 \times 9}$$

It is immediate to check that the Frenet-Serret equations can be written as

$$\dot{\mathbf{\Gamma}} = \mathbf{A}\mathbf{\Gamma}$$

Note: The matrix $\mathbf{A}$ is anti-symmetric, that is

$$\mathbf{A}^T = -\mathbf{A}\,.$$

This observation will be crucial in proving the *Fundamental Theorem of Space Curves*, which is stated in the next section.
Alternative Notation: With a little abuse of notation we can also write the Frenet-Serret equations as

$$\dot{\boldsymbol{\Gamma}} = A\boldsymbol{\Gamma}$$

where A is the 3×3 time-dependent matrix

$$A : (a,b) \to \mathbb{R}^{3\times 3}, \quad A := \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix},$$

and where we think $\boldsymbol{\Gamma}$ as a 3 dimensional vector, with each component being a function $\mathbf{t}$, $\mathbf{n}$ and $\mathbf{b}$. Note that the block in poisition (i,j) of $\mathbf{A}$ is obtained by multiplying by I the entry (i,j) of A.

2.8 Fundamental Theorem of Space Curves

The most important consequence of the Frenet-Serret equations is that they allow to fully characterize space curves in terms of curvature and torsion. This is known as the *Fundamental Theorem of Space Curves* which can be informally stated as:

> If we prescribe two functions $\kappa(t) > 0$ and $\tau(t)$, there exists a unit-speed curve $\boldsymbol{\gamma}(t)$ which has curvature $\kappa(t)$ and torsion $\tau(t)$. Moreover γ is the **only curve** with such curvature and torsion, up to rigid motions.

A rigid motion is a rotation about the origin, followed by a translation. Therefore the Theorem is saying that there exists a unique $\boldsymbol{\gamma}$ with curvature κ and torsion τ, up to rotations and translations.

Let us give the analytic definition of rigid motion.

Definition 2.52: Rigid motion

A *rigid motion* of $\mathbb{R}^3$ is a map $M : \mathbb{R}^3 \to \mathbb{R}^3$ of the form

$$M(\mathbf{v}) = R\mathbf{v} + \mathbf{p}, \qquad \mathbf{v} \in \mathbb{R}^3,$$

where $\mathbf{p} \in \mathbb{R}^3$ is a vector and $R \in \mathbb{R}^{3\times 3}$ is a **rotation matrix**, i.e.,

$$R \in \mathrm{SO}(3) := \{R \in \mathbb{R}^{3\times 3} \;:\; R^T R = I, \; \det(R) = 1\}.$$

In the above definition I is the identity matrix in $\mathbb{R}^3$

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

It is also useful to introduce the set of **orthogonal matrices**

$$\mathrm{O}(3) := \{A \in \mathbb{R}^{3\times 3} \;:\; A^T A = I\}$$

Notice that for $A \in \mathrm{O}(3)$ we have

$$\det(A) = \pm 1$$

Therefore rotations are orthogonal matrices with deteminant 1.

Proof. We have

$$1 = \det(I) = \det(A^T A) = \det(A)\det(A^T) = \det(A)^2$$

and therefore $\det(A) = \pm 1$.

The precise statement of the *Fundamental Theorem of Space Curves* is as follows.

Theorem 2.53: Fundamental Theorem of Space Curves

Let $\kappa, \tau \;:\; (a,b) \to \mathbb{R}$ be smooth functions, with $\kappa > 0$. Then:

1. There exists a unit-speed curve $\boldsymbol{\gamma} \;:\; (a,b) \to \mathbb{R}^3$ with curvature $\kappa(t)$ and torsion $\tau(t)$.
2. Suppose that $\tilde{\boldsymbol{\gamma}} \;:\; (a,b) \to \mathbb{R}^3$ is a unit-speed curve whose curvature $\tilde{\kappa}$ and torsion $\tilde{\tau}$ satisfy

$$\tilde{\kappa}(t) = \kappa(t)\,, \quad \tilde{\tau}(t) = \tau(t)\,, \quad \forall\, t \in (a,b)\,.$$

Then there exists a rigid motion $M : \mathbb{R}^3 \to \mathbb{R}^3$ such that

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

In other words, curvature and torsion fully characterize space curves. This result is the 3D counterpart of Theorem 2.22, which said that signed curvature characterizes 2D curves.

The proof of Theorem 2.53 is rather lengthy and technical. We delay it to the end of the chapter, see Section (Section 2.11). For now, let us show a simple application of Theorem 2.53.

Example 2.54: Application of Fundamental Theorem of Curves

In Example 2.47 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

$$\kappa = 1\,, \quad \tau = 0\,.$$

If we plot $\boldsymbol{\gamma}$, we clearly see that $\boldsymbol{\gamma}$ is just obtained by translating and rotating a unit circle, see plot in Figure 2.9. Theorem 2.53 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 2.46 we have proven that curvature and torsion are

$$\tilde{\kappa} = 1 , \quad \tilde{\tau} = 0 .$$

Therefore

$$\kappa = \tilde{\kappa} , \quad \tau = \tilde{\tau} ,$$

and by Theorem 2.53 we conclude that there exist a rigid motion $M : \mathbb{R}^3 \to \mathbb{R}^3$ such that

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

2.9 Applications of Frenet-Serret

We now state and prove two results which directly follow from the Frenet-Serret equations. They state, respectively:

1. A curve has torsion $\tau = 0$ if and only if it is contained in a plane.
2. A curve has constant curvature $\kappa > 0$ and torsion $\tau = 0$ if and only if it is part of a circle.

Before proceeding, we recall the definition plane in $\mathbb{R}^3$.

Remark 2.55: 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:

1. 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 $\mathbf{0}$.
2. 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 2.11.

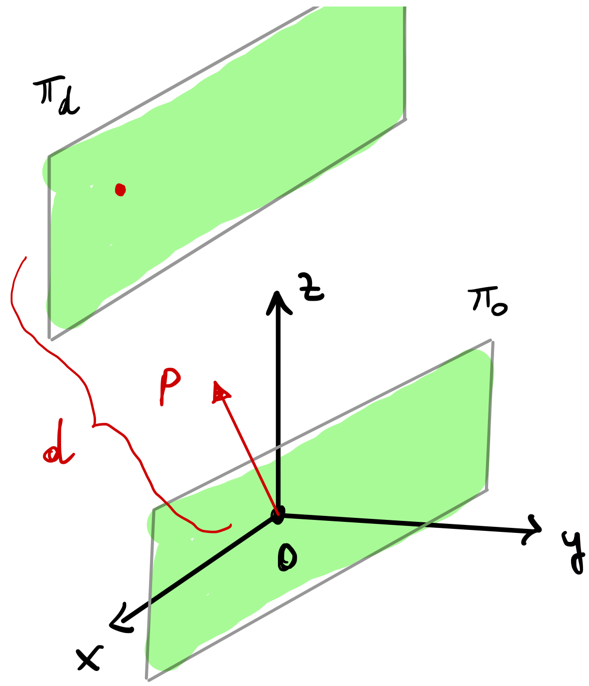


Figure 2.11: The plane $\boldsymbol{\pi}_0$ is the set of points of $\mathbb{R}^3$ orthogonal to $\mathbf{P}$. The plane $\boldsymbol{\pi}_d$ is obtained by translating $\boldsymbol{\pi}_0$ by a quantity d in direction $\mathbf{P}$.

Proposition 2.56

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
$$\tau(t) = 0\,, \quad \forall\, t \in (a, b)\,.$$
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
$$\boldsymbol{\gamma}(t) \cdot \mathbf{P} = d\,, \quad \forall t \in (a, b)\,.$$

Idea of the proof: The third Frenet-Serret equation states that

$$\dot{\mathbf{b}} = -\tau \mathbf{n}$$

Therefore $\tau = 0$ if and only if $\dot{\mathbf{b}} = 0$, which means that $\mathbf{b}$ is constant. Now, $\mathbf{b}$ is orthogonal to the other two vectors $\mathbf{t}, \mathbf{n}$ of the Frenet-Frame. Since $\mathbf{b}$ is constant, this means $\mathbf{t}, \mathbf{n}$ span a constant plane which has $\mathbf{b}$ as normal vector. This tells us $\boldsymbol{\gamma}$ is contained in the plane

$$\boldsymbol{\gamma} \cdot \mathbf{b} = d$$

for suitable $d \in \mathbb{R}$. Let's prove this!

Proof

Without loss of generality we can assume that $\boldsymbol{\gamma}$ is unit-speed.

> Proof. 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 as $\boldsymbol{\gamma}$ is contained. Moreover curvature and torsion are invariante under reparametrization, and so $\tilde{\boldsymbol{\gamma}}$ would still have non-zero curvature and identically zero torsion.

As $\boldsymbol{\gamma}$ is unit-speed, it is well defined the Frenet frame

$$\{\mathbf{t}(t), \mathbf{n}(t), \mathbf{b}(t)\}.$$

Step 1. Suppose $\tau = 0$ for all t. By the third Frenet-Serret equation

$$\dot{\mathbf{b}} = -\tau(t)\mathbf{n} = \mathbf{0},$$

so that $\mathbf{b}(t)$ is constant. As by definition

$$\mathbf{b} = \mathbf{t} \times \mathbf{n},$$

we conclude that the vectors $\mathbf{t}(t)$ 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 2.12. Indeed

$$\frac{d}{dt}(\boldsymbol{\gamma} \cdot \mathbf{b}) = \dot{\boldsymbol{\gamma}} \cdot \mathbf{b} + \ddot{\boldsymbol{\gamma}} \cdot \dot{\mathbf{b}} = 0,$$

where we used that $\dot{\mathbf{b}} = 0$ and that the Frenet frame is orthonormal, i.e.

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{b} = \mathbf{t} \cdot \mathbf{b} = 0.$$

Thus $\boldsymbol{\gamma} \cdot \mathbf{b}$ has zero derivative, meaning it is constant: there exists $d \in \mathbb{R}$ such that

$$\boldsymbol{\gamma}(t) \cdot \mathbf{b} = d, \ \ \forall\, t \in (a, b). \tag{2.29}$$

This shows that $\boldsymbol{\gamma}$ is conatained in a plane orthogonal to $\mathbf{b}$, and the first part of the proof is concluded. *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. The first equation says that $\dot{\boldsymbol{\gamma}}(t)$ is orthogonal to $\mathbf{P}$. By the first Frenet-Serret equation we have

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

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

$$\mathbf{n}(t) \cdot \mathbf{P} = 0\,, \quad \forall\, t \in (a,b)\,.$$

In conclusion, we have shown that $\mathbf{P}$ is orthogonal to both $\dot{\boldsymbol{\gamma}}(t)$ and $\mathbf{n}(t)$. Since $\mathbf{b}(t)$ is orthogonal to both $\dot{\boldsymbol{\gamma}}(t)$ 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

$$\mathbf{b}(t) = \lambda(t)\mathbf{P}\,, \quad \forall\, t \in (a,b)\,.$$

Since $\|\mathbf{b}\| = 1$ and $\mathbf{P}$ is constant, from (2.30) we conclude that $\lambda(t)$ is constant and non-zero. Thus

$$\mathbf{b}(t) = \hat{\lambda}\mathbf{P}\,, \quad \forall\, t \in (a,b)\,, \tag{2.30}$$

for some $\hat{\lambda} \neq 0$. Differentiating (2.30) we obtain

$$\dot{\mathbf{b}}(t) = 0\,, \quad \forall\, t \in (a,b)\,,$$

meaning that the binormal $\mathbf{b}$ is a constant vector. By definition of torsion

$$\tau(t) = -\dot{\mathbf{b}} \cdot \mathbf{n}(t) = 0\,, \quad \forall\, t \in (a,b)\,,$$

as required.

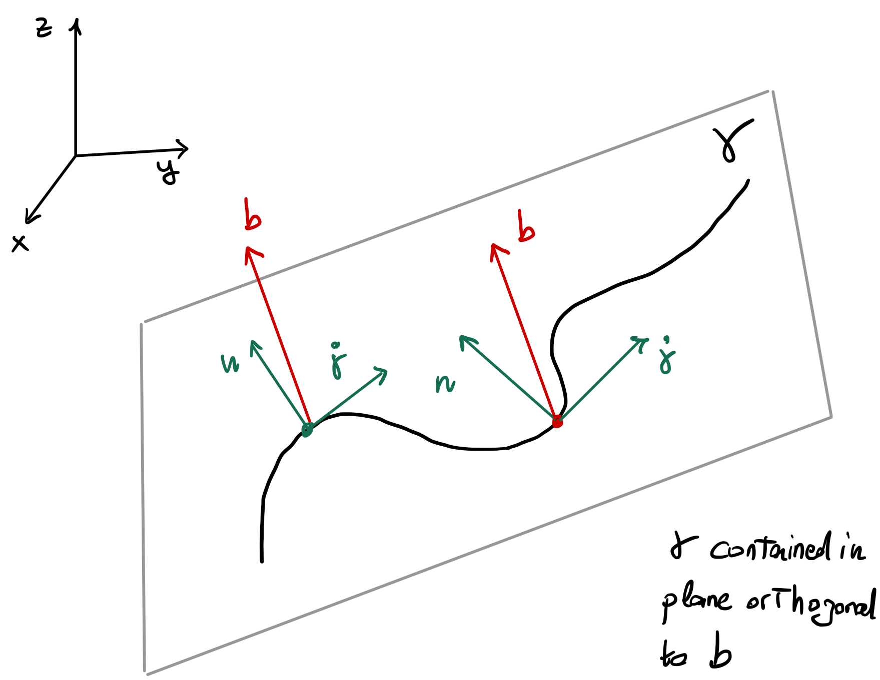


Figure 2.12: If $\mathbf{b}$ is constant, then $\boldsymbol{\gamma}$ lies in the plane spanned by $\dot{\boldsymbol{\gamma}}$ and $\mathbf{n}$. Note that $\mathbf{b}$ is the unit normal to such plane.

As a corollary of Step 1 in the proof of Proposition 2.56 we obtain the following statement.

Corollary 2.57

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be regular and such that $\kappa \neq 0$. Suppose that the torsion

$$\tau(t) = 0\,, \quad \forall\, t \in (a,b)\,.$$

Then the binormal $\mathbf{b}$ is a constant vector, and the image of $\boldsymbol{\gamma}$ is contained in the plane of equation

$$(\mathbf{x} - \boldsymbol{\gamma}(t_0)) \cdot \mathbf{b} = 0\,,$$

where $t_0 \in (a,b)$ is arbitrarily chosen. Note that such plane is orthogonal to the binormal $\mathbf{b}$.

Proof

Following the proof of Step 1 of Proposition 2.56, we get to the conclusion (2.29) that $\boldsymbol{\gamma}$ satisfies

$$\boldsymbol{\gamma}(t) \cdot \mathbf{b} = d\,, \;\; \forall\, t \in (a,b)\,.$$

As the above equation holds for each t, we can fix an arbitrary $t_0 \in (a,b)$ and find that the constant d is

$$d = \boldsymbol{\gamma}(t_0) \cdot \mathbf{b}$$

Hence we obtain

$$(\boldsymbol{\gamma}(t) - \boldsymbol{\gamma}(t_0)) \cdot \mathbf{b} = 0\,, \;\; \forall\, t \in (a,b)\,.$$

The above is says that $\boldsymbol{\gamma}$ is contained in the plane (orthogonal to $\mathbf{b}$) with equation

$$(\mathbf{x} - \boldsymbol{\gamma}(t_0)) \cdot \mathbf{b} = 0$$

Example 2.58

Consider again the curve of Example 2.47

$$\boldsymbol{\gamma}(t) \;:= \left(\frac{4}{5}\cos(t), 1 - \sin(t), -\frac{3}{5}\cos(t) \right)\,,$$

for $t \in \mathbb{R}$. Recall that $\boldsymbol{\gamma}$ is a Circle which has been rotated an translated, as shown in Figure 2.9. We have already proven in Example 2.47 that the curvature and torsion satisfy

$$\kappa = 1\,, \qquad \tau = 0\,.$$

Since $\kappa \neq 0$ and $\tau = 0$, we can apply Corollary 2.57 and infer that $\boldsymbol{\gamma}$ is contained in the plane of equation

$$(\mathbf{x} - \boldsymbol{\gamma}(t_0)) \cdot \mathbf{b} = 0\,, \quad \mathbf{x} \in \mathbb{R}^3\,,$$

where $\mathbf{b}$ is the binormal vector to $\boldsymbol{\gamma}$. As computed in Example 2.47 we have

$$\mathbf{b} = \left(-\frac{3}{5}, 0, -\frac{4}{5} \right)\,.$$

Next, we take a point on the curve. At $t = \pi/2$ we have

$$\boldsymbol{\gamma}(0) = \mathbf{0}$$

The equation of the plane is then

$$-\frac{3}{5}x - \frac{4}{5}z = 0$$

Simplifying, we obtain that $\boldsymbol{\gamma}$ is contained in the plane of equation

$$3x + 4z = 0\,.$$

We now state and prove the second result anticipated at the beginning of this section.

Proposition 2.59

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ 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 $c > 0$.

Proposition 2.59 is actually a consequence of the Fundamental Theorem of Space Curves Theorem 2.53, 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 2.53, 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 2.59, to show yet another explicit application of the Frenet-Serret equations.

Proof

Step 1. Suppose the image of $\boldsymbol{\gamma}$ is contained in a circle of radius $1/c$. Then, up to a rotation and translation, $\boldsymbol{\gamma}$ is parametrized by

$$\boldsymbol{\gamma}(t) = \left(\frac{1}{c}\cos(t), \frac{1}{c}\sin(t), 0 \right)\,.$$

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, it is well defined the Frenet frame

$$\{\mathbf{t}, \mathbf{n}, \mathbf{b}\} \,.$$

Due to the assumptions on κ and τ the Frenet-Serret equations read

$$\begin{aligned} \dot{\mathbf{t}} &= \kappa \mathbf{n} = c\mathbf{n} \\ \dot{\mathbf{n}} &= -\kappa \mathbf{t} + \tau \mathbf{b} = -c\mathbf{t} \\ \dot{\mathbf{b}} &= -\tau \mathbf{n} = \mathbf{0} \end{aligned}$$

In particular $\dot{\mathbf{b}} = \mathbf{0}$ and so $\mathbf{b}$ is a constant vector. As seen in the proof Proposition 2.56, this implies that $\boldsymbol{\gamma}$ is contained in a plane $\boldsymbol{\pi}$ orthogonal to $\mathbf{b}$, see Figure 2.12. As c is constant we get

$$\frac{d}{dt}\left(\boldsymbol{\gamma} + \frac{1}{c}\mathbf{n}\right) = \dot{\boldsymbol{\gamma}} + \frac{1}{c}\dot{\mathbf{n}} = \mathbf{t} - \frac{1}{c}\, c\mathbf{t} = 0 \,,$$

where we used that $\dot{\boldsymbol{\gamma}} = \mathbf{t}$ and 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

$$\|\boldsymbol{\gamma}(t) - \mathbf{p}\| = \left\| -\frac{1}{c}\mathbf{n}(t) \right\| = \frac{1}{|c|} = \frac{1}{c} \,,$$

since $\mathbf{n}$ is a unit vector and $c > 0$. The above shows that $\boldsymbol{\gamma}$ is contained in a sphere of radius $1/c$ and center $\mathbf{p}$. In formulas:

$$\boldsymbol{\gamma}(t) \in \mathcal{S} := \{\mathbf{x} \in \mathbb{R}^3 \,:\, \|\mathbf{x} - \mathbf{p}\| = 1/c\} \,.$$

The intersection of $\mathcal{S}$ with the plane $\boldsymbol{\pi}$ is a circle $\mathscr{C}$ with some radius $R \geq 0$ (note that $\mathscr{C}$ might be a single point, in which case $R = 0$). As we have shown

$$\boldsymbol{\gamma} \in \boldsymbol{\pi} \,, \quad \boldsymbol{\gamma} \in \mathcal{S} \,,$$

we conclude that

$$\boldsymbol{\gamma} \in \boldsymbol{\pi} \cap \mathcal{S} = \mathscr{C} \,. \tag{2.31}$$

Therefore $\boldsymbol{\gamma}$ parametrizes part of $\mathscr{C}$. This immediately implies that $R > 0$.

> Proof. If $R = 0$ then $\mathscr{C}$ is a single point, meaning that $\boldsymbol{\gamma}$ is constant. But then $\dot{\boldsymbol{\gamma}} = 0$ which contradicts the assumption that $\boldsymbol{\gamma}$ is unit speed.

Since $\mathscr{C}$ is a circle of radius $R > 0$, Step 1 of the proof implies that the curvature and torsion of $\boldsymbol{\gamma}$ satisfy

$$\kappa = \frac{1}{R} \,, \quad \tau = 0 \,.$$

As by assumption $\kappa = c$, we conclude that $R = 1/c$. Therefore the circle $\mathscr{C}$ has radius $1/c$ and the thesis follows by (2.31).

2.10 Proof: Curvature and torsion formulas

Another consequence of the Frenet-Serret equations is that they allow us to finally prove the curvature and torsion formulas given in Proposition 2.14 and Proposition 2.44. For reader's convenience we recall these two results.

Proposition 2.60: 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

$$\kappa(t) = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} . \tag{2.32}$$

If $\kappa > 0$ the torsion $\tau(t)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is given by

$$\tau(t) = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} . \tag{2.33}$$

Proof

By assumption $\boldsymbol{\gamma}$ is regular. Denote by $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} \circ s^{-1}$ the arc-length reparametrization. As $\tilde{\boldsymbol{\gamma}}$ is unit-speed, it is well defined its Frenet frame

$$\left\{ \tilde{\mathbf{t}}(s), \tilde{\mathbf{n}}(s), \tilde{\mathbf{b}}(s) \right\} , \quad \tilde{\mathbf{t}} = \dot{\tilde{\boldsymbol{\gamma}}}$$

The Frenet-Serret equations are

$$\begin{aligned} \dot{\tilde{\mathbf{t}}} &= \tilde{\kappa}\, \tilde{\mathbf{n}} \\ \dot{\tilde{\mathbf{n}}} &= -\tilde{\kappa}\, \tilde{\mathbf{t}} + \tilde{\tau}\, \tilde{\mathbf{b}} \\ \dot{\tilde{\mathbf{b}}} &= -\tilde{\tau}\, \tilde{\mathbf{n}} \end{aligned}$$

where $\tilde{\kappa}$ and $\tilde{\tau}$ are the curvature and torsion of $\tilde{\boldsymbol{\gamma}}$.
Part 1. Differentiating $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ s$ we get

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

Differentiating once more

$$\begin{aligned} \ddot{\boldsymbol{\gamma}}(t) &= \frac{d}{dt} \left[\tilde{\mathbf{t}}(s(t))\, \dot{s}(t) \right] \\ &= \dot{\tilde{\mathbf{t}}} \dot{s}^2 + \tilde{\mathbf{t}}\, \ddot{s} \\ &= \tilde{\kappa}\, \tilde{\mathbf{n}}\, \dot{s}^2 + \tilde{\mathbf{t}}\, \ddot{s} \end{aligned} \tag{2.34}$$

where in the last line we used the first Frenet-Serret equation. We are also omitting the dependence on the point for brevity. We compute

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t) &= \tilde{\mathbf{t}}\, \dot{s} \times \left[\tilde{\kappa}\, \tilde{\mathbf{n}}\, \dot{s}^2 + \tilde{\mathbf{t}}\, \ddot{s} \right] \\ &= \dot{s}^3\, \tilde{\kappa}\, \tilde{\mathbf{t}} \times \tilde{\mathbf{n}} + \dot{s}\, \ddot{s}\, \tilde{\mathbf{t}} \times \tilde{\mathbf{t}} \\ &= \dot{s}^3\, \tilde{\kappa}\, \tilde{\mathbf{b}} \end{aligned} \tag{2.35}$$

where we used that $\tilde{\mathbf{b}} = \tilde{\mathbf{t}} \times \tilde{\mathbf{n}}$ by definition, and that $\tilde{\mathbf{t}} \times \tilde{\mathbf{t}} = 0$ by the properties of the cross product. Taking the norms in (2.35) gives

$$\|\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t)\| = \|\boldsymbol{\gamma}(t)\|^3 \ \tilde{\kappa}(s(t))$$

where we used that $\left\|\tilde{\mathbf{b}}\right\| = 1$ and $\dot{s}(t) = \|\boldsymbol{\gamma}(t)\|$. Rearranging we get

$$\tilde{\kappa}(s(t)) = \frac{\|\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t)\|}{\|\boldsymbol{\gamma}(t)\|^3}$$

Recalling that the curvature κ of $\boldsymbol{\gamma}$ at t is defined as

$$\kappa(t) = \tilde{\kappa}(s(t))$$

we conclude (2.32).

Part 2. Differentiating the second line in (2.34) we get

$$\begin{aligned}
\dddot{\boldsymbol{\gamma}}(t) &= \frac{d}{dt}\left[\dot{\tilde{\mathbf{t}}}(s(t))\dot{s}^2(t) + \tilde{\mathbf{t}}(s(t))\,\ddot{s}(t)\right] \\
&= \ddot{\tilde{\mathbf{t}}}\dot{s}^3 + 2\dot{\tilde{\mathbf{t}}}\dot{s}\ddot{s} + \dot{\tilde{\mathbf{t}}}\dot{s}\ddot{s} + \tilde{\mathbf{t}}\,\dddot{s} \\
&= \ddot{\tilde{\mathbf{t}}}\dot{s}^3 + 3\dot{\tilde{\mathbf{t}}}\dot{s}\ddot{s} + \tilde{\mathbf{t}}\,\dddot{s}
\end{aligned}$$

Therefore, using (2.35), we get

$$\begin{aligned}
(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} &= \left[\dot{s}^3\tilde{\kappa}\,\tilde{\mathbf{b}}\right] \cdot \left[\ddot{\tilde{\mathbf{t}}}\dot{s}^3 + 3\dot{\tilde{\mathbf{t}}}\dot{s}\ddot{s} + \tilde{\mathbf{t}}\,\dddot{s}\right] \\
&= \dot{s}^6\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \ddot{\tilde{\mathbf{t}}} + 3\dot{s}^4\,\ddot{s}\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \dot{\tilde{\mathbf{t}}} + \dot{s}^3\,\dddot{s}\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \tilde{\mathbf{t}} \\
&= \dot{s}^6\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \ddot{\tilde{\mathbf{t}}}
\end{aligned} \tag{2.36}$$

where the second term is zero by the first Frenet-Serret equation

$$\tilde{\mathbf{b}} \cdot \dot{\tilde{\mathbf{t}}} = \tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \tilde{\mathbf{n}} = 0\,,$$

as $\tilde{\mathbf{b}} \cdot \tilde{\mathbf{n}} = 0$, and the thrid term is zero because $\tilde{\mathbf{b}} \cdot \tilde{\mathbf{t}} = 0$. To compute $\ddot{\tilde{\mathbf{t}}}$ we differentiate the first Frenet-Serret formula

$$\begin{aligned}
\ddot{\tilde{\mathbf{t}}} &= \frac{d}{ds}\left[\dot{\tilde{\mathbf{t}}}\right] \\
&= \frac{d}{ds}\left[\tilde{\kappa}\,\tilde{\mathbf{n}}\right] \\
&= \dot{\tilde{\kappa}}\,\tilde{\mathbf{n}} + \tilde{\kappa}\,\dot{\tilde{\mathbf{n}}}
\end{aligned}$$

Substituting in (2.36) we get

$$\begin{aligned}
(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} &= \dot{s}^6\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \ddot{\tilde{\mathbf{t}}} \\
&= \dot{s}^6\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \left[\dot{\tilde{\kappa}}\,\tilde{\mathbf{n}} + \tilde{\kappa}\,\dot{\tilde{\mathbf{n}}}\right] \\
&= \dot{s}^6\,\tilde{\kappa}\,\dot{\tilde{\kappa}}\,\tilde{\mathbf{b}} \cdot \tilde{\mathbf{n}} + \dot{s}^6\,\tilde{\kappa}^2\,\tilde{\mathbf{b}} \cdot \dot{\tilde{\mathbf{n}}} \\
&= \dot{s}^6\,\tilde{\kappa}^2\,\tilde{\mathbf{b}} \cdot \dot{\tilde{\mathbf{n}}}
\end{aligned}$$

where we used that $\tilde{\mathbf{b}} \cdot \tilde{\mathbf{n}} = 0$. Using the second Frenet-Serret equation we get

$$
\begin{aligned}
(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} & = \dot{s}^6 \, \tilde{\kappa}^2 \, \tilde{\mathbf{b}} \cdot \dot{\tilde{\mathbf{n}}} \\
& = \dot{s}^6 \, \tilde{\kappa}^2 \, \tilde{\mathbf{b}} \cdot \left[-\tilde{\kappa} \, \tilde{\mathbf{t}} + \tilde{\tau} \, \tilde{\mathbf{b}} \right] \\
& = -\dot{s}^6 \, \tilde{\kappa}^3 \, \tilde{\mathbf{b}} \cdot \tilde{\mathbf{t}} + \dot{s}^6 \, \tilde{\kappa}^2 \, \tilde{\tau} \, \tilde{\mathbf{b}} \cdot \tilde{\mathbf{b}} \\
& = \dot{s}^6 \, \tilde{\kappa}^2 \, \tilde{\tau}
\end{aligned}
$$

where we used that $\tilde{\mathbf{b}} \cdot \tilde{\mathbf{t}} = 0$ and $\tilde{\mathbf{b}} \cdot \tilde{\mathbf{b}} = 1$. Recalling that $\dot{s} = \|\dot{\boldsymbol{\gamma}}\|$, and using the already proven formula (2.32), we get

$$
\begin{aligned}
(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} & = \dot{s}^6 \, \tilde{\kappa}^2 \, \tilde{\tau} \\
& = \|\dot{\boldsymbol{\gamma}}\|^6 \, \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2}{\|\dot{\boldsymbol{\gamma}}\|^6} \, \tilde{\tau} \\
& = \|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2 \, \tilde{\tau}
\end{aligned}
$$

Rerranging we get

$$
\tilde{\tau}(s(t)) = \frac{(\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t)) \cdot \dddot{\boldsymbol{\gamma}}(t)}{\|\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t)\|^2}
$$

Recalling that the torsion τ of $\boldsymbol{\gamma}$ at t is defined as

$$
\tau(t) = \tilde{\tau}(s(t))
$$

we conclude (2.33).

2.11 Proof: Fundamental Theorem of Space Curves

In this section we prove the *Fundamental Theorem of Space Curves* Theorem 2.53. For reader's convenience we recall the statement.

Theorem 2.61: Fundamental Theorem of Space Curves

Let $\kappa, \tau : (a,b) \to \mathbb{R}$ be smooth functions, with $\kappa > 0$. Then:

1. There exists a unit-speed curve $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ with curvature $\kappa(t)$ and torsion $\tau(t)$.
2. Suppose that $\tilde{\boldsymbol{\gamma}} : (a,b) \to \mathbb{R}^3$ is a unit-speed curve whose curvature $\tilde{\kappa}$ and torsion $\tilde{\tau}$ satisfy

$$
\tilde{\kappa}(t) = \kappa(t) \,, \quad \tilde{\tau}(t) = \tau(t) \,, \quad \forall \, t \in (a,b) \,.
$$

Then there exists a rigid motion $M : \mathbb{R}^3 \to \mathbb{R}^3$ such that

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

The proof relies on the following classical existence Theorem for linear systems of ODEs. For a proof see Page 162 in [7].

Theorem 2.62: Existence and uniqueness for linear ODE systems

Assume given a point $t_0 \in (a,b)$, a vector $\mathbf{u}_0 \in \mathbb{R}^n$ and two functions

$$A: (a,b) \to \mathbb{R}^{n \times n}, \qquad f: (a,b) \to \mathbb{R}^n$$

of class C^k. There exists a unique function

$$\mathbf{u}: (a,b) \to \mathbb{R}^n$$

of class C^{k+1} which solves the Cauchy problem

$$\begin{cases} \dot{\mathbf{u}} = A\mathbf{u} + f \\ \mathbf{u}(t_0) = \mathbf{u}_0 \end{cases}$$

We will also need the following 4 Propositions. The first Proposition states that orthogonal matrices preserve scalar product and length.

Proposition 2.63: Orthogonal matrices preserve scalar product and length

Let $A \in \mathrm{O}(3)$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. Then

$$A\mathbf{v} \cdot A\mathbf{w} = \mathbf{v} \cdot \mathbf{w} \tag{2.37}$$

and also

$$\|A\mathbf{v}\| = \|\mathbf{v}\| \tag{2.38}$$

Proof

For $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$

$$A\mathbf{v} \cdot A\mathbf{w} = A^T A\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w}$$

where we used the properties of scalar product and that $A^T A = I$. In particular the above implies

$$\|A\mathbf{v}\| = \sqrt{A\mathbf{v} \cdot A\mathbf{v}} = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \|\mathbf{v}\|$$

concluding the proof.

We now inverstigate how the cross product behaves under linear transformations, i.e., how the cross product of two vector changes under matrix multiplication.

Proposition 2.64: Linear transformations of cross product

Let $A \in \mathbb{R}^{3\times 3}$ be invertible and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. Then

$$A\mathbf{v} \times A\mathbf{w} = \det(A) \left(A^{-1}\right)^T (\mathbf{v} \times \mathbf{w}) \tag{2.39}$$

In particular for $R \in \mathrm{SO}(3)$ we have

$$R\mathbf{v} \times R\mathbf{w} = R(\mathbf{v} \times \mathbf{w}) \tag{2.40}$$

Proof

Part 1. Recall that the inverse of a matrix A is computed by

$$A^{-1} = \frac{1}{\det A} (\operatorname{cof} A)^T$$

where $\operatorname{cof} M$ is the matrix of cofactors. By linearity of the cross product we only need to verify (2.39) on the vectors of the orthonormal basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Let us verify (2.39) for $\mathbf{v} = \mathbf{e}_1$ and $\mathbf{w} = \mathbf{e}_2$. Writing $A\mathbf{e}_1$ and $A\mathbf{e}_2$ in coordinates

$$A\mathbf{e}_1 = \sum_{i=1}^3 m_{i1}\mathbf{e}_i \,, \quad A\mathbf{e}_2 = \sum_{i=1}^3 m_{i2}\mathbf{e}_i$$

for some coefficients $m_{i1}, m_{i2} \in \mathbb{R}$. By the formula for computing the vector product (2.5) and definition of cofactor matrix

$$\begin{aligned}
A\mathbf{e}_1 \times A\mathbf{e}_2 &= \sum_{i=1}^3 m_{i1}\mathbf{e}_i \times \sum_{j=1}^3 m_{j2}\mathbf{e}_j \\
&= \sum_{i<j} \begin{vmatrix} m_{i1} & m_{i2} \\ m_{j1} & m_{j2} \end{vmatrix} \mathbf{e}_i \times \mathbf{e}_j \\
&= \sum_{i=1}^3 (\operatorname{cof} \boldsymbol{A})_{i3}\mathbf{e}_i \\
&= (\operatorname{cof} \boldsymbol{A})\mathbf{e}_3 \\
&= (\det A) \left(A^{-1}\right)^T \mathbf{e}_3 \\
&= (\det A) \left(A^{-1}\right)^T (\mathbf{e}_1 \times \mathbf{e}_2)
\end{aligned}$$

Calculations for the other cases are similar.
Part 2. For $R \in \mathrm{SO}(3)$ it holds $\det(R) = 1$. Moreover $R^T R = I$, so that

$$R^{-1} = R^T \quad \Longrightarrow \quad \left(R^{-1}\right)^T = R$$

Therefore (2.40) follows from (2.39).

We need to clarify how unit-speed curves and Frenet frame behave under rigid motions.

Proposition 2.65: Frenet Frame under rigid motions

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be unit-speed and $M \colon \mathbb{R}^3 \to \mathbb{R}^3$ be a rigid motion, i.e. such that

$$M(\mathbf{v}) = R\,\mathbf{v} + \mathbf{p}$$

for some $R \in \mathrm{SO}(3)$ and $\mathbf{p} \in \mathbb{R}^3$. Define the curve

$$\tilde{\boldsymbol{\gamma}} \colon (a,b) \to \mathbb{R}^3 \,, \qquad \tilde{\boldsymbol{\gamma}}(t) = M(\boldsymbol{\gamma}(t)) \,.$$

Then $\tilde{\boldsymbol{\gamma}}$ is unit-speed. Moreover the Frenet frame of $\tilde{\boldsymbol{\gamma}}$ is obtained by rotating the Frenet frame of $\boldsymbol{\gamma}$ by R

$$\tilde{\mathbf{t}} = R\,\mathbf{t} \,, \qquad \tilde{\mathbf{n}} = R\,\mathbf{n} \,, \qquad \tilde{\mathbf{b}} = R\,\mathbf{b} \,, \tag{2.41}$$

Proof

Differentiating $\tilde{\boldsymbol{\gamma}} = M(\boldsymbol{\gamma}) = R\,\boldsymbol{\gamma} + \mathbf{p}$ gives

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) = R\,\dot{\boldsymbol{\gamma}}(t) \,, \qquad \ddot{\tilde{\boldsymbol{\gamma}}}(t) = R\,\ddot{\boldsymbol{\gamma}}(t)$$

Taking the norms in $\dot{\tilde{\boldsymbol{\gamma}}} = R\dot{\boldsymbol{\gamma}}$ gives

$$\|\dot{\tilde{\boldsymbol{\gamma}}}\| = \|R\dot{\boldsymbol{\gamma}}\| = \|\dot{\boldsymbol{\gamma}}\| = 1$$

where we used that rotations preseve norms, see (2.38), and the assumption of $\boldsymbol{\gamma}$ unit-speed. This concludes the proof that $\tilde{\boldsymbol{\gamma}}$ is unit-speed.
Let us now prove (2.41). The relation $\dot{\tilde{\boldsymbol{\gamma}}} = R\,\dot{\boldsymbol{\gamma}}$ reads

$$\tilde{\mathbf{t}} = R\,\mathbf{t} \,,$$

which gives the first equation in (2.41). Since $\ddot{\tilde{\boldsymbol{\gamma}}} = R\,\ddot{\boldsymbol{\gamma}}$, by (2.38) we deduce

$$\|\ddot{\tilde{\boldsymbol{\gamma}}}\| = \|R\,\ddot{\boldsymbol{\gamma}}\| = \|\ddot{\boldsymbol{\gamma}}\|$$

Therefore, by definition of principal normal,

$$\tilde{\mathbf{n}} = \frac{\ddot{\tilde{\boldsymbol{\gamma}}}}{\|\ddot{\tilde{\boldsymbol{\gamma}}}\|} = \frac{R\,\ddot{\boldsymbol{\gamma}}}{\|\ddot{\boldsymbol{\gamma}}\|} = R\,\mathbf{n} \,,$$

obtaining the second equation in (2.41). Finally, by definition of binormal and the first two equations in (2.41),

$$\tilde{\mathbf{b}} = \tilde{\mathbf{t}} \times \tilde{\mathbf{n}} = R\,\mathbf{t} \times R\,\mathbf{n} = R\,(\mathbf{t} \times \mathbf{n}) = R\,\mathbf{b}$$

where in the third equality we used (2.40). The proof is concluded.

The last Proposition concerns the evolution of orthonormal systems of vectors.

Proposition 2.66: Evolution of orthonormal systems of vectors

Let $A \colon (a,b) \to \mathbb{R}^{3\times 3}$ be smooth and anti-symmetric, that is,

$$A^T(t) = -A(t)\,, \quad \forall\, t \in (a,b)\,.$$

Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \colon (a,b) \to \mathbb{R}^3$ be smooth functions satisfying the following ODE

$$\begin{pmatrix} \dot{\mathbf{u}}_1 \\ \dot{\mathbf{u}}_2 \\ \dot{\mathbf{u}}_3 \end{pmatrix} = A \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix} \tag{2.42}$$

Suppose that for some $t_0 \in (a,b)$ the vectors

$$\mathbf{u}_1(t_0)\,, \quad \mathbf{u}_2(t_0)\,, \quad \mathbf{u}_3(t_0)\,,$$

are orthonormal. Then the vectors

$$\mathbf{u}_1(t)\,, \quad \mathbf{u}_2(t)\,, \quad \mathbf{u}_3(t)\,,$$

are orthonormal for all values of $t \in (a,b)$.

Proof

For each pair i,j define

$$\lambda_{ij} := \mathbf{u}_i \cdot \mathbf{u}_j$$

Further, define

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Note that $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are orthonormal for all t if and only if

$$\lambda_{ij}(t) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.43}$$

for all t. Hence the proof is concluded if we show (2.43).
In order to do so, first note that the ODE in (2.42) reads

$$\dot{\mathbf{u}}_i = \sum_{k=1}^{3} a_{ik} \mathbf{u}_k\,, \qquad i = 1,2,3 \tag{2.44}$$

Differentiating λ_{ij} yields

$$\begin{aligned}\dot\lambda_{ij} &= \frac{d}{dt}\mathbf{u}_i \cdot \mathbf{u}_j \\ &= \dot{\mathbf{u}}_i \cdot \mathbf{u}_j + \mathbf{u}_i \cdot \dot{\mathbf{u}}_j \\ &= \sum_{k=1}^3 a_{ik}\mathbf{u}_k \cdot \mathbf{u}_j + \sum_{k=1}^3 a_{jk}\mathbf{u}_i \cdot \mathbf{u}_k \\ &= \sum_{k=1}^3 a_{ik}\lambda_{kj} + \sum_{k=1}^3 a_{jk}\lambda_{ik}\end{aligned}$$

where in the last two equalities we used (2.44) and the definition of λ_{ij}. The above calculation shows that λ_{ij} solves the ODE

$$\dot\lambda_{ij} = \sum_{k=1}^3 a_{ik}\lambda_{kj} + \sum_{k=1}^3 a_{jk}\lambda_{ik} \tag{2.45}$$

We claim that δ_{ij} solves (2.45). Indeed the LHS is $\dot\delta_{ij} = 0$, while the RHS is

$$\sum_{k=1}^3 a_{ik}\delta_{kj} + \sum_{k=1}^3 a_{jk}\delta_{ik} = a_{ij} + a_{ji} = 0$$

where we used that $a_{ij} = -a_{ij}$ because $A^T = -A$ by assumption. Thus δ_{ij} solves (2.45). Moreover we notice that at $t = t_0$

$$\lambda_{ij}(t_0) = \mathbf{u}_i(t_0) \cdot \mathbf{u}_j(t_0) = \delta_{ij}$$

because the vectors $\mathbf{u}_1(t_0), \mathbf{u}_2(t_0), \mathbf{u}_3(t_0)$ are orthonormal by assumption. Since λ_{ij} is also a solution to (2.45), by the uniqueness of solutions to ODE systems in Theorem 2.62, we conclude that (2.43) holds for all t. The proof is concluded.

We are finally ready to prove the Fundamental Theorem of Space Curves.

Proof: Proof of Theorem 2.61

Suppose given two smooth functions $\kappa, \tau \colon (a,b) \to \mathbb{R}$ with $\kappa > 0$.

Part 1. Existence of $\boldsymbol{\gamma}$.

Consider the Frenet-Serret system of ODEs

$$\begin{pmatrix} \dot{\mathbf{t}} \\ \dot{\mathbf{n}} \\ \dot{\mathbf{b}} \end{pmatrix} = A \begin{pmatrix} \mathbf{t} \\ \mathbf{n} \\ \mathbf{b} \end{pmatrix} \tag{2.46}$$

where the matrix A is

$$A = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} .$$

This is a linear system of 9 equations in 9 unknowns (the coordinates of $\mathbf{t}, \mathbf{n}, \mathbf{b}$). Therefore Theorem 2.62 guarantess the existence of a smooth solution $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ such that

$$\mathbf{t}(t_0) = \mathbf{e}_1 \,, \quad \mathbf{n}(t_0) = \mathbf{e}_2 \,, \quad \mathbf{b}(t_0) = \mathbf{e}_3 \tag{2.47}$$

for some fixed $t_0 \in (a, b)$, where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are the vectors of the standard basis of $\mathbb{R}^3$. Since the matrix A is anti-symmetric, and the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is orthonormal, by Proposition 2.66 we deduce that

$$(\mathbf{t}, \mathbf{n}, \mathbf{b}) \;\text{ are orthonormal for all }\; t \in (a, b)$$

As both $\mathbf{b}$ and $\mathbf{t} \times \mathbf{n}$ are unit vectors orthogonal to $\mathbf{t}$ and $\mathbf{n}$, we conclude that

$$\mathbf{b} = \lambda \, \mathbf{t} \times \mathbf{n}$$

for some continuous function λ such that $\lambda(t) = \pm 1$. Substituting $t = t_0$ and recalling (2.47) gives

$$\mathbf{b}(t_0) = \lambda(t_0) \, \mathbf{t}(t_0) \times \mathbf{n}(t_0) \implies \mathbf{e}_3 = \lambda(t_0) \, \mathbf{e}_1 \times \mathbf{e}_2$$

As $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$, we conclude that $\lambda(t_0) = 1$. By continuity of λ we then have $\lambda \equiv 1$, so that

$$\mathbf{b} = \mathbf{t} \times \mathbf{n} \tag{2.48}$$

Define the curve $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ by

$$\boldsymbol{\gamma}(t) := \int_{t_0}^{t} \mathbf{t}(u) du$$

We now make a few observations:

1. By the Fundamental Theorem of Calculus we have

   $$\dot{\boldsymbol{\gamma}} = \mathbf{t}$$

   showing that $\mathbf{t}$ is the tangent vector to $\boldsymbol{\gamma}$.

2. In particular $\dot{\boldsymbol{\gamma}}$ is unit-speed, since $\|\mathbf{t}\| = 1$.

3. Using the first equation in (2.46) gives

   $$\ddot{\boldsymbol{\gamma}} = \dot{\mathbf{t}} = \kappa \, \mathbf{n}$$

   which shows that κ is the curvature of $\boldsymbol{\gamma}$, and $\mathbf{n}$ the principal normal.

4. By (2.48) we deduce that $\mathbf{b}$ is the binormal to $\boldsymbol{\gamma}$

5. Using the third equation in (2.46) gives
$$\dot{\mathbf{b}} = -\tau\, \mathbf{n}$$
showing that τ is the torsion of $\boldsymbol{\gamma}$

We have therefore constructed a unit-speed curve $\boldsymbol{\gamma}$ with curvature κ and torsion τ.
Part 2. Uniqueness up to rigid motions.
Suppose that $\tilde{\boldsymbol{\gamma}} : (a,b) \to \mathbb{R}^3$ is a unit-speed curve with curvature $\tilde{\kappa} = \kappa$ and torsion $\tilde{\tau} = \tau$. Denote by

$$(\mathbf{t}, \mathbf{n}, \mathbf{b}) , \qquad (\tilde{\mathbf{t}}, \tilde{\mathbf{n}}, \tilde{\mathbf{b}})$$

the Frenet frames of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ respectively. Since the above vectors are orthonormal for each t, there exists a rotation $R \in \mathrm{SO}(3)$ such that

$$\tilde{\mathbf{t}}(t_0) = R\, \mathbf{t}(t_0) , \quad \tilde{\mathbf{n}}(t_0) = R\, \mathbf{n}(t_0) , \quad \tilde{\mathbf{b}}(t_0) = R\, \mathbf{b}(t_0) \tag{2.49}$$

Notice that R can always be found (for fixed time!) since the vectors are orthonormal. Define

$$\mathbf{p} := \tilde{\boldsymbol{\gamma}}(t_0) - R\, \boldsymbol{\gamma}(t_0)$$

We now define the rigid motion
$$M(\mathbf{v}) := R\, \mathbf{v} + \mathbf{p}$$
By construction
$$M(\boldsymbol{\gamma}(t_0)) = \tilde{\boldsymbol{\gamma}}(t_0) \tag{2.50}$$
Define the new curve
$$\hat{\boldsymbol{\gamma}}(t) := M(\boldsymbol{\gamma}(t))$$
By Proposition 2.65 we know that $\hat{\boldsymbol{\gamma}}$ is unit-speed, given that $\boldsymbol{\gamma}$ is unit-speed. Moreover the Frenet frame of $\hat{\boldsymbol{\gamma}}$ satisfies
$$\hat{\mathbf{t}} = R\, \mathbf{t} , \quad \hat{\mathbf{n}} = R\, \mathbf{n} , \quad \hat{\mathbf{b}} = R\, \mathbf{b} \tag{2.51}$$
From the above relations we deduce that $(\hat{\mathbf{t}}, \hat{\mathbf{n}}, \hat{\mathbf{b}})$ solves the Frenet-Serret equations at (2.46). Since $\tilde{\boldsymbol{\gamma}}$ has curvature κ and torsion τ, we also know that $(\tilde{\mathbf{t}}, \tilde{\mathbf{n}}, \tilde{\mathbf{b}})$ solves the Frenet-Serret equations at (2.46). By evaluating (2.51) at t_0, and comparing with (2.49), we also have that

$$\hat{\mathbf{t}}(t_0) = \tilde{\mathbf{t}}(t_0) , \quad \hat{\mathbf{n}}(t_0) = \tilde{\mathbf{n}}(t_0) , \quad \hat{\mathbf{b}}(t_0) = \tilde{\mathbf{b}}(t_0)$$

By applying the uniqueness in Theorem 2.62 to the ODE system (2.46), we conclude that

$$\hat{\mathbf{t}} = \tilde{\mathbf{t}} , \quad \hat{\mathbf{n}} = \tilde{\mathbf{n}} , \quad \hat{\mathbf{b}} = \tilde{\mathbf{b}}$$

for all times t. In particular, since $\tilde{\mathbf{t}} = \dot{\tilde{\boldsymbol{\gamma}}}$ and $\hat{\mathbf{t}} = \dot{\hat{\boldsymbol{\gamma}}}$, we infer

$$\hat{\mathbf{t}} = \tilde{\mathbf{t}} \quad \Longrightarrow \quad \dot{\hat{\boldsymbol{\gamma}}} = \dot{\tilde{\boldsymbol{\gamma}}}$$

and therefore there exists a constant $c \in \mathbb{R}^3$ such that

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

By (2.50) and definition of $\hat{\boldsymbol{\gamma}}$, it holds

$$\hat{\boldsymbol{\gamma}}(t_0) = M(\boldsymbol{\gamma}(t_0)) = \tilde{\boldsymbol{\gamma}}(t_0)$$

from which we deduce that $c = 0$. We have therefore proven that

$$\tilde{\boldsymbol{\gamma}} = \hat{\boldsymbol{\gamma}}$$

Recalling the definition of $\hat{\boldsymbol{\gamma}}$ we conclude that

$$\tilde{\boldsymbol{\gamma}}(t) = M(\boldsymbol{\gamma}(t))$$

proving uniqueness up to rigid motions.

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 $\mathcal{T}$ a collection of subsets of X. We say that $\mathcal{T}$ is a **topology** on X if the following 3 properties hold:

- (A1) We have $\emptyset, X \in \mathcal{T}$,
- (A2) If $\{A_i\}_{i\in I}$ is an arbitrary family of elements of $\mathcal{T}$, then
$$\bigcup_{i\in I} A_i \in \mathcal{T} .$$
- (A3) If $A, B \in \mathcal{T}$ then
$$A \cap B \in \mathcal{T} .$$

Further, we say:

- The pair $(X, \mathcal{T})$ is a **topological space**.
- The elements of X are called **points**.
- The sets in the topology $\mathcal{T}$ are called **open sets**.

Remark 3.2

The intersection property of $\mathcal{T}$, Property (A3) in Definition 3.1, is equivalent to the following:

- (A3') If $A_1, \ldots, A_M \in \mathcal{T}$ for some $M \in \mathbb{N}$, then
$$\bigcap_{n=1}^{M} A_n \in \mathcal{T} .$$

  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.

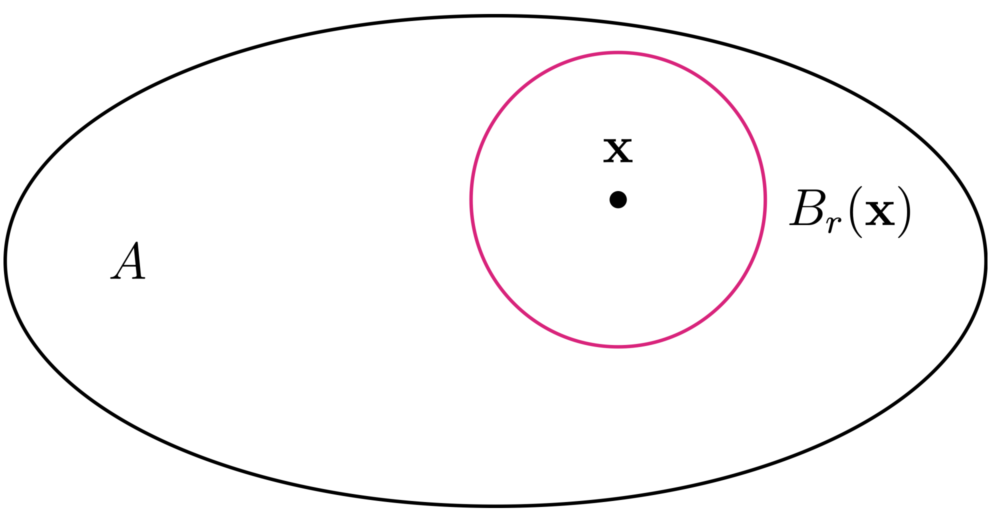


Figure 3.1: The set $A \subseteq \mathbb{R}^n$ is open if for every $\mathbf{x} \in A$ there exists $r > 0$ such that $B_r(\mathbf{x}) \subseteq A$.

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 $\mathbb{R}^n$.

Proof: Well-posedness of Definition 3.6

Let us check that $\mathcal{T}$ is a topology on $\mathbb{R}^n$:

- (A1) We have $\emptyset, \mathbb{R}^n \in \mathcal{T}$: 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 $A_i \in \mathcal{T}$ 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 $A \in \mathcal{T}$.
- (A3) Let $A, B \in \mathcal{T}$. 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 $A \cap B \in \mathcal{T}$.

This proves that $\mathcal{T}$ is a topology on $\mathbb{R}^n$.

Let us make a basic, but 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.

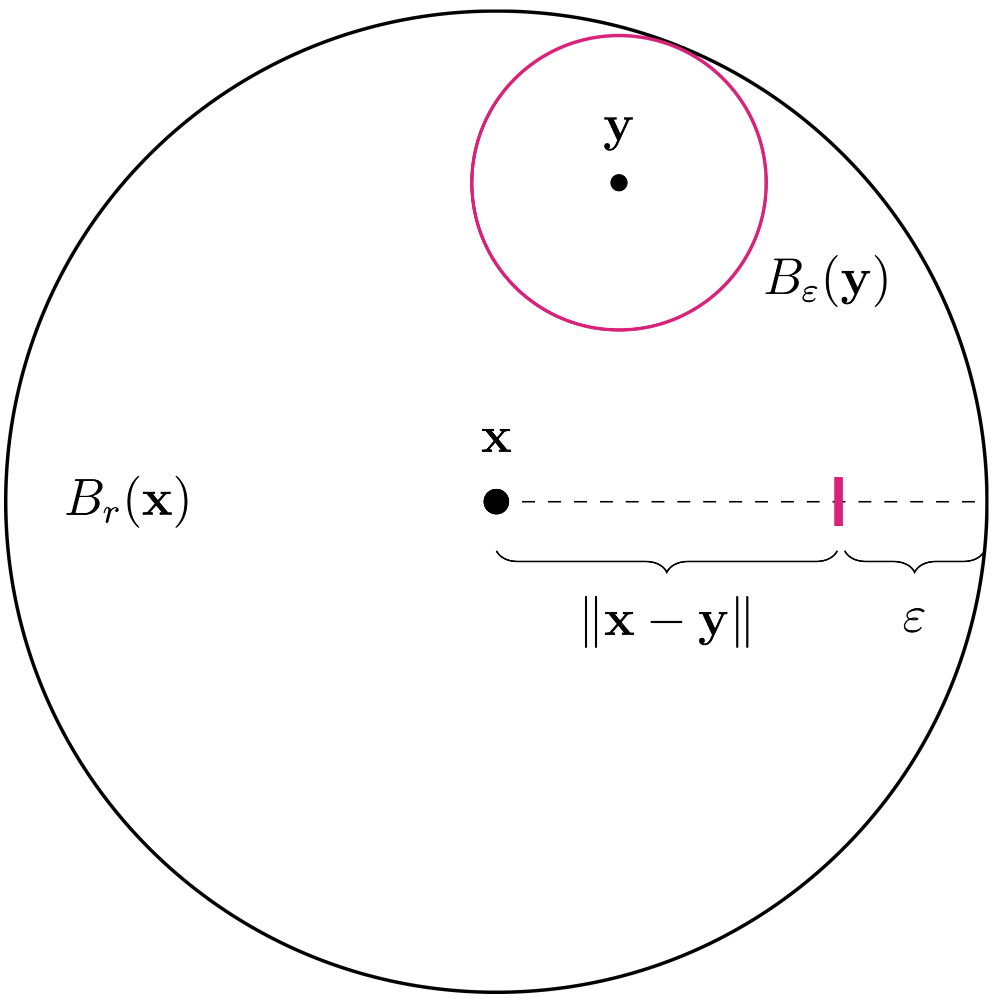


Figure 3.2: The ball $B_\varepsilon(\mathbf{y})$ is contained in $B_r(\mathbf{x})$ if $\varepsilon := r - \|\mathbf{x} - \mathbf{y}\|$.

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.

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, as 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
$$C_1 \cup C_2 \text{ is closed} \quad \iff \quad C_1^c \cap C_2^c \text{ is open}.$$
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 ring 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
$$f(x_1, \dots, x_n) = \lambda_1 x_1^{k_1} + \dots + \lambda_n x_n^{k_n},$$
where $\lambda_1, \dots, \lambda_n$ are given elements of $\mathbb{K}$ and $k_1, \dots, k_n \in \mathbb{N}$. For a collection of polynomials $I \subset \mathbb{K}[x_1, \dots, x_n]$ define
$$V(I) := \{(a_1, \dots, a_n) \in \mathbb{K}^n \, : \, f(a_1, \dots, a_n) = 0 \,, \ \forall f \in I\}.$$
The set $V(I)$ is called an **algebraic set**. 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.
The collection $\mathscr{C}$ is known as the **Zariski topology** on the space $\mathbb{K}^n$. This topology provides a natural framework for studying **Affine Varieties** – generalized surfaces obtained by gluing together algebraic

sets of the form $V(I)$. The area of mathematics studying these objects is known as Algebraic Geometry. For more information, see this Wikipedia page and this paper.
An example of affine variety is the **Quadrifolium**, which is the curve defined by the polar coordinates equation $r = \sin(2\theta)$, see Figure 3.3. It can be easily seen that the Quadrifolium is an affine variety in $\mathbb{R}^2$, which can be described by using just one algebraic set, namely $V((x^2 + y^2)^3 - 4x^2y^2)$.

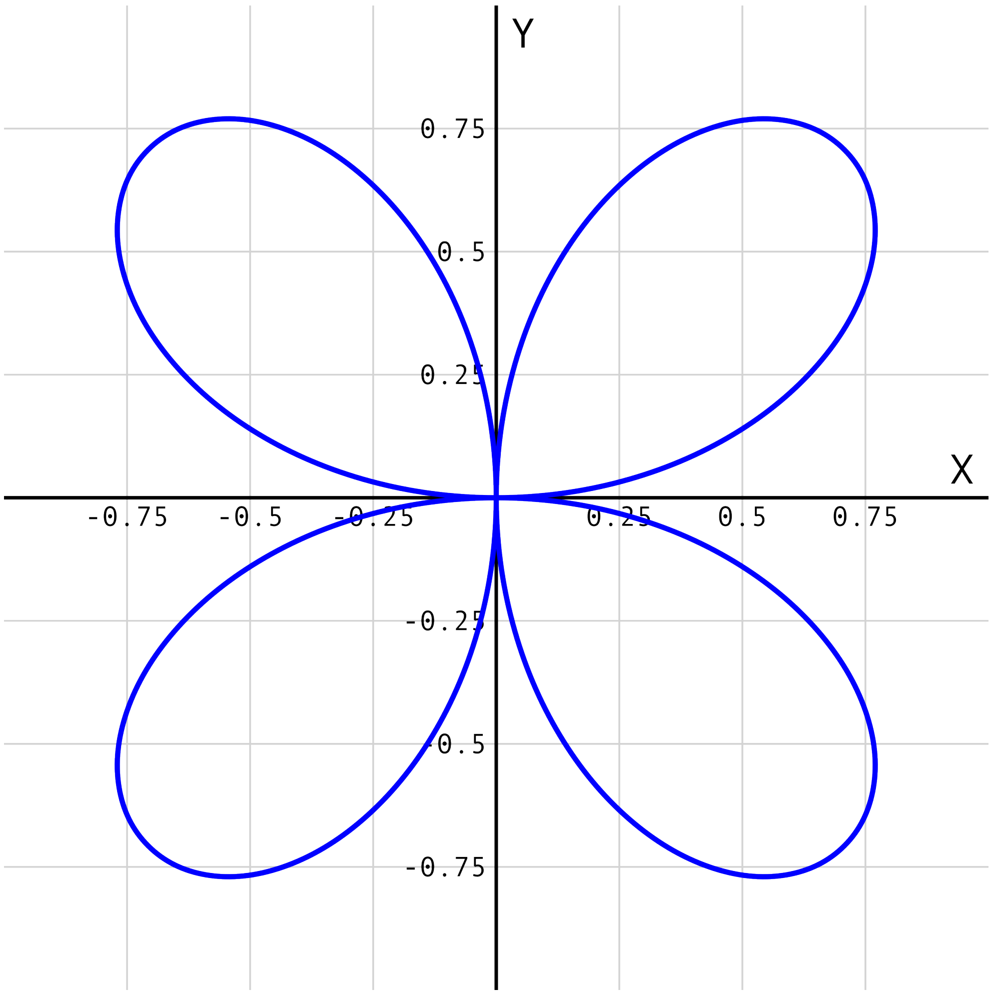


Figure 3.3: The Quadrifolium is an affine variety with algebraic set $V((x^2 + y^2)^3 - 4x^2y^2)$.

3.2 Comparing topologies

Consider the situation where you have two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ on the same set X. We would like to have some notions of comparison between $\mathcal{T}_1$ and $\mathcal{T}_2$.

Definition 3.11: Finer and coarser topology

Let X be a set and let $\mathscr{T}_1, \mathscr{T}_2$ be topologies on X. Suppose that

$$\mathscr{T}_2 \subseteq \mathscr{T}_1 .$$

We say that:

- $\mathscr{T}_1$ is **finer** than $\mathscr{T}_2$.
- $\mathscr{T}_2$ is **coarser** than $\mathscr{T}_1$.

If it holds

$$\mathscr{T}_2 \subsetneq \mathscr{T}_1 ,$$

we say that:

- $\mathscr{T}_1$ is **strictly finer** than $\mathscr{T}_2$.
- $\mathscr{T}_2$ is **strictly coarser** than $\mathscr{T}_1$.

We say that $\mathscr{T}_1$ and $\mathscr{T}_2$ are the **same** topology if

$$\mathscr{T}_1 = \mathscr{T}_2 .$$

Example 3.12

Let X be a set and consider the trivial and discrete topologies

$$\mathscr{T}_{\text{trivial}} = \{\emptyset, X\} , \quad \mathscr{T}_{\text{discrete}} = \{A \, : \, A \subseteq X\} .$$

Then

$$\mathscr{T}_{\text{trivial}} \subsetneq \mathscr{T}_{\text{discrete}} ,$$

so that $\mathscr{T}_{\text{discrete}}$ is strictly finer than $\mathscr{T}_{\text{trivial}}$.

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 $\mathscr{T}_{\text{cofinite}}$ of subsets of $\mathbb{R}$

$$\mathscr{T}_{\text{cofinite}} := \{U \subseteq \mathbb{R} \, : \, U^c \text{ is finite, or } U^c = \mathbb{R}\} .$$

Then $(\mathbb{R}, \mathscr{T}_{\text{cofinite}})$ is a topological space, and $\mathscr{T}_{\text{cofinite}}$ is called the **cofinite topology**. We have that

$$\mathscr{T}_{\text{cofinite}} \subsetneq \mathscr{T}_{\text{euclidean}} .$$

Exercise: Show that $\mathscr{T}_{\text{cofinite}}$ is a topology on $\mathbb{R}$ and that $\mathscr{T}_{\text{cofinite}} \subsetneq \mathscr{T}_{\text{euclidean}}$.

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 $(X, \mathcal{T})$ 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:

$$\forall U \in \mathcal{T} \text{ s.t. } x_0 \in U\,,\ \exists N = N(U) \in \mathbb{N} \text{ s.t. } x_n \in U\,,\ \forall n \geq N\,. \tag{3.4}$$

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 $(X, \mathcal{T})$ be topological space, with $\mathcal{T}$ the trivial topology, that is,

$$\mathcal{T} = \{\emptyset, X\}\,.$$

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 $(X, \mathscr{T})$ be topological space, with $\mathscr{T}$ the discrete topology, that is,

$$\mathscr{T} = \{A \, : \, A \subseteq X\} .$$

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 $U \in \mathscr{T}$. 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 $U \in \mathscr{T}$ 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 $d \colon X \times X \to \mathbb{R}$ 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 3.6. 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 $d \colon X \times X \to \mathbb{R}$ 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

$$\mathscr{T}_d = \mathscr{T}_{\text{discrete}}$$

where $\mathscr{T}_{\text{discrete}}$ is the **discrete topology** on X.

Exercise: Prove that (X,d) is a metric space and $\mathscr{T}_d = \mathscr{T}_{\text{discrete}}$.

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 $\mathscr{T}_d$ the topology induced by d. Then:

- For all $x \in X$, $r > 0$ we have $B_r(x) \subseteq \mathscr{T}_d$.
- $U \in \mathscr{T}_d$ 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

$$\mathscr{T}_{d_1} = \mathscr{T}_{d_2}\,.$$

The following proposition gives a sufficent condition for the equivalence of two metrics.

Proposition 3.29

Let X be a set and d_1, d_2 be metrics on X. Suppose that there exists a constant $\alpha > 0$ such that

$$\frac{1}{\alpha} d_2(x,y) \le d_1(x,y) \le \alpha \, d_2(x,y) , \quad \forall \, x, y \in X .$$

Then d_1 and d_2 are equivalent metrics.

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 3.4.

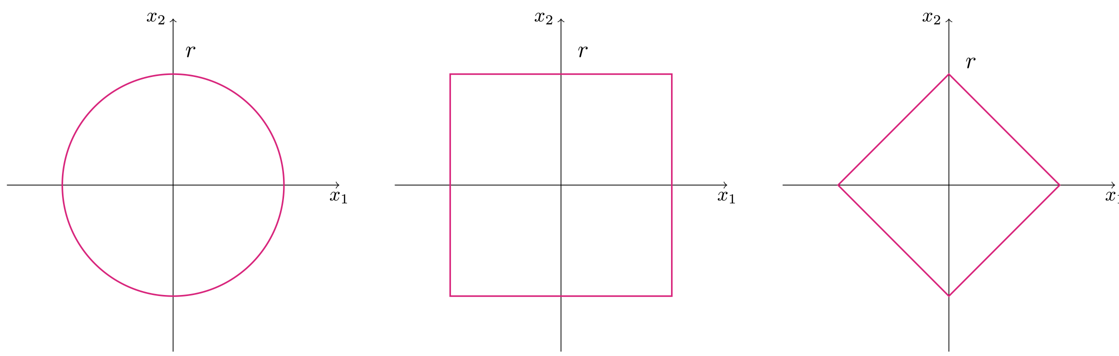


Figure 3.4: Balls $B_r(0)$ for the metrics d_2, d_∞, d_1 in $\mathbb{R}^2$.

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 $\mathcal{T}_d$ 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 $\mathcal{T}_d$.
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 $(X, \mathcal{T})$ be a topological space and $A \subseteq X$ a set. The **interior** of A is the set

$$\operatorname{Int} A := \bigcup_{\substack{U \subseteq A \\ U \in \mathcal{T}}} U\,.$$

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 $V \in \mathscr{T}$ and $V \subseteq A$. Therefore

$$V \subseteq \bigcup_{\substack{U \subseteq A \\ U \in \mathscr{T}}} U = \operatorname{Int} A \,.$$

4. Suppose that A is open. Then

$$A \subseteq \bigcup_{\substack{U \subseteq A \\ U \in \mathscr{T}}} U = \operatorname{Int} A \,.$$

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 $(X, \mathscr{T})$ 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:

$$x_0 \notin \overline{A} \quad \iff \quad \exists \, U \in \mathscr{T} \text{ s.t. } x_0 \in U \,, \;\; U \cap A = \emptyset \,.$$

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 $(X, \mathcal{T})$ 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 $(X, \mathcal{T})$ 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

1. The converse of Proposition 3.42 is false in general, that is,

$$\overline{A} \not\subset L(A)\,.$$

We show a counterexample to the above in Example 3.43.

2. The relation

$$\overline{A} = L(A)\,.$$

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

$$L(A) = A = [-\infty, 0]$$

In particular this shows

$$\overline{A} \not\subset L(A)$$

since $\overline{A} = \mathbb{R}$.

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 $B_{1/n}(x_0) \in \mathscr{T}_d$ 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 $(X, \mathscr{T})$ be a topological space and $A \subseteq X$ a set. We say that A is **dense** in X if

$$A \cap U \neq \emptyset\,, \quad \forall\, U \in \mathscr{T}\,,\ U \neq \emptyset\,.$$

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 $(X, \mathcal{T}_d)$ is Hausdorff.

In general, every metrizable space is Hausdorff.

Definition 3.52: Metrizable space

Let $(X, \mathcal{T})$ be a topological space. We say that the topology $\mathcal{T}$ is metrizable if there exists a metric d on X such that

$$\mathcal{T} = \mathcal{T}_d,$$

with $\mathcal{T}_d$ the topology induced by d.

Corollary 3.53

Let $(X, \mathcal{T})$ be a metrizable space. Then X is Hausforff.

Proof

Since $(X, \mathcal{T})$ is metrizable, there exists a metric d on X such that

$$\mathcal{T} = \mathcal{T}_d.$$

By Proposition 3.51 we know that $(X, \mathcal{T}_d)$ is Hausdorff. Hence $(X, \mathcal{T})$ 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 $(X, \mathcal{T})$ be a topological space with $\mathcal{T}$ 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 $U, V \in \mathcal{T}$ such that

$$x \in U \,, \quad y \in V \,, \quad U \cap V = \emptyset \,.$$

Recall that

$$\mathcal{T} = \{\emptyset, X\} \,.$$

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 $\mathcal{T}$ of subsets of $\mathbb{R}$

$$\mathcal{T} := \{U \subseteq \mathbb{R} \, : \, U^c \text{ is finite, or } U^c = \mathbb{R}\} \,.$$

Then $(\mathbb{R}, \mathcal{T})$ is a topological space which is not Hausdorff. The topology $\mathcal{T}$ is called the **cofinite topology**.

Exercise: Show that $(\mathbb{R}, \mathcal{T})$ is not Hausdorff.

Example 3.56

Consider the following family $\mathcal{T}$ of subsets of $\mathbb{R}$

$$\mathcal{T} := \{U = (-\infty, a) \, : \, -\infty \leq a \leq \infty\} \,.$$

Then $(\mathbb{R}, \mathcal{T})$ is a topological space which is not Hausdorff.

We start by showing that $(\mathbb{R}, \mathcal{T})$ is a topological space. We need to check the properties of topologies:

- (A1) We have that
$$(\infty, \infty) = \emptyset \in \mathcal{T} \,, \quad (-\infty, \infty) = \mathbb{R} \in \mathcal{T} \,.$$
- (A2) Suppose that $A_i \in \mathcal{T}$ 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 $A \in \mathcal{T}$. 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 $A, B \in \mathcal{T}$. 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 $U \in \mathcal{T}$.

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 $f : X \to Y$ 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 $\mathscr{T}_1, \mathscr{T}_2$ be topologies on X. Define the identity map

$$\mathrm{Id}_X : (X, \mathscr{T}_1) \to (X, \mathscr{T}_2), \quad \mathrm{Id}_X(x) := x .$$

They are equivalent:

1. Id_X is continuous from $(X, \mathscr{T}_1)$ to $(X, \mathscr{T}_2)$.
2. $\mathscr{T}_1$ is finer than $\mathscr{T}_2$

$$\mathscr{T}_2 \subseteq \mathscr{T}_1 .$$

Indeed, Id_X is continuous if and only if

$$\mathrm{Id}_X^{-1}(V) \in \mathscr{T}_1 , \quad \forall V \in \mathscr{T}_2 .$$

But $\mathrm{Id}_X^{-1}(V) = V$, so that the above reads

$$V \in \mathscr{T}_1 , \quad \forall V \in \mathscr{T}_2 ,$$

which is equivalent to $\mathscr{T}_2 \subseteq \mathscr{T}_1$.

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 $f \colon X \to Y$ 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 $f \colon X \to Y$. 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

$$x_n \to x_0 \;\text{ in } X \quad \Longrightarrow \quad f(x_n) \to f(x_0) \;\text{ in } Y.$$

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

$$x_n \to x_0 \;\text{ in } X \quad \Longrightarrow \quad f(x_n) \to f(x_0) \;\text{ in } Y.$$

for all sequences $\{x_n\} \subset X$, the function f might **not** be continuous. A counterexample is given in Example 3.70 below.

For the above to hold, it is necessary for the topologies on X and Y to be first countable, as for example is the case for metrizable topologies, see Proposition 3.69 below.

Proposition 3.69

Let (X, d_X) and (Y, d_Y) be metric spaces. Let $f : X \to Y$ and suppose that for all convergent sequences $\{x_n\} \subseteq X$, the sequence $\{f(x_n)\}$ is convergent in Y. Then f is continuous.

Proof

Suppose by contradiction f is not continuous at some point $x_0 \in X$. Then there exists $\varepsilon_0 > 0$ such that, for all $\delta > 0$ it holds

$$d_Y(f(x), f(x_0)) > \varepsilon_0 \,, \quad d_X(x, x_0) < \delta \,.$$

We can therefore choose $\delta = 1/n$ and construct a sequence $\{x_n\} \subseteq X$ such that

$$d_Y(f(x_n), f(x_0)) > \varepsilon_0 \,, \quad d_X(x_n, x_0) < \frac{1}{n} \,, \quad \forall\, n \in \mathbb{N} \,.$$

Therefore $x_n \to x_0$ in X. Define the sequence

$$y_n := \begin{cases} x_n & \text{if } n \text{ even} \\ x_0 & \text{if } n \text{ odd} \end{cases}$$

As $x_n \to x_0$, we have $y_n \to x_0$. However $\{f(y_n)\}$ does not converge to any point in Y: Indeed $\{f(y_n)\}$ cannot converge to $f(x_0)$, since for n even we have

$$d_Y(f(y_n), f(x_0)) = d_Y(f(x_n), f(x_0)) > \varepsilon_0 \,.$$

Also $\{f(y_n)\}$ cannot converge to a point $y \neq f(x_0)$, since for n odd

$$d_Y(f(y_n), y) = d_Y(f(x_0), y) > 0 \,.$$

Hence, we have produced a sequence $\{y_n\}$ which is convergent, but such that $\{f(y_n)\}$ does not converge. This contradicts our assumption. Hence f must be continuous.

Example 3.70

Consider $\mathbb{R}$ with the co-countable topology:

$$\mathcal{T}_{\text{cc}} := \{A \subseteq \mathbb{R} \,:\, A^c = \mathbb{R} \text{ or } A^c \text{ countable}\} \,.$$

Sequences in $(\mathbb{R}, \mathcal{T}_{\text{cc}})$ converge if and only if they are eventually constant. Also consider the discrete topology on $\mathbb{R}$, denoted by $\mathcal{T}_{\text{discrete}}$. We have seen that sequences in $(\mathbb{R}, \mathcal{T}_{\text{discrete}})$ converge if and only if they are eventually constant. Consider the identity function

$$f : (\mathbb{R}, \mathcal{T}_{cc}) \to (\mathbb{R}, \mathcal{T}_{\text{discrete}}), \quad f(x) := x .$$

We have that:

- f is not continuous: Indeed $\{x\} \in \mathcal{T}_{\text{discrete}}$ but

  $$f^{-1}(\{x\}) = \{x\} \notin \mathcal{T}_{cc} ,$$

  since $\{x\}^c$ is neither $\mathbb{R}$, nor countable.

- If $\{x_n\}$ is convergent in $\mathcal{T}_{cc}$, then it is eventually constant. Therefore $\{f(x_n)\}$ is eventually constant, and so it is convergent in $\mathcal{T}_{\text{discrete}}$.

Let us make an observation on continuity of compositions.

Proposition 3.71

Let $(X, \mathcal{T}_X), (Y, \mathcal{T}_Y), (Z, \mathcal{T}_Z)$ be topological spaces. Let

$$f : X \to Y , \quad g : Y \to Z ,$$

be given functions. If f and g are continuous, then

$$(g \circ f) : X \to Z$$

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 3.72: Homeomoprhim

Let $(X, \mathscr{T}_X)$, $(Y, \mathscr{T}_Y)$ be topological space. A function $f : X \to Y$ is called an **homeomorphism** if they hold:

1. f is continuous.
2. There exists $g : Y \to X$ 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 3.73: 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 3.73

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 $A_1, A_2 \in \mathcal{S}$. By definition there exist $U_1, U_2 \in \mathcal{T}$ 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 $A_1 \cap A_2 \in \mathcal{S}$, since $U_1 \cap U_2 \in \mathcal{T}$.

If the set Y is open, then sets are open in the subspace topology if and only if they are open in X.

Proposition 3.74

Let $(X, \mathcal{T})$ be a topological space and $Y \in \mathcal{T}$ a subset. Let $A \subset Y$. Then

$$A \in \mathcal{S} \quad \iff \quad A \in \mathcal{T}\,.$$

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 3.75

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 3.76: 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 3.77

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 $\mathscr{T}_d$.

This is true by Propostion 3.27.

3. Let $(X, \mathscr{T})$ with X the discrete topology. Then

$$\mathscr{B} := \{\{x\} \, : \, x \in X\}$$

is a basis for $\mathscr{T}$.

This is true because for any $U \in \mathscr{T}$ we have

$$U = \bigcup_{x \in U} \{x\} \, .$$

Proposition 3.78

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 3.78 are sufficient for generating a topology.

Proposition 3.79

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 $U_i \in \mathscr{T}$ for all $i \in I$. By definition of $\mathscr{T}$ we have

  $$U_i = \bigcup_{k \in K_i} B^i_k$$

  for some family of indices K_i and $B^i_k \in \mathscr{B}$. Therefore

  $$U := \bigcup_{i \in I} U_i = \bigcup_{i \in I, \, k \in K_i} B^i_k \, ,$$

  showing that $U \in \mathscr{T}$.
- (A3) Suppose that $U_1, U_2 \in \mathscr{T}$. Then

  $$U_1 = \bigcup_{i \in I_1} B^1_i \, , \quad U_2 = \bigcup_{i \in I_2} B^2_i$$

  for $B^1_i, B^2_i \in \mathscr{B}$. From the above we have

  $$U_1 \cap U_2 = \bigcup_{i \in I_1, \, k \in I_2} B^1_i \cap B^2_k \, .$$

  From property (B2) we have that for each pair of indices (i, k) the set $B^1_i \cap B^2_k$ is the union of sets in $\mathscr{B}$. Therefore $U_1 \cap U_2$ is union of sets in $\mathscr{B}$, showing that $U_1 \cap U_2 \in \mathscr{T}$.

2. This trivially follows from defintion of $\mathscr{T}$ and definition of basis.

3.11 Product topology

Given two topological spaces $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ 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 3.80

Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be topological spaces. Define the family $\mathscr{B}$ of subsets of $X \times Y$ as

$$\mathscr{B} := \{U \times V \,:\, U \in \mathscr{T}_X\,,\; V \in \mathscr{T}_Y\} \subset X \times Y\,.$$

Then $\mathscr{B}$ satisfies properties (B1) and (B2) from Proposition 3.78.

The proof is an easy check, and is left as an exercise. As $\mathscr{B}$ satisfies (B1)-(B2), by Proposition 3.79 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 3.81: 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 3.82

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 3.83

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 $U \times Y \in \mathcal{T}_{X\times Y}$ since $U \in \mathcal{T}_X$ and $Y \in \mathcal{T}_Y$. 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 3.84

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces and equip $X \times Y$ with the product topology $\mathcal{T}_{X\times Y}$. Let $(Z, \mathcal{T}_Z)$ be a topological space and

$$f \colon Z \to X \times Y$$

a function. They are equivalent:

1. f is continuous.
2. The compositions
$$\pi_X \circ f \colon Z \to X\,, \quad \pi_Y \circ f \colon Z \to Y$$
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 3.85: 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 3.86: Proper subset

Let X be a set. A subset $A \subseteq X$ is **proper** if

$$A \neq \emptyset , \quad A \neq X .$$

Proposition 3.87: 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

$$A^c = X \setminus A = B ,$$

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

$$A^c = X \setminus A = B ,$$

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 3.87 as equivalent definitions of disconnected topological space.

Example 3.88

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 3.89

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 3.90

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 3.91

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 3.91 is that connectedness is a topological invariant, e.g., connectedness is preserved by homeomorphisms.

Corollary 3.92

Let $(X, \mathcal{T}_X)$, $(Y, \mathcal{T}_Y)$ be homeomorhic topological spaces. Then

$$X \text{ is connected} \iff Y \text{ is connected}$$

The proof follows immediately by Theorem 3.91, and is left to the reader as an exercise.

Example 3.93

Let $n \geq 2$. Then $\mathbb{R}^n$ not homeomorphic to $\mathbb{R}$.

Proof. Suppose by contradiction that there exists a homeomorphism

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

Define $p = f(0)$ and the restriction

$$g : \mathbb{R}^n \smallsetminus \{0\} \to \mathbb{R} \smallsetminus \{p\}, \quad g(x) = f(x).$$

Since g is a restriction of a homeomorphism, then g is a homeomorphism. We have that $\mathbb{R}^n \smallsetminus \{0\}$ is connected, as a consequence of
Example 3.90. Hence, by Corollary 3.92, we infer that $\mathbb{R} \smallsetminus \{p\}$ is connected. This is a contradiction, since $\mathbb{R} \smallsetminus \{p\}$ is disconnected, as shown in Example 3.89.

A stronger version of the statement in Example 3.93 holds.

Theorem 3.94: Topological invariance of dimension

Let $n \neq m$. Then $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}^m$.

Unfortunately, the argument in Example 3.93 cannot be extended to prove Theorem 3.94. The bottom line is that connectedness does not give enough information to tell apart $\mathbb{R}^n$ from $\mathbb{R}^m$. The right topological tool to prove Theorem 3.94 is called **homology**, which requires a serious effort to construct/define.

Let us give another example of spaces which are not homeomorphic.

Example 3.95

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 $[0,1] \smallsetminus \{\frac{1}{2}\}$ defines an omeomorphism

$$g \,:\, \left([0,1] \smallsetminus \left\{\frac{1}{2}\right\}\right) \to (\mathbb{S}^1 \smallsetminus \{\mathbf{p}\}) \,, \quad \mathbf{p} := f\left(\frac{1}{2}\right) .$$

The set $[0,1] \smallsetminus \left\{\frac{1}{2}\right\}$ is disconnected, since

$$[0,1] \smallsetminus \{1/2\} = [0,1/2) \cup (1/2,1]$$

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 3.91, we conclude that $\mathbb{S}^1 \smallsetminus \{\mathbf{p}\}$ is connected. Contradiction.

3.13 Intermediate Value Theorem

Another consequence of Theorem 3.91 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 3.96: 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 3.97

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 3.98: 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 3.91 we know that $f(X)$ is connected in $\mathbb{R}$. By Theorem 3.97 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 3.99: 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 3.100

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 3.101

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 $\alpha : [0,1] \to X$ 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.

The converse of the above theorem does not hold. A counterexample is given by the so-called **topologist curve**, which will be examined in Proposition 3.103. Prior to this, we need a basic Lemma.

Lemma 3.102

Let $(X, \mathcal{T})$ be a topological space. Let $A, U \subseteq X$ with A connected and U open and closed. Suppose that $A \cap U \neq \emptyset$, then $A \subseteq U$.

Proof

The following set identities hold for any pair of sets U and A:

$$\begin{aligned} A &= (A \cap U) \cup (A \cap U^c) \\ \emptyset &= (A \cap U) \cap (A \cap U^c) \end{aligned}$$

Now, suppose by contradiction $A \not\subseteq U$. This means $A \cap U^c \neq \emptyset$. By assumption we also have $A \cap U \neq \emptyset$. Moreover the sets $A \cap U$ and $A \cap U^c$ are open for the subspace topology on A, since U and U^c are open in

X. Hence A is the disjoint union of non-empty open sets, showing that A is disconnected. Contradiction. We conclude that $A \subseteq U$.

Proposition 3.103: Topologist curve

Consider $\mathbb{R}^2$ with the Euclidean topology and define the sets

$$X := A \cup B$$

where

$$A := \left\{ \left(t, \sin\left(\frac{1}{t}\right) \right) \, : \, t > 0 \right\}$$
$$B := \{(0,t) \, : \, t \in [-1,1]\}$$

Then X is connected, but not path connected.

Proof

Step 1. X is not path connected.
Let $x \in A$ and $y \in B$. There is no continuous function $\alpha \colon [0,1] \to X$ such that $\alpha(0) = x$ and $\alpha(1) = y$. If such α existed, then we would obtain a continuous extension for $t = 0$ of the function

$$f(t) = \sin\left(\frac{1}{t}\right), \quad x > 0$$

which is not possible. Hence X is not path connected.
Step 2. Preliminary facts.

- A is connected: Define the curve $\boldsymbol{\gamma} \colon (0,\infty) \to \mathbb{R}^2$ by

  $$\boldsymbol{\gamma}(t) := \left(t, \sin\left(\frac{1}{t}\right) \right).$$

  Clearly $\boldsymbol{\gamma}$ is continuous. Since $(0,\infty)$ is connected, by Theorem 3.91 we have that $\boldsymbol{\gamma}((0,\infty)) = A$ is connected.

- B is connected: Indeed B is homeomorphic to the interval $[-1,1]$. Since $[-1,1]$ is connected, by Corollary 3.92 we conclude that B is connected.

- $\overline{A} = X$: This is because each point $y \in B$ is of the form $y = (0, t_0)$ for some $t_0 \in [-1,1]$. By continuity of sin and the Intermediate Value Theorem there exists some $z > 0$ such that

  $$\sin(z) = t_0 .$$

  Therefore $z_n := z + 2n\pi$ satisfies

  $$z_n \to \infty , \quad \sin(z_n) = t_0 , \quad \forall\, n \in \mathbb{N} .$$

Define $s_n := 1/z_n$. Trivially

$$s_n \to 0 \,, \quad \sin\left(\frac{1}{s_n}\right) = t_0 \,, \quad \forall \, n \in \mathbb{N} \,.$$

Therefore we obtain

$$\left(s_n, \sin\left(\frac{1}{s_n}\right)\right) \to (0, t_0) \,.$$

Hence the set B is contained in the set $L(A)$ of limit points of A. Since we are in $\mathbb{R}^2$, we have that $L(A) = \overline{A}$, proving that $B \subseteq \overline{A}$. Thus $\overline{A} = A \cup B = X$.

Step 3. X is connected.
Let $U \subseteq X$ be non-empty, open and closed. If we prove that $U = X$, we conclude that X is connected. Let us proceed.
Since U is non-empty, we can fix a point $x \in U$. We have two possibilities:

- $x \in A$: In this case $A \cap U \neq \emptyset$. Since A is connected and U is open and closed, by Lemma 3.102 we conclude $A \subseteq U$. As U is closed and contains A, then $\overline{A} \subseteq U$. But we have shown that

  $$\overline{A} = X \,,$$

  and therefore $U = X$.

- $x \in B$: Then $U \cap B \neq \emptyset$. Since B is connected and U is open and closed, we can invoke Lemma 3.102 and conclude that $B \subseteq U$. Since $(0,0) \in B$, it follows that

  $$(0,0) \in U \,.$$

  As U is open in X, and X has the subspace topology induced by the inclusion $X \subseteq \mathbb{R}^2$, there exists an open set W of $\mathbb{R}^2$ such that

  $$U = X \cap W \,.$$

  Therefore $(0,0) \in W$. As W is open in $\mathbb{R}^2$, there exists a radius $\varepsilon > 0$ such that

  $$B_\varepsilon(0,0) \subseteq W \,.$$

  Hence

  $$X \cap B_\varepsilon(0,0) \subseteq X \cap W = U \,.$$

  The ball $B_\varepsilon(0,0)$ contains points of A, and therefore

  $$A \cap U \neq \emptyset \,.$$

  Since A is connected and U is open and closed, we can again use Lemma 3.102 and obtain that $A \subseteq U$. Since we already had $B \subseteq U$, and since $U \subseteq X = A \cup B$, we conclude hence $U = X$.

Therefore $U = X$ in all possible cases, showing that X is connected.

To conclude, we observe that connectedness and path-connectedness are equivalent in open sets of $\mathbb{R}^n$.

Theorem 3.104

Let $A \subset \mathbb{R}^n$ be open for the Euclidean topology. Then A is connected if and only if it is path-connected.

The proof of this theorem is a bit delicate, and we decided to omit it.

4 Surfaces

Curves are 1D objects in $\mathbb{R}^3$, parametrized via functions $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$. There is only one available direction in which to move on a curve:

- $t \mapsto \boldsymbol{\gamma}(t)$ moves forward on the curve
- $t \mapsto \boldsymbol{\gamma}(-t)$ moves backward on the curve

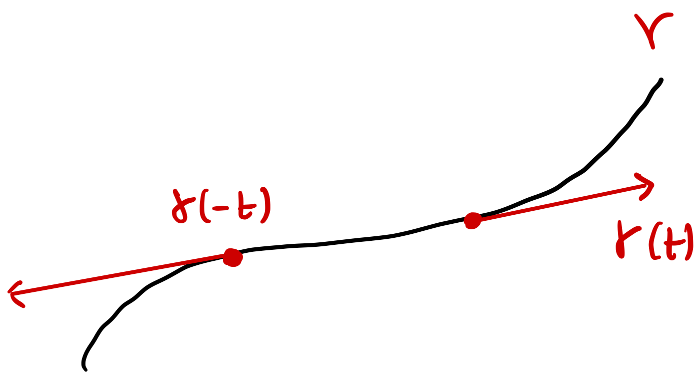


Figure 4.1: Sketch of a curve $\boldsymbol{\gamma}$.

Surfaces are 2D objects in $\mathbb{R}^3$. There are two directions in which one can move on a surface.

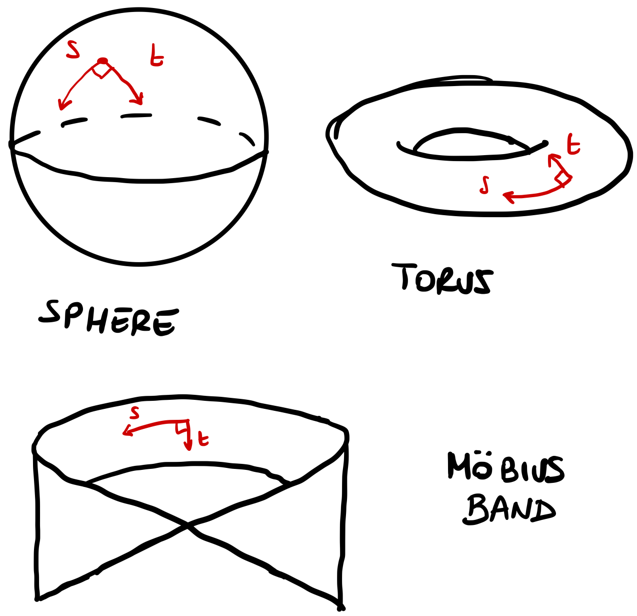


Figure 4.2: Sketch of a surfaces: Sphere, Torus, Möbius band.

Question 4.1

How to dercribe a surface mathematically?

A curve $\Gamma \subseteq \mathbb{R}^3$ can be described with one function $\boldsymbol{\gamma} : (a, b) \to \Gamma$. The idea is that Γ looks locally like $\mathbb{R}$.

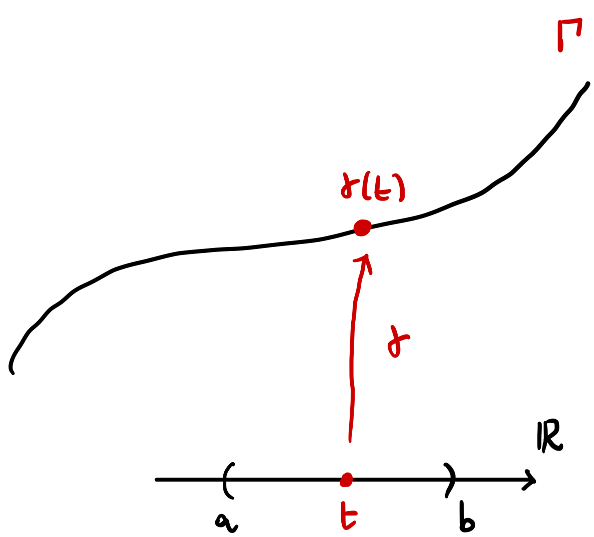


Figure 4.3: A curve Γ can be described by a function $\boldsymbol{\gamma} : (a, b) \to \Gamma$.

How do we represent a surface? Suppose given a function $\boldsymbol{\sigma} : U \to \mathbb{R}^3$, with $U \subseteq \mathbb{R}^2$ open set. Denote by $\mathcal{S} := \boldsymbol{\sigma}(U)$ the image of U through $\boldsymbol{\sigma}$. We say that $\mathcal{S}$ is a surface and $\boldsymbol{\sigma}$ is a **chart**. Unofortunately, not all surfaces can be described with just one chart: in most cases one needs to piece together many local **charts** $\boldsymbol{\sigma}_i : U_i \to \mathcal{S}$, with $U_i \subseteq \mathbb{R}^2$ open. The charts $\boldsymbol{\sigma}_i$ represent $\mathcal{S}$ if they cover the whole surface:

$$\mathcal{S} = \bigcup_i \boldsymbol{\sigma}_i(U_i) \,.$$

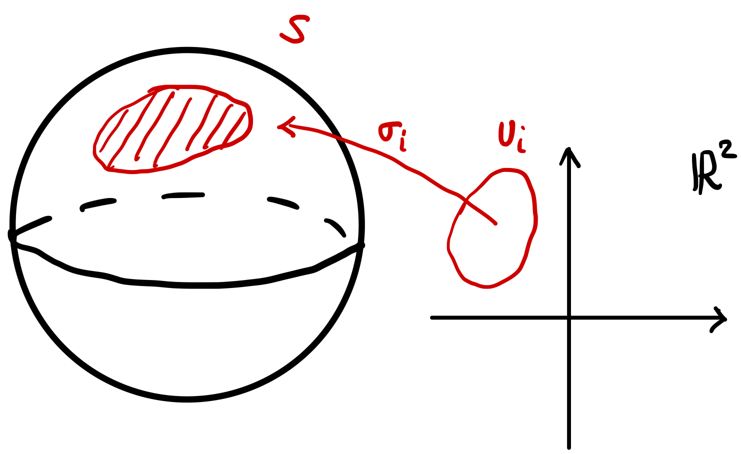


Figure 4.4: A surface $\mathcal{S}$ can be described by a family of charts $\boldsymbol{\sigma}_i : U_i \to \mathcal{S}$ with $U_i \subseteq \mathbb{R}^2$ open set.

Before proceeding with the formal definition of surface, we collect some preliminary definitions and results.

4.1 Preliminaries

Before proceeding with the formal definition of surface, we need to establish some basic notation and terminology regarding linear algebra, the topology of $\mathbb{R}^n$, and calculus for smooth maps from $\mathbb{R}^n$ into $\mathbb{R}^m$.

4.1.1 Linear algebra

Definition 4.2: Bilinear form

Let V be a vector space and $B : V \times V \to \mathbb{R}$. We say that:

- B is **bilinear** if

$$B(\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2, \mathbf{w}) = \lambda_1 B(\mathbf{v}_1, \mathbf{w}) + \lambda_2 B(\mathbf{v}_2, \mathbf{w}),$$
$$B(\mathbf{w}, \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2) = \lambda_1 B(\mathbf{w}, \mathbf{v}_1) + \lambda_2 B(\mathbf{w}, \mathbf{v}_2).$$

  for all $\mathbf{v}_i, \mathbf{w} \in V$, $\lambda_i \in \mathbb{R}$.

- B is **symmetric** if

$$B(\mathbf{v}, \mathbf{w}) = B(\mathbf{w}, \mathbf{v})$$

  for all $\mathbf{v}, \mathbf{w} \in V$.

A bilinear map B is called **bilinear form** on V.

Notation

Let V be a vector space with basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. Then, for a vector $\mathbf{v} \in V$ there exist coefficients $\lambda_1, \dots, \lambda_n$ such that

$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \dots + \lambda_n \mathbf{v}_n .$$

We denote the vector of coefficients of $\mathbf{v}$ by the column vector

$$\mathbf{x} := (\lambda_1, \dots, \lambda_n)^T \in \mathbb{R}^n .$$

The coefficients of a vector $\mathbf{w}$ are denoted by

$$\mathbf{y} := (\mu_1, \dots, \mu_n)^T .$$

Notice that we are using different letters to denote abstract vectors $\mathbf{v}, \mathbf{w} \in V$, and their components $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Bilinear forms can be represented by a matrix.

Remark 4.3: Matrix representation for bilinear forms

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for the vector space V. Given a bilinear form $B \colon V \times V \to \mathbb{R}$ we define the matrix

$$M := \big(B(\mathbf{v}_i, \mathbf{v}_j)\big)_{i,j=1}^n \in \mathbb{R}^{n \times n} .$$

Then

$$B(\mathbf{v}, \mathbf{w}) = \mathbf{x}^T M \mathbf{y} .$$

Proof. We can write $\mathbf{v}$ and $\mathbf{w}$ in cordinates as

$$\mathbf{v} = \sum_{i=1}^n \lambda_i \mathbf{v}_i , \quad \mathbf{w} = \sum_{i=1}^n \mu_i \mathbf{v}_i ,$$

for suitable coefficients $\lambda_i, \mu_i \in \mathbb{R}$. Using bilinearity of B we get

$$\begin{aligned} B(\mathbf{v}, \mathbf{w}) &= B\left(\sum_{i=1}^n \lambda_i \mathbf{v}_i, \sum_{j=1}^n \mu_j \mathbf{v}_j\right) \\ &= \sum_{i,j=1}^n \lambda_i \mu_j B(\mathbf{v}_i, \mathbf{v}_j) \\ &= \mathbf{x}^T M \mathbf{y} \,. \end{aligned}$$

Definition 4.4: Quadratic form

Let V be a vector space and $B : V \times V \to \mathbb{R}$ be a bilinear form. The **quadratic** form associated to B is the map

$$Q : V \to \mathbb{R} \,, \quad Q(\mathbf{v}) := B(\mathbf{v}, \mathbf{v}) \,.$$

A symmetric bilinear form is uniquely determinded by its quadratic form, as stated in the following proposition.

Proposition 4.5

Let $B : V \times V \to \mathbb{R}$ be a symmetric bilinear form and $Q : V \to \mathbb{R}$ the associated quadratic form. Then

$$B(u, v) = \frac{1}{2} \left(Q(\mathbf{v} + \mathbf{w}) - Q(\mathbf{v}) - Q(\mathbf{w}) \right) \,.$$

for all $\mathbf{v}, \mathbf{w} \in V$.

The proof is an easy check, and is left as an exercise.

Definition 4.6: Inner product

Let V be a vector space. An inner product on V is a symmetric bilinear form $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that

$$\langle \mathbf{v}, \mathbf{v} \rangle > 0 \,, \quad \forall \, \mathbf{v} \in V \,.$$

Moreover:

- The **length** of a vector $\mathbf{v} \in V$ with respect to B is defined as

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \,.$$

- Two vectors $\mathbf{v}, \mathbf{w} \in V$ are **orthogonal** if

$$\langle \mathbf{v}, \mathbf{w} \rangle = 0 \,.$$

Example 4.7

Let $V = \mathbb{R}^n$ and consider the euclidean scalar product

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i w_i \,,$$

where $\mathbf{v} = (v_1, \dots, v_n)$, $\mathbf{w} = (w_1, \dots, w_n)$. Then

$$\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v} \cdot \mathbf{w}$$

is an inner product on $\mathbb{R}^n$.

Proposition 4.8

Let V be a vector space and $\langle \cdot, \cdot \rangle$ an inner product on V. There exists an **orthonormal** basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V, that is, such that

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

In particular, the matrix M associated to $\langle \cdot, \cdot \rangle$ is the identity.

Definition 4.9: Linear map

Let V, W be vector spaces and $L : V \to W$. We say that L is **linear** if

$$L(\lambda \mathbf{v} + \mu \mathbf{w}) = \lambda L(\mathbf{v}) + \mu L(\mathbf{w})$$

for all $\mathbf{v}, \mathbf{w} \in V$ and $\lambda, \mu \in \mathbb{R}$.

Remark 4.10: Matrix representation of linear maps

Let V, W be vector spaces and $L : V \to W$ be a linear map. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis of V and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ be a basis of W. Then there exists a matrix $M \in \mathbb{R}^{m \times n}$ such that

$$L\mathbf{v} = M\mathbf{x} \,, \quad \forall \, \mathbf{v} \in V \,.$$

Specifically, $M \in \mathbb{R}^{n \times n}$ is called the matrix associated to L with respect to the basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V and $\{\mathbf{w}_1 \dots, \mathbf{w}_m\}$ of W, and is defined by

$$M := \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} ,$$

where the coefficients a_{ij} are such that

$$L(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + \dots + a_{mj}\mathbf{w}_m = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i \,.$$

In other words, the columns of M are given by the coordinates of the vectors $L(\mathbf{v}_i)$ with respect to the basis $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$.

Definition 4.11: Eigenvalues and eigenvectors

Let V be a vector space and $L: V \to V$ a linear map. We say that $\lambda \in \mathbb{R}$ is an eigenvalue of L if

$$L(\mathbf{v}) = \lambda \mathbf{v}$$

for some $\mathbf{v} \in V$ with $\mathbf{v} \neq 0$. Such $\mathbf{v}$ is called **eigenvector** of L associated to the eigenvalue λ.

Definition 4.12: Self-adjoint map

Let V be a vector space, $\langle \cdot, \cdot \rangle$ an inner product and $L: V \to V$ a linear map. We say that L is **self-adjoint** if

$$\langle \mathbf{v}, L(\mathbf{w}) \rangle = \langle L(\mathbf{v}), \mathbf{w} \rangle \,, \quad \forall\, \mathbf{v},\, \mathbf{w} \in V \,.$$

Theorem 4.13: Spectral Theorem

Let V be a vector space, $\langle \cdot, \cdot \rangle$ an inner product, and $L: V \to V$ a self-adjoint linear map. There exist an orthonormal basis of V

$$\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \,,$$

where $\mathbf{v}_i$ are eigenvectors of L, that is,

$$L\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

for some eigevalue $\lambda_i \in \mathbb{R}$. In particular, the matrix of L with respect to the basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is diagonal:

$$M = \operatorname{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \,.$$

There is also a matrix version of the spectral theorem. To state it, we need to introduce some terminology.

Definition 4.14

Let $A \in \mathbb{R}^{n \times n}$ be a matrix. We say that:

- A is **symmetric** if
$$A^T = A \,.$$
- A is **orthogonal** if
$$A^T A = I \,,$$
where I is the identity matrix.

Remark 4.15

Let $L \colon V \to V$ be linear and $A \in \mathbb{R}^{n \times n}$ be the matrix associated to L with respect to any basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V. They are equivalent:

- L is self-adjoint,
- A is symmetric.

Definition 4.16: Matrix eigenvalues

Let $A \in \mathbb{R}^{n \times n}$ be a matrix. An **eigenvalue** of A is a number $\lambda \in \mathbb{R}$ such that

$$A\mathbf{v} = \lambda \mathbf{v} \,,$$

for some $\mathbf{v} \in \mathbb{R}^n$ with $\mathbf{v} \neq 0$. The vector $\mathbf{v}$ is called an **eigenvector** of A with eigenvalue λ.

Remark 4.17

Let $A \in \mathbb{R}^{n \times n}$. The eigenvalues of λ of A can be computed by solving the **characteristic equation**

$$P(\lambda) = 0 \,,$$

where P is the **characteristic polynomial** of A, defined by

$$P(\lambda) := \det(A - \lambda I) \,.$$

Remark 4.18

Let $L \colon V \to V$ be a linear map and A the associated matrix with respect to any basis of V. Then

$$L(\mathbf{v}) = A\mathbf{x} \,, \qquad \forall \, \mathbf{v} \in V \,,$$

where $\mathbf{x} \in \mathbb{R}^n$ is the vector of coordinates of $\mathbf{v}$. They are equivalent:

- λ is an eigenvalue of L of eigenvector $\mathbf{v}$,

- λ is an eigenvalue of A of eigenvector $\mathbf{x}$.

Theorem 4.19: Spectral Theorem for matrices

Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Consider $\mathbb{R}^n$ equipped with the euclidean scalar product. There exist an orthonormal basis of V

$$\{\mathbf{v}_1, \dots, \mathbf{v}_n\},$$

where $\mathbf{v}_i$ are eigenvectors of A, that is,

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

for some eigevalue $\lambda_i \in \mathbb{R}$. Moreover

$$A = PDP^T,$$

where

$$\begin{aligned} P &:= (\mathbf{v}_1 | \dots | \mathbf{v}_n) \\ D &:= \operatorname{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}. \end{aligned}$$

Remark 4.20

The corresponedence between Theorem 1.13 and Theorem 1.19 is as follows. Let $A \in \mathbb{R}^{n \times n}$ be symmetric and $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ be any orthonormal basis of the vector space V. Define the linear map $L : V \to V$ such that

$$L(\mathbf{v}_j) = \sum_{i=1}^{n} a_{ij} \mathbf{w}_i, \quad \forall j = 1, \dots, n.$$

In this way A is the matrix associated to L with respect to the basis $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$. Then L is self-adjoint. Moreover L and A have the same eigenvalues. By the Spectral Theorem there exists an orthonormal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V such that the matrix of L with respect to such basis, say D, is diagonal. Then

$$A = PDP^T$$

where P is the matrix of change of basis between $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, that is, $P = (p_{ij})$ where

$$\mathbf{w}_j = \sum_{i=1}^{n} p_{ij} \mathbf{v}_i.$$

4.1.2 Topology of $\mathbb{R}^n$

The Euclidean norm on $\mathbb{R}^n$ is denoted by

$$\|\mathbf{x}\| := \sqrt{\sum_{i=1}^n x_i^2}\,, \quad \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n\,.$$

The Euclidean norm induces the distance

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}\,.$$

Definition 4.21: Euclidean Topology

The pair $(\mathbb{R}^n, d)$ is a metric space. The topology induced by the metric d is called the Euclidean topology, denoted by $\mathscr{T}$. In this chapter we will always assume that $\mathbb{R}^n$ is equipped with the Euclidean topology $\mathscr{T}$.

Definition 4.22: Open Sets

A set $U \subseteq \mathbb{R}^n$ is open if for all $\mathbf{x} \in U$ there exists $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{x}) \subseteq U$, where

$$B_\varepsilon(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n \;:\; \|\mathbf{x} - \mathbf{y}\| < \varepsilon\}$$

is the open ball of radius $\varepsilon > 0$ and centered at $\mathbf{x}$. In this case we denote $U \in \mathscr{T}$, with $\mathscr{T}$ the Euclidean topology in $\mathbb{R}^n$.

Definition 4.23: Closed Sets

A set $V \subseteq \mathbb{R}^n$ is closed if $V^c := \mathbb{R}^n \setminus U$ is open.

Example 4.24

- The n-dimensional unit sphere
  $$\mathbb{S}^n = \{\mathbf{x} \in \mathbb{R}^{n+1} \;:\; \|x\| = 1\}$$
  is not open in $\mathbb{R}^{n+1}$, since for any $\mathbf{x} \in \mathbb{S}^n$ we have
  $$B_\varepsilon(\mathbf{x}) \nsubseteq \mathbb{S}^n\,.$$

- The n-dimensional unit cube
  $$C := \{\mathbf{x} \in \mathbb{R}^n \;:\; |x_1| + \ldots + |x_n| < 1\}$$
  is open in $\mathbb{R}^n$, since one can always find $\varepsilon > 0$ small enough so that
  $$B_\varepsilon(\mathbf{x}) \nsubseteq C\,.$$

- The set
$$V := \{\mathbf{x} \in \mathbb{R}^n \, : \, |x_1| + \ldots + |x_n| \geq 1\}$$
is closed, since $V^c = C$ is the unit cube, which is open.

Definition 4.25: Subspace Topology

Given a subset $A \subseteq \mathbb{R}^n$ the subspace topology on A is the family of sets
$$\mathcal{T}_A := \{U \subseteq A \, : \, \exists \, W \in \mathcal{T} \text{ s.t. } U = A \cap W\}\,.$$
If $U \in \mathcal{T}_A$ we say that U is open in A.

4.1.3 Smooth functions

We recall some basic facts about smooth functions from $\mathbb{R}^n$ into $\mathbb{R}^m$. For a vector valued function $f \colon \mathbb{R}^n \to \mathbb{R}^m$ we denote its components by
$$f = (f_1, \ldots, f_m)\,.$$

Definition 4.26: Continuous Function

Let $f \colon U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open. We say that f is continuous at $\mathbf{x} \in U$ if $\forall \, \varepsilon > 0, , \exists \, \delta > 0$ such that
$$\|\mathbf{x} - \mathbf{y}\| < \delta \quad \implies \quad \|f(\mathbf{x}) - f(\mathbf{y})\| < \varepsilon\,.$$
We say that f is continuous in U if it is continuous for all $\mathbf{x} \in U$.

Remark 4.27

Let $f \colon U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$, with U, V open. We have that f is continuous if and only if $f^{-1}(A)$ is open in U, for all A open in V.

Definition 4.28: Homeomorphism

Let $f \colon U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$ with U, V open. We say that f is a homeomorphism if f is continuous and there exists inverse $f^{-1} \colon V \to U$ continuous.

Definition 4.29: Differentiable Function

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open. We say that f is differentiable at $\mathbf{x} \in U$ if there exists a linear map $df_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\varepsilon \to 0} \frac{f(\mathbf{x} + \varepsilon \mathbf{h}) - f(\mathbf{x}) - \varepsilon\, df_{\mathbf{x}}(\mathbf{h})}{\varepsilon} = 0 ,$$

for all $\mathbf{h} \in \mathbb{R}^n$, where the limit is taken in $\mathbb{R}^m$. The map $df_{\mathbf{x}}$ is called the **differential** of f at $\mathbf{x}$.

We denote by $\{\mathbf{e}_i\}_{i=1}^n$ the standard basis of $\mathbb{R}^n$.

Definition 4.30: Partial Derivative

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open be differentiable. The partial derivative of f at $\mathbf{x} \in U$ in direction $\mathbf{e}_i$ is given by

$$\frac{\partial f}{\partial x_i} := \lim_{\varepsilon \to 0} \frac{f(\mathbf{x} + \varepsilon \mathbf{e}_i) - f(\mathbf{x})}{\varepsilon} .$$

Definition 4.31: Jacobian Matrix

The differential $df_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. As such, it must have a matrix representation with respect to the Euclidean basis. Such representation matrix is called **Jacobian**, and is given by:

$$Jf(x) := \left(\frac{\partial f_i}{\partial x_j} \right)_{i,j} \in \mathbb{R}^{m \times n} .$$

If $m = n$ then $Jf \in \mathbb{R}^{n \times n}$ is a square matrix and we can compute its determinant, denoted by

$$\det(Jf) .$$

Definition 4.32: Multi-index notation

For a multi-index

$$\alpha := (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$$

we denote by

$$|\alpha| := \sum_{i=1}^n |\alpha_i|$$

the length of the multi-index.

Definition 4.33: Smooth Function

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open. We say that f is smooth if the derivatives

$$\frac{\partial^{|\alpha|} f}{d\mathbf{x}^\alpha} := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f$$

exist for each multi-index $\alpha \in \mathbb{N}^n$. Note that in this case all the derivatives of f are automatically continuous.

Notation: Gradient and partial derivatives

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be smooth. We denote the partial derivatives by

$$\partial_{x_i} f := \frac{\partial f}{\partial x_i}, \quad \partial_{x_i x_j} f := \frac{\partial^2 f}{\partial x_i \partial x_j}, \quad \partial_{x_i x_j x_k} f := \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}.$$

For $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ smooth we denote the **gradient** by

$$\nabla f(\mathbf{x}) = \left(f_{x_1}(\mathbf{x}), \dots, f_{x_n}(\mathbf{x}) \right).$$

Example 4.34

The functions $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$f(x, y) := \cos(x) y, \quad g(x, y) := (x^2, y^2, x - y)$$

are both smooth.

Definition 4.35: Diffeomorphism

Let $f : U \to V$ with $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^n$ open. We say that f is a **diffeomorphism** between U and V if f is smooth and there exists smooth inverse $f^{-1} : V \to U$.

We recall, without proof, the Inverse Function Theorem. Please note that in the statement the function f is defined from $\mathbb{R}^n$ into $\mathbb{R}^n$.

Theorem 4.36: Inverse Function Theorem

Let $f : U \to \mathbb{R}^n$ with $U \subseteq \mathbb{R}^n$ open. Suppose f is a smooth function and

$$\det Jf(\mathbf{x}_0) \neq 0,$$

for some $\mathbf{x}_0 \in U$. Then there exist open sets $U_0, V \subseteq \mathbb{R}^n$ such that $\mathbf{x}_0 \in U_0$, $f(\mathbf{x}_0) \in V$ and $f : U_0 \to V$ is a diffeomorphism.

Warning

Even if
$$\det Jf(\mathbf{x}) \neq 0\,,$$
for all $\mathbf{x} \in U$, it is not guaranteed that f is a diffeomorphism between U and $f(U)$.

Non-vanishing Jacobian determinant is a necessary condition for being a diffeomorphism.

Proposition 4.37

Let $f : U \to \mathbb{R}^n$ with $U \subseteq \mathbb{R}^n$ open. Suppose f is a diffeomorphism on U. Then
$$\det Jf(\mathbf{x}) \neq 0\,, \quad \forall\, \mathbf{x} \in U\,.$$

Example 4.38

We have already encountered Proposition 4.37 in the scalar case when we were studying curves. Indeed, if $f : U \subset \mathbb{R} \to \mathbb{R}$ is a diffeomorphism, then $Jf(x) = \dot{f}(x)$. Hence
$$\det Jf(x) = \dot{f}(x)\,,$$
and we recover the familiar result
$$\dot{f}(x) \neq 0\,, \quad \forall\, x \in U\,.$$

Example 4.39

Define $f : \mathbb{R}^2 \to \mathbb{R}^2$ by
$$f(x,y) := (\cos(x)\sin(y), \sin(x)\sin(y))\,.$$
Then
$$Jf(x,y) = \begin{pmatrix} -\sin(x)\sin(y) & \cos(x)\cos(y) \\ \cos(x)\sin(y) & \sin(x)\cos(y) \end{pmatrix}.$$
and
$$\begin{aligned} \det Jf(x,y) &= -\sin^2(x)\cos(y)\sin(y) - \cos^2(x)\cos(y)\sin(y) \\ &= -\sin(y)\cos(y) \\ &= -\frac{1}{2}\sin(2y)\,. \end{aligned}$$
Therefore
$$\det Jf(x,y) \neq 0 \quad \iff \quad y \neq \frac{n\pi}{2}\,,\ n \in \mathbb{N}\,.$$
Hence f is a diffeomorphism away from the lines
$$L_n := \left\{ \left(x, \frac{n\pi}{2} \right) \ : \ x \in \mathbb{R} \right\}.$$

4.2 Definition of Surface

We give our main definition of surface in $\mathbb{R}^3$.

Definition 4.40: Surface

Let $\mathcal{S} \subseteq \mathbb{R}^3$ be a connected set. We say that $\mathcal{S}$ is a **surface** if for every point $\mathbf{p} \in \mathcal{S}$ there exist an open set $U \subseteq \mathbb{R}^2$ and a smooth map

$$\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S}$$

such that

- $\mathbf{p} \in \boldsymbol{\sigma}(U)$
- $\boldsymbol{\sigma}(U)$ is open in $\mathcal{S}$
- $\boldsymbol{\sigma}$ is a homeomorphism between U and $\boldsymbol{\sigma}(U)$

The homeomorphism $\boldsymbol{\sigma}$ is called a **surface chart** at $\mathbf{p}$.

Remark 4.41

- A surface chart $\boldsymbol{\sigma}$ is a map
$$\boldsymbol{\sigma} : U \to \mathbb{R}^3 ,$$
with $U \subseteq \mathbb{R}^2$ open. Therefore smoothness of $\boldsymbol{\sigma}$ is intended in the classical sense.

- Given a chart $\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U)$, the set U is open in $\mathbb{R}^2$ while $\boldsymbol{\sigma}(U)$ is open in $\mathcal{S}$ with the subspace topology. This means that there exists $W \subseteq \mathbb{R}^3$ open such that
$$\boldsymbol{\sigma}(U) = W \cap \mathcal{S} .$$

- The omeomorphism condition is saying that $\boldsymbol{\sigma}(U) \subseteq \mathcal{S}$ looks locally (around $\mathbf{p}$) like an open set $U \subseteq \mathbb{R}^2$.

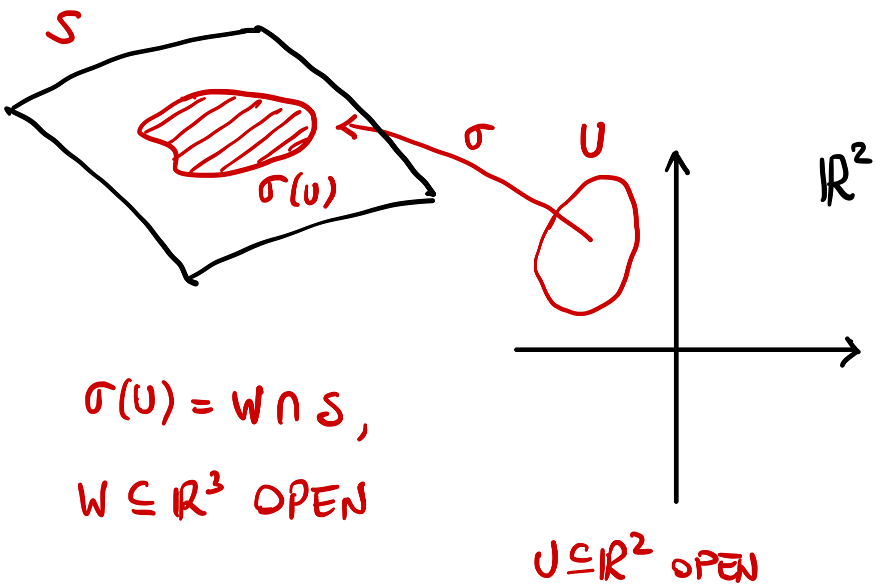


Figure 4.5: Sketch of the surface $\mathcal{S}$ and chart $\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S}$. The set $U \subseteq \mathbb{R}^2$ is open in $\mathbb{R}^2$ and $\boldsymbol{\sigma}(U)$ is open in $\mathcal{S}$. This means there exists W open in $\mathbb{R}^3$ such that $\boldsymbol{\sigma}(U) = \mathcal{S} \cap W$.

Notation

- Points in U will be denoted with the pair (u, v).
- Partial derivatives of a chart $\boldsymbol{\sigma} = \boldsymbol{\sigma}(u, v)$ will be denoted by

$$\boldsymbol{\sigma}_u := \frac{\partial \boldsymbol{\sigma}}{\partial u}, \quad \boldsymbol{\sigma}_v := \frac{\partial \boldsymbol{\sigma}}{\partial v}.$$

  Similar notations are adopted for higher order derivatives, e.g.,

$$\begin{aligned} \boldsymbol{\sigma}_{uu} &:= \frac{\partial^2 \boldsymbol{\sigma}}{\partial u^2}, & \boldsymbol{\sigma}_{uv} &:= \frac{\partial^2 \boldsymbol{\sigma}}{\partial u \partial v}, \\ \boldsymbol{\sigma}_{vu} &:= \frac{\partial^2 \boldsymbol{\sigma}}{\partial v \partial u}, & \boldsymbol{\sigma}_{vv} &:= \frac{\partial^2 \boldsymbol{\sigma}}{\partial v^2}, \end{aligned}$$

- Components of $\boldsymbol{\sigma}$ will be denoted by

$$\boldsymbol{\sigma} = (\sigma^1, \sigma^2, \sigma^3).$$

Definition 4.42: Atlas of a surface

Let $\mathcal{S}$ be a surface. Assume there exists a collection of **charts**

$$\mathscr{A} = \{\boldsymbol{\sigma}_i\}_{i\in I} \,, \qquad \boldsymbol{\sigma}_i \,:\, U_i \to \boldsymbol{\sigma}(U_i) \subseteq \mathcal{S} \,,$$

where I is a suitable family of indices. The family $\mathscr{A}$ is an **atlas** of $\mathcal{S}$ if

$$\mathcal{S} = \bigcup_{i\in I} \boldsymbol{\sigma}_i(U_i) \,.$$

Example 4.43: 2D Plane in $\mathbb{R}^3$

Planes in $\mathbb{R}^3$ are surfaces with atlas made by one chart. This is because a plane $\pi \subseteq \mathbb{R}^3$ is described by the equation

$$\pi = \{\mathbf{x} \in \mathbb{R}^3 \,:\, \mathbf{x} \cdot \mathbf{w} = \lambda\} \,.$$

Let

- $\mathbf{p}, \mathbf{q} \in \mathbb{R}^3$ be linearly independent, and orthogonal to $\mathbf{w}$.
- $\mathbf{a} \in \pi$ be any point in the plane.

If $\mathbf{x} \in \pi$ then $\mathbf{x} - \mathbf{a}$ is parallel to the plane. In particular $\mathbf{x} - \mathbf{a}$ can be written as linear combination of the vectors $\mathbf{p}$ and $\mathbf{q}$. Hence π can be equivalently represented as

$$\pi = \{\mathbf{a} + u\mathbf{p} + v\mathbf{q} \,:\, u, v \in \mathbb{R}\} \,.$$

Define the map

$$\boldsymbol{\sigma} \,:\, \mathbb{R}^2 \to \pi \,, \quad \boldsymbol{\sigma}(u,v) := \mathbf{a} + u\mathbf{p} + v\mathbf{q} \,.$$

We have:

- $\boldsymbol{\sigma}$ is smooth.
- $\mathbb{R}^2$ is obviously open.
- $\boldsymbol{\sigma}(\mathbb{R}^2)$ is open in π, since $\boldsymbol{\sigma}(\mathbb{R}^2) = \pi$.
- The inverse of $\boldsymbol{\sigma}$ is

$$\boldsymbol{\sigma}^{-1} \,:\, \pi \to \mathbb{R}^2 \,, \quad \boldsymbol{\sigma}^{-1}(\mathbf{x}) = ((\mathbf{x} - \mathbf{a}) \cdot \mathbf{p}, (\mathbf{x} - \mathbf{a}) \cdot \mathbf{q}) \,.$$

- As $\boldsymbol{\sigma}^{-1}$ is continuous, then $\boldsymbol{\sigma}$ is a homeomorphism between $\mathbb{R}^2$ and π.

Therefore $\boldsymbol{\sigma}$ is a chart for π. Since

$$\boldsymbol{\sigma}(\mathbb{R}^2) = \pi \,,$$

we have that $\{\boldsymbol{\sigma}\}$ is an atlas for π, and hence π is a surface.

$$\sigma(u,v) = \underline{a} + u\,\underline{p} + v\,\underline{q}$$

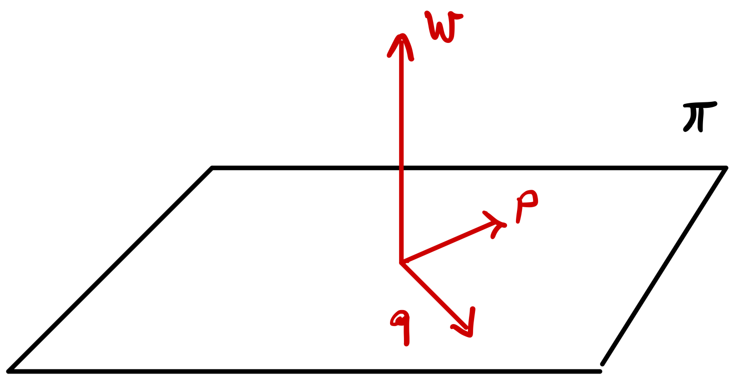


Figure 4.6: A plane π is a surface with atlas containing a single chart $\boldsymbol{\sigma} : \mathbb{R}^2 \to \pi$.

Example 4.44: Unit cylinder

Consider the infinite unit cylinder

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 = 1\}.$$

$\mathcal{S}$ is a surface with an atlas consisting of two charts:

$$\boldsymbol{\sigma}_i : U_i \to \mathbb{R}^3 , \quad \boldsymbol{\sigma}_i(u, v) := (\cos(u), \sin(u), v)$$

for $i = 1, 2$, where

$$U_1 := \left(0, \frac{3\pi}{2}\right) \times \mathbb{R} , \quad U_2 := \left(\pi, \frac{5\pi}{2}\right) \times \mathbb{R} .$$

Indeed:

- $\boldsymbol{\sigma}_i$ is smooth.
- U_i is clearly open in $\mathbb{R}^2$.
- One can check that $\boldsymbol{\sigma}_i(U_i)$ is open in $\mathcal{S}$.
- $\boldsymbol{\sigma}_i$ is a homeomorphism of U_i in $\boldsymbol{\sigma}(U_i)$.
- $\{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$ is an atlas for $\mathcal{S}$, since

$$\mathcal{S} = \boldsymbol{\sigma}_1(U_1) \cup \boldsymbol{\sigma}_2(U_2) .$$

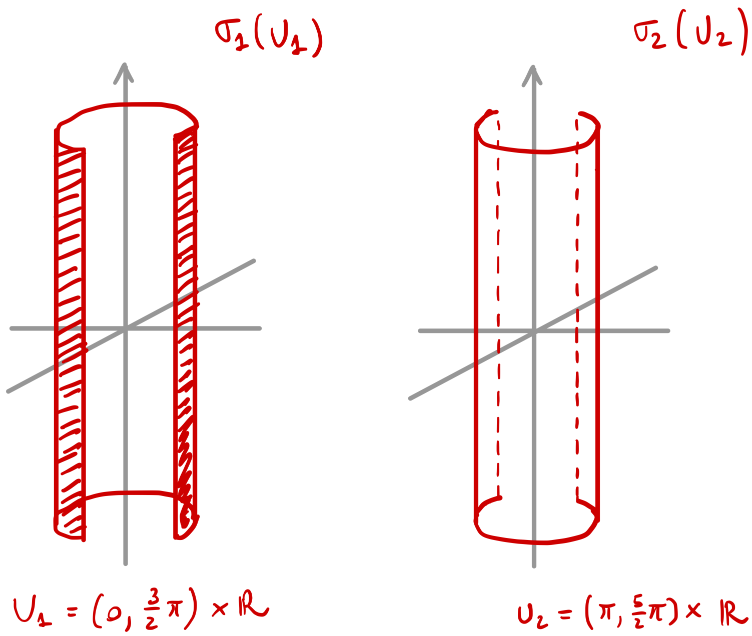


Figure 4.7: Unit cylinder $\mathcal{S}$ is a surface with atlas $\mathscr{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$. Depicted are the images $\boldsymbol{\sigma}_1(U_1)$ and $\boldsymbol{\sigma}_2(U_2)$.

Important

Consider again the unit cylinder

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 = 1\}.$$

Define the map

$$\boldsymbol{\sigma} : U \to \mathbb{R}^3 , \quad \boldsymbol{\sigma}(u, v) := (\cos(u), \sin(u), v)$$

where

$$U := [0, 2\pi] \times \mathbb{R}.$$

Clearly we have

$$\boldsymbol{\sigma}(U) = \mathcal{S} .$$

However $\{\boldsymbol{\sigma}\}$ is not an atlas for $\mathcal{S}$, since $\boldsymbol{\sigma}$ is not a chart. This is because $\boldsymbol{\sigma}$ is not invertible, as for example

$$\boldsymbol{\sigma}(0, 0) = \boldsymbol{\sigma}(2\pi, 0) .$$

Therefore $\boldsymbol{\sigma}$ cannot be an omeomorphism between U and $\mathcal{S}$.

Example 4.45: Graph of a function

Let $U \subseteq \mathbb{R}^2$ be open and $f \colon U \to \mathbb{R}$ be smooth. The graph of f is the set

$$\Gamma_f := \{(u,v,f(u,v)) \,:\, (u,v) \in U\}\,.$$

We have that Γ_f is a surface with atlas given by

$$\mathscr{A} = \{\boldsymbol{\sigma}\}$$

where $\boldsymbol{\sigma} \colon U \to \Gamma_f$ is

$$\boldsymbol{\sigma}(u,v) := (u,v,f(u,v))\,.$$

Let us check that Γ_f is a surface:

- $\boldsymbol{\sigma}$ is smooth since f is smooth.
- U is open in $\mathbb{R}^2$ by assumption.
- $\boldsymbol{\sigma}(U) = \Gamma_f$, and therefore $\boldsymbol{\sigma}(U)$ is open in Γ_f.
- The inverse of $\boldsymbol{\sigma}$ is given by $\tilde{\boldsymbol{\sigma}} \colon \Gamma_f \to U$ defined as

  $$\tilde{\boldsymbol{\sigma}}(u,v,f(u,v)) := (u,v)\,.$$

  Clearly $\tilde{\boldsymbol{\sigma}}$ is continuous.
- Therefore $\boldsymbol{\sigma}$ is a homeomorphism of U into Γ_f.
- $\mathscr{A} = \{\boldsymbol{\sigma}\}$ is an atlas for Γ_f, since

  $$\Gamma_f = \boldsymbol{\sigma}(U)\,.$$

Let us conclude the section with an example of a set which is not a surface.

Example 4.46: Circular cone

Consider the circular cone

$$\mathscr{S} := \{(x,y,z) \in \mathbb{R}^3 \,:\, x^2 + y^2 = z^2\}\,.$$

Then $\mathscr{S}$ is not a surface. This is essentially consequence of the fact that

$$\mathscr{S} \setminus \{\mathbf{0}\}$$

is a disconnected set.
To see that $\mathscr{S}$ is not a surface, suppose there exists an atlas $\{\boldsymbol{\sigma}_i\}$ of $\mathscr{S}$

$$\boldsymbol{\sigma}_i \colon U_i \to \boldsymbol{\sigma}_i(U_i) \subseteq \mathscr{S}\,.$$

In particular there exists a chart $\boldsymbol{\sigma}$ such that

$$\mathbf{0} \in \boldsymbol{\sigma}(U)\,.$$

Let $\mathbf{x}_0 \in U$ be the point such that

$$\boldsymbol{\sigma}(\mathbf{x}_0) = \mathbf{0}\,.$$

Since U is open in $\mathbb{R}^2$, there exists $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{x}_0) \subseteq U$. Since $\boldsymbol{\sigma}$ is a homeomorphism, we deduce that

$$\boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0))$$

is open in $\mathcal{S}$. Hence there exists an open set W in $\mathbb{R}^3$ such that

$$\boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0)) = \boldsymbol{\sigma}(U) \cap W \,.$$

As $\mathbf{0} \in \boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0))$, we conclude that $\mathbf{0} \in W$. Since W is open in $\mathbb{R}^3$, there exists $\delta > 0$ such that

$$B_\delta(\mathbf{0}) \subseteq W \,.$$

In particular we deduce that

$$B_\delta(\mathbf{0}) \cap \boldsymbol{\sigma}(U) \subseteq \boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0)) \,.$$

Hence $\boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0))$ contains points of both $\mathcal{S}^-$ and $\mathcal{S}^+$, with

$$\mathcal{S}^- := \mathcal{S} \cap \{z < 0\} \,, \quad \mathcal{S}^+ := \mathcal{S} \cap \{z > 0\} \,.$$

This implies that

$$V := \boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0)) \smallsetminus \{\mathbf{0}\}$$

is disconnected, with disconnection given by

$$V = (V \cap \mathcal{S}^-) \cup (V \cap \mathcal{S}^+) \,.$$

However V is homeomorphic to

$$B_\varepsilon(\mathbf{x}_0) \smallsetminus \{\mathbf{x}_0\} \,,$$

which is instead connected. Contradiction. Hence $\mathcal{S}$ is not a surface.

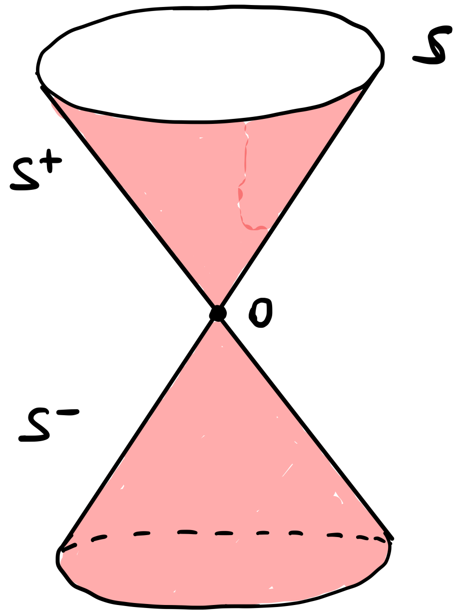


Figure 4.8: The circular cone is not a surface. This is because $\mathcal{S} \setminus \{\mathbf{0}\}$ is disconnected.

4.3 Regular Surfaces

We have defined a regular curve to be a map $\boldsymbol{\gamma} \colon (a, b) \to \mathbb{R}^n$ such that

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

This allowed us to define tangent vectors and, eventually, Frenet frame.

We want to do something similar for surfaces: We look for a condition that eventually will allow us to define tangent planes. This is why we introduce **regular charts** and **regular surfaces**.

Definition 4.47: Regular Chart

Let $U \subseteq \mathbb{R}^2$ be open. A map

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}(u, v) \colon U \to \mathbb{R}^3$$

is called a **regular chart** 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 $(u, v) \in U$.

The following gives more insight into the regularity condition.

Proposition 4.48

Let $U \subseteq \mathbb{R}^2$ be open and consider a map

$$\boldsymbol{\sigma} \colon U \to \mathbb{R}^3 .$$

They are equivalent:

1. $\boldsymbol{\sigma}$ is a regular chart.
2. The differential $d\boldsymbol{\sigma}_{\mathbf{x}} \colon \mathbb{R}^2 \to \mathbb{R}^3$ is injective for all $\mathbf{x} \in U$.
3. The Jacobian matrix
$$J\boldsymbol{\sigma}(u,v) = \begin{pmatrix} \sigma_u^1 & \sigma_v^1 \\ \sigma_u^2 & \sigma_v^2 \\ \sigma_u^3 & \sigma_v^3 \end{pmatrix}$$
has rank 2 for all $(u,v) \in U$.
4. It holds
$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq 0 \quad \forall\, (u,v) \in U .$$

Proof

Part 1. Equivalence of Point 1 and Point 4.
By the properties of vector product, we have that

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq 0 \quad \forall (u,v) \in U$$

if and only if $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent for all $(u,v) \in U$.
Part 2. Equivalence of Point 2 and Point 3.
The differential $d\boldsymbol{\sigma}_{\mathbf{x}} \colon \mathbb{R}^2 \to \mathbb{R}^3$ is represented in matrix form by the Jacobian

$$J\boldsymbol{\sigma}(u,v) = \begin{pmatrix} \sigma_u^1 & \sigma_v^1 \\ \sigma_u^2 & \sigma_v^2 \\ \sigma_u^3 & \sigma_v^3 \end{pmatrix}$$

By standard linear algebra results, $J\boldsymbol{\sigma}$ has rank 2 if and only if $d\boldsymbol{\sigma}$ is injective.
Part 3. Equivalence of Point 1 and Point 3.
A 3×2 matrix has rank 2 if and only if its columns are linearly independent. Since the columns of $J\boldsymbol{\sigma}$ are $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$, we conclude that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent.

We are now ready to define regular surfaces.

Definition 4.49: Regular surface

Let $\mathcal{S}$ be a surface. Let

$$\mathcal{A} = \{\boldsymbol{\sigma}_i\}_{i \in I} ,$$

be an atlas for $\mathcal{S}$. We say that:

- $\mathcal{A}$ is a **regular atlas** if the map $\boldsymbol{\sigma}_i$ is a regular chart for all $i \in I$.

- $\mathcal{S}$ is a **regular surface** if there exists a regular atlas for $\mathcal{S}$.

Example 4.50: 2D Plane in $\mathbb{R}^3$

Let $\mathbf{a}, \mathbf{p}, \mathbf{q} \in \mathbb{R}^3$, with $\mathbf{p}$ and $\mathbf{q}$ orthogonal. We have shown that the plane

$$\pi = \{\mathbf{a} + u\mathbf{p} + v\mathbf{q} \ : \ u, v \in \mathbb{R}\}$$

is a surface with atlas $\mathscr{A} = \{\boldsymbol{\sigma}\}$, where

$$\boldsymbol{\sigma} : \mathbb{R}^2 \to \pi \,, \quad \boldsymbol{\sigma}(u,v) := \mathbf{a} + u\mathbf{p} + v\mathbf{q} \,.$$

Then π is a regular surface, because $\boldsymbol{\sigma}$ is a regular chart. To see this, compute

$$\boldsymbol{\sigma}_u = \mathbf{p} \,, \quad \boldsymbol{\sigma}_v = \mathbf{q} \,.$$

Since $\mathbf{p}$ and $\mathbf{q}$ are orthogonal, then they are linearly independent. Thus $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent, and $\boldsymbol{\sigma}$ is a regular chart.

Example 4.51: Unit cylinder

Consider the infinite unit cylinder

$$\mathcal{S} = \{(x,y,z) \in \mathbb{R}^3 \ : \ x^2 + y^2 = 1\} \,.$$

We have seen that $\mathcal{S}$ is a surface with atlas $\mathscr{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$ where we define

$$\boldsymbol{\sigma} : \mathbb{R}^2 \to \mathbb{R}^3 \,, \quad \boldsymbol{\sigma}(u,v) := (\cos(u), \sin(u), v)$$

and

$$\boldsymbol{\sigma}_1 := \boldsymbol{\sigma}|_{U_1} \,, \qquad \boldsymbol{\sigma}_2 := \boldsymbol{\sigma}|_{U_2} \,,$$
$$U_1 := \left(0, \frac{3\pi}{2}\right) \times \mathbb{R} \,, \qquad U_2 := \left(\pi, \frac{5\pi}{2}\right) \times \mathbb{R} \,.$$

We have that $\mathcal{S}$ is a regular surface, since the atlas $\mathscr{A}$ is regular. Indeed:

$$\boldsymbol{\sigma}_u = (-\sin(u), \cos(u), 0) \,, \quad \boldsymbol{\sigma}_v = (0,0,1) \,,$$

and therefore

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (\cos(u), \sin(u), 0) \,, \quad \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = 1 \,.$$

This implies

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq 0 \,, \quad \forall \, (u,v) \in \mathbb{R}^2 \,,$$

showing that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent. Therefore $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are regular charts, being restrictions of $\boldsymbol{\sigma}$.

Example 4.52: Graph of a function

Let $U \subseteq \mathbb{R}^2$ be open and $f : U \to \mathbb{R}$ be smooth. The graph of f is the set

$$\Gamma_f := \{(u, v, f(u, v)) : (u, v) \in U\}.$$

We have seen that Γ_f is surface with atlas given by $\mathscr{A} = \{\boldsymbol{\sigma}\}$, where $\boldsymbol{\sigma} : U \to \Gamma_f$ is

$$\boldsymbol{\sigma}(u, v) := (u, v, f(u, v)).$$

We have that Γ_f is regular, since $\mathscr{A}$ is a regular atlas. Indeed,

$$\boldsymbol{\sigma}_u = (1, 0, f_u), \quad \boldsymbol{\sigma}_v = (0, 1, f_v),$$

and so

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (-f_u, -f_v, 1) \neq \mathbf{0},$$

since the last component never vanishes. Therefore $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent and $\boldsymbol{\sigma}$ is a regular chart.

Example 4.53: Unit sphere

Consider the unit sphere in $\mathbb{R}^3$

$$\mathbb{S}^2 := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

We have that $\mathbb{S}^2$ is a regular surface, with regular atlas

$$\mathscr{A} = \{\boldsymbol{\sigma}_i\}_{i=1}^6,$$

defined as follows: Let

$$U := \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 < 1\}$$

be the unit open ball in $\mathbb{R}^2$ and define $\boldsymbol{\sigma}_i : U \to \mathbb{R}^3$ by

$$\begin{aligned}
\boldsymbol{\sigma}_1(u, v) &= \left(u, v, \sqrt{1 - u^2 - v^2}\right) \\
\boldsymbol{\sigma}_2(u, v) &= \left(u, v, -\sqrt{1 - u^2 - v^2}\right) \\
\boldsymbol{\sigma}_3(u, v) &= \left(u, \sqrt{1 - u^2 - v^2}, v\right) \\
\boldsymbol{\sigma}_4(u, v) &= \left(u, -\sqrt{1 - u^2 - v^2}, v\right) \\
\boldsymbol{\sigma}_5(u, v) &= \left(\sqrt{1 - u^2 - v^2}, u, v\right) \\
\boldsymbol{\sigma}_6(u, v) &= \left(-\sqrt{1 - u^2 - v^2}, u, v, \right)
\end{aligned}$$

Exercise: Check that $\mathbb{S}^2$ is a regular surface.

Remark 4.54: Spherical coordinates

The equivalent of polar coordinates in dimension 3 are spherical coordinates. A point $(x, y, z) \in \mathbb{R}^3 \smallsetminus \{\mathbf{0}\}$ can be represented in spherical coordinates by

$$\begin{aligned} x &= \rho \cos(\theta) \cos(\phi) \\ y &= \rho \cos(\theta) \sin(\phi) \\ z &= \rho \sin(\theta) \end{aligned}$$

where

$$\rho := \sqrt{x^2 + y^2 + z^2}\,, \quad \phi \in [0, 2\pi]\,, \quad \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\,,$$

with the angles ϕ and θ as in Figure Figure 4.9.

It is clear that $z = \rho \sin(\theta)$, by basic trigonometry. To compute x and y, we note that the segment joining $\mathbf{0}$ to $\mathbf{p}$ has length

$$L = \rho \cos \theta\,.$$

Therefore we get

$$\begin{aligned} x &= L \cos(\phi) = \rho \cos(\theta) \cos(\phi) \\ y &= L \sin(\phi) = \rho \cos(\theta) \sin(\phi) \end{aligned}$$

concluding.

Example 4.55: Unit sphere in spherical coordinates

Consider again the unit sphere in $\mathbb{R}^3$

$$\mathbb{S}^2 := \{(x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 + z^2 = 1\}\,.$$

We want to give an alternative atlas for $\mathbb{S}^2$ based on spherical coordinates. To this end, define

$$U := \left\{(\theta, \phi) \in \mathbb{R}^2 \,:\, -\frac{\pi}{2} < \theta < \frac{\pi}{2}\,,\ 0 < \phi < 2\pi\right\}$$

and $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ by

$$\boldsymbol{\sigma}(\theta, \phi) := (\cos(\theta) \cos(\phi), \cos(\theta) \sin(\phi), \sin(\theta))\,.$$

We have:

- $\boldsymbol{\sigma}$ is smooth.
- U is open in $\mathbb{R}^2$.
- Moreover
$$\boldsymbol{\sigma}(U) = \mathbb{S}^2 \smallsetminus \{(x, 0, z) \in \mathbb{R}^3 \,:\, x \geq 0\}\,,$$
as seen also in the left picture in Figure 4.10.

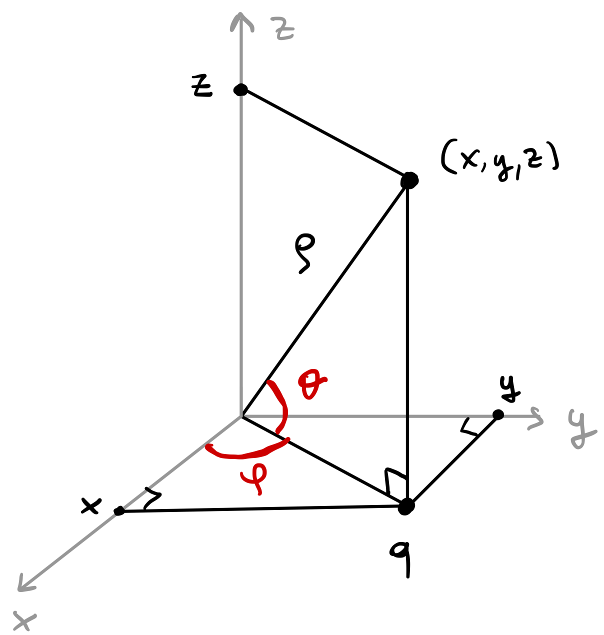


Figure 4.9: Spherical coordinates in $\mathbb{R}^3$.

- The set $\boldsymbol{\sigma}(U)$ is evidently open in $\mathbb{S}^2$.
- It is easy to check that $\boldsymbol{\sigma}$ is invertible, with continuous inverse.
- Thus $\boldsymbol{\sigma}$ is a homeomorphism from U into $\boldsymbol{\sigma}(U)$.

Let us check that $\boldsymbol{\sigma}$ is a regular chart:

$$
\begin{aligned}
\boldsymbol{\sigma}_\theta &= (-\sin(\theta)\cos(\phi), -\sin(\theta)\sin(\phi), \cos(\theta)) \\
\boldsymbol{\sigma}_\phi &= (-\cos(\theta)\sin(\phi), \cos(\theta)\cos(\phi), 0)\,.
\end{aligned}
$$

Therefore

$$
\boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\phi = (-\cos^2(\theta)\cos(\phi), -\cos^2(\theta)\sin(\phi), -\sin(\theta)\cos(\theta))\,,
$$

from which

$$
\left\| \boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\phi \right\| = |\cos(\theta)|\,.
$$

Since $(\theta, \phi) \in U$, we have $\theta \in (-\pi/2, \pi/2)$, and so

$$
\left\| \boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\phi \right\| = |\cos(\theta)| \neq 0\,,
$$

showing that $\boldsymbol{\sigma}_\theta$ and $\boldsymbol{\sigma}_\phi$ are linearly independent, and $\boldsymbol{\sigma}$ is regular.
Since $\boldsymbol{\sigma}(U) \neq \mathbb{S}^2$, the chart $\boldsymbol{\sigma}$ does not form an atlas. We need a second chart. An option is to define $\tilde{\boldsymbol{\sigma}} : U \to \mathbb{R}^3$ by

$$\tilde{\boldsymbol{\sigma}} := (-\cos(\theta)\cos(\phi), -\sin(\theta), -\cos(\theta)\sin(\phi)) .$$

Notice that $\tilde{\boldsymbol{\sigma}}$ is obtained by rotating $\boldsymbol{\sigma}$ by π about the z-axis and by $\pi/2$ about the y-axis, as seen in the right picture in Figure 4.10. It is an exercise to check that $\tilde{\boldsymbol{\sigma}}$ is a regular chart.
Since we have

$$\tilde{\boldsymbol{\sigma}}(U) = \mathbb{S}^2 \setminus \{(x, y, 0) \in \mathbb{R}^3 \ : \ x \leq 0\} ,$$

it is immediate to see that

$$\mathbb{S}^2 = \boldsymbol{\sigma}(U) \cup \tilde{\boldsymbol{\sigma}}(U) .$$

Hence

$$\mathscr{A} := \{\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}\}$$

is a regular atlas for $\mathbb{S}^2$.

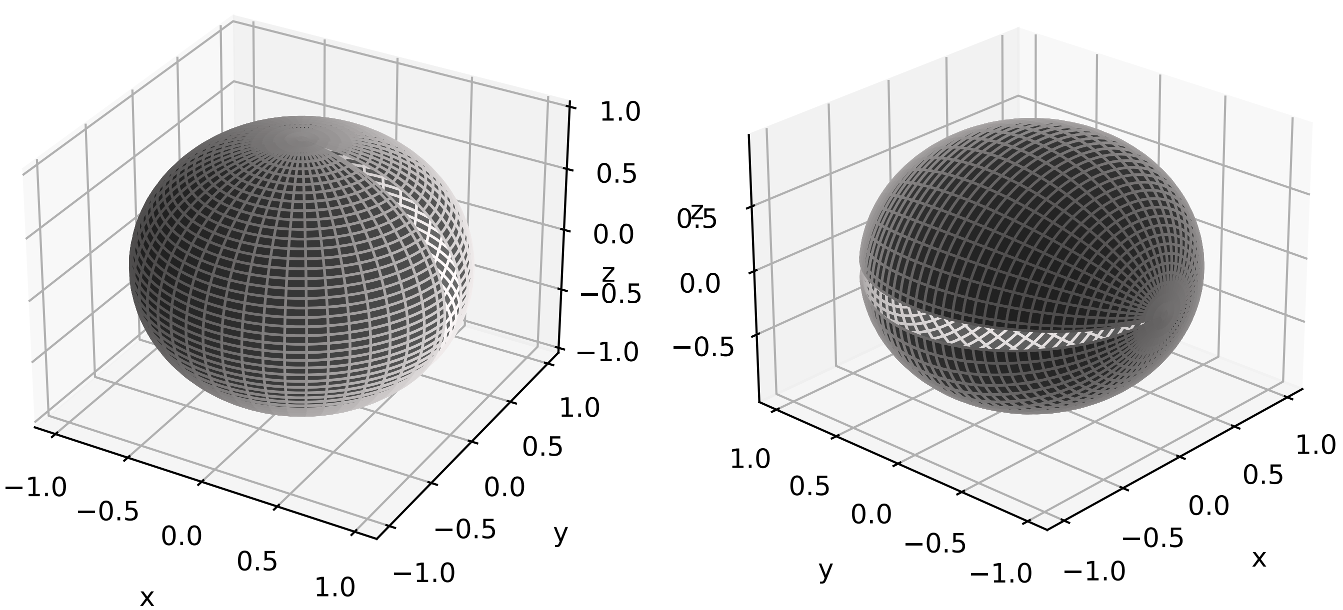


Figure 4.10: Image of the charts of the sphere from the above example.

Let us make an example of a non-regular surface.

Example 4.56

The surface parametrized by

$$\boldsymbol{\sigma}(u, v) = (u, v^2, v^3) , \quad \forall (u, v) \in \mathbb{R}^2$$

is not regular. This is because

$$\boldsymbol{\sigma}_u = (1,0,0)\,, \quad \boldsymbol{\sigma}_v = (0,2v,3v^2)$$

and therefore

$$\boldsymbol{\sigma}_v(u,0) = (0,0,0)\,,$$

showing that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly dependent along the line

$$L = \{(u,0) \,:\, u \in \mathbb{R}\}\,.$$

Hence $\boldsymbol{\sigma}$ is not a regular chart.
Looking at Figure Figure 4.11, it is clear that $\mathcal{S}$ is not regular, since $\mathcal{S}$ has a cusp along the line $\boldsymbol{\sigma}(L)$.

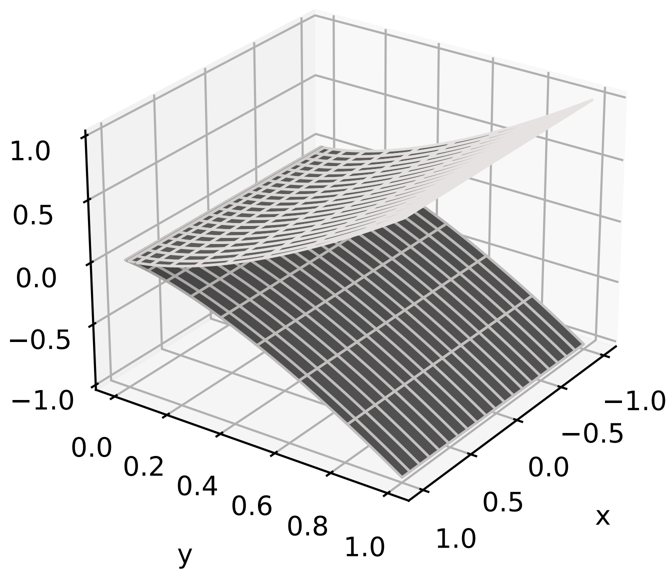


Figure 4.11: Example of non-regular surface.

4.4 Level surfaces

Definition 4.57: Level surface

Let $V \subseteq \mathbb{R}^3$ be an open set and $f \colon V \to \mathbb{R}$ be smooth. The **level surface** associated with f is the set

$$\mathcal{S}_f := f^{-1}(0) = \{(x,y,z) \in V \,:\, f(x,y,z) = 0\}\,.$$

We now give a result concerning regularity of level surfaces. The proof, rather technical, is based on the Implicit Function Theorem and can be found in Proposition 3.1.25 of [1]. We decide to omit it.

Theorem 4.58

Let $V \subseteq \mathbb{R}^3$ be an open set and $f : V \to \mathbb{R}$ be smooth. Consider the level surface

$$\mathcal{S}_f = \{(x, y, z) \in V \,:\, f(x, y, z) = 0\} .$$

Suppose that

$$\nabla f(x, y, z) \neq 0 \,, \quad \forall \, (x, y, z) \in V .$$

Then $\mathcal{S}_f$ is a regular surface.

Example 4.59

We want to determine if the set defined by the equation

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 = 1\}$$

is a regular surface. Note that $\mathcal{S}$ is a unit cylinder: From Example 4.51 we already know that $\mathcal{S}$ is a regular surface.
Let us prove that $\mathcal{S}$ is regular by using Theorem 4.58. To this end, define the open set

$$V := \mathbb{R}^3 \setminus \{(0, 0, z) \,:\, z \in \mathbb{R}\} .$$

Note that V is obtained by removing the z-axis from $\mathbb{R}^3$. Also define the function $f : \mathbb{R}^3 \to \mathbb{R}$ by

$$f(x, y, z) := x^2 + y^2 - 1 .$$

We have

$$\nabla f(x, y, z) = (2x, 2y, 0) \neq 0 \,, \quad \forall \, (x, y, z) \in V .$$

Since

$$\mathcal{S} = \mathcal{S}_f \,,$$

by Theorem 4.58 we conclude that $\mathcal{S}$ is a regular surface.

Example 4.60: Circular cone

We saw that the circular cone

$$\mathcal{S} := \{(x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 = z^2\} .$$

is not a surface. However the positive sheet

$$\mathcal{S}^+ := \{(x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 = z^2 \,, \, z > 0\} .$$

is a regular surface, see Figure 4.12 Indeed, define the open set

$$V := \{(x, y, z) \in \mathbb{R}^3 \,:\, z > 0\}$$

and the function $f : V \to \mathbb{R}$ by

$$f(x, y, z) := x^2 + y^2 - z^2 .$$

We have

$$\nabla f(x, y, z) = (2x, 2y, -2z) \neq 0 , \quad \forall (x, y, z) \in V .$$

Since

$$\mathcal{S}^+ = \mathcal{S}_f ,$$

by Theorem 4.58 we conclude that $\mathcal{S}$ is a regular surface.
As a side note, a regular atlas for $\mathcal{S}^+$ is given by $\mathcal{A} = \{\boldsymbol{\sigma}\}$ where $\boldsymbol{\sigma} : \mathbb{R}^2 \to \mathbb{R}^3$ is defined by

$$\boldsymbol{\sigma}(u, v) := (u, v, \sqrt{u^2 + v^2}) .$$

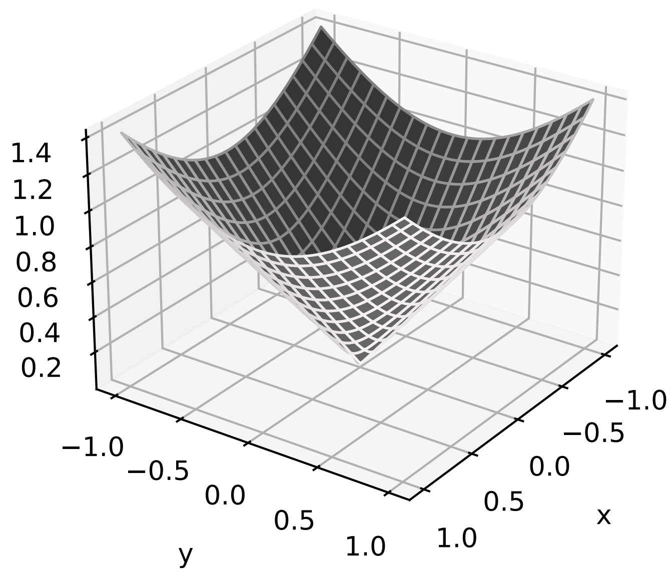


Figure 4.12: Positive sheet of circular cone.

4.5 Reparametrizations

We have defined the reparametrization of curves. In a similar way, one can reparametrize surface charts.

Definition 4.61

Suppose that $U, \widetilde{U} \subseteq \mathbb{R}^2$ are open sets and

$$\boldsymbol{\sigma} : U \to \mathbb{R}^3 , \quad \tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3 ,$$

are surface charts. We say that $\tilde{\boldsymbol{\sigma}}$ is a **reparametrization** of $\boldsymbol{\sigma}$ if there exists a diffeomorphism

$$\Phi : \widetilde{U} \to U \,,$$

such that

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi \,,$$

that is,

$$\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v}) = \boldsymbol{\sigma}(\Phi(\tilde{u}, \tilde{v})) \,, \quad \forall \ (\tilde{u}, \tilde{v}) \in \widetilde{U} \,.$$

We call Φ a **reparametrization map**.

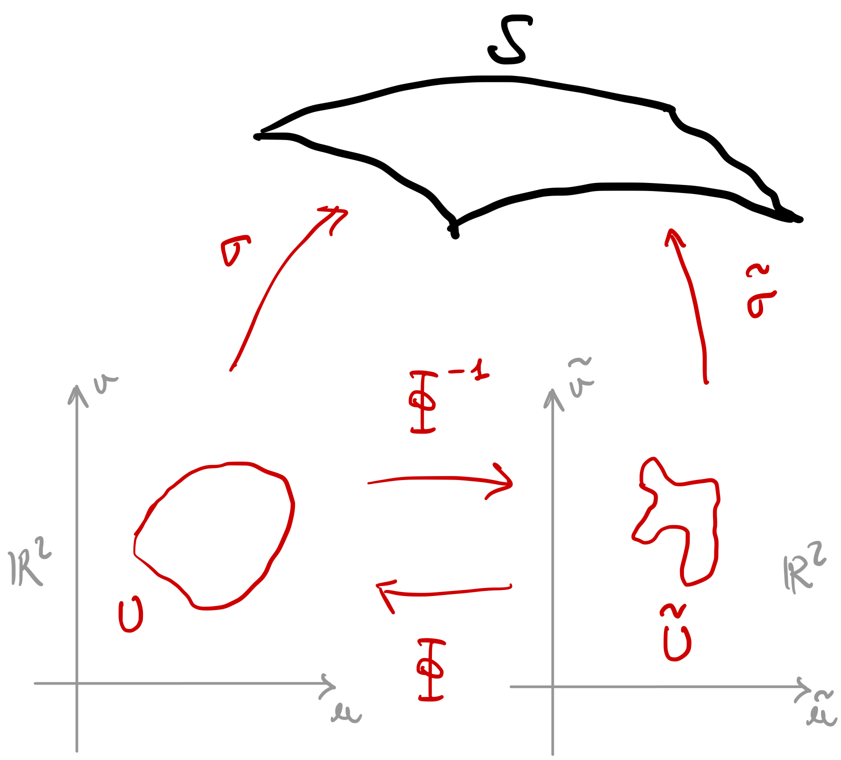


Figure 4.13: Schematic illustration of surface chart $\boldsymbol{\sigma}$ and reparametrization $\tilde{\boldsymbol{\sigma}}$.

We will show that reparametrizations of regular charts are regular. To prove this, first we need to recall the chain rule for multivariable functions.

Remark 4.62: Chain rule

Suppose that $U, \widetilde{U} \subseteq \mathbb{R}^2$ are open sets,

$$f : U \to \mathbb{R}^3$$

is smooth, and

$$\Phi : \widetilde{U} \to U$$

is a diffeomorphism. Define $\tilde{f} : \widetilde{U} \to \mathbb{R}^3$ by composition:

$$\tilde{f} := f \circ \Phi .$$

Explicitly, the above means

$$\tilde{f}(\tilde{u}, \tilde{v}) = f(\Phi(\tilde{u}, \tilde{v})) , \quad \forall \, (\tilde{u}, \tilde{v}) \in \widetilde{U} .$$

We denote the components of $f, \tilde{f}$ and Φ by

$$\tilde{f} = (\tilde{f}^1, \tilde{f}^2, \tilde{f}^3) , \quad f = (f^1, f^2, f^3) , \quad \Phi = (\Phi^1, \Phi^2) .$$

The Jacobians are

$$J\tilde{f} = \begin{pmatrix} \tilde{f}^1_{\tilde{u}} & \tilde{f}^1_{\tilde{v}} \\ \tilde{f}^2_{\tilde{u}} & \tilde{f}^2_{\tilde{v}} \\ \tilde{f}^3_{\tilde{u}} & \tilde{f}^3_{\tilde{v}} \end{pmatrix} , \quad Jf = \begin{pmatrix} f^1_u & f^1_v \\ f^2_u & f^2_v \\ f^3_u & f^3_v \end{pmatrix} , \quad J\Phi = \begin{pmatrix} \Phi^1_{\tilde{u}} & \Phi^1_{\tilde{v}} \\ \Phi^2_{\tilde{u}} & \Phi^2_{\tilde{v}} \end{pmatrix} .$$

The chain rule states that

$$J\tilde{f}(\tilde{u}, \tilde{v}) = Jf(\Phi(\tilde{u}, \tilde{v})) \, J\Phi(\tilde{u}, \tilde{v}) .$$

By expanding the above identity we obtain the chain rule in vectorial form

$$\begin{aligned} \tilde{f}_{\tilde{u}}(\tilde{u}, \tilde{v}) &= f_u(\Phi(\tilde{u}, \tilde{v}))\Phi^1_{\tilde{u}}(\tilde{u}, \tilde{v}) + f_v(\Phi(\tilde{u}, \tilde{v}))\Phi^2_{\tilde{u}}(\tilde{u}, \tilde{v}) \\ \tilde{f}_{\tilde{v}}(\tilde{u}, \tilde{v}) &= f_u(\Phi(\tilde{u}, \tilde{v}))\Phi^1_{\tilde{v}}(\tilde{u}, \tilde{v}) + f_v(\Phi(\tilde{u}, \tilde{v}))\Phi^2_{\tilde{v}}(\tilde{u}, \tilde{v}) \end{aligned}$$

The above is quite cumbersome. Hence we introduce compact notation for reparametrizations and chain rule. Specifically, we denote the components of the diffeomorphism Φ by

$$\begin{aligned} \Phi^1 \quad &\rightsquigarrow \quad (\tilde{u}, \tilde{v}) \mapsto u(\tilde{u}, \tilde{v}) \\ \Phi^2 \quad &\rightsquigarrow \quad (\tilde{u}, \tilde{v}) \mapsto v(\tilde{u}, \tilde{v}) \end{aligned}$$

Accordingly, the Jacobian of Φ is denoted as:

$$J\Phi = \begin{pmatrix} \Phi^1_{\tilde{u}} & \Phi^1_{\tilde{v}} \\ \Phi^2_{\tilde{u}} & \Phi^2_{\tilde{v}} \end{pmatrix} \quad \rightsquigarrow \quad \begin{pmatrix} \dfrac{\partial u}{\partial \tilde{u}} & \dfrac{\partial u}{\partial \tilde{v}} \\ \dfrac{\partial v}{\partial \tilde{u}} & \dfrac{\partial v}{\partial \tilde{v}} \end{pmatrix} .$$

Hence, the chain rule in vectorial form reads

$$\begin{aligned} \tilde{f}_{\tilde{u}} &= f_u \frac{\partial u}{\partial \tilde{u}} + f_v \frac{\partial v}{\partial \tilde{u}} \\ \tilde{f}_{\tilde{v}} &= f_u \frac{\partial u}{\partial \tilde{v}} + f_v \frac{\partial v}{\partial \tilde{v}} \end{aligned}$$

We will now prove that the reparametrization of a regular chart is regular.

Proposition 4.63

Suppose that $U, \widetilde{U} \subseteq \mathbb{R}^2$ are open sets and

$$\boldsymbol{\sigma} : U \to \mathbb{R}^3$$

is a regular chart. Assume given a diffeomorphism

$$\Phi : \widetilde{U} \to U .$$

The reparametrization $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3$ defined by

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi$$

is a regular chart, and it holds

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \det J\Phi \; (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)$$

Proof

Since $\boldsymbol{\sigma}$ is a regular chart we have that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent. Hence

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq 0 .$$

To see that $\tilde{\boldsymbol{\sigma}}$ is regular it is sufficient to prove that

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \neq 0 . \tag{4.1}$$

By chain rule we have

$$\begin{aligned} \tilde{\boldsymbol{\sigma}}_{\tilde{u}} &= \boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{u}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{u}} \\ \tilde{\boldsymbol{\sigma}}_{\tilde{v}} &= \boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{v}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{v}} \end{aligned}$$

By the properties of vector product we get

$$\begin{aligned} \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} &= \left(\boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{u}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{u}} \right) \times \left(\boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{v}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{v}} \right) \\ &= \frac{\partial u}{\partial \tilde{u}} \frac{\partial u}{\partial \tilde{v}} \, (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_u) + \frac{\partial u}{\partial \tilde{u}} \frac{\partial v}{\partial \tilde{v}} \, (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) \\ &+ \frac{\partial v}{\partial \tilde{u}} \frac{\partial u}{\partial \tilde{v}} \, (\boldsymbol{\sigma}_v \times \boldsymbol{\sigma}_u) + \frac{\partial v}{\partial \tilde{u}} \frac{\partial v}{\partial \tilde{v}} \, (\boldsymbol{\sigma}_v \times \boldsymbol{\sigma}_v) \\ &= \left(\frac{\partial u}{\partial \tilde{u}} \frac{\partial v}{\partial \tilde{v}} - \frac{\partial v}{\partial \tilde{u}} \frac{\partial u}{\partial \tilde{v}} \right) (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) \\ &= \det \begin{pmatrix} \frac{\partial u}{\partial \tilde{u}} & \frac{\partial u}{\partial \tilde{v}} \\ \frac{\partial v}{\partial \tilde{u}} & \frac{\partial v}{\partial \tilde{v}} \end{pmatrix} (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) \\ &= \det J\Phi \; (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v) . \end{aligned}$$

Since Φ is a diffeomorphism, we have that

$$\det J\Phi \neq 0 \,,$$

from which we conclude (4.1).

4.6 Transition maps

Consider the situation in which two regular charts have overlapping image.
It is natural to ask wether these maps are reparametrizations of each other on the overlapping region, see Figure 4.14. If such reparametrization exists, it is called a transition map.

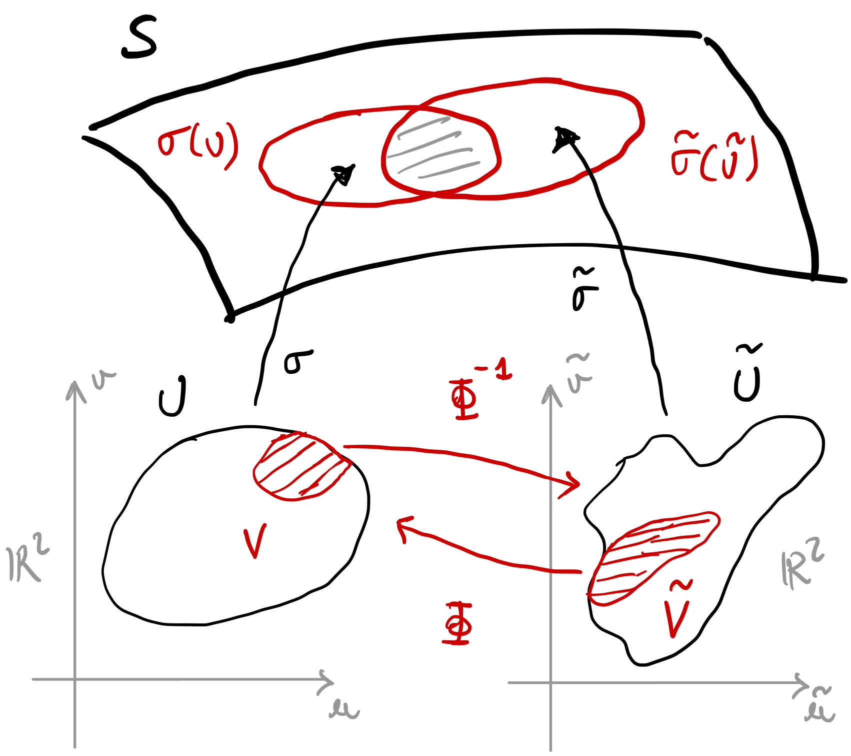


Figure 4.14: If the two regular charts $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have overlapping image, then they are reparametrization of each other, through a transition map Φ.

Definition 4.64: Transition map

Let $\mathcal{S}$ be a regular surface and

$$\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S} \,, \quad \tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \tilde{\boldsymbol{\sigma}}(\widetilde{U}) \subseteq \mathcal{S}$$

be regular charts. Assume that the images of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ overlap, that is,

$$I := \boldsymbol{\sigma}(U) \cap \tilde{\boldsymbol{\sigma}}(\widetilde{U}) \neq \emptyset .$$

The set I is open in $\mathcal{S}$, since it is intersection of open sets. Define the sets

$$V := \boldsymbol{\sigma}^{-1}(I) \subseteq U , \quad \widetilde{V} := \tilde{\boldsymbol{\sigma}}^{-1}(I) \subseteq \widetilde{U} ,$$

The sets V and $\widetilde{V}$ are open and by construction

$$\boldsymbol{\sigma}(V) = \tilde{\boldsymbol{\sigma}}(\widetilde{V}) = I .$$

Therefore they are well defined the restrictions

$$\boldsymbol{\sigma}|_V : V \to I , \quad \tilde{\boldsymbol{\sigma}}|_{\widetilde{V}} : \widetilde{V} \to I ,$$

which are homeomorphisms. The homeomorphism

$$\Phi : \widetilde{V} \to V , \quad \Phi := \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}}$$

is called a **transition map** from $\boldsymbol{\sigma}$ to $\tilde{\boldsymbol{\sigma}}$.

The theorem below states that transition maps between regular charts are diffeomorphisms. The proof is slightly technical and is based on the Implicit Function Theorem. We decide to omit it. The interested reader can find a proof at Page 117 of [6].

Theorem 4.65

Let $\mathcal{S}$ be a regular surface. The transition maps between regular charts are diffeomorphisms.

We can now use Theorem 4.65 to show that transition maps are reparametrizations.

Proposition 4.66

Let $\mathcal{S}$ be a regular surface and

$$\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S} , \quad \tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \tilde{\boldsymbol{\sigma}}(\widetilde{U}) \subseteq \mathcal{S}$$

be regular charts. Assume that the images of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ overlap, that is,

$$\boldsymbol{\sigma}(U) \cap \tilde{\boldsymbol{\sigma}}(\widetilde{U}) \neq \emptyset .$$

Then there exist open sets

$$V \subseteq U , \quad \widetilde{V} \subseteq \widetilde{U} ,$$

and a diffeomorphism

$$\Phi : \widetilde{V} \to V$$

such that $\tilde{\boldsymbol{\sigma}}|_{\widetilde{V}}$ is a reparametrization of $\boldsymbol{\sigma}|_V$, that is,

$$\tilde{\boldsymbol{\sigma}}|_{\widetilde{V}} = (\boldsymbol{\sigma}|_V) \circ \Phi \,.$$

Proof

Define

$$I := \boldsymbol{\sigma}(U) \cap \tilde{\boldsymbol{\sigma}}(\widetilde{U}) \neq \emptyset \,.$$

Note that this set is open in $\mathcal{S}$, being intersection of open sets. Set

$$V := \boldsymbol{\sigma}^{-1}(I) \,, \quad \widetilde{V} := \tilde{\boldsymbol{\sigma}}^{-1}(I) \,.$$

The sets V and $\widetilde{V}$ are open, since $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are homeomorphisms, and hence are continuous. By construction we have

$$\boldsymbol{\sigma}(V) = \tilde{\boldsymbol{\sigma}}(\widetilde{V}) = I \,.$$

Therefore they are well defined the restrictions

$$\boldsymbol{\sigma}|_V : V \to I \,, \quad \tilde{\boldsymbol{\sigma}}|_{\widetilde{V}} : \widetilde{V} \to I \,,$$

which are homeomorphisms. Consider the transition map

$$\Phi : \widetilde{V} \to V \,, \quad \Phi := \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}} \,.$$

By Theorem 4.65 we know that Φ is a diffeomorphism. Hence

$$\tilde{\boldsymbol{\sigma}}|_{\widetilde{V}} = (\boldsymbol{\sigma}|_V) \circ \Phi \,,$$

with Φ diffeomorphism, showing that $\tilde{\boldsymbol{\sigma}}|_{\widetilde{V}}$ is a reparametrization of $\boldsymbol{\sigma}|_V$.

Important

Proposition 4.66 allows us to define properties of surfaces using charts, as long as we check that the property in question does not depend on reparametrization, and hence on the choice of chart.

4.7 Functions between surfaces

We would like to define a concept of smooth function

$$f : \mathcal{S}_1 \to \mathcal{S}_2 \,,$$

where $\mathcal{S}_1$ and $\mathcal{S}_2$ are regular surfaces. So far, we only know how to define smooth functions from $\mathbb{R}^n$ into $\mathbb{R}^m$. The idea is to use surface charts to define smooth functions between surfaces.

Definition 4.67

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be regular surfaces and let

$$f : \mathcal{S}_1 \to \mathcal{S}_2$$

be a map. We say that:

- f is smooth at $\mathbf{p} \in \mathcal{S}_1$ if there exist charts $\boldsymbol{\sigma}_i : U_i \to \mathcal{S}_i$ for $i = 1, 2$ such that

  $$\mathbf{p} \in \boldsymbol{\sigma}_1(U_1), \quad f(\mathbf{p}) \in \boldsymbol{\sigma}_2(U_2)$$

  and

  $$(\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1) : U_1 \to U_2$$

  is smooth.
- f is smooth if it is smooth for each $\mathbf{p} \in \mathcal{S}_1$.
- f is a diffeomorphism if f is smooth and invertible, with smooth inverse.

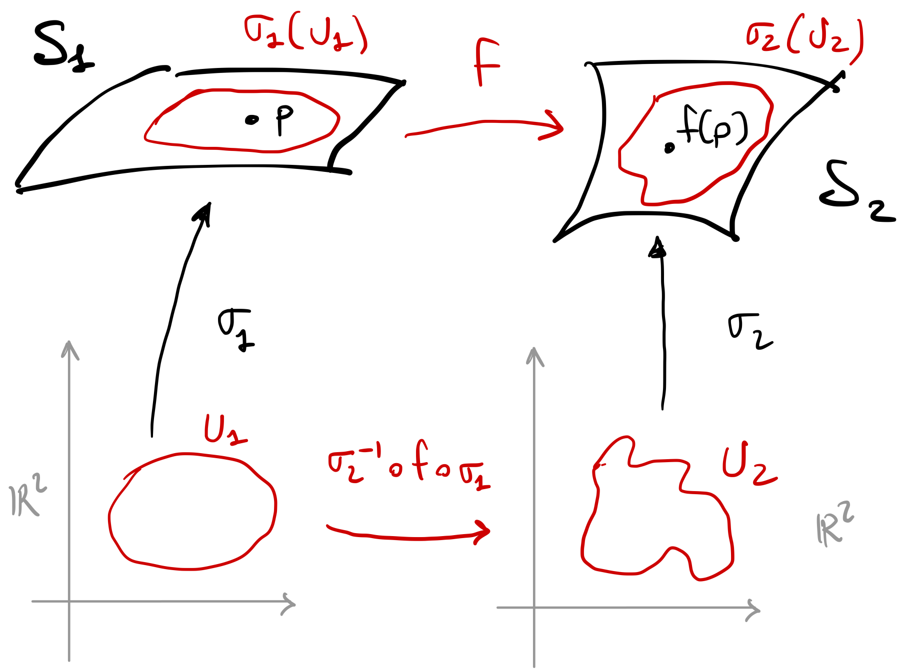


Figure 4.15: Sketch function f smooth at $\mathbf{p}$ between the surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$.

Remark 4.68

- Definition 4.67 makes sense because $\boldsymbol{\sigma}_2^{-1}$ exists.
- The map $\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1$ is only defined for $\mathbf{x} \in U_1$ such that

$$f(\boldsymbol{\sigma}_1(\mathbf{x})) \in \boldsymbol{\sigma}_2(U_2)\,.$$

- The function $\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1$ maps from $\mathbb{R}^2$ into $\mathbb{R}^2$, therefore differentiability is intended in the classical sense.
- Definition 4.67 does not depend on the choice of charts $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$

  Indeed, suppose that $\tilde{\boldsymbol{\sigma}}_i : \widetilde{U}_i \to \mathcal{S}_i$ are charts such that

$$\mathbf{p} \in \tilde{\boldsymbol{\sigma}}_1(\widetilde{U}_1)\,, \quad f(\mathbf{p}) \in \tilde{\boldsymbol{\sigma}}_2(\widetilde{U}_2)\,.$$

  In particular we have

$$\boldsymbol{\sigma}_i(U_i) \cap \tilde{\boldsymbol{\sigma}}_i(\widetilde{U}_i) \neq \emptyset\,.$$

  As $\mathcal{S}_1$ and $\mathcal{S}_2$ are regular surfaces, by Theorem 4.65 there exist open sets

$$V_i \subseteq U_i\,, \quad \widetilde{V}_i \subseteq \widetilde{U}_i\,,$$

  and transition maps

$$\Phi_i : \widetilde{V}_i \to V_i$$

  which are diffeomorphisms and satisfy

$$\tilde{\boldsymbol{\sigma}}_i = \boldsymbol{\sigma}_i \circ \Phi_i\,.$$

  Hence

$$\begin{aligned}
\tilde{\boldsymbol{\sigma}}_2^{-1} \circ f \circ \tilde{\boldsymbol{\sigma}}_1 &= \tilde{\boldsymbol{\sigma}}_2^{-1} \circ (\boldsymbol{\sigma}_2 \circ \boldsymbol{\sigma}_2^{-1}) \circ f \circ (\boldsymbol{\sigma}_1 \circ \boldsymbol{\sigma}_1^{-1}) \circ \tilde{\boldsymbol{\sigma}}_1 \\
&= (\tilde{\boldsymbol{\sigma}}_2^{-1} \circ \boldsymbol{\sigma}_2) \circ (\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1) \circ (\boldsymbol{\sigma}_1^{-1} \circ \tilde{\boldsymbol{\sigma}}_1) \\
&= \Phi_2^{-1} \circ (\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1) \circ \Phi_1^{-1}\,.
\end{aligned}$$

  Since Φ_i^{-1} and $\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1$ are smooth, we conclude that

$$\tilde{\boldsymbol{\sigma}}_2^{-1} \circ f \circ \tilde{\boldsymbol{\sigma}}_1$$

  is smooth. Hence Definition 4.67 does not depend on the choice of charts.

Proposition 4.69

If $f : \mathscr{S}_1 \to \mathscr{S}_2$ and $g : \mathscr{S}_2 \to \mathscr{S}_3$ are smooth maps (resp. diffeomorphisms) between surfaces, then the composition

$$(g \circ f) : \mathscr{S}_1 \to \mathscr{S}_3$$

is smooth (resp. a diffeomorphisms).

Proof

Fix $\mathbf{p} \in \mathscr{S}_1$ and choose charts

$$\boldsymbol{\sigma}_i : U_i \to \mathscr{S}_i$$

such that

$$\mathbf{p} \in \boldsymbol{\sigma}_1(U_1), \quad f(\mathbf{p}) \in \boldsymbol{\sigma}_2(U_2), \quad g(f(\mathbf{p})) \in \boldsymbol{\sigma}_3(U_3).$$

Since f and g are smooth we have that the maps

$$\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1, \quad \boldsymbol{\sigma}_3^{-1} \circ g \circ \boldsymbol{\sigma}_2,$$

are smooth. Hence

$$\boldsymbol{\sigma}_3^{-1} \circ (g \circ f) \circ \boldsymbol{\sigma}_1 = (\boldsymbol{\sigma}_3^{-1} \circ g \circ \boldsymbol{\sigma}_2) \circ (\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1)$$

is smooth, ending the proof.

Definition 4.70

Let $\mathscr{S}_1$ and $\mathscr{S}_2$ be regular surfaces. We say that $\mathscr{S}_1$ and $\mathscr{S}_2$ are diffeomorphic if there exists $f : \mathscr{S}_1 \to \mathscr{S}_2$ diffeomorphism.

The key ideas around diffeomorphisms are:

1. Two diffeomorphic surfaces are essentially the same.

   Indeed, it is immediate to show that being diffeomorphic is an equivalence relation on the set of regular surfaces.

2. Two diffeomorphic surfaces have essentially the same charts, as shown in the next proposition.

Proposition 4.71

Let $\mathscr{S}$ and $\widetilde{\mathscr{S}}$ be regular surfaces and $f : \mathscr{S} \to \widetilde{\mathscr{S}}$ be a diffeomorphism. If $\boldsymbol{\sigma} : U \to \mathscr{S}$ is a regular chart for $\mathscr{S}$ at $\mathbf{p}$, then

$$\tilde{\boldsymbol{\sigma}} := f \circ \boldsymbol{\sigma} : U \to \widetilde{\mathscr{S}}$$

is a regular chart for $\widetilde{\mathscr{S}}$ at $f(\mathbf{p})$.

Proof

Let $\boldsymbol{\sigma}_2 : U_2 \to \widetilde{\mathcal{S}}$ be a regular chart for $\widetilde{\mathcal{S}}$ at $f(\mathbf{p})$. By definition of diffeomorphism between surfaces, the map

$$\Phi := \boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma} : U \to U_2$$

is a diffeomorphism. Therfore

$$(f \circ \boldsymbol{\sigma})(u, v) = \boldsymbol{\sigma}_2 \left(\Phi(u, v)\right)$$

with Φ diffeomorphism, meaning that $f \circ \boldsymbol{\sigma}$ is a reparametrization of $\boldsymbol{\sigma}_2$. Since $\boldsymbol{\sigma}_2$ is regular, by Proposition 4.63 we deduce that $f \circ \boldsymbol{\sigma}$ is regular.

We conclude with the definition of local diffeomorphism between surfaces.

Definition 4.72: Local diffeomorphism

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be regular surfaces. A smooth map $f : \mathcal{S}_1 \to \mathcal{S}_2$ is called a **local diffeomorphism** if for each point $\mathbf{p} \in \mathcal{S}_1$ there exists an open set $V \subseteq \mathcal{S}_1$ with $\mathbf{p} \in V$, such that $f(V) \subseteq \mathcal{S}_2$ is open and

$$f : V \to f(V)$$

is a diffeomorphism between surfaces.

The above definition is well posed since open subsets of surfaces are themselves surfaces.

4.8 Tangent space

We have seen that tangent vectors to regular curves allow to define the Frenet Frame, curvature and torsion. Eventually, these quantities are sufficient to characterize a curve. It turns out that also regular surfaces admit tangent vectors. To avoid clumsy terminology, we make the following assumption.

Assumption 4.73

From now on, all the surfaces will be regular and all the charts will be regular.

Definition 4.74: Tangent vectors and tangent space

Let $\mathcal{S}$ be a surface and $\mathbf{p} \in \mathcal{S}$. A **tangent vector** to $\mathcal{S}$ at $\mathbf{p}$ is any vector $\mathbf{v} \in \mathbb{R}^3$ such that

$$\mathbf{v} = \dot{\boldsymbol{\gamma}}(0) \,,$$

where $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ is a smooth curve such that

$$\boldsymbol{\gamma}(-\varepsilon, \varepsilon) \subseteq \mathcal{S} \,, \quad \boldsymbol{\gamma}(0) = \mathbf{p} \,,$$

where $\varepsilon > 0$. The **tangent space** of $\mathcal{S}$ at $\mathbf{p}$ is the set

$$T_{\mathbf{p}}\mathcal{S} := \{\mathbf{v} \in \mathbb{R}^3 \, : \, \mathbf{v} \text{ tangent vector of } \mathcal{S} \text{ at } \mathbf{p}\} \,.$$

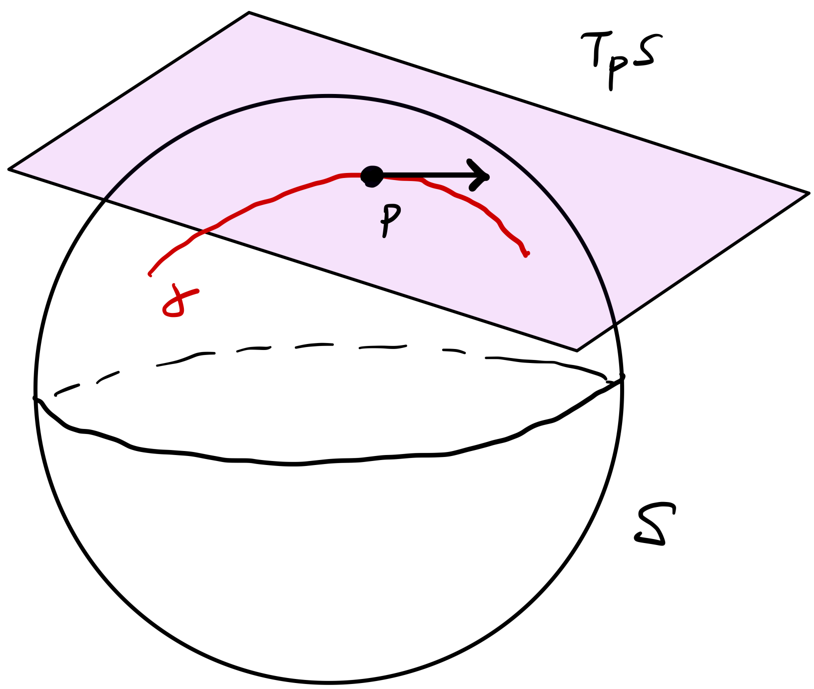


Figure 4.16: Tangent space $T_{\mathbf{p}}\mathcal{S}$ of surface $\mathcal{S}$ at the point $\mathbf{p}$. A tangent vector $\mathbf{v}$ coincides with $\dot{\boldsymbol{\gamma}}(0)$ for some $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}$ such that $\boldsymbol{\gamma}(0) = \mathbf{p}$.

Let us start with the most basic example: We want to compute the tangent space to an open set in $\mathbb{R}^2$.

Example 4.75

Let $U \subseteq \mathbb{R}^2$ be open and $\mathbf{p} \in U$. Then

$$T_{\mathbf{p}}U = \mathbb{R}^2 \,.$$

Proof. Let $\mathbf{v} \in T_{\mathbf{p}}U$. By definition there exists a smooth curve

$$\gamma : (-\varepsilon, \varepsilon) \to U$$

such that $\boldsymbol{\gamma}(0) = \mathbf{p}$ and $\dot{\boldsymbol{\gamma}}(0) = \mathbf{v}$. Since $U \subseteq \mathbb{R}^2$, it follows that $\boldsymbol{\gamma}$ is a plane curve, so that

$$\mathbf{v} = \dot{\boldsymbol{\gamma}}(0) \in \mathbb{R}^2 \,.$$

Conversely, let $\mathbf{v} \in \mathbb{R}^2$. Since $\mathbf{p} \in U$ and U is open, there exists $\varepsilon > 0$ such that $B_\varepsilon(p) \subseteq U$. Define the curve

$$\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathbb{R}^3 \,, \quad \boldsymbol{\gamma}(t) := \mathbf{p} + t\mathbf{v} \,.$$

By construction

$$\boldsymbol{\gamma}(-\varepsilon, \varepsilon) \subseteq B_\varepsilon(\mathbf{p}) \subseteq U \,, \quad \boldsymbol{\gamma}(0) = \mathbf{p} \,, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v} \,,$$

showing that $\mathbf{v} \in T_{\mathbf{p}}U$.

In the above example we have seen that $T_{\mathbf{p}}U = \mathbb{R}^2$. This property holds in general for $T_{\mathbf{p}}\mathcal{S}$ with $\mathcal{S}$ regular surface. Before proving this fact, we need a lemma.

Lemma 4.76

Let $\mathcal{S}$ be regular and $\mathbf{p} \in \mathcal{S}$. Let $\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S}$ be a regular chart at $\mathbf{p}$, with

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p} \,.$$

We have:

1. Suppose $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ is a smooth curve such that

   $$\boldsymbol{\gamma}(-\varepsilon, \varepsilon) \subseteq \boldsymbol{\sigma}(U) \,, \quad \boldsymbol{\gamma}(0) = \mathbf{p} \,.$$

   Then there exist smooth functions

   $$u, v : (-\varepsilon, \varepsilon) \to \mathbb{R}$$

   such that

   $$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)) \,, \quad \forall\, t \in (-\varepsilon, \varepsilon) \,,$$

   and

   $$u(0) = u_0 \,, \quad v(0) = v_0 \,.$$

2. Conversely, assume $u, v : (-\varepsilon, \varepsilon) \to \mathbb{R}$ are smooth functions such that

$$u(0) = u_0 \,, \quad v(0) = v_0 \,.$$

Then

$$\boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(u(t), v(t))$$

is a smooth curve such that

$$\boldsymbol{\gamma}(-\varepsilon, \varepsilon) \subseteq \mathcal{S} \,, \quad \boldsymbol{\gamma}(0) = \mathbf{p} \,.$$

Proof

Denote the coordinates of $\boldsymbol{\sigma}$ by

$$\boldsymbol{\sigma}(u, v) = (f(u, v), g(u, v), h(u, v)) \,.$$

The differential of $\boldsymbol{\sigma}$ is

$$d\boldsymbol{\sigma} = \begin{pmatrix} f_u & f_v \\ g_u & g_v \\ h_u & h_v \end{pmatrix} .$$

Since $\boldsymbol{\sigma}$ is regular, by definition $d\boldsymbol{\sigma}$ has rank-2 at (u_0, v_0). This means that at least one of the 3 minors

$$\begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix} , \quad \begin{pmatrix} f_u & f_v \\ h_u & h_v \end{pmatrix} , \quad \begin{pmatrix} g_u & g_v \\ h_u & h_v \end{pmatrix} .$$

is invertible. WLOG assume the first is invertible (the proof in case the other two are invertible is similar.) Define the map

$$F : U \subseteq \mathbb{R}^2 \to \mathbb{R}^2 \,, \quad F(u, v) = (f(u, v), g(u, v)) \,.$$

We have

$$dF = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix} ,$$

which is invertible at (u_0, v_0) by assumption. Hence, by the Inverse Function Theorem, there exist

- $W \subseteq U \subseteq \mathbb{R}^2$ open set with $(u_0, v_0) \in W$,
- $V \subseteq \mathbb{R}^2$ open set with $F(u_0, v_0) \in V$,

such that

$$F : W \to V$$

is a diffeomorphism. Hence

$$F^{-1} : V \to W$$

is smooth. Since $\boldsymbol{\gamma}(-\varepsilon, \varepsilon) \subseteq \boldsymbol{\sigma}(U)$, it is well defined the composition

$$F^{-1} \circ \boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to W \subseteq U \,.$$

In particular, there exist two functions u, v (the components of $F^{-1} \circ \boldsymbol{\gamma}$) such that

$$(F^{-1} \circ \boldsymbol{\gamma})(t) = (u(t), v(t)) \tag{4.2}$$

Note that u, v are smooth, since $F^{-1} \circ \boldsymbol{\gamma}$ is smooth, being F^{-1} and $\boldsymbol{\gamma}$ smooth. As $\boldsymbol{\gamma}(0) = \mathbf{p}$, by definition of F we have

$$(u(0), v(0)) = (F^{-1} \circ \boldsymbol{\gamma})(0) = F^{-1}(\mathbf{p}) = (u_0, v_0)\,,$$

showing that

$$u(0) = u_0\,, \quad v(0) = v_0\,.$$

Moreover, applying $\boldsymbol{\sigma}$ to both sides of (4.2) yields

$$\boldsymbol{\sigma}(u(t), v(t)) = \boldsymbol{\sigma}((F^{-1} \circ \boldsymbol{\gamma}))(t) = \boldsymbol{\gamma}(t)\,,$$

as we wanted to show.
The converse statement is trivial.

We are now ready to characterize $T_{\mathbf{p}}\mathcal{S}$ when $\mathcal{S}$ is a regular surface.

Theorem 4.77

Let $\mathcal{S}$ be a (regular) surface and $\mathbf{p} \in \mathcal{S}$. Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a chart at $\mathbf{p}$. Denote by $(u_0, v_0) \in U$ a point such that

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}\,.$$

Then

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} := \{\lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v \ : \ \lambda, \mu \in \mathbb{R}\}\,,$$

where $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are evaluated at (u_0, v_0). In particular

$$T_{\mathbf{p}}\mathcal{S} = \mathbb{R}^2\,.$$

Proof

Let $\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S}$ be a chart at p. If we show that

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$$

then we deduce

$$T_{\mathbf{p}}\mathcal{S} = \mathbb{R}^2\,,$$

since $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent.
Step 1. Suppose $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. By definition there exists a smooth curve $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}$ such that

$$\boldsymbol{\gamma}(0) = \mathbf{p}\,, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v}\,.$$

By continuity, we can take ε small enough so that

$$\boldsymbol{\gamma}(-\varepsilon,\varepsilon) \subseteq \boldsymbol{\sigma}(U)\,.$$

By Lemma 4.76 there exist smooth functions $u, v \colon (-\varepsilon,\varepsilon) \to \mathbb{R}$ such that

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))\,, \quad \forall\, t \in (-\varepsilon,\varepsilon)\,,$$

and

$$u(0) = u_0\,, \quad v(0) = v_0\,.$$

Therefore, by chain rule,

$$\dot{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}_u(u(t), v(t))\,\dot{u}(t) + \boldsymbol{\sigma}_v(u(t), v(t))\,\dot{v}(t)\,.$$

Evaluating the above at $t = 0$ yields

$$\begin{aligned} \mathbf{v} &= \dot{\boldsymbol{\gamma}}(0) \\ &= \boldsymbol{\sigma}_u(u(0), v(0))\,\dot{u}(0) + \boldsymbol{\sigma}_v(u(0), v(0))\,\dot{v}(0) \\ &= \boldsymbol{\sigma}_u(u_0, v_0)\,\dot{u}(0) + \boldsymbol{\sigma}_v(u_0, v_0)\,\dot{v}(0)\,, \end{aligned}$$

which shows

$$\mathbf{v} \in \operatorname{span}\{\boldsymbol{\sigma}_u(u_0, v_0), \boldsymbol{\sigma}_v(u_0, v_0)\}\,.$$

Step 2. Suppose that

$$\mathbf{v} \in \operatorname{span}\{\boldsymbol{\sigma}_u(u_0, v_0), \boldsymbol{\sigma}_v(u_0, v_0)\}\,.$$

Then there exist $\lambda, \mu \in \mathbb{R}$ such that

$$\mathbf{v} = \lambda\boldsymbol{\sigma}_u(u_0, v_0) + \mu\boldsymbol{\sigma}_v(u_0, v_0)\,.$$

Define the curve

$$\boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(u_0 + \lambda t, v_0 + \mu t)\,, \quad t \in (-\varepsilon,\varepsilon)\,.$$

We have

$$\boldsymbol{\gamma}(0) = \boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}\,.$$

Therefore, for ε sufficiently small, we have

$$\boldsymbol{\gamma}(-\varepsilon,\varepsilon) \subseteq \boldsymbol{\sigma}(U)\,.$$

By chain rule

$$\dot{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}_u(u_0 + \lambda t, v_0 + \mu t)\lambda + \boldsymbol{\sigma}_v(u_0 + \lambda t, v_0 + \mu t)\mu\,,$$

and therefore

$$\dot{\boldsymbol{\gamma}}(0) = \boldsymbol{\sigma}_u(u_0, v_0)\lambda + \boldsymbol{\sigma}_v(u_0, v_0)\mu = \mathbf{v}\,.$$

This proves that $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$, ending the proof.

Therefore $T_{\mathbf{p}}\mathcal{S}$ is always two-dimensional. This justifies the following definition.

Definition 4.78: Tangent plane

Let $\mathcal{S}$ be a regular surface and $\mathbf{p} \in \mathcal{S}$. The set

$$T_{\mathbf{p}}\mathcal{S}$$

is called the **tangent plane** to $\mathcal{S}$ at $\mathbf{p}$.

Remark 4.79

By definition $T_{\mathbf{p}}\mathcal{S}$ is a vector subspace of $\mathbb{R}^3$. As such, it holds that

$$\mathbf{0} \in T_{\mathbf{p}}\mathcal{S} .$$

To see this, take the curve $\boldsymbol{\gamma}(t) \equiv \mathbf{p}$. Then $\boldsymbol{\gamma}(0) = \mathbf{p}$ and $\dot{\boldsymbol{\gamma}}(0) = \mathbf{0}$, showing that $\mathbf{0} \in T_{\mathbf{p}}\mathcal{S}$.

Therefore $T_{\mathbf{p}}\mathcal{S}$ is a plane through the origin, no matter where the point $\mathbf{p} \in \mathcal{S}$ is located. In pictures, such as Figure 4.16, we draw the tangent plane as if it was touching the surfaces at the point $\mathbf{p}$, and still denote it by $T_{\mathbf{p}}\mathcal{S}$. This is a slight abuse of notation: to be precise, the plane drawn is

$$\mathbf{p} + T_{\mathbf{p}}\mathcal{S} ,$$

which is the **affine tangent plane** through $\mathbf{p} \in \mathcal{S}$.

It is possible to give a cartesian equation for the tangent plane

$$T_{\mathbf{p}}\mathcal{S}$$

and for the affine tangent plane

$$\mathbf{p} + T_{\mathbf{p}}\mathcal{S} .$$

Proposition 4.80: Equation of tangent plane

Let $\mathcal{S}$ be a regular surface and $\mathbf{p} \in \mathcal{S}$. Let $\boldsymbol{\sigma}$ be a regular chart at $\mathbf{p}$, with

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p} = (x_0, y_0, z_0) .$$

Let

$$\mathbf{n} := \boldsymbol{\sigma}_u(u_0, v_0) \times \boldsymbol{\sigma}_v(u_0, v_0)$$

We have:

1. The equation of the tangent plane $T_{\mathbf{p}}\mathcal{S}$ is given by

$$\mathbf{n}_1 x + \mathbf{n}_2 y + \mathbf{n}_3 z = 0 , \quad \forall\, (x, y, z) \in \mathbb{R}^3 ,$$

where $\mathbf{n} = (\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3)$.

2. The equation of the affine tangent plane $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is given by

$$\mathbf{n}_1(x - x_0) + \mathbf{n}_2(y - x_0) + \mathbf{n}_3(z - z_0) = 0\,, \quad \forall\,(x, y, z) \in \mathbb{R}^3\,.$$

Proof

By Theorem 4.77 we know that

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u(u_0, v_0), \boldsymbol{\sigma}_v(u_0, v_0)\}\,.$$

By the properties of cross product, the vector $\mathbf{n}$ is orthogonal to both $\boldsymbol{\sigma}_u(u_0, v_0)$ and $\boldsymbol{\sigma}_v(u_0, v_0)$. Therefore it is orthogonal to $T_{\mathbf{p}}\mathcal{S}$. The equation for $T_{\mathbf{p}}\mathcal{S}$ is then

$$(x, y, z) \cdot \mathbf{n} = 0\,, \forall\,(x, y, z) \in \mathbb{R}^3\,.$$

In particular, the equation for the affine tangent plane $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is

$$(x, y, z) \cdot \mathbf{n} = k\,, \quad \forall\,(x, y, z) \in \mathbb{R}^3\,,$$

for some $k \in \mathbb{R}$. To compute k, it is sufficient to evaluate the above equation at $\mathbf{p}$, since $\mathbf{p}$ belongs to $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$. We obtain

$$k = \mathbf{p} \cdot \mathbf{n}\,.$$

Hence the equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is

$$(x - x_0, y - y_0, z - z_0) \cdot \mathbf{n} = 0\,, \quad \forall\,(x, y, z) \in \mathbb{R}^3\,,$$

ending the proof.

Example 4.81

Consider the surface $\mathcal{S}$ defined by the chart

$$\boldsymbol{\sigma}(u, v) := \left(\sqrt{1 - v}\cos(u), \sqrt{1 - v}\sin(u), v\right)\,.$$

We want to compute the equation for the tangent plane $T_{\mathbf{p}}\mathcal{S}$, and for the affine tangent plane $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$. First, we need to check that $\boldsymbol{\sigma}$ is regular. We have

$$\begin{aligned}
\boldsymbol{\sigma}_u &= \left(-\sqrt{1 - v}\sin(u), \sqrt{1 - v}\cos(u), 0\right) \\
\boldsymbol{\sigma}_v &= \left(\frac{1}{2}(1 - v)^{-1/2}\cos(u), \frac{1}{2}(1 - v)^{-1/2}\sin(u), 1\right)
\end{aligned}$$

As the last component of $\boldsymbol{\sigma}_u$ is 0 and the last component of $\boldsymbol{\sigma}_v$ is 1, we conclude that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent. Thus $\boldsymbol{\sigma}$ is regular.
Suppose $\mathbf{p} \in \mathcal{S}$ is such that

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$$

for some $(u_0, v_0) \in \mathbb{R}^2$. By Theorem 4.77 we have

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u(u_0,v_0), \boldsymbol{\sigma}_v(u_0,v_0)\}\,.$$

To find the equation of $T_{\mathbf{p}}\mathcal{S}$ we compute:

$$\begin{aligned}
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v & = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\sqrt{1-v}\sin(u) & \sqrt{1-v}\cos(u) & 0 \\ \frac{1}{2}(1-v)^{-1/2}\cos(u) & \frac{1}{2}(1-v)^{-1/2}\sin(u) & 1 \end{vmatrix} \\
& = \left(\sqrt{1-v}\cos(u), \sqrt{1-v}\sin(u), -\frac{1}{2}\right)
\end{aligned}$$

For

$$(u_0, v_0) = \left(\frac{\pi}{4}, 0\right)$$

we have

$$\mathbf{p} = \boldsymbol{\sigma}(u_0, v_0) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0\right),$$

and therefore

$$\mathbf{n} = (\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v)(u_0, v_0) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, -\frac{1}{2}\right).$$

The equation for $T_{\mathbf{p}}\mathcal{S}$ is therefore

$$(x,y,z)\cdot \mathbf{n} = 0\,, \quad \forall\,(x,y,z) \in \mathbb{R}^3\,.$$

The above reads

$$\frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}y - \frac{1}{2}z = 0\,.$$

The equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is instead

$$\frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}y - \frac{1}{2}z = k\,,$$

for some $k \in \mathbb{R}$. To compute k, note that $\mathbf{p} \in \mathbf{p} + T_{\mathbf{p}}\mathcal{S}$, and therefore

$$\frac{\sqrt{2}}{2}\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}\frac{\sqrt{2}}{2} = k \quad \implies \quad k = 1\,.$$

The equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is then

$$\frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}y - \frac{1}{2}z = 1\,.$$

Remark 4.82: Tangent space and derivations

The definition of tangent plane depends on the fact that $\mathcal{S}$ is contained in $\mathbb{R}^3$. This is a serious drawback in many applications, as the surface $\mathcal{S}$ does not necessarily need to be Euclidean. There is a way to get rid of such dependence, and give an *intrinsic* definition of tangent plane, depending only on the point $\mathbf{p}$ and the surface $\mathcal{S}$.

The basic idea is as follows: If $U \subseteq \mathbb{R}^2$ is open and $\mathbf{p} \in U$, then $T_{\mathbf{p}}U = \mathbb{R}^2$. We can associate to any point $\mathbf{v} \in T_{\mathbf{p}}U$ a directional derivative acting on smooth functions $f : U \to \mathbb{R}$:

$$\mathbf{v} = (v_1, v_2) \mapsto \left.\frac{\partial}{\partial v}\right|_p = v_1 \left.\frac{\partial}{\partial x_1}\right|_p + v_2 \left.\frac{\partial}{\partial x_2}\right|_p$$

The above directional derivative is called a **derivation**.

The point is that derivations do not need to be defined through vectors, but can be defined as follows: D is a **derivation** if

- $D : C^\infty(U) \to \mathbb{R}$ is a linear operator, where $C^\infty(U)$ is the set of smooth functions $f : U \to \mathbb{R}$,
- D satisfies the Leibniz rule

$$D(fg) = f(\mathbf{p})D(g) + g(\mathbf{p})D(f), \quad \forall f, g \in C^\infty(U).$$

The tangent plane at p can then be defined as

$$T_{\mathbf{p}}U = \{D \text{ derivation at } \mathbf{p}\}.$$

Therefore

$$T_{\mathbf{p}}U \subseteq (C^\infty(U))^*,$$

the dual space of smooth functions.

It is possible to do such construction directly on $\mathcal{S}$, by introducing the concepts of:

- **germ** of a function
- **algebra** of derivations, acting on germs

An in depth discussion can be found in Chapter 3.4 of [1].

4.9 Differential of smooth functions

Let $f : U \to V$ with $U, V \subseteq \mathbb{R}^2$ open. Suppose f is smooth. By definition, the differential of f at $\mathbf{p} \in U$ is a linear map

$$df_{\mathbf{p}} : \mathbb{R}^2 \to \mathbb{R}^2$$

which approximates f locally at $\mathbf{p}$. We have seen in Example 4.75 that

$$T_{\mathbf{p}}U = \mathbb{R}^2$$

Therefore we can interpret $df_{\mathbf{p}}$ as a map between tangent planes:

$$df_{\mathbf{p}} : T_{\mathbf{p}}U \to T_{\mathbf{p}}U .$$

This reasoning suggests how to define the differential of a smooth map between surfaces: If $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ is smooth, we could define its differential at $\mathbf{p} \in \mathcal{S}$ as a linear map

$$df_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\widetilde{\mathcal{S}} .$$

How is the above map defined explicitly? To answer this question, we need a lemma.

Lemma 4.83

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ a smooth map. For $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$, let $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}$ be such that

$$\boldsymbol{\gamma}(0) = \mathbf{p} , \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v} .$$

Define

$$\tilde{\boldsymbol{\gamma}} := f \circ \boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \widetilde{\mathcal{S}} .$$

Then $\tilde{\boldsymbol{\gamma}}$ is a smooth curve into $\mathbb{R}^3$ and

$$\tilde{\mathbf{v}} \in T_{f(\mathbf{p})}\widetilde{\mathcal{S}} , \quad \tilde{\mathbf{v}} := \dot{\tilde{\boldsymbol{\gamma}}}(0) .$$

Proof

Note that

$$\tilde{\boldsymbol{\gamma}} = i \circ f \circ \boldsymbol{\gamma} ,$$

with $i : \widetilde{\mathcal{S}} \to \mathbb{R}^3$ inclusion map. Since $i, f, \boldsymbol{\gamma}$ are smooth, we conclude that $\tilde{\boldsymbol{\gamma}} : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ is smooth. Moreover

$$\tilde{\boldsymbol{\gamma}}(0) = f(\boldsymbol{\gamma}(0)) = f(\mathbf{p}) ,$$

and therefore

$$\tilde{\mathbf{v}} := \dot{\tilde{\boldsymbol{\gamma}}}(0) \in T_{f(\mathbf{p})}\widetilde{\mathcal{S}} ,$$

by definition of tangent space.

Definition 4.84: Differential of smooth function

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ a smooth map. The differential $df_{\mathbf{p}}$ of f at $\mathbf{p}$ is defined as the map

$$df_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\widetilde{\mathcal{S}} , \quad df_{\mathbf{p}}(\mathbf{v}) := \tilde{\mathbf{v}} ,$$

where $\tilde{\mathbf{v}}$ is as in Lemma 4.83.

Remark 4.85: Computing $d_{\mathbf{p}} f$

The differential of f can be computed using the definition. Specifically, let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$ and let $\boldsymbol{\gamma}$ be a curve on $\mathcal{S}$ such that

$$\boldsymbol{\gamma}(0) = \mathbf{p}\,, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v}$$

Then

$$d_{\mathbf{p}} f(\mathbf{v}) = (f \circ \gamma)'(0)$$

The problem with this calculation is that one has to produce a curve $\boldsymbol{\gamma}$ with $\dot{\boldsymbol{\gamma}}(0) = \mathbf{v}$, in order to compute $d_{\mathbf{p}} f$ at $\mathbf{v}$. As you can imagine, finding one curve for each vector is far from ideal.

Example 4.86

Let $\mathcal{S} = (0, 2\pi) \times \mathbb{R}$ and

$$\widetilde{\mathcal{S}} = \{(x, y, z) \in \mathbb{R}^3 \,:\, x^2 + y^2 = 1\}$$

Note that $\mathcal{S}$ is a subset of $\mathbb{R}^2$ (and in particular a regular surface), while $\widetilde{\mathcal{S}}$ is a unit cylinder. Define the map

$$f \,:\, S \to \widetilde{\mathcal{S}}\,, \quad f(u, v) := (\cos u, \sin u, v)$$

Clearly $f \in \widetilde{\mathcal{S}}$ since $\cos^2 u + \sin^2 u = 1$. We want to compute the differential of f

$$d_{\mathbf{p}} f \,:\, T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\widetilde{\mathcal{S}}$$

Since $\mathcal{S}$ is a subset of $\mathbb{R}^2$, we have that $T_{\mathbf{p}}\mathcal{S}$ is just $\mathbb{R}^2$. Let $\mathbf{v} = (a, b) \in \mathbb{R}^2$. We need to find a curve $\boldsymbol{\gamma} \,:\, (-\varepsilon, \varepsilon) \to \mathcal{S}$ such that

$$\boldsymbol{\gamma}(0) = \mathbf{p}\,, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v} = (a, b) \tag{4.3}$$

Since $\mathcal{S}$ is a subset of $\mathbb{R}^2$, $\boldsymbol{\gamma}$ can be chosen as the straight line through $\mathbf{p}$ of direction $\mathbf{v}$, e.g.,

$$\boldsymbol{\gamma}(t) := \mathbf{p} + t\mathbf{v} = (u_0 + ta, v_0 + tb)\,,$$

where we denoted $\mathbf{p} = (u_0, v_0)$. Clearly $\boldsymbol{\gamma}$ satisfies (4.3). We have

$$\begin{aligned}
(f \circ \boldsymbol{\gamma})(t) &= f(u_0 + ta, v_0 + tb) \\
&= (\cos(u_0 + ta), \sin(u_0 + ta), v_0 + tb) \\
(f \circ \boldsymbol{\gamma})'(t) &= (-a \sin(u_0 + ta), a \cos(u_0 + ta), b)
\end{aligned}$$

Therefore the differential at $\mathbf{p}$ is

$$d_{\mathbf{p}} f(\mathbf{v}) = (f \circ \boldsymbol{\gamma})'(0) = (-a \sin(u_0), a \cos(u_0), b)\,.$$

We need to show that the definition of differential is well-posed, i.e., that $d_{\mathbf{p}} f(\mathbf{v})$ depends only on $\mathbf{p}$, f, $\mathbf{v}$: This is because there are infinitely many curves $\boldsymbol{\gamma}$ passing trough $\mathbf{p}$ and such that $\dot{\boldsymbol{\gamma}}(0) = \mathbf{v}$, and a priori $d_{\mathbf{p}} f(\mathbf{v})$ could depend on which curve is chosen.

We prove well-posedness of $df_{\mathbf{p}}$ in the next Proposition. We also show that the map $df_{\mathbf{p}}$ is linear, and we provide the matrix representation of $df_{\mathbf{p}}$. The matrix representation, in particular, allows to compute $d_{\mathbf{p}}f$ without the need to construct suitable curves.

Proposition 4.87

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ a smooth map. Denote the differential of f by

$$df_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\widetilde{\mathcal{S}}\,.$$

We have:

1. $df_{\mathbf{p}}(\mathbf{v})$ does not depend on the choice of $\boldsymbol{\gamma}$, but rather only on $f, \mathbf{p}, \mathbf{v}$.
2. $df_{\mathbf{p}}$ is linear, that is,
$$df_{\mathbf{p}}(\lambda\mathbf{v} + \mu\mathbf{w}) = \lambda df_{\mathbf{p}}(\mathbf{v}) + \mu df_{\mathbf{p}}(\mathbf{w})\,,$$
for all $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ and $\lambda, \mu \in \mathbb{R}$.
3. Let
$$\boldsymbol{\sigma} : U \to \mathcal{S}\,, \quad \tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \widetilde{\mathcal{S}}\,,$$
be regular charts at $\mathbf{p}$ and $f(\mathbf{p})$, respectively. Denote by
$$(u, v) \mapsto (\alpha(u, v), \beta(u, v))$$
the components of the smooth map
$$\Psi := \tilde{\boldsymbol{\sigma}}^{-1} \circ f \circ \boldsymbol{\sigma} : U \to \widetilde{U}\,.$$
In particular it holds
$$\tilde{\boldsymbol{\sigma}}(\alpha(u, v), \beta(u, v)) = f(\boldsymbol{\sigma}(u, v))\,, \quad \forall\,(u, v) \in U\,.$$
The matrix of the linear map $df_{\mathbf{p}}$ with respect to the basis
$$\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} \;\text{ on }\; T_{\mathbf{p}}\mathcal{S}\,, \quad \{\tilde{\boldsymbol{\sigma}}_{\tilde{u}}, \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\} \;\text{ on }\; T_{f(\mathbf{p})}\widetilde{\mathcal{S}}\,,$$
is given by the Jacobian of the map Ψ, that is,
$$d_{\mathbf{p}}f = J\Psi = \begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix}\,.$$

Point 3 in the above proposition seems complicated, but it is saying something really simple:

1. Let $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ be a smooth function between surfaces. Consider local charts $\boldsymbol{\sigma}$ at $\mathbf{p}$ and $\tilde{\boldsymbol{\sigma}}$ at $f(\mathbf{p})$. By definition of smooth map, the real map
$$\Psi : \mathbb{R}^2 \to \mathbb{R}^2\,, \qquad \Psi := \tilde{\boldsymbol{\sigma}}^{-1} \circ f \circ \boldsymbol{\sigma}$$
is smooth.

2. The matrix of the differential $d_{\mathbf{p}} f$ with respect to the basis

$$\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} \;\text{ on }\; T_{\mathbf{p}}\mathcal{S}\,, \quad \{\tilde{\boldsymbol{\sigma}}_{\tilde{u}}, \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\} \;\text{ on }\; T_{f(\mathbf{p})}\widetilde{\mathcal{S}}\,,$$

is just the Jacobian of Ψ.

Proof

Let $\mathbf{p} \in \mathcal{S}$ and $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$ be fixed. In order to prove the thesis, we need compute $d_{\mathbf{p}} f$. To this end, let $\boldsymbol{\sigma} : U \to \mathcal{S}$ be a chart at $\mathbf{p}$. Denote by $(u_0, v_0) \in U$ the point such that

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}\,.$$

Let $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \widetilde{\mathcal{S}}$ a chart at $f(\mathbf{p})$. Since f is smooth, the map

$$\Psi : U \to \widetilde{U}\,, \quad \Psi := \tilde{\boldsymbol{\sigma}}^{-1} \circ f \circ \boldsymbol{\sigma}$$

is smooth. Denote by

$$(u, v) \mapsto (\alpha(u, v), \beta(u, v))$$

the smooth components of Ψ. By definition of Ψ it holds

$$\tilde{\boldsymbol{\sigma}}(\alpha(u, v), \beta(u, v)) = f(\boldsymbol{\sigma}(u, v))\,, \quad \forall\,(u, v) \in U\,. \tag{4.4}$$

Denote by

$$\mathbf{v} = (\lambda, \mu)$$

the components of $\mathbf{v}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$. Define the scalar functions

$$u(t) := u_0 + \lambda t\,, \quad v(t) := v_0 + \mu t\,,$$

and the curve

$$\boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(u(t), v(t))\,.$$

Clearly $\boldsymbol{\gamma}$ is a regular curve on $\mathcal{S}$, and it holds

$$\boldsymbol{\gamma}(0) = \boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}\,, \quad \dot{\boldsymbol{\gamma}}(0) = (\lambda, \mu) = \mathbf{v}$$

By (4.4) we have

$$\begin{aligned}
(f \circ \boldsymbol{\gamma})(t) &= f(\boldsymbol{\gamma}(t)) \\
&= f(\boldsymbol{\sigma}(u(t), v(t))) \\
&= \tilde{\boldsymbol{\sigma}}(\alpha(u(t), v(t)), \beta(u(t), v(t)))
\end{aligned}$$

By chain rule we obtain

$$\begin{aligned}
(f \circ \boldsymbol{\gamma})'(t) &= \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \frac{d}{dt}\alpha(u(t), v(t)) + \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \frac{d}{dt}\beta(u(t), v(t)) \\
&= \tilde{\boldsymbol{\sigma}}_{\tilde{u}}[\alpha_u \dot{u}(t) + \alpha_v \dot{v}(t)] + \tilde{\boldsymbol{\sigma}}_{\tilde{v}}[\beta_u \dot{u}(t) + \beta_v \dot{v}(t)]
\end{aligned}$$

Recalling that $\dot{u}(0) = \lambda$ and $\dot{v}(0) = \mu$, we get

$$(f \circ \boldsymbol{\gamma})'(t) = \tilde{\boldsymbol{\sigma}}_{\tilde{u}}[\lambda \alpha_u + \mu \alpha_v] + \tilde{\boldsymbol{\sigma}}_{\tilde{v}}[\lambda \beta_u + \mu \beta_v]$$

As by definition of differential

$$d_{\mathbf{p}} f(\mathbf{v}) = (f \circ \boldsymbol{\gamma})'(0)$$

we have obtained

$$d_{\mathbf{p}} f(\mathbf{v}) = \tilde{\boldsymbol{\sigma}}_{\tilde{u}}[\lambda \alpha_u + \mu \alpha_v] + \tilde{\boldsymbol{\sigma}}_{\tilde{v}}[\lambda \beta_u + \mu \beta_v] \tag{4.5}$$

We can now draw our 3 conclusions:

1. The RHS of (4.5) depends only on λ, μ (the components of $\mathbf{v}$), f (via the components α, β of Ψ), and the point $\mathbf{p}$. In particular $d_{\mathbf{p}} f(\mathbf{v})$ does not depend on the choice of $\boldsymbol{\gamma}$, and the definition is well-posed.

2. The RHS of (4.5) is linear in the components λ, μ of $\mathbf{v}$. In particular $d_{\mathbf{p}} f(\mathbf{v})$ is linear in $\mathbf{v}$.

3. The coordinates of $\boldsymbol{\sigma}_u$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ are $(\lambda, \mu) = (1, 0)$. Using (4.5), we get

$$d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) = \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \alpha_u + \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \beta_u$$

Similarly, the coordinates of $\boldsymbol{\sigma}_v$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ are $(\lambda, \mu) = (0, 1)$. Therefore

$$d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) = \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \alpha_v + \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \beta_v$$

This shows that the matrix of the linear application $d_{\mathbf{p}} f$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ on $T_{\mathbf{p}}\mathcal{S}$ and the basis $\{\tilde{\boldsymbol{\sigma}}_{\tilde{u}}, \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\}$ on $T_{f(\mathbf{p})}\widetilde{\mathcal{S}}$ is

$$\begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix} = J\Psi \,.$$

Example 4.88: Computing the matrix of $d_{\mathbf{p}} f$

Consider the unit cylinder

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \, : \, x^2 + y^2 = 1\} \,.$$

and the map

$$f \colon \mathcal{S} \to \mathbb{R}^2 \,, \quad f(x, y, z) = (y, xz, 0)$$

We want to compute the matrix of $d_{\mathbf{p}} f$. To this end, a chart of $\mathcal{S}$ is given by

$$\boldsymbol{\sigma}(u, v) = (\cos u, \sin u, v) \,, \quad (u, v) \in U := (0, 2\pi) \times \mathbb{R} \,.$$

A chart of $\widetilde{\mathcal{S}} = \mathbb{R}^2$ is given by

$$\tilde{\boldsymbol{\sigma}}(\tilde{u}, \tilde{v}) = (\tilde{u}, \tilde{v}, 0) \,, \quad (\tilde{u}, \tilde{v}) \in \tilde{U} = \mathbb{R}^2 \,.$$

We need to compute the map

$$\Psi : U \to \widetilde{U} , \quad \Psi := \tilde{\boldsymbol{\sigma}}^{-1} \circ f \circ \boldsymbol{\sigma} .$$

Clearly we have

$$\tilde{\boldsymbol{\sigma}}^{-1}(\tilde{u}, \tilde{v}, 0) = (\tilde{u}, \tilde{v}) .$$

Therefore

$$\begin{aligned} \Psi(u, v) &= \tilde{\boldsymbol{\sigma}}^{-1} (f(\boldsymbol{\sigma}(u, v))) \\ &= \tilde{\boldsymbol{\sigma}}^{-1} (f(\cos u, \sin u, v)) \\ &= \tilde{\boldsymbol{\sigma}}^{-1} (\sin(u), \cos(u)v, 0) \\ &= (\sin(u), \cos(u)v) \end{aligned}$$

Therefore

$$\begin{aligned} \partial_u \Psi^1 &= \cos(u) , & \partial_v \Psi^1 &= 0 \\ \partial_u \Psi^2 &= -\sin(u)v , & \partial_v \Psi^2 &= \cos(u) \end{aligned}$$

The matrix of $d_{\mathbf{p}} f$ is hence

$$d_{\mathbf{p}} f = J\Psi = \begin{pmatrix} \cos(u) & 0 \\ -\sin(u)v & \cos(u) \end{pmatrix} .$$

The differential satisfies the following useful properties.

Proposition 4.89

The following hold:

1. If $\mathcal{S}$ is a regular surface and $\mathbf{p} \in \mathcal{S}$, the differential at $\mathbf{p}$ of the identity map

$$I : \mathcal{S} \to \mathcal{S} , \quad I(x) := x ,$$

is the identity map

$$I : T_{\mathbf{p}}(\mathcal{S}) \to T_{\mathbf{p}}(\mathcal{S}) , \quad I(v) := v .$$

2. If $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ are regular surfaces and

$$f : \mathcal{S}_1 \to \mathcal{S}_2 , \quad g : \mathcal{S}_2 \to \mathcal{S}_3 ,$$

are smooth maps, then

$$d_{\mathbf{p}}(g \circ f) = d_{f(\mathbf{p})} g \circ d_{\mathbf{p}} f ,$$

for all $\mathbf{p} \in T_{\mathbf{p}} \mathcal{S}_1$.

3. If $\mathcal{S}_1$, $\mathcal{S}_2$ are regular surfaces and
$$f \colon \mathcal{S}_1 \to \mathcal{S}_2 \,,$$
is a diffeomorphism, then the differential
$$d_{\mathbf{p}} \colon T_{\mathbf{p}}\mathcal{S}_1 \to T_{f(\mathbf{p})}\mathcal{S}_2$$
is invertible for all $\mathbf{p} \in \mathcal{S}_1$.

For a proof see Proposition 4.4.5 in [6]. The above proposition says that the differential of diffeomorphism is invertible. The converse statement is true locally.

Theorem 4.90

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be regular surfaces. Suppose that
$$f \colon \mathcal{S}_1 \to \mathcal{S}_2$$
is smooth. They are equivalent:

1. f is a local diffeomorphism.
2. The differential $d_{\mathbf{p}} f \colon T_{\mathbf{p}}\mathcal{S}_1 \to T_{f(\mathbf{p})}\mathcal{S}_2$ is invertible for all $\mathbf{p} \in \mathcal{S}_1$.

The proof is based on the Inverse Function Theorem, see Proposition 4.4.6 in [6].

4.10 Examples of Surfaces

4.10.1 Level surfaces

We have already seen level surfaces. Let us recall the defintion.

Definition 4.91: Level surface

Let $V \subseteq \mathbb{R}^3$ be an open set and $f \colon V \to \mathbb{R}$ be smooth. The **level surface** associated with f is the set
$$\mathcal{S}_f := f^{-1}(0) = \{(x,y,z) \in V \ : \ f(x,y,z) = 0\} \,.$$

The following Theorem gives a sufficient condition for $\mathcal{S}_f$ to be a regular surface.

Theorem 4.92

Let $V \subseteq \mathbb{R}^3$ be an open set and $f : V \to \mathbb{R}$ be smooth. Suppose that

$$\nabla f(x,y,z) \neq 0\,, \quad \forall\,(x,y,z) \in V\,.$$

Then $\mathcal{S}_f$ is a regular surface.

Let us give a characterization of the tangent plane to $\mathcal{S}_f$.

Proposition 4.93

Let $V \subseteq \mathbb{R}^3$ be an open set and $f : V \to \mathbb{R}$ be smooth. Suppose that

$$\nabla f(x,y,z) \neq 0\,, \quad \forall\,(x,y,z) \in V\,.$$

Then $\nabla f(\mathbf{p})$ is orthogonal to $T_{\mathbf{p}}\mathcal{S}_f$. In particular, the equation of $T_{\mathbf{p}}\mathcal{S}_f$ is given by

$$\partial_x f(\mathbf{p})x + \partial_y f(\mathbf{p})y + \partial_z f(\mathbf{p})z = 0\,, \quad \forall\,(x,y,z) \in \mathbb{R}^3\,.$$

The equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}_f$ is given by

$$\partial_x f(\mathbf{p})(x-x_0) + \partial_y f(\mathbf{p})(y-y_0) + \partial_z f(\mathbf{p})(z-z_0) = 0\,, \forall\,(x,y,z) \in \mathbb{R}^3\,,$$

where $\mathbf{p} = (x_0, y_0, z_0)$.

Proof

Let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}_f$. By definition there exists a smooth curve

$$\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}_f \subseteq \mathbb{R}^3$$

such that

$$\boldsymbol{\gamma}(0) = \mathbf{p}\,, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v}\,.$$

Since $\boldsymbol{\gamma}(t) \in \mathcal{S}_f$, we have that

$$f(\boldsymbol{\gamma}(t)) = 0\,, \quad \forall\, t \in (-\varepsilon, \varepsilon)\,.$$

By chain rule we get

$$\nabla f(\boldsymbol{\gamma}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) = 0\,, \quad \forall\, t \in (-\varepsilon, \varepsilon)\,.$$

Evaluating the above at $t = 0$ yields

$$0 = \nabla f(\boldsymbol{\gamma}(0)) \cdot \dot{\boldsymbol{\gamma}}(0) = \nabla f(\mathbf{p}) \cdot \mathbf{v}\,,$$

showing that $\mathbf{v}$ is orthogonal to $\nabla f(\mathbf{p})$. Since $\mathbf{v}$ is arbitrary, we conclude that $\nabla f(\mathbf{p})$ is orthogonal to $T_{\mathbf{p}}\mathcal{S}_f$. In particular, the equation for $T_{\mathbf{p}}\mathcal{S}_f$ is

$$\nabla f(\mathbf{p}) \cdot (x,y,z) = 0\,, \quad \forall\,(x,y,z) \in \mathbb{R}^3\,.$$

Therefore the equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is given by

$$\nabla f(\mathbf{p}) \cdot (x, y, z) = k \,, \quad \forall \, (x, y, z) \in \mathbb{R}^3 \,,$$

for some $k \in \mathbb{R}$. Since $\mathbf{p} \in \mathbf{p} + T_{\mathbf{p}}\mathcal{S}$, we can substitute

$$(x, y, z) = (x_0, y_0, z_0) = \mathbf{p}$$

in the above equation to obtain

$$k = \nabla f(\mathbf{p}) \cdot (x_0, y_0, z_0) \,.$$

Hence the equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is

$$\nabla f(\mathbf{p}) \cdot (x - x_0, y - y_0, z - z_0) = 0 \,, \quad \forall \, (x, y, z) \in \mathbb{R}^3 \,.$$

4.10.2 Quadrics

Quadrics are level surfaces

$$S_f = \left\{ (x, y, z) \in \mathbb{R}^3 \;:\; f(x, y, z) = 0 \right\} \,,$$

where

$$\begin{aligned} f(x, y, z) = & a_1 x^2 + a_2 y^2 + a_3 z^2 + 2a_4 xy + 2a_5 xz + 2a_6 yz + \\ & + b_1 x + b_2 y + b_3 z + c \,, \end{aligned}$$

for some coefficients $a_i, b_i, c \in \mathbb{R}$. Let

$$A = \begin{pmatrix} a_1 & a_4 & a_6 \\ a_4 & a_2 & a_5 \\ a_6 & a_5 & a_3 \end{pmatrix} \in \mathbb{R}^{3 \times 3} \,,$$

and

$$\mathbf{x} = (x, y, z)^T \,, \quad \mathbf{b} = (b_1, b_2, b_3)^T \,.$$

Then f can be represented by the quadratic form

$$f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} + \mathbf{b} \cdot \mathbf{x} + c \,.$$

The expression $f = 0$ is called a **quadric equation**.

As stated in the following theorem, there are 14 quadrics in total. Out of these:

- 9 are *interesting* surfaces,
- 3 are planes,
- 1 is a line,
- 1 is a point.

Theorem 4.94

Suppose $\mathcal{S}$ is a level surface defined by a quadric equation. Then, up to rigid motions, $\mathcal{S}$ can be described by one of the following equations:

1. Ellipsoid: $\dfrac{x^2}{p^2} + \dfrac{y^2}{q^2} + \dfrac{z^2}{r^2} = 1.$
2. Hyperboloid of one sheet: $\dfrac{x^2}{p^2} + \dfrac{y^2}{q^2} - \dfrac{z^2}{r^2} = 1$
3. Hyperboloid of two sheets: $\dfrac{x^2}{p^2} - \dfrac{y^2}{q^2} - \dfrac{z^2}{r^2} = 1$
4. Elliptic Paraboloid: $\dfrac{x^2}{p^2} + \dfrac{y^2}{q^2} = z$
5. Hyperbolic Paraboloid: $\dfrac{x^2}{p^2} - \dfrac{y^2}{q^2} = z$
6. Quadric Cone: $\dfrac{x^2}{p^2} + \dfrac{y^2}{q^2} - \dfrac{z^2}{r^2} = 0$
7. Elliptic Cylinder: $\dfrac{x^2}{p^2} + \dfrac{y^2}{q^2} = 1$
8. Hyperbolic Cylinder: $\dfrac{x^2}{p^2} - \dfrac{y^2}{q^2} = 1$
9. Parabolic Cylinder: $\dfrac{x^2}{p^2} = y$
10. Plane: $x = 0$
11. Two parallel planes: $x^2 = p^2$
12. Two intersecting planes: $\dfrac{x^2}{p^2} - \dfrac{y^2}{q^2} = 0$
13. Straight line: $\dfrac{x^2}{p^2} + \dfrac{y^2}{q^2} = 0$
14. Single point: $\dfrac{x^2}{p^2} + \dfrac{y^2}{q^2} + \dfrac{z^2}{r^2} = 0$

The proof of Theorem 4.94 follows by diagonalizing the symmetric matrix A, and by studying the eigenvalues, see Theorem 5.5.2 in [6].

Example 4.95

The sphere is described by

$$S = \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 + z^2 = 1\} \,.$$

This is an ellipsoid with

$$p = q = r = 1 \,.$$

In particular we can write the sphere as the quadric equation:

$$\mathbf{x}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x} = 1 \,.$$

Example 4.96

Consider the level surface

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \ : \ f(x, y, z) = 0\}$$

with

$$f(x, y, z) = x^2 + 2y^2 - 4z^2 + 2xy + yz - 6xz + 1 = 0 \,.$$

Therefore $\mathcal{S}$ is a quadric. The matrix associated to f is

$$A = \begin{pmatrix} 1 & 1 & -3 \\ 1 & 2 & 1/2 \\ -3 & 1/2 & -4 \end{pmatrix} .$$

Diagonalizing the matrix A we obtain $A = PDP^{-1}$, with P matrix of eigenvectors and

$$D = \begin{pmatrix} -5.51 & 0 & 0 \\ 0 & 1.55 & 0 \\ 0 & 0 & 2.96 \end{pmatrix} .$$

Therefore, up to changing basis via the matrix P, S can be described by the quadric equation

$$5.51\tilde{x}^2 - 1.55\tilde{y}^2 - 2.96\tilde{z}^2 = 1 \,,$$

showing that S is a Hyperboloid of two sheets.

4.10.3 Ruled surfaces

A ruled surface is a surface obtained as union of straight lines, called the rulings of the surface. By using curves, ruled surfaces can be defined in the following way.

Definition 4.97: Ruled surface

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be a smooth curve and $\mathbf{a} : (a,b) \to \mathbb{R}^3$ a vector, such that $\dot{\boldsymbol{\gamma}}(t)$ and $\mathbf{a}(t)$ are linearly independent for all $t \in (a,b)$. A **ruled surface** is a surface with chart

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\mathbf{a}(u) .$$

We say that:

- $\boldsymbol{\gamma}$ is the **base curve**
- The lines $v \mapsto v\mathbf{a}(u)$ are the **rulings**

Proposition 4.98

A ruled surface $\mathcal{S}$ is regular if v is sufficiently small.

Proof

A chart for $\mathcal{S}$ is

$$\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}}(u) + v\dot{\mathbf{a}}(u) , \quad \boldsymbol{\sigma}_v = \mathbf{a}(u) ,$$

with $\dot{\boldsymbol{\gamma}}$ and $\mathbf{a}$ linerly independent. Thus $\dot{\boldsymbol{\gamma}}(u) + v\dot{\mathbf{a}}(u)$ and $\mathbf{a}$ are linearly independent for v sufficiently small.

The same base curve can yield multiple ruled surfaces. For example, if $\boldsymbol{\gamma}$ is a circle, we can obtain both the unit cylinder and the Möbius band.

Example 4.99: Unit Cylinder

As seen in Example 4.51, the cylinder is a surface with atlas $\mathscr{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$, where $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are suitable restriction of

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \cos(u), v) , \quad (u,v) \in [0, 2\pi) \times \mathbb{R} .$$

We have

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\mathbf{a}(u) ,$$

with

$$\boldsymbol{\gamma}(u) := (\cos(u), \cos(u), 0) , \quad \mathbf{a} = (0,0,1) .$$

Hence the unit cylinder is a ruled surface, see Figure 4.17.

Example 4.100: Möbius band

The Möbius band is a ruled surface with chart

$$\boldsymbol{\sigma} = \boldsymbol{\gamma}(u) + v\mathbf{a}(u) , \quad u \in (0, 2\pi), \ v \in \left(-\frac{1}{2}, \frac{1}{2}\right) ,$$

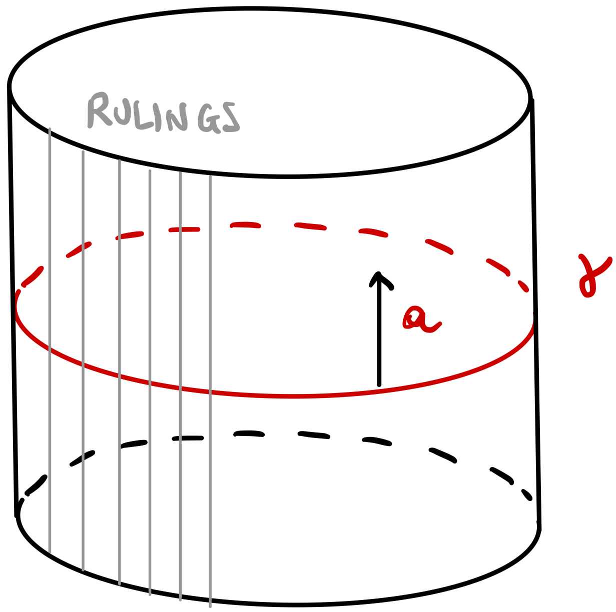


Figure 4.17: Unit cylinder is a ruled surface with base curve $\boldsymbol{\gamma}$ and rulings given by vertical lines.

where

$$\boldsymbol{\gamma}(u) = (\cos(u), \sin(u), 0)$$

is the unit circle and

$$\mathbf{a} = \left(-\sin\left(\frac{u}{2}\right)\cos(u), -\sin\left(\frac{u}{2}\right)\sin(u), \cos\left(\frac{u}{2}\right)\right)$$

is a vector which does a full rotation while going around the unit circle $\boldsymbol{\gamma}$. This is shown in Figure 4.18.

4.10.4 Surfaces of Revolution

Surfaces of revolution are obtained by rotating a curve about the z-axis.

Definition 4.101: Surface of revolution

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be a smooth curve in the (x,z)-plane, that is,

$$\boldsymbol{\gamma}(u) = (f(u), 0, g(u)) .$$

Suppose that $f > 0$. The surface obtained by rotating $\boldsymbol{\gamma}$ about the z-axis is called **surface of revolution**. A chart for $\mathcal{S}$ is given by

$$\boldsymbol{\sigma}(u,v) := (f(u)\cos(v), f(u)\sin(v), g(u)) , \; u \in (a,b) , \; v \in [0, 2\pi) .$$

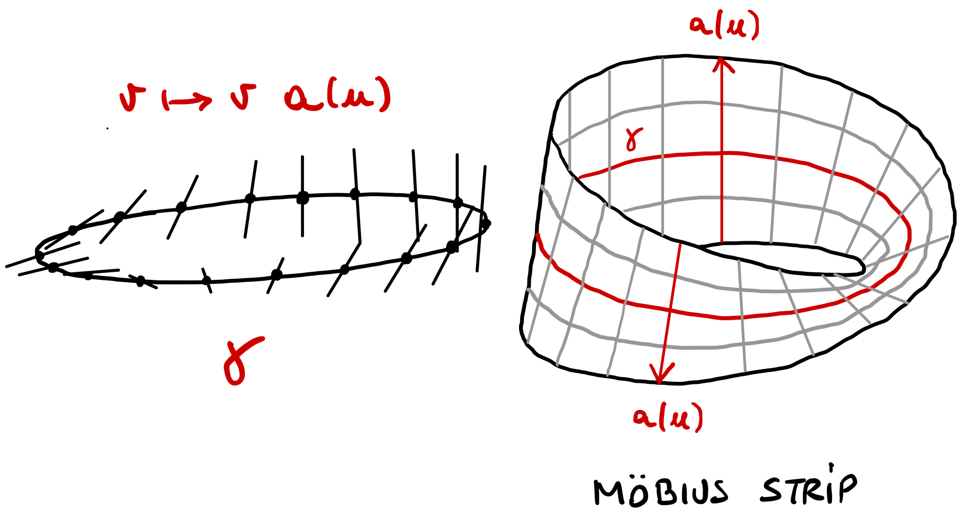


Figure 4.18: The Möbius band is a ruled surface with base curve $\boldsymbol{\gamma}$ and rulings given by rotating vertical lines.

Proposition 4.102

A surface of revolution is regular if and only if $\boldsymbol{\gamma}$ is regular.

Proof

We have

$$\begin{aligned}\boldsymbol{\sigma}_u &= \left(\dot{f}(u)\cos(v), \dot{f}(u)\sin(v), \dot{g}(u)\right) , \\ \boldsymbol{\sigma}_v &= (-f(u)\sin(v), f(u)\cos(v), 0) .\end{aligned}$$

Therefore

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \left(f\dot{g}\cos(v), -\dot{f}g\sin(v), f\dot{f}\right)$$

and

$$\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|^2 = f^2\left(\dot{f}^2 + \dot{g}^2\right) = f^2 \|\boldsymbol{\gamma}\|^2 .$$

Recall that $f > 0$ by definition, so that $f^2 \neq 0$. Therefore $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent if and only if $\boldsymbol{\gamma}$ is regular.

Example 4.103: Catenoid

The catenoid is the surface of revolution obtained by rotating the catenary about the z-axis, see Figure 4.19. Recall that the catenary function is defined by

$$f(u) = \cosh(u)\,.$$

Therefore the catenoid is obtained by rotating

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

A chart for the catenoid is given by

$$\boldsymbol{\sigma}(u, v) = (\cosh(u)\cos(v), \cosh(u)\sin(v), u)\,,$$

where $u \in \mathbb{R}$ and $v \in [0, 2\pi)$. Note that $f > 0$ and

$$\dot{\boldsymbol{\gamma}} = (\sinh(u), 0, 1)\ , \quad \|\dot{\boldsymbol{\gamma}}\|^2 = 1 + \sinh(u)^2 \geq 1\,.$$

Therefore $\boldsymbol{\gamma}$ is regular. By Proposition 4.102 we conclude that the catenoid is a regular surface.

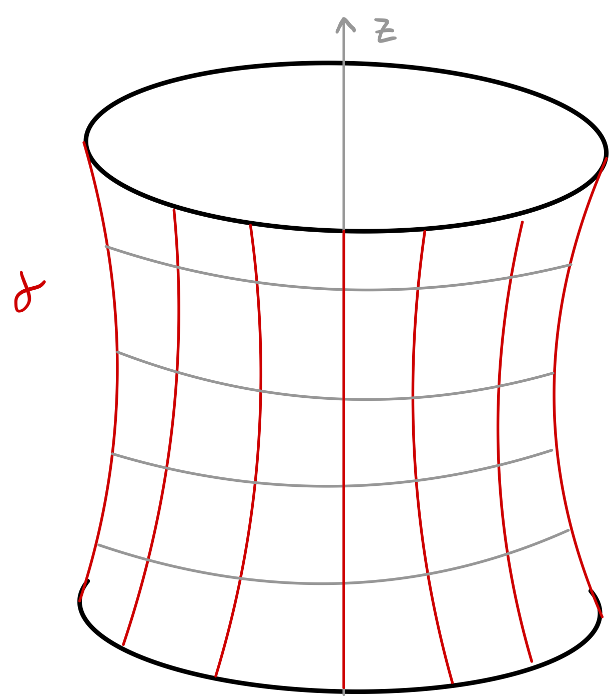


Figure 4.19: The Catenoid is the surface of revolution obtained by rotating the catenary about the z-axis.

4.11 First fundamental form

In this section we introduce the first **fundamental form** of a surface. This will allow us to compute:

- Angle between tangent vectors
- Lengths of tangent vectors
- Area of surface regions

4.11.1 Length and angles between tangent vectors

Let $\mathcal{S}$ be a surface and consider two points $\mathbf{p}, \mathbf{q} \in \mathcal{S}$. The euclidean distance between $\mathbf{p}$ and $\mathbf{q}$ is

$$\|\mathbf{p} - \mathbf{q}\| \,.$$

However this measures the length of the straight segment which connects $\mathbf{p}$ to $\mathbf{q}$. We are interested in measuring the distance of $\mathbf{p}$ and $\mathbf{q}$ on $\mathcal{S}$. A way to measure such distance is the following: Suppose

$$\boldsymbol{\gamma} \colon (t_0, t_1) \to \mathcal{S}$$

is a smooth curve such that

$$\boldsymbol{\gamma}(t_0) = \mathbf{p} \,, \quad \boldsymbol{\gamma}(t_1) = \mathbf{q} \,.$$

The distance between $\mathbf{p}$ and $\mathbf{q}$ on $\mathcal{S}$ is the length of $\boldsymbol{\gamma}$, i.e.,

$$\int_{t_0}^{t_1} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt \,.$$

Since $\boldsymbol{\gamma}(t) \in \mathcal{S}$, by definition we have

$$\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{x}}S \,, \quad \mathbf{x} := \boldsymbol{\gamma}(t) \,.$$

Therefore, computing $\|\dot{\boldsymbol{\gamma}}(t)\|$ is equivalent to computing the length of tangent vectors to $\mathcal{S}$. This motivates the definition of first fundamental form.

Definition 4.104: First fundamental form

Let $\mathcal{S}$ be a regular surface and $\mathbf{p} \in \mathcal{S}$. The **first fundamental form** of $\mathcal{S}$ at $\mathbf{p}$ is the bilinear symmetric map

$$I_{\mathbf{p}} \colon T_{\mathbf{p}}\mathcal{S} \times T_{\mathbf{p}}\mathcal{S} \to \mathbb{R} \,, \quad I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) := \mathbf{v} \cdot \mathbf{w} \,.$$

Three observations:

1. The first fundamental form of $\mathcal{S}$ at $\mathbf{p}$ is the map obtained by restricting the scalar product of $\mathbb{R}^3$ to $T_{\mathbf{p}}\mathcal{S}$.
2. Note that
$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{v}) = \|\mathbf{v}\|^2 \,,$$
so that $I_{\mathbf{p}}$ can be used to compute the length of tangent vectors.

3. The definition of $I_{\mathbf{p}}$ does not depend on a chosen chart, since $T_{\mathbf{p}}\mathcal{S}$ can be defined without using charts.

To use the first fundamental form in practice, we need to express $I_{\mathbf{p}}$ in terms of local charts. To this end, we first define the coordinates functions du and dv on $T_{\mathbf{p}}S$.

Definition 4.105: Coordinate functions on tangent plane

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart of $\mathcal{S}$. For each $\mathbf{p} \in \boldsymbol{\sigma}(U)$ we have

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\},$$

where $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are evaluated at the point $(u_0, v_0) \in U$ such that

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}.$$

Therefore, for each $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$, there exist $\lambda, \mu \in \mathbb{R}$ such that

$$\mathbf{v} = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v.$$

The **coordinate functions** on $T_{\mathbf{p}}\mathcal{S}$ are the linear maps

$$du, dv : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}, \quad du(\mathbf{v}) := \lambda, \quad dv(\mathbf{v}) := \mu.$$

Definition 4.106: First fundamental form of a chart

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart of $\mathcal{S}$. Define the functions

$$E, F, G : U \to \mathbb{R}$$

by setting

$$E := \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u, \quad F := \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v, \quad G := \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v.$$

Let $\mathbf{p} \in \boldsymbol{\sigma}(U)$ and denote by $(u_0, v_0) \in U$ the point such that

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}.$$

The **first fundamental form** of $\boldsymbol{\sigma}$ at $\mathbf{p}$ is the quadratic form

$$\mathcal{F}_1 : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$$

defined by

$$\mathcal{F}_1(\mathbf{v}) := E\, du^2(\mathbf{v}) + 2F\, du(\mathbf{v})\, dv(\mathbf{v}) + G\, dv^2(\mathbf{v}), \tag{4.6}$$

for all $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$, where E, F, G are evaluated at (u_0, v_0).

We usually omit the dependence on $\mathbf{v}$ in (4.6), and write

$$\mathcal{F}_1 = E\, du^2 + 2F\, du\, dv + G\, dv^2.$$

The quadratic form $\mathscr{F}_1$ is related to $I_{\mathbf{p}}$ in the following way.

Proposition 4.107

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart of $\mathcal{S}$, and $\mathbf{p} \in \boldsymbol{\sigma}(U)$. Then

$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) = (du(\mathbf{v}), dv(\mathbf{v})) \begin{pmatrix} E & F \\ F & G \end{pmatrix} (du(\mathbf{w}), dv(\mathbf{w}))^T,$$

for all $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$. In particular, $\mathscr{F}_1$ is the quadratic form associated to the symmetric bilinear form $I_{\mathbf{p}}$, that is,

$$\mathscr{F}_1(\mathbf{v}) = I_{\mathbf{p}}(\mathbf{v}, \mathbf{v}), \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}.$$

Proof

By Theorem 4.77 we have

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}.$$

Therefore, for $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$, there exist $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{R}$ such that

$$\mathbf{v} = \lambda_1 \boldsymbol{\sigma}_u + \mu_1 \boldsymbol{\sigma}_v, \quad \mathbf{w} = \lambda_2 \boldsymbol{\sigma}_u + \mu_2 \boldsymbol{\sigma}_v.$$

We have

$$\begin{aligned} I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) &= \mathbf{v} \cdot \mathbf{w} \\ &= \lambda_1 \lambda_2 \, \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v + (\lambda_1 \mu_2 + \lambda_2 \mu_1) \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v + \mu_1 \mu_2 \, \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v \\ &= E \, du(\mathbf{v}) du(\mathbf{w}) + F \, (du(\mathbf{v}) \, dv(\mathbf{w}) + du(\mathbf{w}) dv(\mathbf{v})) \\ &\qquad + G \, dv(\mathbf{v}) dv(\mathbf{w}) \\ &= (du(\mathbf{v}), dv(\mathbf{v})) \begin{pmatrix} E & F \\ F & G \end{pmatrix} (du(\mathbf{w}), dv(\mathbf{w}))^T. \end{aligned}$$

The fact that

$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{v}) = \mathscr{F}_1(\mathbf{v})$$

follows from the first part of the statement and definition of $\mathscr{F}_1$.

Remark 4.108: Linear algebra interpretation

Using linear algebra, Proposition 4.107 has a clear interpretation, as follows. $I_{\mathbf{p}}$ is a symmetric bilinear form on the vector space $T_{\mathbf{p}}\mathcal{S}$. Fixing the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ for $T_{\mathbf{p}}\mathcal{S}$, we can represent $I_{\mathbf{p}}$ via the matrix

$$\begin{aligned} M &:= \begin{pmatrix} I_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u) & I_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v) \\ I_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_u) & I_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v) \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u & \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v \\ \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_u & \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v \end{pmatrix} \\ &= \begin{pmatrix} E & F \\ F & G \end{pmatrix}, \end{aligned}$$

where we used that $\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_u$.

Notation

With a little abuse of notation, we also denote by $\mathscr{F}_1$ the 2×2 matrix

$$\mathscr{F}_1 := \begin{pmatrix} E & F \\ F & G \end{pmatrix}.$$

Remark 4.109: First fundamental form and reparametrizations

The first fundamental form $I_{\mathbf{p}}$ depends only on the surface $\mathcal{S}$ and the point $\mathbf{p}$. Instead the representation of $I_{\mathbf{p}}$

$$\mathscr{F}_1 = E\,du^2 + 2F\,dudv + G\,dv^2$$

depends on the choice of chart $\boldsymbol{\sigma} : U \to \mathbb{R}^3$. Indeed suppose that $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3$ is a reparametrization of $\boldsymbol{\sigma}$, that is,

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi,$$

where $\Phi : \widetilde{U} \to U$ is a diffeomorphism. Recall that we denote the components Φ^1 and Φ^2 of Φ by

$$(\tilde{u}, \tilde{v}) \mapsto u(\tilde{u}, \tilde{v}), \quad (\tilde{u}, \tilde{v}) \mapsto v(\tilde{u}, \tilde{v}),$$

respectively. The Jacobian of Φ is then

$$J\Phi = \begin{pmatrix} \frac{\partial u}{\partial \tilde{u}} & \frac{\partial u}{\partial \tilde{v}} \\ \frac{\partial v}{\partial \tilde{u}} & \frac{\partial v}{\partial \tilde{v}} \end{pmatrix}.$$

Denote the first fundamental form of $\tilde{\boldsymbol{\sigma}}$ by

$$\widetilde{\mathscr{F}}_1 = \widetilde{E}\,d\tilde{u}^2 + 2\widetilde{F}\,d\tilde{u}d\tilde{v} + \widetilde{G}\,d\tilde{v}^2.$$

The linear maps du, dv and $d\tilde{u}, d\tilde{v}$ are related by

$$du = \frac{\partial u}{\partial \tilde{u}}\,d\tilde{u} + \frac{\partial u}{\partial \tilde{v}}\,d\tilde{v}, \quad dv = \frac{\partial v}{\partial \tilde{u}}\,d\tilde{u} + \frac{\partial v}{\partial \tilde{v}}\,d\tilde{v} \tag{4.7}$$

Moreover the matrices of $\mathscr{F}_1$ and $\widetilde{\mathscr{F}}_1$ are related by

$$\begin{pmatrix} \widetilde{E} & \widetilde{F} \\ \widetilde{F} & \widetilde{G} \end{pmatrix} = (J\Phi)^T \begin{pmatrix} E & F \\ F & G \end{pmatrix} J\Phi \,. \tag{4.8}$$

The proof of the above statements follows by basic linear algebra: The pairs $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u\}$ and $\{\tilde{\boldsymbol{\sigma}}_{\tilde{u}}, \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\}$ are bases for the vector space $T_{\mathbf{p}}\mathcal{S}$. The change of basis matrix is given exactly by $J\Phi$. Therefore formulas (4.7) and (4.8) are consequence of change of basis results for linear maps and bilinear forms, respectively.

Let us compute the first fundamental form of a plane and of a cylinder.

Example 4.110: Plane

Let $\mathbf{a}, \mathbf{p}, \mathbf{q} \in \mathbb{R}^3$. Suppose that $\mathbf{p}$ and $\mathbf{q}$ are orthonormal vectors, that is,

$$\|\mathbf{p}\| = \|\mathbf{q}\| = 1 \,, \quad \mathbf{p} \cdot \mathbf{q} = 0 \,.$$

Consider the plane with chart

$$\boldsymbol{\sigma}(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q} \,, \quad (u, v) \in \mathbb{R}^2 \,.$$

The first fundamental form of $\boldsymbol{\sigma}$ is

$$\mathscr{F}_1 = du^2 + dv^2 \,.$$

We have

$$\boldsymbol{\sigma}_u = \mathbf{p} \,, \quad \boldsymbol{\sigma}_v = \mathbf{q}$$

and therefore

$$\begin{aligned} E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = \|\mathbf{p}\|^2 = 1 \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = \mathbf{p} \cdot \mathbf{q} = 0 \\ G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = \|\mathbf{q}\|^2 = 1 \end{aligned}$$

Then the first fundamental form is

$$\mathscr{F}_1 = E\, du^2 + 2F\, du\, dv + G\, dv^2 = du^2 + dv^2 \,.$$

Two remarks concerning Example 4.108 :

- The above example should not be surprising, since distances on a plane are the same as Euclidean distances, given that straight segments are contained in the plane.

- If we drop the assumption of $\mathbf{p}$ and $\mathbf{q}$ being orthonormal, then
$$\mathscr{F}_1 = \|\mathbf{p}\|^2 \, du^2 + \mathbf{p}\cdot\mathbf{q}\, du\, dv + \|\mathbf{q}\|^2 \, dv^2 .$$
Again, this is not surprising, due to Remark 4.109.

Example 4.111: Unit cylinder

Consider the unit cylinder with chart
$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v), \quad (u,v) \in (0,2\pi) \times \mathbb{R} .$$
The first fundamental form of $\boldsymbol{\sigma}$ is
$$\mathscr{F}_1 = du^2 + dv^2 .$$

We have
$$\boldsymbol{\sigma}_u = (-\sin(u), \cos(u), 0), \quad \boldsymbol{\sigma}_v = (0,0,1),$$
and therefore
$$\begin{aligned} E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\ G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1 \end{aligned}$$

Then the first fundamental form is
$$\mathscr{F}_1 = E\, du^2 + 2F\, du\, dv + G\, dv^2 = du^2 + dv^2 .$$

Remark 4.112

We have seen that a plane and the unit cylinder have the same first fundamental form
$$\mathscr{F}_1 = du^2 + dv^2 .$$
Therefore lengths and angles are the same on the two surfaces.

4.11.2 Length of curves

Let us show how the first fundamental form allows to compute the length of curves with values on surfaces.

Proposition 4.113

Let $\mathcal{S}$ be a regular surface with chart $\boldsymbol{\sigma} : U \to \mathbb{R}^3$. Suppose
$$\boldsymbol{\gamma} : (t_0, t_1) \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S}$$

is a smooth curve. Then
$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)),$$
for some smooth functions $u, v : (t_0, t_1) \to \mathbb{R}$ and
$$\int_{t_0}^{t_1} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt = \int_{t_0}^{t_1} \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2} \, dt,$$
where $\dot{u}, \dot{v}$ are computed at t, and E, F, G are computed at $(u(t), v(t))$.

Proof

Since $\boldsymbol{\gamma}$ takes values into $\boldsymbol{\sigma}(U)$, by Lemma 4.76 there exist smooth functions u, v such that
$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)), \quad \forall\, t \in (t_0, t_1).$$
By chain rule we have
$$\dot{\boldsymbol{\gamma}}(t) = \dot{u}(t)\boldsymbol{\sigma}_u(u(t), v(t)) + \dot{v}(t)\boldsymbol{\sigma}_v(u(t), v(t)).$$
The above means that the coefficients of $\dot{\boldsymbol{\gamma}}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$ are $\dot{u}$ and $\dot{v}$, respectively. Then
$$du(\dot{\boldsymbol{\gamma}}) = \dot{u}, \quad dv(\dot{\boldsymbol{\gamma}}) = \dot{v}.$$
Since $\dot{\boldsymbol{\gamma}}$ is a tagent vector, by Proposition 4.107 we get
$$\begin{aligned} \|\dot{\boldsymbol{\gamma}}(t)\|^2 &= \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} \\ &= I_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}}) \\ &= E\, du(\dot{\boldsymbol{\gamma}})^2 + 2F\, du(\dot{\boldsymbol{\gamma}}) dv(\boldsymbol{\gamma}) + G\, dv(\dot{\boldsymbol{\gamma}})^2 \\ &= E\, \dot{u}^2 + 2F\, \dot{u}\dot{v} + G\, \dot{v}^2, \end{aligned}$$
concluding the proof.

Example 4.114: Cone

Consider the cone with chart
$$\boldsymbol{\sigma}(u, v) = (u\cos(v), u\sin(v), u),$$
where $u > 0$ and $v \in [0, 2\pi]$.

1. The first fundamental form of $\boldsymbol{\sigma}$ is
$$\mathcal{F}_1 = 2\, du^2 + u^2\, dv^2.$$

2. Let $\boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(t,t)$. The length of $\boldsymbol{\gamma}$ is

$$\int_{\pi/2}^{\pi} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt = \int_{\pi/2}^{\pi} \sqrt{2+t^2} \, dt \, .$$

We have

$$\boldsymbol{\sigma}_u = (\cos(v), \sin(v), 1) \, , \quad \boldsymbol{\sigma}_v = (-u\sin(v), u\cos(v), 0) \, .$$

Therefore

$$\begin{aligned} E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = \cos^2(v) + \sin^2(v) + 1 = 2 \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = -u\cos(v)\sin(v) + u\cos(v)\sin(v) = 0 \\ G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = u^2\sin^2(v) + u^2\cos^2(v) = u^2 \end{aligned}$$

The first fundamental form of $\boldsymbol{\sigma}$ is

$$\mathscr{F}_1 = 2\, du^2 + u^2\, dv^2 \, .$$

Concering the curve $\boldsymbol{\gamma}$, by definition we have

$$\boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(t,t) \, ,$$

so that

$$u(t) = t \, , \quad v(t) = t \, .$$

In particular

$$\dot{u} = 1 \, , \quad \dot{v} = 1$$

and

$$\begin{aligned} E(u(t), v(t)) &= E(t,t) = 2 \\ F(u(t), v(t)) &= F(t,t) = 0 \\ G(u(t), v(t)) &= G(t,t) = t^2 \, . \end{aligned}$$

By Proposition 4.113 we have

$$\begin{aligned} \int_{\pi/2}^{\pi} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt &= \int_{\pi/2}^{\pi} \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2} \, dt \\ &= \int_{\pi/2}^{\pi} \sqrt{2+t^2} \, dt \, . \end{aligned}$$

4.11.3 Local isometries

We have seen that a plane $\boldsymbol{\pi}$ and a cylinder $\mathscr{C}$ have the same first fundamental form. This means that scalar product on the two surfaces is the same, as is the length of curves. In this case we say that $\boldsymbol{\pi}$ and $\mathscr{C}$ are locally isometric. Let us give a general definition of such concept.

Definition 4.115: Local isometry

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces. A local diffeomorphism $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ is a **local isometry** if for all $\mathbf{p} \in \mathcal{S}$ the differential $d_{\mathbf{p}} f : T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\widetilde{\mathcal{S}}$ satisfies

$$\mathbf{v} \cdot \mathbf{w} = d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{w}) , \quad \forall \, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S} .$$

We say that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are **locally isometric** if there exists a local isometry $f : \mathcal{S} \to \widetilde{\mathcal{S}}$.

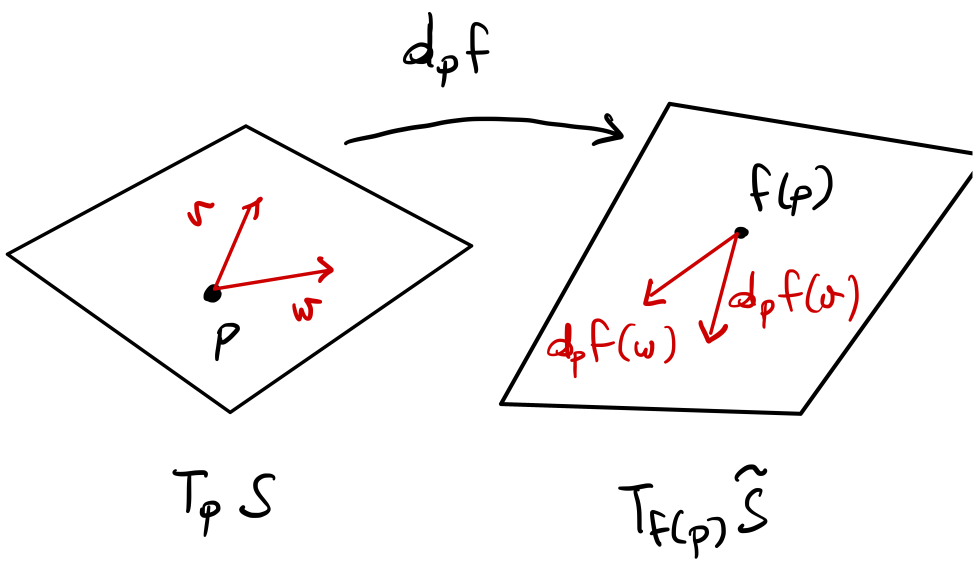


Figure 4.20: Sketch of local isometry f between $\mathcal{S}$ and $\widetilde{\mathcal{S}}$. The scalar product between tangent vectors $\mathbf{v}$ and $\mathbf{w}$ is preserved by $d_{\mathbf{p}} f$.

Notation

For brevity we denote

$$\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v} \cdot \mathbf{w} , \quad \langle \mathbf{v}, \mathbf{w} \rangle_f := d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{w}) ,$$

and also

$$\|\mathbf{v}\| := \sqrt{\langle v, v \rangle} , \quad \|\mathbf{v}\|_f := \sqrt{\langle v, v \rangle_f} .$$

Remark 4.116

A local diffeomorphism $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ is a local isometry if and only if

$$\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle_f \,, \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S} \,.$$

The proof follows from the elementary identity

$$\mathbf{v} \cdot \mathbf{w} = \frac{1}{2} \left((\mathbf{v} + \mathbf{w}) \cdot (\mathbf{v} + \mathbf{w}) - \mathbf{v} \cdot \mathbf{v} - \mathbf{w} \cdot \mathbf{w} \right) \,,$$

which holds for all $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ (and more in general in arbitrary vector spaces with inner product).

Local isometries preserve the length of curves, as shown in the following proposition.

Proposition 4.117

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ be a local diffeomorphism. They are equivalent:

1. f is a local isometry.
2. Let $\boldsymbol{\gamma}$ be a curve in $\mathcal{S}$ and consider the curve $\tilde{\boldsymbol{\gamma}} = f \circ \boldsymbol{\gamma}$ on $\widetilde{\mathcal{S}}$. Then $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same length.

Proof

Part 1. Suppose $\boldsymbol{\gamma} : (t_0, t_1) \to \mathcal{S}$ is a smooth curve. Consider the smooth curve $\tilde{\boldsymbol{\gamma}} := f \circ \boldsymbol{\gamma} : (t_0, t_1) \to \widetilde{\mathcal{S}}$. Set $\mathbf{p} := \boldsymbol{\gamma}(t)$. By definition of differential of a function between surfaces, we have

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) = d f_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}(t)) \,.$$

Using that f is a local isometry gives:

$$\begin{aligned}
\left\| \dot{\tilde{\boldsymbol{\gamma}}}(t) \right\|^2 & = \dot{\tilde{\boldsymbol{\gamma}}}(t) \cdot \dot{\tilde{\boldsymbol{\gamma}}}(t) \\
& = d f_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}(t)) \cdot d f_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}(t)) \\
& = \dot{\boldsymbol{\gamma}}(t) \cdot \dot{\boldsymbol{\gamma}}(t) \\
& = \left\| \dot{\boldsymbol{\gamma}}(t) \right\|^2
\end{aligned}$$

Therefore $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same length:

$$\int_{t_0}^{t_1} \left\| \dot{\tilde{\boldsymbol{\gamma}}}(t) \right\| \, dt = \int_{t_0}^{t_1} \left\| \dot{\boldsymbol{\gamma}}(t) \right\| \, dt \,.$$

Part 2. We need to prove that f is a local isometry. Thanks to Remark 4.116, it is sufficient to show that

$$d f_{\mathbf{p}}(\mathbf{v}) \cdot d f_{\mathbf{p}}(\mathbf{v}) = \mathbf{v} \cdot \mathbf{v} \,, \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}(\mathcal{S}) \,. \tag{4.9}$$

Therefore, let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$ be arbitrary. By definition of tangent plane, there exists a curve $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}$ such that

$$\boldsymbol{\gamma}(0) = \mathbf{p}\,, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v}\,.$$

Define the curve $\tilde{\boldsymbol{\gamma}} := f \circ \boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \widetilde{\mathcal{S}}$. By assumption $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same length, that is,

$$\int_{-\varepsilon}^{\varepsilon} \sqrt{\dot{\tilde{\boldsymbol{\gamma}}}(t) \cdot \dot{\tilde{\boldsymbol{\gamma}}}(t)}\, dt = \int_{-\varepsilon}^{\varepsilon} \sqrt{\dot{\boldsymbol{\gamma}}(t) \cdot \dot{\boldsymbol{\gamma}}(t)}\, dt\,.$$

Since the above is true for each $\varepsilon > 0$, and the functions integrated are continuous, we infer

$$\dot{\tilde{\boldsymbol{\gamma}}}(0) \cdot \dot{\tilde{\boldsymbol{\gamma}}}(0) = \dot{\boldsymbol{\gamma}}(0) \cdot \dot{\boldsymbol{\gamma}}(0)\,.$$

Recall that by definition of differential we have

$$d f_{\mathbf{p}}(\mathbf{v}) = \dot{\tilde{\boldsymbol{\gamma}}}(0)\,.$$

Therefore

$$\begin{aligned} d f_{\mathbf{p}}(\mathbf{v}) \cdot d f_{\mathbf{p}}(\mathbf{v}) &= \dot{\tilde{\boldsymbol{\gamma}}}(0) \cdot \dot{\tilde{\boldsymbol{\gamma}}}(0) \\ &= \dot{\boldsymbol{\gamma}}(0) \cdot \dot{\boldsymbol{\gamma}}(0) \\ &= \mathbf{v} \cdot \mathbf{v}\,. \end{aligned}$$

As $\mathbf{v}$ was arbitrary, we conclude (4.9).

We have seen that local isometries preserve the length of curves. Recall that the first fundamental form of $\mathcal{S}$ is defined by

$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot \mathbf{w}\,, \quad \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}\,.$$

A natural question is how the first fundamental form changes under local isometries. The result is that local isometries preserve the first fundamental form.

Theorem 4.118

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ be a local diffeomorphism. They are equivalent:

1. f is a local isometry.
2. Let $\boldsymbol{\sigma} : U \to \mathcal{S}$ be a regular chart of $\mathcal{S}$ and consider the chart of $\widetilde{\mathcal{S}}$ given by

$$\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma} : U \to \widetilde{\mathcal{S}}\,.$$

   Then $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same first fundamental form, that is,

$$E = \widetilde{E}\,, \quad F = \widetilde{F}\,, \quad G = \widetilde{G}\,,$$

where

$$E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u \,, \quad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v \,, \quad G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v \,,$$
$$\widetilde{E} = \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u \,, \quad \widetilde{F} = \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v \,, \quad \widetilde{G} = \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v \,.$$

Note that E, F, G and $\widetilde{E}, \widetilde{F}, \widetilde{G}$ are defined on the same set U. Therefore equality is intended pointwise.

Proof

Part 1. Suppose that f is a local isometry, that is,

$$\mathbf{v} \cdot \mathbf{w} = d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{w}) \,, \quad \forall\, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}} \mathcal{S} \,.$$

Let $\boldsymbol{\sigma}$ be a chart for $\mathcal{S}$ at $\mathbf{p}$. Define $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$. By Proposition 4.71, $\tilde{\boldsymbol{\sigma}}$ is a regualar chart of $\widetilde{\mathcal{S}}$ at $f(\mathbf{p})$. Now, recall the statement of Proposition 4.87: if

$$\tilde{\boldsymbol{\sigma}}(\alpha(u, v), \beta(u, v)) = f(\boldsymbol{\sigma}(u, v)) \,,$$

for some smooth maps

$$\alpha, \beta : U \to \widetilde{U} \,,$$

then the matrix of $d_{\mathbf{p}} f$ with respect to the basis

$$\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} \quad \text{of} \quad T_{\mathbf{p}} \mathcal{S} \,, \quad \{\tilde{\boldsymbol{\sigma}}_u, \tilde{\boldsymbol{\sigma}}_v\} \quad \text{of} \quad T_{f(\mathbf{p})} \widetilde{\mathcal{S}} \,,$$

is given by

$$d_{\mathbf{p}} f = \begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix} .$$

In our case, we have $U = \widetilde{U}$ and

$$\tilde{\boldsymbol{\sigma}}(u, v) = f(\boldsymbol{\sigma}(u, v)) \,,$$

so that

$$\alpha(u, v) = u \,, \quad \beta(u, v) = v \,.$$

Therefore

$$d_{\mathbf{p}} f = \begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} ,$$

which means that

$$d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) = 1 \cdot \tilde{\boldsymbol{\sigma}}_u + 0 \cdot \tilde{\boldsymbol{\sigma}}_v = \tilde{\boldsymbol{\sigma}}_u$$
$$d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) = 0 \cdot \tilde{\boldsymbol{\sigma}}_u + 1 \cdot \tilde{\boldsymbol{\sigma}}_v = \tilde{\boldsymbol{\sigma}}_v$$

Using that f is a local isometry gives

$$\begin{aligned} E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u &= d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) \cdot d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) \\ &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u = \widetilde{E}\,. \end{aligned}$$

Simlarly, we obtain also

$$\begin{aligned} F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v &= d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) \cdot d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) \\ &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v = \widetilde{F}\,, \end{aligned}$$

and

$$\begin{aligned} G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v &= d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) \cdot d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) \\ &= \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v = \widetilde{G}\,, \end{aligned}$$

showing that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same first fundamental form.
Part 2. Define $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ and suppose that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same first fundamental form. In particular they hold

$$\begin{aligned} \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u \\ \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v \\ \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v &= \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v \end{aligned}$$

As discussed above, since $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$, by Proposition 4.87 we get

$$d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) = \tilde{\boldsymbol{\sigma}}_u\,, \quad d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) = \tilde{\boldsymbol{\sigma}}_v\,.$$

Let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. Since $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis for $T_{\mathbf{p}}\mathcal{S}$ we get

$$\mathbf{v} = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v$$

for some $\lambda, \mu \in \mathbb{R}$. Therefore

$$\begin{aligned} d_{\mathbf{p}} f(\mathbf{v}) &= d_{\mathbf{p}} f(\lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v) \\ &= \lambda\, d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) + \mu\, d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) \\ &= \lambda \tilde{\boldsymbol{\sigma}}_u + \mu \tilde{\boldsymbol{\sigma}}_v\,. \end{aligned}$$

Hence

$$\begin{aligned} \mathbf{v} \cdot \mathbf{v} &= (\lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v) \cdot (\lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v) \\ &= \lambda^2 (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v) + 2\lambda\mu (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v) + \mu^2 (\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v) \\ &= \lambda^2 (\tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u) + 2\lambda\mu (\tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v) + \mu^2 (\tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v) \\ &= (\lambda \tilde{\boldsymbol{\sigma}}_u + \mu \tilde{\boldsymbol{\sigma}}_v) \cdot (\lambda \tilde{\boldsymbol{\sigma}}_u + \mu \tilde{\boldsymbol{\sigma}}_v) \\ &= d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{v})\,, \end{aligned}$$

showing that

$$\mathbf{v} \cdot \mathbf{v} = d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{v})\,, \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}\,.$$

By Remark 4.116 we conclude that f is a local isometry.

4.11.4 Angles on surfaces

We want to define the notion of angle between tangent vectors.

Definition 4.119: Angle between tangent vectors

Let $\mathcal{S}$ be a regular surface and $\mathbf{p} \in \mathcal{S}$. The angle between two vectors $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ is defined as the number θ such that

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \, \|\mathbf{w}\|} .$$

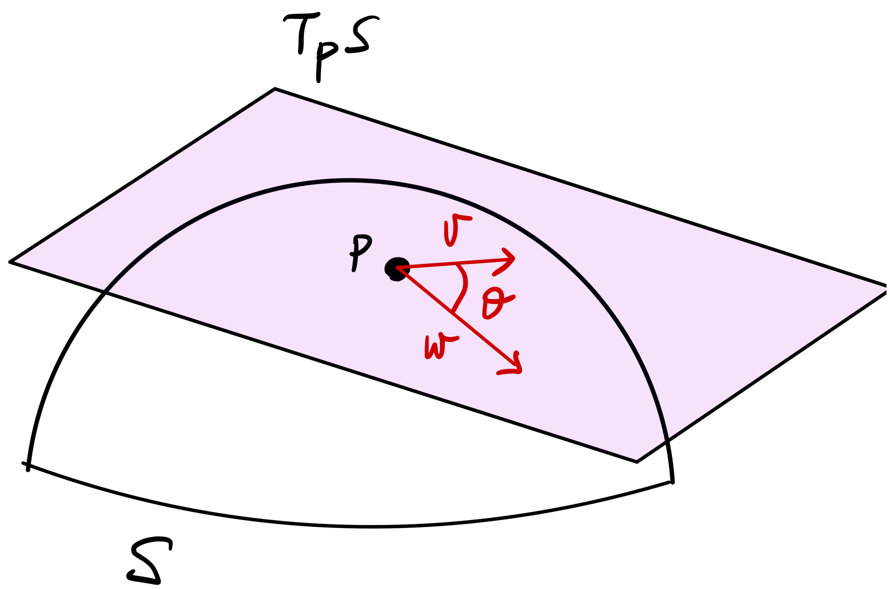


Figure 4.21: Sketch of angle θ between two vectors $\mathbf{v}, \mathbf{w}$ in $T_{\mathbf{p}}\mathcal{S}$.

The angle between tangent vectors can be computed in terms of local charts.

Proposition 4.120

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\sigma}$ a regular chart at $\mathbf{p}$. Let $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$. Then

$$\cos(\theta) = \frac{E\lambda\tilde{\lambda} + F(\lambda\tilde{\mu} + \tilde{\lambda}\mu) + G\mu\tilde{\mu}}{(E\lambda^2 + 2F\lambda\mu + G\mu^2)^{1/2}(E\tilde{\lambda}^2 + 2F\tilde{\lambda}\tilde{\mu} + G\tilde{\mu}^2)^{1/2}} ,$$

where $\lambda, \mu, \tilde{\lambda}, \tilde{\mu} \in \mathbb{R}$ are such that

$$\mathbf{v} = \lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v , \quad \mathbf{w} = \tilde{\lambda}\boldsymbol{\sigma}_u + \tilde{\mu}\boldsymbol{\sigma}_v .$$

Proof

By definition the angle between $\mathbf{v}$ and $\mathbf{w}$ is

$$\cos(\theta) = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \, \|\mathbf{w}\|} . \tag{4.10}$$

The vectors $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ form a basis of $T_{\mathbf{p}}\mathcal{S}$. Therefore

$$\mathbf{v} = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v , \quad \mathbf{w} = \tilde{\lambda} \boldsymbol{\sigma}_u + \tilde{\mu} \boldsymbol{\sigma}_v .$$

for some $\lambda, \mu, \tilde{\lambda}, \tilde{\mu} \in \mathbb{R}$. Hence, the coordinates of $\mathbf{v}$ and $\mathbf{w}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ are

$$\mathbf{v} = (\lambda, \mu) , \quad \mathbf{w} = (\tilde{\lambda}, \tilde{\mu}) .$$

By Proposition 4.107 we get

$$\begin{aligned}
\mathbf{v} \cdot \mathbf{w} & = I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) \\
& = (\lambda, \mu) \begin{pmatrix} E & F \\ F & G \end{pmatrix} (\tilde{\lambda}, \tilde{\mu})^T \\
& = E\lambda\tilde{\lambda} + F(\lambda\tilde{\mu} + \tilde{\lambda}\mu) + G\mu\tilde{\mu} .
\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v} = E\lambda^2 + 2F\lambda\mu + G\mu^2 \\
\|\mathbf{w}\|^2 = \mathbf{w} \cdot \mathbf{w} = E\tilde{\lambda}^2 + 2F\tilde{\lambda}\tilde{\mu} + G\tilde{\mu}^2 .
\end{aligned}$$

Substituting in (4.10) we conclude.

4.11.5 Angle between curves

Since tangent vectors are derivatives of curves with values in $\mathcal{S}$, it also makes sense to define the angle between two intersecting curves.

Definition 4.121: Angle between curves

Let $\mathcal{S}$ be a regular surface and suppose to have two curves

$$\boldsymbol{\gamma} : (a, b) \to \mathcal{S} , \quad \tilde{\boldsymbol{\gamma}} : (\tilde{a}, \tilde{b}) \to \mathcal{S}$$

such that

$$\boldsymbol{\gamma}(t_0) = \mathbf{p} , \quad \tilde{\boldsymbol{\gamma}}(\tilde{t}_0) = \mathbf{p} .$$

Then

$$\dot{\boldsymbol{\gamma}}(t_0) , \dot{\tilde{\boldsymbol{\gamma}}}(\tilde{t}_0) \in T_{\mathbf{p}}\mathcal{S} .$$

The angle θ between $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ is the angle between $\dot{\boldsymbol{\gamma}}(t_0)$ and $\dot{\tilde{\boldsymbol{\gamma}}}(\tilde{t}_0)$, that is,

$$\cos(\theta) = \frac{\dot{\boldsymbol{\gamma}} \cdot \dot{\tilde{\boldsymbol{\gamma}}}}{\|\dot{\boldsymbol{\gamma}}\| \, \|\dot{\tilde{\boldsymbol{\gamma}}}\|} \,,$$

where $\tilde{\boldsymbol{\gamma}}$ is evaluated at t_0 and $\dot{\tilde{\boldsymbol{\gamma}}}$ at $\tilde{t}_0$.

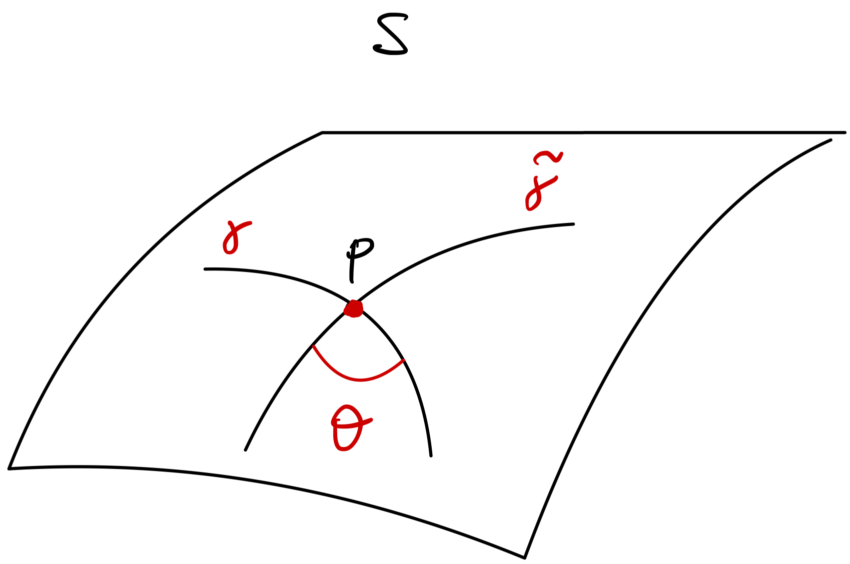


Figure 4.22: Sketch of angle θ between two curves $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ on $\mathcal{S}$.

Proposition 4.122

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\sigma}$ a regular chart at $\mathbf{p}$. Suppose given two curves

$$\boldsymbol{\gamma} \colon (a,b) \to \mathcal{S} \,, \quad \tilde{\boldsymbol{\gamma}} \colon (\tilde{a},\tilde{b}) \to \mathcal{S}$$

such that

$$\boldsymbol{\gamma}(t_0) = \mathbf{p} \,, \quad \tilde{\boldsymbol{\gamma}}(\tilde{t}_0) = \mathbf{p} \,.$$

The angle between $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ is

$$\cos(\theta) = \frac{E\dot{u}\dot{\tilde{u}} + F(\dot{u}\dot{\tilde{v}} + \dot{\tilde{u}}\dot{v}) + G\dot{v}\dot{\tilde{v}}}{(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2}(E\dot{\tilde{u}}^2 + 2F\dot{\tilde{u}}\dot{\tilde{v}} + G\dot{\tilde{v}}^2)^{1/2}} \,,$$

where $u, v, \tilde{u}, \tilde{v}$ are smooth functions such that

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)) \,, \quad \tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}(\tilde{u}(t), \tilde{v}(t)) \,.$$

Proof

By definition the angle between $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ is

$$\cos(\theta) = \frac{\dot{\boldsymbol{\gamma}} \cdot \dot{\tilde{\boldsymbol{\gamma}}}}{\|\dot{\boldsymbol{\gamma}}\| \, \|\dot{\tilde{\boldsymbol{\gamma}}}\|} . \tag{4.11}$$

As $\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}$ are smooth curves with values in $\mathcal{S}$, by Lemma 4.76 there exist smooth functions $u, v, \tilde{u}, \tilde{v}$ such that

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)) , \quad \tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}(\tilde{u}(t), \tilde{v}(t)) .$$

Differentiating the above expressions we obtain

$$\dot{\boldsymbol{\gamma}} = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v , \quad \dot{\tilde{\boldsymbol{\gamma}}} = \dot{\tilde{u}}\boldsymbol{\sigma}_u + \dot{\tilde{v}}\boldsymbol{\sigma}_v .$$

Therefore the coordinates of $\dot{\boldsymbol{\gamma}}$ and $\dot{\tilde{\boldsymbol{\gamma}}}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$ are

$$\dot{\boldsymbol{\gamma}} = (\dot{u}, \dot{v}) , \quad \dot{\tilde{\boldsymbol{\gamma}}} = (\dot{\tilde{u}}, \dot{\tilde{v}}) .$$

By Proposition 4.107 we get

$$\begin{aligned} \dot{\boldsymbol{\gamma}} \cdot \dot{\tilde{\boldsymbol{\gamma}}} &= I_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}, \dot{\tilde{\boldsymbol{\gamma}}}) \\ &= (\dot{u}, \dot{v}) \begin{pmatrix} E & F \\ F & G \end{pmatrix} (\dot{\tilde{u}}, \dot{\tilde{v}})^T \\ &= E\dot{u}\dot{\tilde{u}} + F(\dot{u}\dot{\tilde{v}} + \dot{\tilde{u}}\dot{v}) + G\dot{v}\dot{\tilde{v}} . \end{aligned}$$

Similarly, we obtain

$$\begin{aligned} \|\dot{\boldsymbol{\gamma}}\|^2 &= \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2 \\ \|\dot{\tilde{\boldsymbol{\gamma}}}\|^2 &= \dot{\tilde{\boldsymbol{\gamma}}} \cdot \dot{\tilde{\boldsymbol{\gamma}}} = E\dot{\tilde{u}}^2 + 2F\dot{\tilde{u}}\dot{\tilde{v}} + G\dot{\tilde{v}}^2 . \end{aligned}$$

Substituting in (4.11) we conclude.

4.11.6 Conformal maps

Local isometries are maps which preserve the **scalar product** between tangent vectors. We want to consider maps which preserve the **angle** between tangent vectors. These will be called **conformal maps**.

Definition 4.123: Conformal map

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces. A local diffeomorphism $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ is a **conformal mapping** if for all $\mathbf{p} \in \mathcal{S}$ and $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ is holds

$$\theta = \tilde{\theta} ,$$

with θ, $\tilde{\theta}$ the angles between $\mathbf{v}, \mathbf{w}$ and $d_{\mathbf{p}} f(\mathbf{v}), d_{\mathbf{p}} f(\mathbf{w})$, respectively.

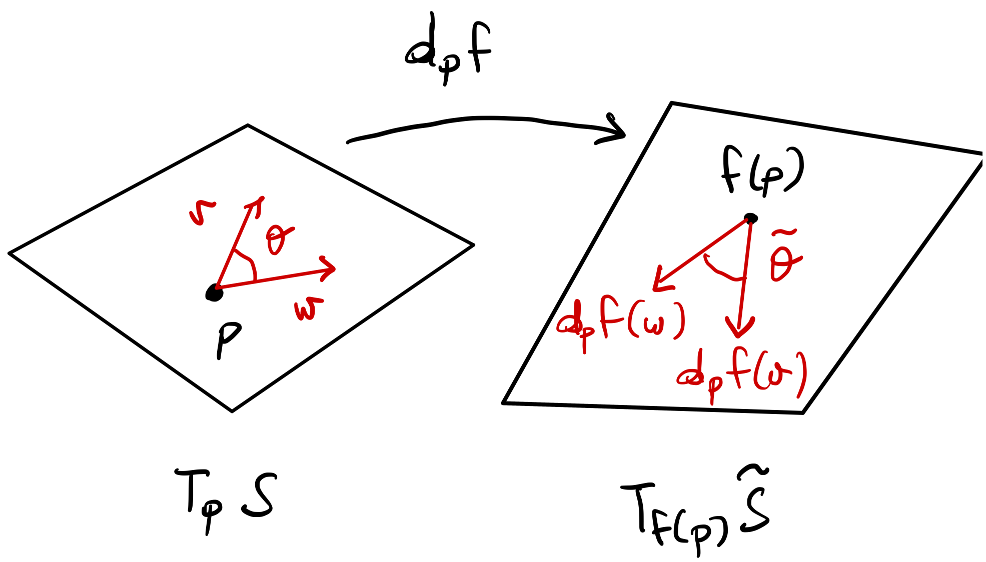


Figure 4.23: Sketch of conformal map f between $\mathcal{S}$ and $\widetilde{\mathcal{S}}$. The angles between tangent vectors are preserved by $d_{\mathbf{p}} f$.

Remark 4.124

We have that f is a conformal map if and only if

$$\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \, \|\mathbf{w}\|} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle_f}{\|\mathbf{v}\|_f \, \|\mathbf{w}\|_f} , \quad \forall \, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}} \mathcal{S} .$$

This follows immediately by the definition of angle between tangent vectors.

Proposition 4.125

Let f be a local isometry. Then f is a conformal map.

Proof

By definition of local isometry we have

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle_f \,, \quad \forall \, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S} \,.$$

In particular we have

$$\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle_f = \|\mathbf{v}\|_f^2 \,,$$

for all $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. Therefore

$$\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \, \|\mathbf{w}\|} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle_f}{\|\mathbf{v}\|_f \, \|\mathbf{w}\|_f} \,,$$

showing that f is a conformal map.

Therefore every local isometry is a conformal map. The converse is false, as we will show in Example 4.129 below. Before giving the example, let us provide a characterization of conformal maps in terms of the first fundamental form.

Theorem 4.126

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f \colon \mathcal{S} \to \widetilde{\mathcal{S}}$ a local diffeomorphism. They are equivalent:

1. f is a conformal map.
2. There exists a function $\lambda \colon \mathcal{S} \to \mathbb{R}$ such that

$$\langle \mathbf{v}, \mathbf{w} \rangle_f = \lambda(\mathbf{p}) \, \langle \mathbf{v}, \mathbf{w} \rangle \,, \quad \forall \, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S} \,.$$

Proof

Step 1. Suppose f is a conformal map, so that

$$\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \, \|\mathbf{w}\|} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle_f}{\|\mathbf{v}\|_f \, \|\mathbf{w}\|_f} \,, \quad \forall \, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S} \,. \tag{4.12}$$

Let $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$ be an orthonormal basis for $T_{\mathbf{p}}\mathcal{S}$, that is,

$$\langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle = 0 \,, \quad \|\boldsymbol{\alpha}_1\| = \|\boldsymbol{\alpha}_2\| = 1 \,.$$

Define

$$\begin{aligned} \lambda(\mathbf{p}) &:= \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 \rangle_f = \|\boldsymbol{\alpha}_1\|_f^2 \,, \\ \mu(\mathbf{p}) &:= \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle_f \,, \\ \nu(\mathbf{p}) &:= \langle \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_2 \rangle_f = \|\boldsymbol{\alpha}_2\|_f^2 \,. \end{aligned}$$

By (4.12) we have

$$\frac{\langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle}{\|\boldsymbol{\alpha}_1\| \|\boldsymbol{\alpha}_2\|} = \frac{\langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle_f}{\|\boldsymbol{\alpha}_1\|_f \|\boldsymbol{\alpha}_2\|_f} \,.$$

Since $\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 = 0$, from the above we get

$$\mu(\mathbf{p}) = \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle_f = 0\,.$$

Moreover, since $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are orthonormal, the angle between $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ is $\theta = \pi/4$. By definition of angle between vectors, we infer

$$\frac{\sqrt{2}}{2} = \cos(\theta) = \frac{\langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 \rangle}{\|\boldsymbol{\alpha}_1\| \|\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_1\|}\,.$$

On the other hand, using (4.12) we get

$$\frac{\langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 \rangle}{\|\boldsymbol{\alpha}_1\| \|\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_1\|} = \frac{\langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 \rangle_f}{\|\boldsymbol{\alpha}_1\|_f \|\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2\|_f}\,.$$

The numerator of the right hand side satisfies

$$\begin{aligned} \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 \rangle_f & = \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 \rangle_f + \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle_f \\ & = \lambda(\mathbf{p}) + \mu(\mathbf{p}) \\ & = \lambda(\mathbf{p})\,, \end{aligned}$$

since $\mu(\mathbf{p}) = 0$. Concerning the denominator, we have

$$\begin{aligned} \|\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2\|_f^2 & = \|\boldsymbol{\alpha}_1\|_f^2 + \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle_f + \|\boldsymbol{\alpha}_2\|_f^2 \\ & = \lambda(\mathbf{p}) + \mu(\mathbf{p}) + \nu(\mathbf{p}) \\ & = \lambda(\mathbf{p}) + \nu(\mathbf{p})\,, \end{aligned}$$

since $\mu(\mathbf{p}) = 0$. Putting together the last 4 groups of equations, we obtain

$$\frac{\sqrt{2}}{2} = \frac{\lambda}{\lambda^{1/2}(\lambda + \nu)^{1/2}}\,.$$

Rearraging the above equation yields

$$\lambda(\mathbf{p}) = \nu(\mathbf{p})\,.$$

Now let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. Since $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2\}$ is a basis for $T_{\mathbf{p}}\mathcal{S}$, there exist $v_1, v_2 \in \mathbb{R}$ such that

$$\mathbf{v} = v_1 \boldsymbol{\alpha}_1 + v_2 \boldsymbol{\alpha}_2\,.$$

Therefore

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v} \rangle & = v_1^2 \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 \rangle + 2 v_1 v_2 \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle + v_2^2 \langle \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_2 \rangle \\ & = v_1^2 + v_2^2\,, \end{aligned}$$

where we used that $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are orthonormal. On the other hand,

$$\begin{aligned}\langle \mathbf{v},\mathbf{v}\rangle_f &= v_1^2\,\langle \boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1\rangle_f + 2v_1v_2\,\langle \boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\rangle_f + v_2^2\,\langle \boldsymbol{\alpha}_2,\boldsymbol{\alpha}_2\rangle_f \\ &= v_1^2\,\lambda(\mathbf{p}) + 2v_1v_2\,\mu(\mathbf{p}) + v_2^2\,\nu(\mathbf{p}) \\ &= \lambda(\mathbf{p})\,(v_1^2+v_2^2)\,,\end{aligned}$$

where we used that $\lambda(\mathbf{p}) = \nu(\mathbf{p})$ and $\mu(\mathbf{p}) = 0$. Thus

$$\langle \mathbf{v},\mathbf{v}\rangle_f = \lambda(\mathbf{p})\,(v_1^2+v_2^2) = \lambda(\mathbf{p})\,\langle \mathbf{v},\mathbf{v}\rangle\,,$$

for all $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. Since $\langle\cdot,\cdot\rangle$ and $\langle\cdot,\cdot\rangle_f$, by arguing as in Remark 4.116 we conclude that

$$\langle \mathbf{v},\mathbf{w}\rangle_f = \lambda(\mathbf{p})\,\langle \mathbf{v},\mathbf{w}\rangle$$

for all $\mathbf{v},\mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$.
Step 2. Suppose that there exists a function $\lambda : \mathcal{S} \to \mathbb{R}$ such that

$$\langle \mathbf{v},\mathbf{w}\rangle_f = \lambda(\mathbf{p})\,\langle \mathbf{v},\mathbf{w}\rangle\,, \quad \forall\, \mathbf{v},\mathbf{w} \in T_{\mathbf{p}}\mathcal{S}\,.$$

In particular, we have

$$\|\mathbf{v}\|_f = \sqrt{\lambda(\mathbf{p})}\|\mathbf{v}\|\,, \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}\,.$$

Then

$$\frac{\langle \mathbf{v},\mathbf{w}\rangle_f}{\|\mathbf{v}\|_f\|\mathbf{w}\|_f} = \frac{\lambda(\mathbf{p})\,\langle \mathbf{v},\mathbf{w}\rangle}{\sqrt{\lambda(\mathbf{p})}\|\mathbf{v}\|\sqrt{\lambda(\mathbf{p})}\|\mathbf{w}\|} = \frac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|\|\mathbf{w}\|}\,,$$

showing that f is a conformal map.

Corollary 4.127

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ be a local diffeomorphism. They are equivalent:

1. f is a conformal map.
2. Let $\boldsymbol{\sigma} : U \to \mathcal{S}$ be a regular chart of $\mathcal{S}$ and consider the chart of $\widetilde{\mathcal{S}}$ given by

$$\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma} : U \to \widetilde{\mathcal{S}}\,.$$

   There exists $\lambda : U \to \mathbb{R}$ such that

$$\widetilde{\mathcal{F}}_1 = \lambda(u,v)\mathcal{F}_1\,, \quad \forall\,(u,v) \in U\,,$$

   where $\mathcal{F}_1$ and $\widetilde{\mathcal{F}}_1$ are the first fundamental forms of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$, respectively.

The follows by using Theorem 4.126, and by adapting the argument in the proof of Theorem 4.118.

Example 4.128: Conformal maps are not local isometries

Consider the plane $\mathcal{S}$ with chart

$$\boldsymbol{\sigma}(u,v) := (u,v,0)\,.$$

Let $\widetilde{\mathcal{S}}$ be the sphere with parametrization

$$\tilde{\boldsymbol{\sigma}}(u,v) := (\operatorname{sech}(u)\cos(v), \operatorname{sech}(u)\sin(v), \tanh(u))\ .$$

We have

$$\boldsymbol{\sigma}_u = (1,0,0)\,, \quad \boldsymbol{\sigma}_v = (0,1,0)\,,$$

so that

$$\begin{aligned} E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\ G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1 \end{aligned}$$

Therefore the first fundamental form of $\mathcal{S}$ is

$$\mathcal{F}_1 = du^2 + dv^2\,.$$

Using the identitities

$$\begin{aligned} \frac{d}{du}(\operatorname{sech}(u)) &= -\operatorname{sech}(u)\tanh(u)\,, \\ \frac{d}{du}(\tanh(u)) &= \operatorname{sech}^2(u)\,, \end{aligned}$$

we obtain

$$\begin{aligned} \tilde{\boldsymbol{\sigma}}_u &= (-\operatorname{sech}(u)\tanh(u)\cos(v), -\operatorname{sech}(u)\tanh(u)\sin(v), \operatorname{sech}^2(u)) \\ \tilde{\boldsymbol{\sigma}}_v &= (-\operatorname{sech}(u)\sin(v), \operatorname{sech}(u)\cos(v), 0) \end{aligned}$$

By recalling that

$$\operatorname{sech}^2(u) + \tanh^2(u) = 1\,,$$

we compute

$$\begin{aligned} \widetilde{E} &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u = \operatorname{sech}^2(u)(\tanh^2(u) + \operatorname{sech}^2(u)) = \operatorname{sech}^2(u) \\ \widetilde{F} &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v = 0 \\ \widetilde{G} &= \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v = \operatorname{sech}^2(u)(\cos^2(v) + \sin^2(v)) = \operatorname{sech}^2(u) \end{aligned}$$

Hence the first fundamental form of $\widetilde{\mathcal{S}}$ is

$$\widetilde{\mathcal{F}}_1 = \operatorname{sech}^2(u)\,\left(du^2 + dv^2\right)\ .$$

Now, consider the map $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ defined by

$$f(u, v, 0) = \tilde{\boldsymbol{\sigma}}(u, v) .$$

In particular f satisfies

$$f(\boldsymbol{\sigma}(u, v)) = \tilde{\boldsymbol{\sigma}}(u, v) .$$

We have:

- f is not a local isometry.

  If f was a local isometry, by Theorem 4.118 we would conclude that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ have the same first fundamental form. However

  $$\mathcal{F}_1 = du^2 + dv^2 \neq \operatorname{sech}^2(u) \left(du^2 + dv^2\right) = \widetilde{\mathcal{F}}_1 .$$

- f is a conformal map.

  The first fundamental forms of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ satisfy

  $$\widetilde{\mathcal{F}}_1 = \lambda(u, v)\, \mathcal{F}_1 , \quad \lambda(u, v) := \operatorname{sech}(u) .$$

  Therefore f is a conformal map by Corollary 4.127.

4.11.7 Conformal parametrizations

We conclude this section with the definition of **conformally flat surface** and **conformal parametrization**.

Definition 4.129: Conformally flat surface and conformal parametrization

Let $\mathcal{S}$ be a regular surface and

$$\boldsymbol{\sigma} : U \to \mathcal{S}$$

be a regular chart of $\mathcal{S}$. We say that $\mathcal{S}$ is **conformally flat** and $\boldsymbol{\sigma}$ is a **conformal parametrization** if the first fundamental form of $\boldsymbol{\sigma}$ satisfies

$$\mathcal{F}_1 = \lambda(u, v)(du^2 + dv^2)$$

for some smooth function $\lambda : U \to \mathbb{R}$.

Definition 4.129 is motivated by the following Theorem: It states that angles on conformally flat surfaces look like angles on a plane.

Theorem 4.130

Let $\mathcal{S}$ be a regular surface and

$$\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S}$$

be a regular chart of $\mathcal{S}$. Define the plane $\boldsymbol{\pi}$ charted by

$$\tilde{\boldsymbol{\sigma}}(u,v) = (u,v,0)\,, \quad \forall\,(u,v) \in U\,.$$

1. They are equivalent:
   - $\boldsymbol{\sigma}$ is a conformal parametrization.
   - There exists a conformal map $f : \pi \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S}$.
2. A conformal parametrization $\boldsymbol{\sigma}$ preserves angles between vectors, in the following sense: Suppose $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ are curves in $\mathbb{R}^2$ such that
$$\boldsymbol{\gamma}_1(t_0) = \boldsymbol{\gamma}_2(t_0)\,.$$
Consider the corresponding curves on $\mathcal{S}$ given by
$$\tilde{\boldsymbol{\gamma}}_1 := \boldsymbol{\sigma} \circ \boldsymbol{\gamma}_1\,, \quad \tilde{\boldsymbol{\gamma}}_2 = \boldsymbol{\sigma} \circ \boldsymbol{\gamma}_2\,.$$
If
$$\dot{\boldsymbol{\gamma}}_1(t_0)\,, \dot{\boldsymbol{\gamma}}_2(t_0) \;\text{ form an angle }\; \theta\,,$$
then
$$\dot{\tilde{\boldsymbol{\gamma}}}_1(t_0)\,, \dot{\tilde{\boldsymbol{\gamma}}}_2(t_0) \;\text{ form an angle }\; \theta\,.$$

Proof

Proof of Point 1. Define the diffeomorphism $f : \boldsymbol{\pi} \to \mathcal{S}$ by

$$f(u,v,0) = \boldsymbol{\sigma}(u,v)\,.$$

In particular

$$f(\tilde{\boldsymbol{\sigma}}(u,v)) = \boldsymbol{\sigma}(u,v)\,.$$

By Corollary 4.127 we have that f is a conformal map if and only if there exists $\lambda : \boldsymbol{\pi} \to \mathbb{R}$ such that

$$\mathcal{F}_1 = \lambda(u,v)\widetilde{\mathcal{F}}_1\,,$$

where $\mathcal{F}_1$ and $\widetilde{\mathcal{F}}_1$ are the first fundamental forms of $\mathcal{S}$ and $\boldsymbol{\pi}$, respectively. Since $\boldsymbol{\pi}$ is a plane, the first fundamental form is given by

$$\widetilde{\mathcal{F}}_1 = du^2 + dv^2\,.$$

Therefore

$$\mathcal{F}_1 = \lambda(u,v)\left(du^2 + dv^2\right)\,,$$

showing that $\boldsymbol{\sigma}$ is a conformal parametrization.

Proof of Point 2. Suppose $\boldsymbol{\sigma}$ is a conformal parametrization. By the proof of Point 1 we have that

$$f : \pi \to \mathcal{S}, \quad f(u,v,0) = \boldsymbol{\sigma}(u,v),$$

is a conformal map. Since $T_{\mathbf{p}}\pi = \mathbb{R}^2$ and $f = \boldsymbol{\sigma}$, it follows by the definition of differential and f being conformal that the angle between $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ is the same as the angle between $\tilde{\boldsymbol{\gamma}}_1$ and $\tilde{\boldsymbol{\gamma}}_2$.

Example 4.131: Unit cylinder

The cylinder $\mathcal{S}$ charted by

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v)$$

is conformally flat, since the first fundamental form of $\boldsymbol{\sigma}$ is

$$\mathcal{F}_1 = du^2 + dv^2 .$$

Therefore $\boldsymbol{\sigma}$ is a conformal parametrization of $\mathcal{S}$.

Example 4.132: Shpere

Consider the parametrization of the sphere

$$\boldsymbol{\sigma}(u,v) = (\operatorname{sech}(u)\cos(v), \operatorname{sech}(u)\sin(v), \tanh(u)) .$$

In Example 4.129 we have seen that the first fundamental form of $\boldsymbol{\sigma}$ is

$$\mathcal{F}_1 = \operatorname{sech}(u)\,(du^2 + dv^2) .$$

Therefore $\boldsymbol{\sigma}$ is a conformal parametrization of the sphere.

4.12 Second fudamental form

The first fundamental form allows to measure distances on a surface. However it does not give any information on how curved a surface is: For example, we saw that a plane and a cylinder have the same first fundamental form

$$\mathcal{F}_1 = du^2 + dv^2 .$$

However the plane is flat, while the cylinder curves. We would like to find a measure of curvature which allows us to tell these two surfaces apart.

4.12.1 Unit normal and orientability

Before talking about curvatures, we need to clarify what we mean by normal vector to a surface and orientability. Let $\mathcal{S}$ be a regular surface and $\mathbf{p} \in \mathcal{S}$. The tangent plane $T_{\mathbf{p}}\mathcal{S}$ passes through the origin. Therefore $T_{\mathbf{p}}\mathcal{S}$ is completely determined by giving a unit vector $\mathbf{N}$ perpendicular to it:

$$T_{\mathbf{p}}\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^3 \ : \ \mathbf{x} \cdot \mathbf{N} = 0\}.$$

In this case we write

$$\mathbf{N} \perp T_{\mathbf{p}}\mathcal{S} ,$$

to denote that $\mathbf{N}$ is **perpendicular** to $T_{\mathbf{p}}\mathcal{S}$. Clearly, also $-\mathbf{N}$ is a unit vector, and

$$(-\mathbf{N}) \perp T_{\mathbf{p}}\mathcal{S} .$$

Question 4.133

Which unit normal should we choose between $\mathbf{N}$ and $-\mathbf{N}$?

There is no right answer to the above question. One way to proceed is the following.

Remark 4.134

Suppose that $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ is a regular chart for $\mathcal{S}$. Let $\mathbf{p} \in \boldsymbol{\sigma}(U)$. Then

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}.$$

Therefore we can choose the unit normal to $T_{\mathbf{p}}\mathcal{S}$ as

$$\mathbf{N}_{\boldsymbol{\sigma}} := \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} .$$

Clearly $\mathbf{N}_{\boldsymbol{\sigma}}$ has unit norm. Moreover

$$\mathbf{N}_{\boldsymbol{\sigma}} \cdot \boldsymbol{\sigma}_u = 0 , \quad \mathbf{N}_{\boldsymbol{\sigma}} \cdot \boldsymbol{\sigma}_v = 0$$

by the properties of cross product, showing that $\mathbf{N}_{\boldsymbol{\sigma}}$ is perpendicular to $T_{\mathbf{p}}\mathcal{S}$.
There is however an issue: $\mathbf{N}_{\boldsymbol{\sigma}}$ is not independent on the choice of chart $\boldsymbol{\sigma}$. Indeed, suppose that $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3$ is a reparametrization of $\boldsymbol{\sigma}$, that is,

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi , \quad \Phi : \widetilde{U} \to U ,$$

with Φ diffeomorphism. As shown in the proof of Proposition 4.63, we have

$$\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} = \det(J\Phi) \, \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v .$$

Hence

$$\mathbf{N}_{\tilde{\boldsymbol{\sigma}}} = \frac{\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}}}{\|\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\|} = \frac{\det J\Phi}{|\det J\Phi|} \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = \pm \mathbf{N}_{\boldsymbol{\sigma}} .$$

Therefore the sign on the right hand side depends on the sign of the Jacobian determinant of the transition map Φ.

The above remark motivates the following definitions.

Definition 4.135: Standard unit normal of a chart

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ a regular chart. The **standard unit normal** of $\boldsymbol{\sigma}$ is the smooth function

$$\mathbf{N}_{\boldsymbol{\sigma}} : U \to \mathbb{R}^3 , \quad \mathbf{N}_{\boldsymbol{\sigma}} := \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} .$$

Definition 4.136: Charts with same orientation

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\sigma} : U \to \mathbb{R}^3, \tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3$ regular charts such that

$$\boldsymbol{\sigma}(U) \cap \tilde{\boldsymbol{\sigma}}(\widetilde{U}) \neq \emptyset .$$

Denote by Φ the transition map between $\tilde{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}$. We say that:

1. $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ determine the **same orientation** if

   $$\det J\Phi > 0 ,$$

   where Φ is defined.

2. $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ determine **opposite orientations** if

   $$\det J\Phi < 0 ,$$

   where Φ is defined.

Example 4.137

Let $\mathbf{a}, \mathbf{p}, \mathbf{q} \in \mathbb{R}^3$ and suppose that $\mathbf{p}$ and $\mathbf{q}$ are linearly independent. The plane spanned by $\mathbf{p}, \mathbf{q}$ and passing through $\mathbf{a}$ can be parametrized by

$$\boldsymbol{\sigma}(u, v) := \mathbf{a} + \mathbf{p}u + \mathbf{q}v , \quad \forall\, (u, v) \in \mathbb{R}^2 .$$

An alternative parametrization is given by

$$\tilde{\boldsymbol{\sigma}}(u, v) := \mathbf{a} + \mathbf{q}u + \mathbf{p}v , \quad \forall\, (u, v) \in \mathbb{R}^2 .$$

We have

$$\boldsymbol{\sigma}_u = \mathbf{p} , \quad \boldsymbol{\sigma}_v = \mathbf{q} ,$$

and therefore

$$\mathbf{N}_{\boldsymbol{\sigma}} = \frac{\mathbf{p} \times \mathbf{q}}{\|\mathbf{p} \times \mathbf{q}\|} .$$

Similarly, we have

$$\mathbf{N}_{\tilde{\boldsymbol{\sigma}}} = \frac{\mathbf{q} \times \mathbf{p}}{\|\mathbf{q} \times \mathbf{p}\|} = \frac{-\mathbf{p} \times \mathbf{q}}{\|\mathbf{p} \times \mathbf{q}\|} ,$$

showing that

$$\mathbf{N}_{\boldsymbol{\sigma}} = -\mathbf{N}_{\tilde{\boldsymbol{\sigma}}} \, .$$

Hence $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ determine opposite orientations.

If a surface can be covered by charts with the same orientation, it is called orientable.

Definition 4.138: Orientable surface

Let $\mathscr{S}$ be a regular surface. Then:

1. An atlas $\mathscr{A} = \{\boldsymbol{\sigma}_i\}_{i \in I}$ is **oriented** if the following property holds:

$$\boldsymbol{\sigma}_i(U_i) \cap \boldsymbol{\sigma}_j(U_j) \neq \emptyset \qquad \Longrightarrow \qquad \det J\Phi > 0 \, ,$$

   where Φ is the transition map between $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}_j$.

2. $\mathscr{S}$ is **orientable** if there exists an oriented atlas $\mathscr{A}$.

3. If an oriented atlas $\mathscr{A}$ is assigned, we say that $\mathscr{S}$ is **oriented** by $\mathscr{A}$.

Example 4.139

All the surfaces we encountered in these Lecture Notes are orientable, except for the Möbius band in Example 4.100. Details about the non-orientability of the Möbius band can be found in Example 4.5.3 in [6].

Example 4.140

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart. Then

$$\mathscr{S}_{\boldsymbol{\sigma}} := \boldsymbol{\sigma}(U)$$

is a regular surface with atlas $\mathscr{A} = \{\boldsymbol{\sigma}\}$. Therefore $\mathscr{S}_{\boldsymbol{\sigma}}$ is orientable.

This is because we have only one chart. Therefore any transition map Φ will be the identity, so that $\det J\Phi = 1 > 0$.

Warning: Orientability is a global property

The above example is saying that orientability is a global property: To determine wether a surface $\mathscr{S}$ is orientable, we need to examine the transition maps for the entire atlas $\mathscr{A}$. This is because a single local parametrization $\boldsymbol{\sigma}(U) \subseteq \mathscr{S}$ is always orientable.

Remark 4.141

Let $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ be regular charts with transition map Φ. We have seen in Remark 4.134 that

$$\mathbf{N}_{\tilde{\boldsymbol{\sigma}}} = \frac{\det J\Phi}{|\det J\Phi|} \mathbf{N}_{\boldsymbol{\sigma}} .$$

If $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ determine the same orientation, then

$$\det J\Phi > 0 ,$$

which implies

$$\mathbf{N}_{\tilde{\boldsymbol{\sigma}}} = \mathbf{N}_{\boldsymbol{\sigma}} .$$

Hence, if $\mathcal{S}$ is an orientable surface, one can define a unit normal vector at each point of $\mathcal{S}$, without ambiguity.

Definition 4.142: Unit normal of a surface

Let $\mathcal{S}$ be a regular surface. A **unit normal** of $\mathcal{S}$ is a smooth function $\mathbf{N} \colon \mathcal{S} \to \mathbb{R}^3$ such that

$$\mathbf{N}(\mathbf{p}) \perp T_{\mathbf{p}}\mathcal{S} , \quad \|\mathbf{N}(\mathbf{p})\| = 1 , \quad \forall \, \mathbf{p} \in \mathcal{S} .$$

Warning

We require the function $\mathbf{p} \mapsto \mathbf{N}(\mathbf{p})$ to be globally defined on $\mathcal{S}$ and smooth.

Proposition 4.143

Let $\mathcal{S}$ be a regular surface. They are equivalent:

1. $\mathcal{S}$ is orientable.
2. There exists a unit normal $\mathbf{N} \colon \mathcal{S} \to \mathbb{R}^3$.

The proof follows from the above arguments. For details, we refer the reader to Proposition 4.3.7 in [1].

In view of the above propostion, for an oriented surface there is a natural choice of unit normal, which we call **standard unit normal** of $\mathcal{S}$.

Definition 4.144: Standard unit normal of a surface

Let $\mathcal{S}$ be a regular surface oriented by the atlas $\mathcal{A}$. The **standard unit normal** to $\mathcal{S}$ is the map $\mathbf{N} \colon \mathcal{S} \to \mathbb{R}^3$ such that

$$\mathbf{N} \circ \boldsymbol{\sigma} = \mathbf{N}_{\boldsymbol{\sigma}} ,$$

for each chart $\boldsymbol{\sigma} \in \mathscr{A}$, where

$$\mathbf{N}_{\boldsymbol{\sigma}} : U \to \mathbb{R}^3 , \quad \mathbf{N}_{\boldsymbol{\sigma}} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|}$$

is the standard unit normal of $\boldsymbol{\sigma}$.

Notation

In the following we will denote by $\mathbf{N}$ both the standard unit normal of $\mathscr{S}$ and of a chart.

4.12.2 Definition of Second fundamental form

We can now start our discussion about curvature of surfaces. We can make a similar argument to the one we made for curves: If $\boldsymbol{\gamma}$ is a unit speed curve, the curvature of $\boldsymbol{\gamma}$ is defined as

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

The quantity $\kappa(t)$ gave us a measure of how much $\boldsymbol{\gamma}$ is deviating from a straight line. Similarly, we would like to quantify how much a surface $\mathscr{S}$ is deviating from the tangent plane $T_{\mathbf{p}}\mathscr{S}$. Recall that

$$T_{\mathbf{p}}\mathscr{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} ,$$

where $\boldsymbol{\sigma}$ is a regular chart of $\mathscr{S}$ at $\mathbf{p}$. The standard unit normal of $\boldsymbol{\sigma}$ is

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} ,$$

which is orthogonal to $T_{\mathbf{p}}\mathscr{S}$. Let $(u_0, v_0) \in \mathbb{R}^2$ be the point such that

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p} .$$

As the scalar quantities Δu and Δv vary, the point

$$\boldsymbol{\sigma}(u_0 + \Delta u, v_0 + \Delta v) \in \mathscr{S}$$

deviates from the tangent plane $T_{\mathbf{p}}\mathscr{S}$. Since $\mathbf{N}$ is orthogonal to $T_{\mathbf{p}}\mathscr{S}$, the deviation is given by

$$\delta := [\boldsymbol{\sigma}(u_0 + \Delta u, v_0 + \Delta v) - \boldsymbol{\sigma}(u_0, v_0)] \cdot \mathbf{N} ,$$

as shown in Figure 4.24.

Using Taylor's formula we get

$$\begin{aligned}\boldsymbol{\sigma}(u_0 + \Delta u, v_0 + \Delta v) = \boldsymbol{\sigma}(u_0, v_0) + \boldsymbol{\sigma}_u(u_0, v_0)\, \Delta u + \boldsymbol{\sigma}_v(u_0, v_0)\, \Delta v \\ + \frac{1}{2}\left(\boldsymbol{\sigma}_{uu}(u_0, v_0)(\Delta u)^2 + 2\boldsymbol{\sigma}_{uv}(u_0, v_0)\Delta u \Delta v + \boldsymbol{\sigma}_{vv}(u_0, v_0)(\Delta v)^2\right) \\ + R(\Delta u, \Delta v) ,\end{aligned}$$

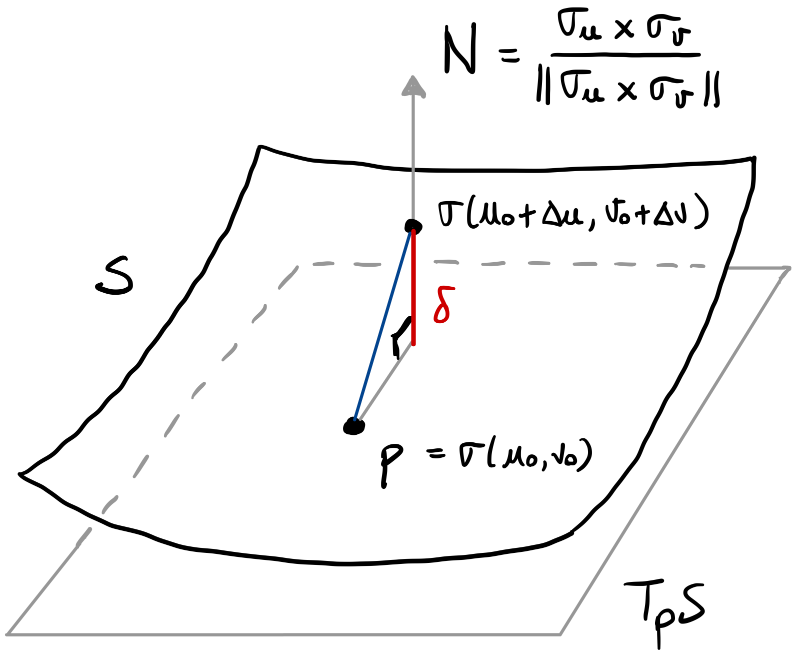


Figure 4.24: The point $\boldsymbol{\sigma}(u_0 + \Delta u, v_0 + \Delta v)$ on $\mathcal{S}$ deviates from $T_{\mathbf{p}}\mathcal{S}$ by a quantity δ.

where $R(\Delta u, \Delta v)$ is a remainder such that

$$\lim_{\Delta \to 0} \frac{R(\Delta u, \Delta v)}{\Delta} = 0 \,, \quad \Delta := (\Delta u)^2 + (\Delta v)^2 \,.$$

Since $\mathbf{N}$ is orthogonal to $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$, if we multiply the above Taylor expansion by $\mathbf{N}$, and ignore the remainder, we obtain

$$\delta = \frac{1}{2} \left(L(\Delta u)^2 + 2M\Delta u \Delta v + N(\Delta v)^2 \right) \,,$$

where we set

$$L := \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} \,, \quad M := \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} \,, \quad N := \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} \,.$$

The expression

$$\mathcal{F}_2 := L\, du^2 + 2M\, dudv + N\, dv^2$$

is called the **second fundamental form** of $\mathcal{S}$. Therefore $\mathcal{F}_2$ measures how much the surface $\mathcal{S}$ deviates from being a plane. Let us make this definition precise.

Definition 4.145: Second fundamental form of a chart

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart of $\mathcal{S}$. Denote the standard unit normal of $\boldsymbol{\sigma}$ by

$$\mathbf{N} : U \to \mathbb{R}^3 \,, \quad \mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} \,.$$

Define the functions

$$L, M, N : U \to \mathbb{R}$$

by setting

$$L := \boldsymbol{\sigma}_{uu} \cdot \mathbf{N}, \quad M := \boldsymbol{\sigma}_{uv} \cdot \mathbf{N}, \quad N := \boldsymbol{\sigma}_{vv} \cdot \mathbf{N}.$$

Let $\mathbf{p} \in \boldsymbol{\sigma}(U)$ and denote by $(u_0, v_0) \in U$ the point such that

$$\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}.$$

The **second fundamental form** of $\boldsymbol{\sigma}$ at $\mathbf{p}$ is the quadratic form

$$\mathscr{F}_2 : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$$

defined by

$$\mathscr{F}_2(\mathbf{v}) := L\, du^2(\mathbf{v}) + 2M\, du(\mathbf{v})\, dv(\mathbf{v}) + N\, dv^2(\mathbf{v}), \tag{4.13}$$

for all $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. Here L, M, N are evaluated at (u_0, v_0), and du, dv are the coordinate functions as in Definition 4.105.

Notation

With a little abuse of notation, we also denote by $\mathscr{F}_2$ the 2×2 matrix

$$\mathscr{F}_2 = \begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

Remark 4.146: Second fundamental form and reparametrizations

The second fundamental form

$$\mathscr{F}_2 = L\, du^2 + 2M\, dudv + N\, dv^2$$

depends on the choice of chart $\boldsymbol{\sigma} : U \to \mathbb{R}^3$. Indeed, let us adopt the same notations as Remark 4.109. Suppose that $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3$ is a reparametrization of $\boldsymbol{\sigma}$ with

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi,$$

where $\Phi : \widetilde{U} \to U$ is a diffeomorphism. Denote the second fundamental form of $\tilde{\boldsymbol{\sigma}}$ by

$$\widetilde{\mathscr{F}}_2 = \tilde{L}\, d\tilde{u}^2 + 2\widetilde{M}\, d\tilde{u}d\tilde{v} + \widetilde{N}\, d\tilde{v}^2.$$

The matrices of $\mathscr{F}_2$ and $\widetilde{\mathscr{F}}_2$ are related by

$$\begin{pmatrix} \tilde{L} & \widetilde{M} \\ \widetilde{M} & \widetilde{N} \end{pmatrix} = \pm (J\Phi)^T \begin{pmatrix} L & M \\ M & N \end{pmatrix} J\Phi, \tag{4.14}$$

where (4.14) holds with $+$ if $\det J\Phi > 0$ and $-$ if $\det J\Phi < 0$.

Formula (4.14) holds by a change of variable argument. The sign depends on the sign of $\det J\Phi$ because

$$\widetilde{\mathbf{N}} = \frac{\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}}}{\|\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\|} = \frac{\det J\Phi}{|\det J\Phi|} \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = \pm \mathbf{N},$$

as shown in Remark 4.134.

Let us show that a plane and a cylinder have different second fundamental forms.

Example 4.147: Plane

Let $\mathbf{a}, \mathbf{p}, \mathbf{q} \in \mathbb{R}^3$. Suppose that $\mathbf{p}$ and $\mathbf{q}$ are orthonormal vectors, that is,

$$\|\mathbf{p}\| = \|\mathbf{q}\| = 1 , \quad \mathbf{p} \cdot \mathbf{q} = 0 .$$

Consider the plane with chart

$$\boldsymbol{\sigma}(u, v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q} , \quad (u, v) \in \mathbb{R}^2 .$$

Prove that the second fundamental form of $\boldsymbol{\sigma}$ is

$$\mathscr{F}_2 = 0 .$$

This reflects the intuition that a plane is flat, and therefore there is no *curvature.*

We have

$$\boldsymbol{\sigma}_u = \mathbf{p} , \quad \boldsymbol{\sigma}_v = \mathbf{q} .$$

The principal unit normal is

$$\mathbf{N} = \frac{\mathbf{p} \times \mathbf{q}}{\|\mathbf{p} \times \mathbf{q}\|} ,$$

while the second derivatives are

$$\boldsymbol{\sigma}_{\mathbf{u}} = \boldsymbol{\sigma}_{\mathbf{u}} = \boldsymbol{\sigma}_{\mathbf{u}} = \mathbf{0} .$$

Therefore

$$\begin{aligned} L &= \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} = 0 \\ M &= \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = 0 \\ N &= \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} = 0 \end{aligned}$$

and the second fundamental form is

$$\mathscr{F}_2 = L\, du^2 + 2M\, du\, dv + N\, dv^2 = 0 .$$

Example 4.148: Unit cylinder

Consider the unit cylinder with chart

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v)\,, \quad (u,v) \in (0, 2\pi) \times \mathbb{R}\,.$$

Prove that the second fundamental form of $\boldsymbol{\sigma}$ is

$$\mathscr{F}_2 = -du^2\,.$$

This reflects the intuition that the cylinder curves only when moving in the v-direction. In such direction we are moving on a circle of radius 1, therefore we expect the curvature to be -1.

We have

$$\boldsymbol{\sigma}_u = (-\sin(u), \cos(u), 0)\,, \quad \boldsymbol{\sigma}_v = (0,0,1)\,,$$

and also

$$\boldsymbol{\sigma}_{uu} = (-\cos(u), -\sin(u), 0)\,, \quad \boldsymbol{\sigma}_{uv} = \boldsymbol{\sigma}_{vv} = \mathbf{0}\,.$$

We have also

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\sin(u) & \cos(u) & 0 \\ 0 & 0 & 1 \end{vmatrix} = (\cos(u), \sin(u), 0)$$

so that

$$\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = \sqrt{\cos^2(u) + \sin^2(u)} = 1\,.$$

The principal unit normal is

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = (\cos(u), \sin(u), 0)\,.$$

We finally compute

$$\begin{aligned} L &= \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} \\ &= (-\cos(u), -\sin(u), 0) \cdot (\cos(u), \sin(u), 0) \\ &= -\cos^2(u) - \sin^2(u) = -1 \\ M &= \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = 0 \\ N &= \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} = 0 \end{aligned}$$

The second fundamental form is

$$\mathscr{F}_2 = L\,du^2 + 2M\,du\,dv + N\,dv^2 = -du^2\,.$$

Remark 4.149

We have seen that a plane and the unit cylinder have the same first fundamental form

$$\mathscr{F}_1 = \widetilde{\mathscr{F}}_1 = du^2 + dv^2 \,,$$

while their second fundamental forms differ: we have

$$\mathscr{F}_2 = 0 \,, \quad \widetilde{\mathscr{F}}_2 = -du^2 \,,$$

respectively.

4.12.3 Gauss and Weingarten maps

Another way to quantify how much a surface $\mathscr{S}$ is curving is by examining the behavior of standard unit normal $\mathbf{N}$. If $\mathscr{S}$ is a plane spanned by vectors $\mathbf{p}$ and $\mathbf{q}$, then its standard unit normal is

$$\mathbf{N} = \frac{\mathbf{p} \times \mathbf{q}}{\|\mathbf{p} \times \mathbf{q}\|} \,,$$

which is constant across $\mathscr{S}$. If $\mathscr{S}$ is a general surface, measuring the variation of $\mathbf{N}$ will tell us how much $\mathscr{S}$ is deviating from being a plane. This is the idea behind the definition of the **Gauss** and **Weingarten** maps.

Remark 4.150

Let $\mathscr{S}$ be oriented and $\mathbf{N} \colon \mathscr{S} \to \mathbb{R}^3$ be the standard unit normal. In particular $\mathbf{N}$ is a smooth map and

$$\mathbf{N}(\mathbf{p}) \perp T_{\mathbf{p}}\mathscr{S} \,, \quad \|\mathbf{N}(\mathbf{p})\| = 1 \,, \quad \forall\, \mathbf{p} \in \mathscr{S} \,.$$

Since $T_{\mathbf{p}}\mathscr{S}$ passes through the origin and $\mathbf{N}$ has norm 1, it follows that

$$\mathbf{N}(\mathbf{p}) \in \mathbb{S}^2 := \{\mathbf{x} \in \mathbb{R}^3 \ : \ \|\mathbf{x}\| = 1\} \,,$$

where $\mathbb{S}^2$ is the unit sphere in $\mathbb{R}^3$. Thus $\mathbf{N} \colon \mathscr{S} \to \mathbb{S}^2$.

Definition 4.151: Gauss map

Let $\mathscr{S}$ be an oriented surface and $\mathbf{N}$ the standard unit normal to $\mathscr{S}$. The **Gauss map** of $\mathscr{S}$ is the map

$$\mathscr{G}_{\mathscr{S}} \colon \mathscr{S} \to \mathbb{S}^2 \,, \quad \mathscr{G}_{\mathscr{S}}(\mathbf{p}) := \mathbf{N}(\mathbf{p}) \,.$$

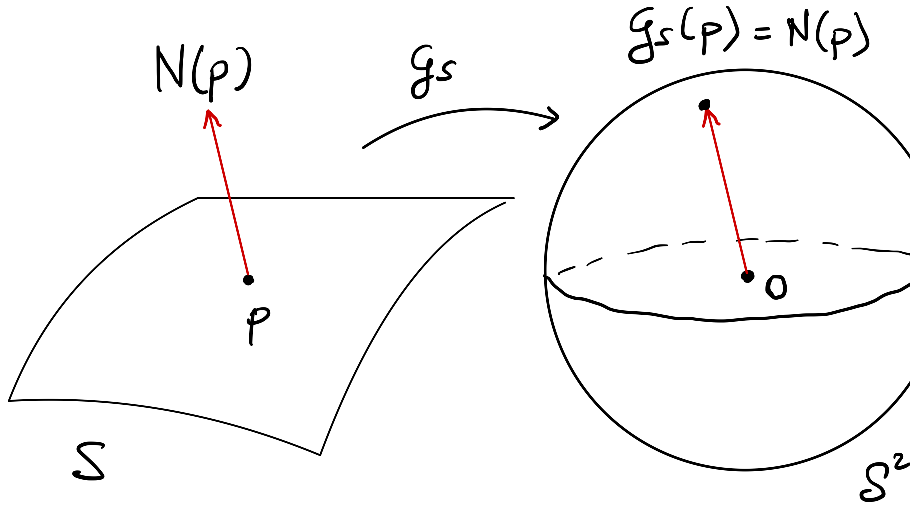


Figure 4.25: The Gauss map $\mathscr{G}_{\mathscr{S}}$ of $\mathscr{S}$ is defined as $\mathscr{G}_{\mathscr{S}}(\mathbf{p}) := \mathbf{N}(\mathbf{p})$. Note that $\mathscr{G}_{\mathscr{S}}(\mathbf{p}) \in \mathbb{S}^2$.

Remark 4.152

The Gauss map of $\mathscr{S}$ is just the standard unit normal of $\mathscr{S}$. By definition of standard unit normal to $\mathscr{S}$ we obtain that

$$\mathscr{G}_{\mathscr{S}} \circ \boldsymbol{\sigma} = \mathbf{N}$$

for all charts $\boldsymbol{\sigma} : U \to \mathbb{R}^3$, where $\mathbf{N} = \mathbf{N}_{\boldsymbol{\sigma}}$ is the standard unit normal to $\boldsymbol{\sigma}$, that is,

$$\mathbf{N} : U \to \mathbb{R}^3 , \quad \mathbf{N} := \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} .$$

Example 4.153

1. Suppose $\mathscr{S}$ is the unit sphere $\mathbb{S}^2$. Then $\mathscr{G}_{\mathscr{S}} : \mathscr{S} \to \mathbb{S}^2$ is the identity, see Figure 4.26.
2. Let $\mathbf{a}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ with $\mathbf{v}$ and $\mathbf{w}$ linearly independent. Let $\mathscr{S}$ be the plane

$$\boldsymbol{\sigma}(u, v) := \mathbf{a} + \mathbf{v}u + \mathbf{w}v , \quad \forall (u, v) \in \mathbb{R}^2 .$$

   The Gauss map of $\mathscr{S}$ is constant:

$$\mathscr{G}_{\mathscr{S}}(\mathbf{p}) = \frac{\mathbf{v} \times \mathbf{w}}{\|\mathbf{v} \times \mathbf{w}\|} ,$$

   for all $\mathbf{p} \in \mathscr{S}$, see Figure 4.27.

3. Let $\mathscr{S}$ be the unit cylinder

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v)\,, \quad (u,v) \in (0,2\pi) \times \mathbb{R}\,.$$

Then

$$\boldsymbol{\sigma}_u = (-\sin(u), \cos(u), 0)\,, \quad \boldsymbol{\sigma}_v = (0,0,1)\,,$$

and

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\sin(u) & \cos(u) & 0 \\ 0 & 0 & 1 \end{vmatrix} = (\cos(u), \sin(u), 0)\,.$$

Therefore

$$\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = 1\,,$$

and

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = (\cos(u), \sin(u), 0)\,.$$

The Gauss map of $\mathscr{S}$ is

$$\mathscr{G}_{\mathscr{S}}(\mathbf{p}) = (\cos(u_0), \sin(u_0), 0)\,,$$

where (u_0, v_0) is such that $\boldsymbol{\sigma}(u_0,v_0) = \mathbf{p}$. Note that $\mathscr{G}_{\mathscr{S}}$ maps $\mathscr{S}$ into the equator of $\mathbb{S}^2$, see Figure 4.28.

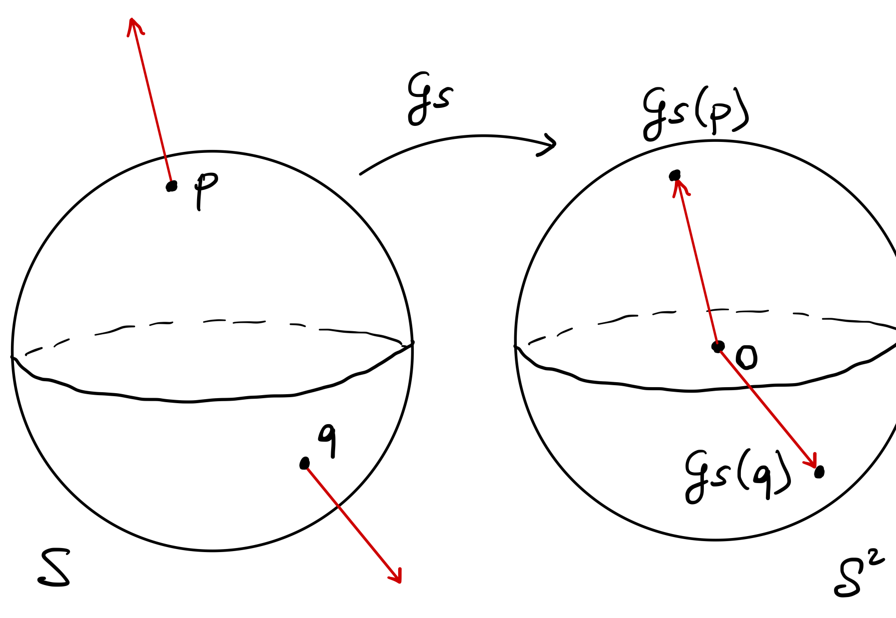


Figure 4.26: The Gauss map $\mathscr{G}_{\mathscr{S}}$ of a sphere is the identity.

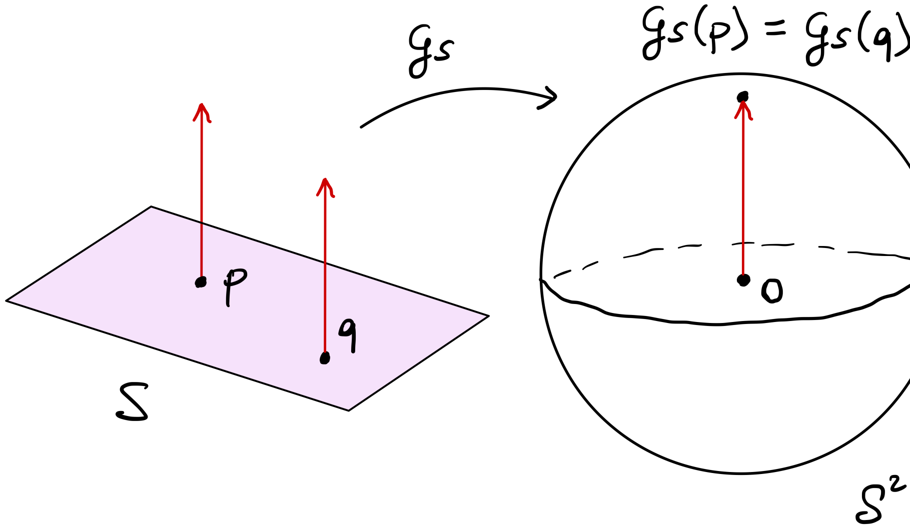


Figure 4.27: The Gauss map $\mathscr{G}_{\mathscr{S}}$ of a plane is constant.

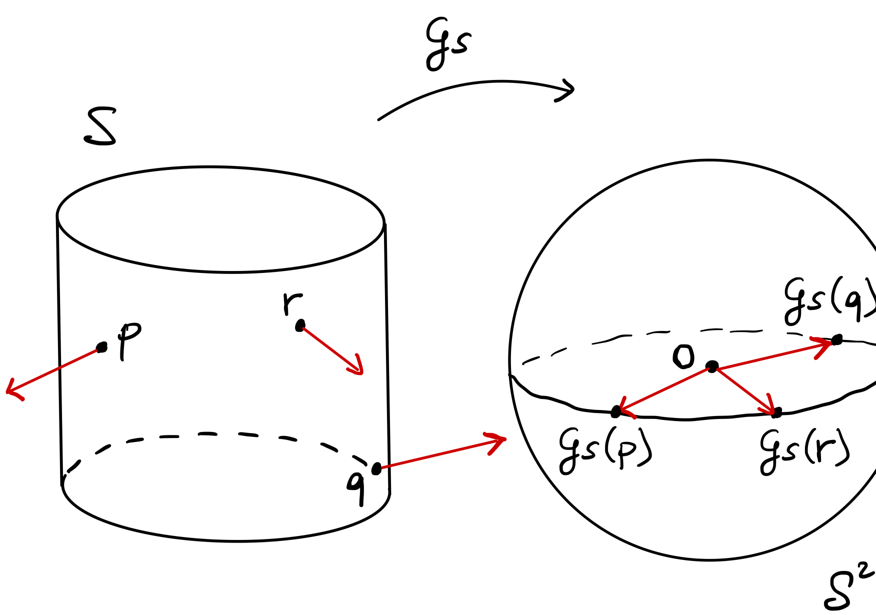


Figure 4.28: If $\mathscr{S}$ is the unit cylinder, the Gauss map $\mathscr{G}_{\mathscr{S}}$ maps $\mathscr{S}$ into the equator of $\mathbb{S}^2$.

Remark 4.154

By definition, the Gauss map is a smooth function between surfaces. Therefore the differential of $\mathscr{G}_{\mathscr{S}}$ is well defined, and

$$d_{\mathbf{p}}\mathscr{G}_{\mathscr{S}} \colon T_{\mathbf{p}}\mathscr{S} \to T_{\mathscr{G}_{\mathscr{S}}(\mathbf{p})}\mathbb{S}^2 ,$$

for all $\mathbf{p} \in \mathscr{S}$. We have that

$$T_{\mathscr{G}_{\mathscr{S}}(\mathbf{p})}\mathbb{S}^2 = T_{\mathbf{p}}\mathscr{S} , \tag{4.15}$$

see Figure 4.29. Therefore

$$d_{\mathbf{p}}\mathscr{G}_{\mathscr{S}} \colon T_{\mathbf{p}}\mathscr{S} \to T_{\mathbf{p}}\mathscr{S} .$$

Proof. The tangent plane $T_{\mathscr{G}_{\mathscr{S}}(\mathbf{p})}\mathbb{S}^2$ passes through the origin and

$$\mathscr{G}(\mathbf{p}) \perp T_{\mathscr{G}_{\mathscr{S}}(\mathbf{p})}\mathbb{S}^2 .$$

By definition $\mathscr{G}(\mathbf{p}) = \mathbf{N}(\mathbf{p})$, and thus

$$\mathbf{N}(\mathbf{p}) \perp T_{\mathscr{G}_{\mathscr{S}}(\mathbf{p})}\mathbb{S}^2 .$$

Since by definition

$$\mathbf{N}(\mathbf{p}) \perp T_{\mathbf{p}}\mathscr{S} ,$$

we infer (4.15).

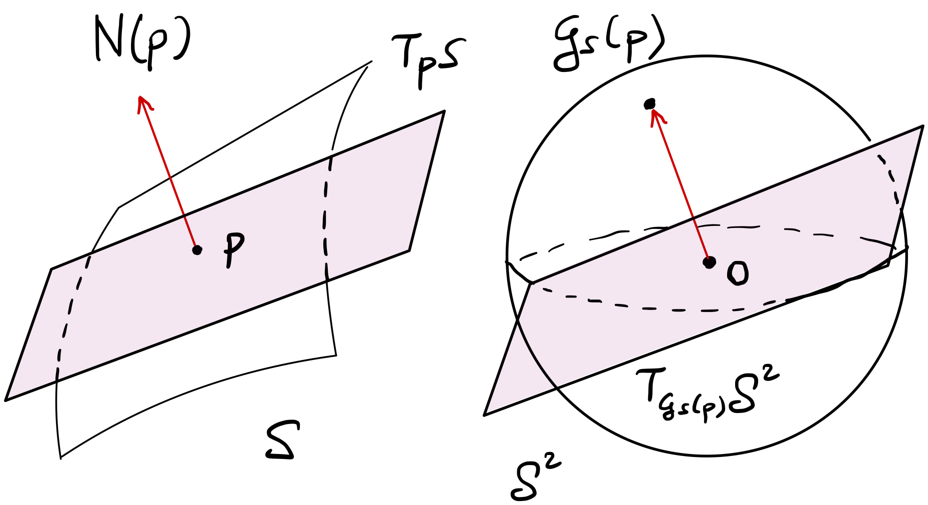


Figure 4.29: We ca identify $T_{\mathscr{G}_{\mathscr{S}}(\mathbf{p})}\mathbb{S}^2$ with $T_{\mathbf{p}}\mathscr{S}$. This is because $\mathscr{G}(\mathbf{p}) \perp T_{\mathscr{G}_{\mathscr{S}}(\mathbf{p})}\mathbb{S}^2$ and $\mathscr{G}(\mathbf{p}) = \mathbf{N}(\mathbf{p})$.

Definition 4.155: Weingarten map

Let $\mathcal{S}$ be an orientable surface and $\mathcal{G} : \mathcal{S} \to \mathbb{S}^2$ its Gauss map. The **Weingarten map** $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ of $\mathcal{S}$ at $\mathbf{p}$ is the negative differential of the Gauss map at $\mathbf{p}$, that is,

$$\mathcal{W}_{\mathbf{p},\mathcal{S}} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S} \,, \quad \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) := -d_{\mathbf{p}}\mathcal{G}(\mathbf{v}) \,,$$

for all $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$.

Important

The Gauss map encodes information on the standard unit normal $\mathbf{N}$ to $\mathcal{S}$. Hence its derivative, the Weingarten map, detects the rate of change of $\mathbf{N}$.

Remark 4.156

The minus sign in the definition of $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ is a convention, just like we defined the torsion to be the scalar τ such that

$$\dot{\mathbf{b}} = -\tau \mathbf{n} \,.$$

The Weingarten map allows us to define a bilnear form on $T_{\mathbf{p}}\mathcal{S}$. We call such bilinear form the **second fundamental form** of $\mathcal{S}$.

Definition 4.157: Second fundamental form of a surface

Let $\mathcal{S}$ be an orientable surface and denote by

$$\mathcal{W}_{\mathbf{p},\mathcal{S}} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}$$

its Weingarten map at $\mathbf{p}$. The **second fundamental form** of $\mathcal{S}$ at $\mathbf{p}$ is the map

$$II_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{S} \times T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$$

defined by

$$II_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) := \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) \cdot \mathbf{w} \,, \quad \forall\, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S} \,.$$

Remark 4.158

The second fudamental form $II_{\mathbf{p}}$ of $\mathcal{S}$ is bilinear.

> Indeed, $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ is linear, being the differential of a smooth map. Hence $II_{\mathbf{p}}$ is bilinear, given that the scalar product is bilinear.

Remark 4.159: Matrix of the second fundamental form

Let $\boldsymbol{\sigma}$ be a chart at $\mathbf{p} \in \mathcal{S}$. Since $II_{\mathbf{p}}$ is a bilinear form on $T_{\mathbf{p}}\mathcal{S}$, it can be represented by the 2×2 matrix

$$A = \begin{pmatrix} II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u) & II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v) \\ II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_u) & II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v) \end{pmatrix},$$

given that $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis for $T_{\mathbf{p}}\mathcal{S}$. In a not so shocking turn of events, it happens that

$$A = \mathcal{F}_2 = \begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

where

$$L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N}, \quad M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N}, \quad N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N}.$$

Therefore, the second fundamental form $II_{\mathbf{p}}$ coincides with the second fundamental form $\mathcal{F}_2$ of the chart $\boldsymbol{\sigma}$. We prove this statement in the next theorem.

Theorem 4.160

Let $\mathcal{S}$ be an orientable surface and $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart. Let $\mathbf{p} \in \boldsymbol{\sigma}(U)$.

1. The second funamental form $II_{\mathbf{p}}$ is a symmetric bilinear map.
2. It holds
$$II_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) = (du(\mathbf{v}), dv(\mathbf{v})) \begin{pmatrix} L & M \\ M & N \end{pmatrix} (du(\mathbf{w}), dv(\mathbf{w}))^T,$$
for all $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$, where
$$L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N}, \quad M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N}, \quad N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N}.$$
3. $\mathcal{F}_2$ is the quadratic form associated to $II_{\mathbf{p}}$, that is,
$$\mathcal{F}_2(\mathbf{v}) = II_{\mathbf{p}}(\mathbf{v}, \mathbf{v}), \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}.$$

To prove Theorem 4.160 we use the following two Lemmas.

Lemma 4.161

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart with standard unit normal $\mathbf{N} : U \to \mathbb{R}^3$. Then

$$\begin{aligned} \mathbf{N}_u \cdot \boldsymbol{\sigma}_u &= -L, \\ \mathbf{N}_u \cdot \boldsymbol{\sigma}_v = \mathbf{N}_v \cdot \boldsymbol{\sigma}_u &= -M, \\ \mathbf{N}_v \cdot \boldsymbol{\sigma}_v &= -N. \end{aligned}$$

Proof

The vectors $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ form a basis for $T_{\mathbf{p}}\mathcal{S}$. Since $\mathbf{N}$ is orthogonal to $T_{\mathbf{p}}\mathcal{S}$ by definition, it follows that

$$\mathbf{N} \cdot \boldsymbol{\sigma}_u = 0\,, \quad \mathbf{N} \cdot \boldsymbol{\sigma}_v = 0\,.$$

Differentiating the above with respect to u and v yields the thesis. For example, we have

$$\frac{\partial}{\partial u}(\mathbf{N} \cdot \boldsymbol{\sigma}_u) = 0\,.$$

On the other hand, by chain rule,

$$\frac{\partial}{\partial u}(\mathbf{N} \cdot \boldsymbol{\sigma}_u) = \mathbf{N}_u \cdot \boldsymbol{\sigma}_u + \mathbf{N} \cdot \boldsymbol{\sigma}_{uu} = \mathbf{N}_u \cdot \boldsymbol{\sigma}_u + L\,,$$

from which we infer

$$\mathbf{N}_u \cdot \boldsymbol{\sigma}_u = -L\,.$$

The rest of the proof follows similarly.

Lemma 4.162

Let $\mathcal{S}$ be an orientable surface and $\mathcal{W}_{\mathbf{p},\mathcal{S}} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}$ be its Weingarten map at $\mathbf{p}$. Let $\boldsymbol{\sigma}$ be a regular chart at $\mathbf{p}$, with $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$. Then

$$\mathcal{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) = -\mathbf{N}_u\,, \quad \mathcal{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v\,,$$

where $\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v, \mathbf{N}_u, \mathbf{N}_v$ are evaluated at (u_0, v_0).

Proof

Since $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ is defined as $-d_{\mathbf{p}}\mathcal{G}_{\mathcal{S}}$, we can compute $\mathcal{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u)$ and $\mathcal{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v)$ by using the definition of differential of a smooth function. To this end, consider the curve

$$\boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(u_0 + t, v_0)\,.$$

We have that $\boldsymbol{\gamma}$ is a smooth curve in $\mathcal{S}$ and

$$\dot{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}_u(u_0 + t, v_0)\,.$$

Therefore

$$\boldsymbol{\gamma}(0) = \boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}\,, \quad \dot{\boldsymbol{\gamma}}(0) = \boldsymbol{\sigma}_u(u_0, v_0)\,.$$

Define

$$\tilde{\boldsymbol{\gamma}}(t) := (\mathcal{G}_{\mathcal{S}} \circ \boldsymbol{\gamma})(t)\,.$$

By Remark 4.152

$$\tilde{\boldsymbol{\gamma}}(t) = \mathcal{G}_{\mathcal{S}}(\boldsymbol{\gamma}(t)) = \mathcal{G}_{\mathcal{S}}(\boldsymbol{\sigma}(u_0 + t, v_0)) = \mathbf{N}(u_0 + t, v_0)\,.$$

Thus

$$\dot{\boldsymbol{\gamma}}(t) = \mathbf{N}_u(u_0 + t, v_0)\,, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{N}_u(u_0, v_0)\,.$$

By definition of differential, we have

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) = -d_{\mathbf{p}}\mathcal{G}_{\mathcal{S}}(\boldsymbol{\sigma}_u) = -\dot{\boldsymbol{\gamma}}(0) = -\mathbf{N}_u(u_0, v_0)\,,$$

as we wanted to prove. To show that

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v(u_0, v_0)\,,$$

it is sufficient to consider the curve

$$\boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(u_0, v_0 + t)\,,$$

and argue similarly. This is left as an exercise.

We can now prove Theorem 4.160

Proof: Proof of Theorem 4.160

By Theorem 4.77 we have

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}\,.$$

Therefore, for $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$, there exist $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{R}$ such that

$$\mathbf{v} = \lambda_1\boldsymbol{\sigma}_u + \mu_1\boldsymbol{\sigma}_v\,, \quad \mathbf{w} = \lambda_2\boldsymbol{\sigma}_u + \mu_2\boldsymbol{\sigma}_v\,.$$

By bilinearity of $II_{\mathbf{p}}$ we infer

$$\begin{aligned} II_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) &= \lambda_1\lambda_2\, II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u) + \lambda_1\mu_2\, II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v) \\ &\quad + \lambda_2\mu_1\, II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_u) + \mu_1\mu_2\, II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v) \\ &= du(\mathbf{v})du(\mathbf{w})\, II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u) + du(\mathbf{v})dv(\mathbf{w})\, II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v) \\ &\quad + dv(\mathbf{v})du(\mathbf{v})\, II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_u) + dv(\mathbf{v})dv(\mathbf{w})\, II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v) \\ &= (du(\mathbf{v}), dv(\mathbf{v})) \begin{pmatrix} II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u) & II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v) \\ II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_u) & II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v) \end{pmatrix} (du(\mathbf{w}), dv(\mathbf{w}))^T\,. \end{aligned}$$

By Lemma 4.162 and Lemma 4.161 we have

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) = -\mathbf{N}_u\,, \quad L = -\mathbf{N}_u \cdot \boldsymbol{\sigma}_u\,.$$

Therefore, using the above and the definition of $II_{\mathbf{p}}$, we get

$$II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u) = \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) \cdot \boldsymbol{\sigma}_u = -\mathbf{N}_u \cdot \boldsymbol{\sigma}_u = L\,.$$

With similar calculations we obtain

$$II_{\mathbf{p}}(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v) = II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_u) = M\,, \quad II_{\mathbf{p}}(\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_v) = N\,,$$

concluding the proof of point 2. In particular this also proves that $II_{\mathbf{p}}$ is symmetric, which is Point 1 of the statement. The fact that

$$II_{\mathbf{p}}(\mathbf{v},\mathbf{v}) = \mathscr{F}_2(\mathbf{v})$$

follows from Point 2 and definition of $\mathscr{F}_2$.

4.12.4 Matrix of Weingarten map

The Weingarten map is a linear map

$$\mathscr{W}_{\mathbf{p},\mathcal{S}} \colon T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S} \,.$$

We would like to find a formula to compute $\mathscr{W}_{\mathbf{p},\mathcal{S}}$. This is easily done: Given a chart $\boldsymbol{\sigma}$ at $\mathbf{p}$, we have that $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis for the vector space $T_{\mathbf{p}}\mathcal{S}$. Therefore there exists a 2×2 matrix A which represents $\mathscr{W}_{\mathbf{p},\mathcal{S}}$, that is,

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) = A\mathbf{v}\,, \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}\,.$$

It turns out that

$$A = \mathscr{F}_1^{-1}\mathscr{F}_2\,,$$

where we recall that

$$\mathscr{F}_1 = \begin{pmatrix} E & F \\ F & G \end{pmatrix}, \quad \mathscr{F}_2 = \begin{pmatrix} L & M \\ M & N \end{pmatrix},$$

where

$$E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u\,, \qquad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v\,, \qquad G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v\,,$$
$$L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N}\,, \qquad M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N}\,, \qquad N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N}\,,$$

and

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|}\,.$$

Let us prove this claim.

Theorem 4.163: Matrix of Weingarten map

Let $\mathcal{S}$ be an orientable surface and $\mathscr{W}_{\mathbf{p},\mathcal{S}} \colon T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}$ be its Weingarten map at $\mathbf{p}$. Let $\boldsymbol{\sigma}$ be a regular chart at $\mathbf{p}$, with $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$. Then

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) = \mathscr{F}_1^{-1}\mathscr{F}_2 \begin{pmatrix} \lambda \\ \mu \end{pmatrix}, \quad \forall\, v \in T_{\mathbf{p}}\mathcal{S}\,,$$

where

$$\mathbf{v} = \lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v\,,$$

with $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ evaluated at (u_0, v_0).

Proof

By Theorem 4.77 we know that $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis of $T_{\mathbf{p}}\mathcal{S}$. Since $\mathscr{W}_{\mathbf{p},\mathcal{S}} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}$ is linear, by standard linear algebra results there exist coefficients $a, b, c, d \in \mathbb{R}$ such that

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \quad \forall \, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S} \,,$$

where

$$\mathbf{v} = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v \,.$$

The coefficients $a, b, c, d \in \mathbb{R}$ can be compute by solving the linear system

$$\begin{aligned} \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) &= a\boldsymbol{\sigma}_u + b\boldsymbol{\sigma}_v \\ \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v) &= c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v \,. \end{aligned}$$

By Lemma 4.162 we have

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) = -\mathbf{N}_u \,, \quad \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v \,,$$

so that we obtain

$$\begin{aligned} -\mathbf{N}_u &= a\boldsymbol{\sigma}_u + b\boldsymbol{\sigma}_v \\ -\mathbf{N}_v &= c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v \,. \end{aligned}$$

Taking the scalar product of the above equations with $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ we get

$$\begin{aligned} -\mathbf{N}_u \cdot \boldsymbol{\sigma}_u &= a(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u) + b(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_u) \\ -\mathbf{N}_u \cdot \boldsymbol{\sigma}_v &= a(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v) + b(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v) \\ -\mathbf{N}_v \cdot \boldsymbol{\sigma}_u &= c(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u) + d(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_u) \\ -\mathbf{N}_v \cdot \boldsymbol{\sigma}_v &= c(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v) + d(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v) \end{aligned}$$

By Lemma 4.161 we have

$$\begin{aligned} \mathbf{N}_u \cdot \boldsymbol{\sigma}_u &= -L \,, & \qquad \mathbf{N}_u \cdot \boldsymbol{\sigma}_v &= -M \,, \\ \mathbf{N}_v \cdot \boldsymbol{\sigma}_u &= -M \,, & \qquad \mathbf{N}_v \cdot \boldsymbol{\sigma}_v &= -N \,. \end{aligned}$$

If in addition we recall the definition of E, F, G, we obtain

$$\begin{aligned} L &= aE + bF \\ M &= aF + bG \\ M &= cE + dF \\ N &= cF + dG \end{aligned}$$

The above equations are equivalent to the matrix multiplication

$$\begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} E & F \\ F & G \end{pmatrix} ,$$

which reads

$$\mathscr{F}_1 A = \mathscr{F}_2 \, .$$

Now, notice that

$$\det \mathscr{F}_1 > 0 \, .$$

Indeed, recall Cauchy-Schwarz inequality:

$$\mathbf{v} \cdot \mathbf{v} \leq \|\mathbf{v}\| \|\mathbf{w}\| \, , \quad \forall \, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3 \, ,$$

where the inequality is strict if and only if $\mathbf{v}$ and $\mathbf{w}$ are linearly independent. Since $\mathscr{S}$ is regular, we have that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent. Therefore by Cauchy-Schwarz we have

$$\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v < \|\boldsymbol{\sigma}_u\| \, \|\boldsymbol{\sigma}_v\| \, ,$$

and so, squaring both sides,

$$(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)^2 < \|\boldsymbol{\sigma}_u\|^2 \, \|\boldsymbol{\sigma}_v\|^2 \, .$$

Hence

$$\begin{aligned} \det(\mathscr{F}_1) &= EG - F^2 \\ &= (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u)(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v) - (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)^2 \\ &= \|\boldsymbol{\sigma}_u\|^2 \, \|\boldsymbol{\sigma}_v\|^2 - (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)^2 > 0 \, . \end{aligned}$$

In particular the matrix $\mathscr{F}_1$ is invertible and thus

$$A = \mathscr{F}_1^{-1} \mathscr{F}_2 \, ,$$

concluding the proof.

Important

A matrix $A \in \mathbb{R}^{n \times n}$ is invertible if and only if $\det(A) \neq 0$. In such case the inverse A^{-1} is computed via the formula

$$A^{-1} = \frac{1}{\det(A)} \operatorname{cof}(A)^T \, ,$$

where $\operatorname{cof}(A)$ is the cofactor matrix of A. For $n = 2$ the above formula reads:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} .$$

If the matrix is diagonal, then

$$\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} = \begin{pmatrix} 1/\lambda & 0 \\ 0 & 1/\mu \end{pmatrix} .$$

Notation

In the following we denote the matrix of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ by the symbol $\mathscr{W}$.

Example 4.164: Helicoid

The Helicoid is charted by

$$\boldsymbol{\sigma}(u,v) = (u\cos(v), u\sin(v), \lambda v)\,, \quad u \in [0,1]\,,\ v \in [0,4\pi)\,,$$

where $\lambda > 0$ is a constant, see Figure 4.30. Prove that the matrix of the Weingarten map is

$$\mathscr{W} = \begin{pmatrix} 0 & -\dfrac{\lambda}{(u^2+\lambda^2)^{1/2}} \\ \dfrac{\lambda}{(u^2+\lambda^2)^{3/2}} & 0 \end{pmatrix}.$$

Solution. We compute

$$\begin{aligned}
\boldsymbol{\sigma}_u &= (\cos(v), \sin(v), 0) \\
\boldsymbol{\sigma}_v &= (-u\sin(v), u\cos(v), \lambda) \\
\boldsymbol{\sigma}_{uu} &= (0,0,0) \\
\boldsymbol{\sigma}_{uv} &= (-\sin(v), \cos(v), 0) \\
\boldsymbol{\sigma}_{vv} &= (-u\cos(v), -u\sin(v), 0)
\end{aligned}$$

from which

$$\begin{aligned}
E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 \\
F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\
G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = u^2 + \lambda^2\,,
\end{aligned}$$

so that the first fundamental form is

$$\mathscr{F}_1 = \begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & u^2+\lambda^2 \end{pmatrix}.$$

Since $\mathscr{F}_1$ is diagonal, the inverse is immediately computed

$$\mathscr{F}_1^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{1}{u^2+\lambda^2} \end{pmatrix}.$$

Moreover

$$\begin{aligned}
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \cos(v) & \sin(v) & 0 \\ -u\sin(v) & u\cos(v) & \lambda \end{vmatrix} \\
&= (\lambda\sin(v), -\lambda\cos(v), u)
\end{aligned}$$

and so

$$\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = \sqrt{u^2+\lambda^2}\,.$$

The standard unit normal to $\boldsymbol{\sigma}$ is

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = \frac{1}{\sqrt{u^2+\lambda^2}}\,(\lambda\sin(v), -\lambda\cos(v), u)\,.$$

Hence

$$\begin{aligned}
L &= \boldsymbol{\sigma}_{uu}\cdot\mathbf{N} = 0\\
M &= \boldsymbol{\sigma}_{uv}\cdot\mathbf{N} = -\frac{\lambda}{\sqrt{u^2+\lambda^2}}\\
N &= \boldsymbol{\sigma}_{vv}\cdot\mathbf{N} = 0
\end{aligned}$$

and the second funamental form $\mathscr{F}_2$ is

$$\mathscr{F}_2 = \begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} 0 & -\dfrac{\lambda}{\sqrt{u^2+\lambda^2}} \\ -\dfrac{\lambda}{\sqrt{u^2+\lambda^2}} & 0 \end{pmatrix}.$$

Finally

$$\begin{aligned}
\mathscr{W} &= \mathscr{F}_1^{-1}\mathscr{F}_2\\
&= \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{1}{u^2+\lambda^2} \end{pmatrix}\begin{pmatrix} 0 & -\dfrac{\lambda}{\sqrt{u^2+\lambda^2}} \\ -\dfrac{\lambda}{\sqrt{u^2+\lambda^2}} & 0 \end{pmatrix}\\
&= \begin{pmatrix} 0 & -\dfrac{\lambda}{(u^2+\lambda^2)^{1/2}} \\ \dfrac{\lambda}{(u^2+\lambda^2)^{3/2}} & 0 \end{pmatrix}.
\end{aligned}$$

Example 4.165

Find the Weingarten matrix of the following surface chart

$$\boldsymbol{\sigma}(u,v) = \left(u-v, u+v, u^2+v^2\right)\,.$$

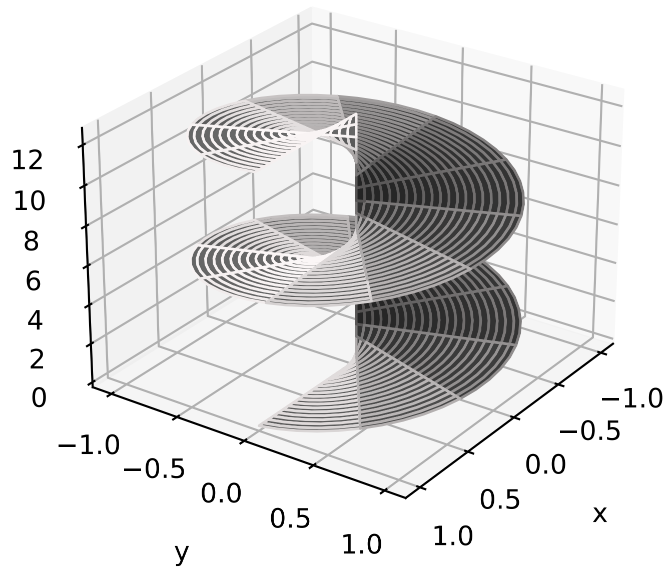


Figure 4.30: Plot of Helicoid.

Solution. Start by computing the first fundamental form:

$$\begin{aligned} \boldsymbol{\sigma}_u &= (1, 1, 2u) \\ \boldsymbol{\sigma}_v &= (-1, 1, 2v) \\ E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 2(1 + 2u^2) \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 4uv \\ G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 2(1 + 2v^2) \end{aligned}$$

so that

$$\mathscr{F}_1 = \begin{pmatrix} E & F \\ F & G \end{pmatrix} = \begin{pmatrix} 2(1+2u^2) & 4uv \\ 4uv & 2(1+2v^2) \end{pmatrix}$$

The determinant of $\mathscr{F}_1$ is

$$\det(\mathscr{F}_1) = 4(1 + 2u^2 + 2v^2)$$

and therefore

$$\begin{aligned} \mathscr{F}_1^{-1} &= \frac{1}{\det(\mathscr{F}_1)} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \\ &= \frac{1}{2(1+2u^2+2v^2)} \begin{pmatrix} 1+2v^2 & -2uv \\ -2uv & 1+2u^2 \end{pmatrix} . \end{aligned}$$

We now need to compute the second fundamental form

$$
\begin{aligned}
\boldsymbol{\sigma}_{uu} &= (0,0,2) \\
\boldsymbol{\sigma}_{uv} &= (0,0,0) \\
\boldsymbol{\sigma}_{vv} &= (0,0,2) \\
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 2u \\ -1 & 1 & 2v \end{vmatrix} \\
&= 2(v-u, -u-v, 1) \\
\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| &= 2\left(1+2u^2+2v^2\right)^{\frac{1}{2}} \\
\mathbf{N} &= \frac{(v-u,-u-v,1)}{\left(1+2u^2+2v^2\right)^{\frac{1}{2}}} \\
L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} &= \frac{2}{\left(1+2u^2+2v^2\right)^{\frac{1}{2}}} \\
M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} &= 0 \\
N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} &= \frac{2}{\left(1+2u^2+2v^2\right)^{\frac{1}{2}}}
\end{aligned}
$$

so that

$$
\begin{aligned}
\mathscr{F}_2 &= \begin{pmatrix} L & M \\ M & N \end{pmatrix} \\
&= \frac{2}{\left(1+2u^2+2v^2\right)^{\frac{1}{2}}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned}
$$

The matrix of the Weingarten map is

$$
\begin{aligned}
\mathscr{W} &= \mathscr{F}_1^{-1} \mathscr{F}_2 \\
&= \frac{1}{(1+2u^2+2v^2)^{\frac{3}{2}}} \begin{pmatrix} 1+2v^2 & -2uv \\ -2uv & 1+2u^2 \end{pmatrix}.
\end{aligned}
$$

4.13 Curvatures

Curvatures of a surface $\mathscr{S}$ are scalars associated to the Weingarten map $\mathscr{W}_{\mathbf{p},\mathscr{S}}$. We will define:

- Gaussian curvature

- Mean curvature
- Principal curvatures
- Normal curvature
- Geodesic curvature

4.13.1 Gaussian and mean curvature

The Weingarten map of $\mathcal{S}$ encodes the rate of change of the standard unit normal $\mathbf{N}$. We use this map to produce scalar values, which we call **curvatures**. The first two curvatures that we consider are called **Gaussian** and **mean** curvatures.

Definition 4.166: Gaussian and mean curvature

Let $\mathcal{S}$ be an orientable surface and let $\mathscr{W}$ denote the matrix of the Weingarten map $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ of $\mathcal{S}$ at $\mathbf{p}$. We define:

- The **Gaussian curvature** of $\mathcal{S}$ at $\mathbf{p}$ as

$$K := \det(\mathscr{W}),$$

- The **mean curvature** of $\mathcal{S}$ at $\mathbf{p}$ as

$$H := \frac{1}{2} \operatorname{trace}(\mathscr{W}),$$

Notation: Trace of a 2×2 matrix

We recall that the **trace** of a 2×2 matrix A is defined as the sum of the diagonal entries, that is,

$$\operatorname{trace} A = a + d, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Remark 4.167

The Gaussian curvature and mean curvature do not depend on the choice of basis of $T_{\mathbf{p}}\mathcal{S}$. Indeed, if $\widetilde{\mathscr{W}}$ is the matrix of the Weingarten map with respect to the basis $\{\tilde{\boldsymbol{\sigma}}_u, \tilde{\boldsymbol{\sigma}}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$, then

$$\det(\mathscr{W}) = \det(\widetilde{\mathscr{W}}), \quad \operatorname{trace}(\mathscr{W}) = \operatorname{trace}(\widetilde{\mathscr{W}}).$$

The above is true by a general linear algebra result: The determinant and trace of a matrix are invariant under change of basis.

Since we have shown that the matrix of the Weingarten map is

$$\mathscr{W} = \mathscr{F}_1^{-1}\mathscr{F}_2,$$

we can express K and H in terms of the first and second fundamental forms.

Proposition 4.168

Let $\mathcal{S}$ be an orientable surface and $\boldsymbol{\sigma}$ a regular chart at $\mathbf{p}$. Then

$$K = \frac{LN - M^2}{EG - F^2}\,, \qquad H = \frac{LG - 2MF - NE}{2(EG - F^2)}\,.$$

Proof

By Theorem 4.163 the matrix of the Weingarten map $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ of $\mathcal{S}$ at $\mathbf{p}$ is given by

$$\mathscr{W} = \mathscr{F}_1^{-1}\mathscr{F}_2\,.$$

We have

$$\det(\mathscr{F}_1) = \begin{vmatrix} E & F \\ F & G \end{vmatrix} = EF - G^2\,,$$
$$\det(\mathscr{F}_2) = \begin{vmatrix} L & M \\ M & N \end{vmatrix} = LN - M^2\,.$$

By the properties of determinant we get

$$\det(\mathscr{F}_1^{-1}) = \frac{1}{\det(\mathscr{F}_1)} = \frac{1}{EF - G^2}\,,$$

and therefore

$$\begin{aligned} K &= \det(\mathscr{W}) = \det\left(\mathscr{F}_1^{-1}\mathscr{F}_2\right) \\ &= \det(\mathscr{F}_1^{-1})\det(\mathscr{F}_2) = \frac{LN - M^2}{EG - F^2}\,. \end{aligned}$$

To compute H we need to find the diagonal entries of $\mathscr{W}$. Since

$$\mathscr{F}_1^{-1} = \frac{1}{EG - F^2}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}$$

we have

$$\mathscr{W} = \frac{1}{EG - F^2}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}\begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

From the above we compute

$$\begin{aligned} w_{11} &= \frac{1}{EG - F^2}(LG - MF) \\ w_{22} &= \frac{1}{EG - F^2}(-MF + EN) \end{aligned}$$

Therefore

$$\begin{aligned} H &= \frac{1}{2} \operatorname{trace} \mathscr{W} \\ &= \frac{1}{2} (w_{11} + w_{22}) \\ &= \frac{LG - 2MF + EN}{2(EG - F^2)} . \end{aligned}$$

Example 4.169: Plane

Consider the plane charted by

$$\boldsymbol{\sigma}(u,v) = \mathbf{a} + \mathbf{p}u + \mathbf{q}v , \quad u \in (0, 2\pi) , \, u, v \in \mathbb{R} .$$

We have already computed in Example 4.108 and Example 4.147 that the first and second fundamental forms of $\boldsymbol{\sigma}$ are

$$\mathscr{F}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \quad \mathscr{F}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} .$$

Therefore the matrix of the Weingarten map is

$$\mathscr{W} = \mathscr{F}_1^{-1} \mathscr{F}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} .$$

Hence the Gaussian curvature is

$$K = \det(\mathscr{W}) = 0 ,$$

while the mean curvature is

$$H = \frac{1}{2} \operatorname{trace} \mathscr{W} = 0 .$$

Example 4.170: Unit cylinder

Consider the unit cylinder charted by

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v) , \quad u \in (0, 2\pi) , \, v \in \mathbb{R} .$$

We have already computed in Example 4.111 and Example 4.148 that the first and second fundamental forms of $\boldsymbol{\sigma}$ are

$$\mathscr{F}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \quad \mathscr{F}_2 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} .$$

Therefore the matrix of the Weingarten map is

$$\begin{aligned}\mathscr{W} &= \mathscr{F}_1^{-1}\mathscr{F}_2 \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

Therefore the Gaussian curvature is

$$K = \det(\mathscr{W}) = 0\,,$$

while the mean curvature is

$$H = \frac{1}{2}\operatorname{trace}\mathscr{W} = -\frac{1}{2}\,.$$

4.13.2 Principal curvatures

Let V be a two-dimensional vector space. For a linear map $L\colon V \to V$ we say that $\lambda \in \mathbb{R}$ is an **eigenvalue** of L with **eigenvector** $\mathbf{v} \in V$ if

$$L(\mathbf{v}) = \lambda\mathbf{v}\,, \quad \mathbf{v} \neq 0\,.$$

Suppose $A \in \mathbb{R}^{2\times 2}$ is the matrix of L with respect to a basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ of V. Denote by

$$\mathbf{x} = (\lambda, \mu)\,, \quad \mathbf{v} = \lambda\mathbf{w}_1 + \mu\mathbf{w}_2\,.$$

the vector of coordinates of $\mathbf{v}$. Then

$$A\mathbf{v} = \lambda\mathbf{v}\,,$$

meaning that λ is an eigenvalue of A with eigenvector $\mathbf{x}$. The eigenvalues of A can be computed by solving the **characteristic equation**

$$P(\lambda) = 0\,, \quad P(\lambda) := \det(A - \lambda I)\,,$$

where P is the **characteristic polynomial** of A. Finally, we recall that $A \in \mathbb{R}^{2\times 2}$ is **diagonalizable** if there exists a diagonal matrix D and an invertible matrix P such that

$$A = P^{-1}DP\,.$$

Theorem 4.171

Let $\mathcal{S}$ be an orientable surface and let $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ be the Weingarten map at $\mathbf{p}$. There exist scalars $\kappa_1, \kappa_2 \in \mathbb{R}$ and an orthonormal basis $\{\mathbf{t}_1, \mathbf{t}_2\}$ of $T_{\mathbf{p}}\mathcal{S}$ such that

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) = \kappa_1\mathbf{t}_1\,, \quad \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) = \kappa_2\mathbf{t}_2\,.$$

Proof

Let $\boldsymbol{\sigma}$ be a chart for $\mathcal{S}$ at $\mathbf{p}$. Then $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis of $T_{\mathbf{p}}\mathcal{S}$. Let $\mathcal{W}$ be the matrix of $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ with respect to such basis. By Theorem 4.163 we have

$$\mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2 .$$

Recall that

$$\mathcal{F}_1^{-1} = \frac{1}{EG - F^2}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix} .$$

Thus $\mathcal{F}_1^{-1}$ is symmetric. Since $\mathcal{F}_2$ is symmetric, and the product of symmetric matrices is symmetric, we conclude that $\mathcal{W}$ is symmetric. Therefore $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ is self-adjoint, see Remark 4.15. The thesis now follows from the Spectral Theorem, see Theorem 1.13.

The matrix version of Theorem 4.171 is given in the following Corollary.

Corollary 4.172

Let $\mathcal{S}$ be orientable, and let $\mathcal{W}$ the matrix of the Weingarten map $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$, where $\boldsymbol{\sigma}$ is a regular chart at $\mathbf{p}$. Let $\kappa_1, \kappa_2, \mathbf{t}_1, \mathbf{t}_2$ be as in Theorem 4.171. Let $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{R}$ be such that

$$\mathbf{t}_1 = \lambda_1 \boldsymbol{\sigma}_u + \mu_1 \boldsymbol{\sigma}_v , \quad \mathbf{t}_2 = \lambda_2 \boldsymbol{\sigma}_u + \mu_2 \boldsymbol{\sigma}_v .$$

and denote

$$\mathbf{x}_1 = (\lambda_1, \mu_1) , \quad \mathbf{x}_2 = (\lambda_2, \mu_2) .$$

They hold:

- The scalars κ_1, κ_2 are eingenvalues of $\mathcal{W}$ of eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$, that is,

$$\mathcal{W}\mathbf{x}_1 = \kappa_1 \mathbf{x}_1 , \quad \mathcal{W}\mathbf{x}_2 = \kappa_2 \mathbf{x}_2 .$$

- The matrix $\mathcal{W}$ is diagonalizable, with

$$\mathcal{W} = P^{-1}DP , \quad D = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} , \quad P = \begin{pmatrix} \lambda_1 & \lambda_2 \\ \mu_1 & \mu_2 \end{pmatrix} .$$

Proof

Recall that $\mathcal{W}$ is the matrix of $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$. Therefore, by definition of $\mathbf{x}_1$, $\mathbf{x}_2$ we get

$$\mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) = \mathcal{W}\mathbf{x}_1 , \quad \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) = \mathcal{W}\mathbf{x}_2 .$$

The thesis follows by Theorem 4.171 and the Spectral Theorem for matrices, see Theorem 1.19.

The eigenvalues and eigenvectors of the weingarten map have a name.

Definition 4.173: Principal curvatures and vectors

Let $\mathcal{S}$ be an orientable surface and $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ be the Weingarten map of $\mathcal{S}$ at $\mathbf{p}$. We define:

- The **principal curvatures** of $\mathcal{S}$ at $\mathbf{p}$ are the eigenvalues κ_1, κ_2 of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$.
- The **principal vectors** corresponding to κ_1 and κ_2 are the eigenvectors $\mathbf{t_1}, \mathbf{t}_2$.

Remark 4.174: Computing principal curvatures and vectors

Corollary 4.172 gives an explicit way to compute the principal curvatures and vectors:

1. Compute the eigenvalues of $\mathscr{W}$. This is done by solving for κ the equation
$$\det(\mathscr{W} - \kappa I) = 0 \,.$$
This gives one of the principal curvatures
$$\kappa_i = \kappa$$
2. Compute the eigenvector(s) related to the eigenvalue κ. This is done by finding scalars λ, μ which solve the linear system
$$(\mathscr{W} - \kappa_i I) \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = 0$$
This gives the eigenvector of $\mathscr{W}$
$$\mathbf{x}_i = (\lambda, \mu)$$
3. The principal vector associated to κ_i is
$$\mathbf{t}_i = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v$$

Remark 4.175: Computing principal curvatures and vectors

If the matrix of the Weingarten map has the form

$$\mathscr{W} = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}$$

then $\mathscr{W}$ is already diagonal. The eigenvalues of $\mathscr{W}$ are κ_1 and κ_2, with eigenvectors

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

Therefore κ_1, κ_2 are the principal curvatures, with principal vectors given by

$$\mathbf{t}_1 = \boldsymbol{\sigma}_u \,, \quad \mathbf{t}_2 = \boldsymbol{\sigma}_v \,.$$

The principal curvatures are related to the Gaussian and mean curvatures.

Proposition 4.176

Let $\mathcal{S}$ be an orientable surface. Then

$$K = \kappa_1 \kappa_2 \,, \quad H = \frac{\kappa_1 + \kappa_2}{2} \,.$$

Proof

By Corollary 4.172 we have

$$\mathscr{W} = P^{-1} D P \,, \quad D = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} .$$

By the properties of determinant

$$\det(AB) = \det(A)\det(B) \,, \quad \forall \, A, B \in \mathbb{R}^{2\times 2} \,.$$

By definition of Gaussian curvature and the above formula we infer

$$\begin{aligned} K &= \det(\mathscr{W}) \\ &= \det\left(P^{-1} D P\right) \\ &= \det(P^{-1}) \det(D) \det(P) \\ &= \det(D) \\ &= \kappa_1 \kappa_2 \,, \end{aligned}$$

where we also used that

$$\det(P^{-1}) = \frac{1}{\det(P)} \,.$$

The trace satisfies

$$\operatorname{trace}(AB) = \operatorname{trace}(BA) \,, \quad \forall \, A, B \in \mathbb{R}^{2\times 2} \,.$$

By definition of mean curvature and the above formula we get

$$\begin{aligned} H &= \frac{1}{2} \operatorname{trace}(\mathscr{W}) \\ &= \frac{1}{2} \operatorname{trace}\left(P^{-1} D P\right) \\ &= \frac{1}{2} \operatorname{trace}\left(P P^{-1} D\right) \\ &= \frac{1}{2} \operatorname{trace}(D) \\ &= \frac{1}{2} (\kappa_1 + \kappa_2) \,, \end{aligned}$$

concluding the proof.

Important

In general κ_1 and κ_2 are hard to compute, as they require solving a second order equation. Instead K and H are easier to compute, as they are directly expressed in terms of the first and second fundamental form coefficients.

Example 4.177: Unit Cylinder

Consider the unit cylinder charted by

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v)\,, \quad u \in (0, 2\pi)\,,\ v \in \mathbb{R}\,.$$

We have already computed in Example 4.171 that the matrix of the Weingarten map is

$$\mathscr{W} = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Since $\mathscr{W}$ is diagonal, the eigenvalues are the diagonal entries of $\mathscr{W}$ and eigenvectors are

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

Therefore the principal curvatures are

$$\kappa_1 = -1\,, \quad \kappa_2 = 0$$

and the principal vectors are

$$\begin{aligned} \mathbf{t}_1 &= \boldsymbol{\sigma}_u = (-\sin(u), \cos(v), 0)\,, \\ \mathbf{t}_2 &= \boldsymbol{\sigma}_v = (0,0,1)\,, \end{aligned}$$

as shown in Figure 4.31.

Example 4.178: Sphere

Consider the chart for the sphere

$$\boldsymbol{\sigma}(u,v) = (\cos(u)\sin(v), \sin(u)\sin(v), \cos(v))$$

Prove that

$$\mathcal{F}_1 = \mathcal{F}_2 = \begin{pmatrix} \sin^2(v) & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathscr{W} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and

$$K = H = \kappa_1 = \kappa_2 = 1\,, \quad \mathbf{t}_1 = \boldsymbol{\sigma}_u\,, \quad \mathbf{t}_2 = \boldsymbol{\sigma}_v\,.$$

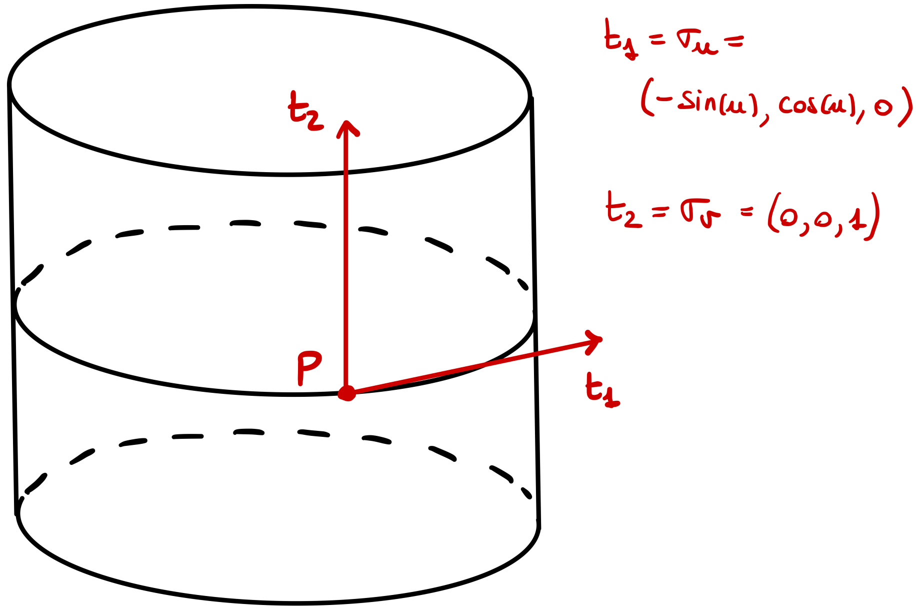


Figure 4.31: Principal vectors of the unit cylinder.

Solution. We compute

$$\begin{aligned}
\boldsymbol{\sigma}_u &= (-\sin(u)\sin(v), \cos(u)\sin(v), 0) \\
\boldsymbol{\sigma}_v &= (\cos(u)\cos(v), \sin(u)\cos(v), -\sin(v)) \\
E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = \sin^2(v) \\
F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\
G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1
\end{aligned}$$

and therefore the first fundamental form is

$$\mathscr{F}_1 = \begin{pmatrix} \sin^2(v) & 0 \\ 0 & 1 \end{pmatrix}.$$

Moreover

$$\begin{aligned}
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\sin(u)\sin(v) & \cos(u)\sin(v) & 0 \\ \cos(u)\cos(v) & \sin(u)\cos(v) & -\sin(v) \end{vmatrix} \\
&= (-\cos(u)\sin^2(v), -\sin(u)\sin^2(v), -\cos(v)\sin(v)) \\
\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| &= |\sin(v)| \\
\mathbf{N} &= (-\cos(u)\sin(v), -\sin(u)\sin(v), -\cos(v)) \\
\boldsymbol{\sigma}_{uu} &= (-\cos(u)\sin(v), -\sin(u)\sin(v), 0) \\
\boldsymbol{\sigma}_{uv} &= (-\sin(u)\cos(v), \cos(u)\cos(v), 0) \\
\boldsymbol{\sigma}_{vv} &= (-\cos(u)\sin(v), -\sin(u)\sin(v), -\cos(v)) \\
L &= \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} = \sin^2(v) \\
M &= \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = 0 \\
N &= \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} = 1
\end{aligned}$$

so that the second fundamental form is

$$\mathscr{F}_2 = \begin{pmatrix} \sin^2(v) & 0 \\ 0 & 1 \end{pmatrix} .$$

In particular the matrix of the Weingarten map is

$$\mathscr{W} = \mathscr{F}_1^{-1}\mathscr{F}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Since $\mathscr{W}$ is diagonal, the principal curvatures are

$$\kappa_1 = \kappa_2 = 1$$

and the principal vectors

$$\mathbf{t}_1 = \boldsymbol{\sigma}_u , \quad \mathbf{t}_2 = \boldsymbol{\sigma}_v .$$

Finally, we have that

$$H = \frac{\kappa_1 + \kappa_2}{2} = 1 , \quad K = \kappa_1\kappa_2 = 1 .$$

Example 4.179: Torus

Consider a circle $\mathscr{C}$ contained in the xz-plane, with center at distance $b > 0$ from the z-axis and radius a, with $0 < a < b$. The torus is obtained by rotating $\mathscr{C}$ around the z-axis. This surface is charted by

$$\boldsymbol{\sigma}(\theta, \phi) = ((a + b\cos(\theta))\cos(\phi), (a + b\cos(\theta))\sin(\phi), b\sin(\theta)) ,$$

where $\theta \in (-\pi/2, \pi/2)$ and $\phi \in (0, 2\pi)$. One can compute that the first and second fundamental forms are

$$\mathcal{F}_1 = \begin{pmatrix} b^2 & 0 \\ 0 & (a + b\cos(\theta))^2 \end{pmatrix}$$

$$\mathcal{F}_2 = \begin{pmatrix} b & 0 \\ 0 & (a + b\cos(\theta))\cos(\theta) \end{pmatrix}.$$

Therefore the matrix of the Weingarten map is

$$\mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2 = \begin{pmatrix} \frac{1}{b} & 0 \\ 0 & \frac{\cos(\theta)}{a + b\cos(\theta)} \end{pmatrix}.$$

Since $\mathcal{W}$ is diagonal, the principal curvatures are

$$\kappa_1 = \frac{1}{b}, \quad \kappa_2 = \frac{\cos(\theta)}{a + b\cos(\theta)},$$

and the principal vectors

$$\mathbf{t}_1 = \boldsymbol{\sigma}_u, \quad \mathbf{t}_2 = \boldsymbol{\sigma}_v.$$

The Gaussian and mean curvature are

$$K = \kappa_1\kappa_2 = \frac{\cos(\theta)}{b\,(a + b\cos(\theta))}$$

$$H = \frac{\kappa_1 + \kappa_2}{2} = \frac{a + 2b\cos(\theta)}{2b\,(a + b\cos(\theta))}$$

4.13.3 Normal and geodesic curvatures

Let $\mathcal{S}$ be a regular surface and consider all the curves $\boldsymbol{\gamma}$ on $\mathcal{S}$ passing through the point $\mathbf{p} \in \mathcal{S}$.

Question 4.180

Which curves through $\mathbf{p}$ have greatest or lowest curvature?

We start our analysis with the following proposition.

Proposition 4.181

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\gamma} : (a, b) \to \mathcal{S}$ be a unit speed curve. Then

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{N}, \mathbf{N} \times \dot{\boldsymbol{\gamma}}\}$$

is an orthornormal basis of $\mathbb{R}^3$ for all $t \in (a,b)$, where $\mathbf{N}$ is the standard unit normal to $\mathcal{S}$ evaluated at $\mathbf{p} = \boldsymbol{\gamma}(t)$.

Proof

By definition

$$\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}}\mathcal{S}\,, \quad \mathbf{p} := \boldsymbol{\gamma}(t)\,,$$

for all $t \in (a,b)$. This means $\dot{\boldsymbol{\gamma}}$ is tangent to $\mathcal{S}$. Thus

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{N} = 0\,.$$

We have $\|\dot{\boldsymbol{\gamma}}\| = 1$ since $\boldsymbol{\gamma}$ is unit speed. Moreover $\|\mathbf{N}\| = 1$ by definition. Since $\dot{\boldsymbol{\gamma}}$ and $\mathbf{N}$ are orthogonal, we also obtain

$$\|\mathbf{N} \times \dot{\boldsymbol{\gamma}}\| = \|\mathbf{N}\| \, \|\dot{\boldsymbol{\gamma}}\| = 1\,,$$

by the properties of vector product. Finally

$$(\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \cdot \mathbf{N} = 0\,, \quad (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \cdot \dot{\boldsymbol{\gamma}} = 0\,,$$

by the properties of vector product.

Important

Notice that the basis

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{N}, \mathbf{N} \times \dot{\boldsymbol{\gamma}}\}$$

does not coincide with the Frenet frame of $\boldsymbol{\gamma}$ in general.

Proposition 4.182

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\gamma} \colon (a,b) \to \mathcal{S}$ be a unit speed curve. Then

$$\ddot{\boldsymbol{\gamma}} = \kappa_n \mathbf{N} + \kappa_g \, (\mathbf{N} \times \dot{\boldsymbol{\gamma}})\,, \tag{4.16}$$

where $\mathbf{N}$ is evaluated at $\mathbf{p} := \boldsymbol{\gamma}(t)$ and κ_n, κ_g are scalars depedent on $\mathbf{p}$. Moreover

$$\kappa_n = \ddot{\boldsymbol{\gamma}} \cdot \mathbf{N}\,, \quad \kappa_g = \ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}})\,, \tag{4.17}$$

$$\kappa^2 = \kappa_n^2 + \kappa_g^2\,, \tag{4.18}$$

$$\kappa_n = \kappa \cos(\phi)\,, \quad \kappa_g = \pm \kappa \sin(\phi)\,, \tag{4.19}$$

where κ is the curvature of $\boldsymbol{\gamma}$ and ϕ is the angle between $\mathbf{N}$ and $\mathbf{n}$, the principal unit normal of $\boldsymbol{\gamma}$.

Proof

Part 1. By Proposition 4.181 we know that

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{N}, \mathbf{N} \times \dot{\boldsymbol{\gamma}}\}$$

is an orthornormal basis of $\mathbb{R}^3$. Hence

$$\ddot{\boldsymbol{\gamma}} = a\dot{\boldsymbol{\gamma}} + b\mathbf{N} + c\left(\mathbf{N} \times \dot{\boldsymbol{\gamma}}\right) ,$$

for some coefficients $a, b, c \in \mathbb{R}$. Since $\boldsymbol{\gamma}$ is unit speed, we have that

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

On the other hand,

$$\dot{\boldsymbol{\gamma}} \cdot \ddot{\boldsymbol{\gamma}} = a(\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}}) + b(\dot{\boldsymbol{\gamma}} \cdot \mathbf{N}) + c\dot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) = a \, ,$$

since $\dot{\boldsymbol{\gamma}}$ is orthogonal to $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$, and

$$\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = \left\|\dot{\boldsymbol{\gamma}}\right\|^2 = 1 \, .$$

Therefore $a = 0$ and

$$\ddot{\boldsymbol{\gamma}} = b\mathbf{N} + c\left(\mathbf{N} \times \dot{\boldsymbol{\gamma}}\right) .$$

Setting $\kappa_n := b$ and $\kappa_g := c$ we conclude (4.16).
Part 2. Taking the scalar product of (4.16) with $\mathbf{N}$ yields

$$\ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} = \kappa_n \left\|\mathbf{N}\right\|^2 + \kappa_g \left(\mathbf{N} \times \dot{\boldsymbol{\gamma}}\right) \cdot \mathbf{N} = \kappa_n \, ,$$

where we used that $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ are orthonormal vectors. Similarly, taking the scalar product of (4.16) with $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ yields the second equation in (4.17).
Part 3. By (4.16) we infer

$$\begin{aligned} \left\|\ddot{\boldsymbol{\gamma}}\right\|^2 & = \kappa_n^2 \left\|\mathbf{N}\right\|^2 + 2\kappa_n\kappa_g \mathbf{N} \cdot \left(\mathbf{N} \times \dot{\boldsymbol{\gamma}}\right) + \kappa_g^2 \left\|\mathbf{N} \times \dot{\boldsymbol{\gamma}}\right\|^2 \\ & = \kappa_n^2 + \kappa_g^2 \, , \end{aligned}$$

where we used that $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ are orthonormal. Since $\kappa(t) = \|\ddot{\boldsymbol{\gamma}}(t)\|$, we get (4.18).
Part 4. Recalling that

$$\ddot{\boldsymbol{\gamma}} = \kappa \mathbf{n} \, ,$$

from the first equation in (4.17) we obtain

$$\begin{aligned} \kappa_n & = \ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} \\ & = \kappa \mathbf{n} \cdot \mathbf{N} \\ & = \kappa \|\mathbf{n}\|^2 \|\mathbf{N}\|^2 \cos(\phi) \\ & = \kappa \cos(\phi) \, , \end{aligned}$$

where we used that $\mathbf{n}$ and $\mathbf{N}$ have unit norm. Hence the first equation in (4.19) is established. By (4.18) we get

$$\begin{aligned}\kappa_g^2 &= \kappa^2 - \kappa_n^2 \\ &= \kappa^2 \cos^2(\phi) - \kappa_n^2 \\ &= \kappa^2(\cos^2(\phi) - 1) \\ &= \kappa^2 \sin^2(\phi)\,,\end{aligned}$$

from which we obtain the second equation in (4.19).

The quantities κ_n and κ_g are the normal and geodesic curvatures of $\boldsymbol{\gamma}$.

Definition 4.183: Normal and geodesic curvature

Let $\mathcal{S}$ be regular and $\boldsymbol{\gamma} \colon (a,b) \to \mathcal{S}$ a unit speed curve. By (4.16) we have

$$\ddot{\boldsymbol{\gamma}} = \kappa_n \mathbf{N} + \kappa_g (\mathbf{N} \times \dot{\boldsymbol{\gamma}})$$

for $\mathbf{N}$ the standard unit normal to $\mathcal{S}$ and scalars $\kappa_n, \kappa_g \in \mathbb{R}$. We call

- κ_n the **normal curvature** of $\boldsymbol{\gamma}$,
- κ_g the **geodesic curvature** of $\boldsymbol{\gamma}$.

The normal curvature κ_n can be computed via the second fundamental form, as shown in the theorem below.

Theorem 4.184

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\gamma} \colon (a,b) \to \mathcal{S}$ a unit speed curve. Denote $\mathbf{p} := \boldsymbol{\gamma}(t)$. We have:

1. The normal curvature κ_n satisfies
$$\kappa_n = II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}})\,.$$
2. Let $\boldsymbol{\sigma}$ be a chart for $\mathcal{S}$ at $\mathbf{p}$. Then
$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$$
for some smooth functions $u, v \colon (a,b) \to \mathbb{R}$, and
$$\kappa_n = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2\,.$$

Proof

Part 1. By definition we have

$$\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}}\mathcal{S}$$

when $\mathbf{p} = \boldsymbol{\gamma}(t)$. Set

$$\tilde{\boldsymbol{\gamma}}(t) := \mathbf{N}(\boldsymbol{\gamma}(t)) . \tag{4.20}$$

By definition of differential we have

$$d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) = \dot{\tilde{\boldsymbol{\gamma}}}(t) . \tag{4.21}$$

Note that

$$\tilde{\boldsymbol{\gamma}}(t) \cdot \dot{\boldsymbol{\gamma}}(t) = 0 ,$$

since $\mathbf{N}$ is normal to $\mathcal{S}$ at $\mathbf{p}$ and $\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}}(\mathcal{S})$. Differentiating the above expression we get

$$\begin{aligned} 0 &= \frac{d}{dt}(\tilde{\boldsymbol{\gamma}}(t) \cdot \dot{\boldsymbol{\gamma}}(t)) \\ &= \tilde{\boldsymbol{\gamma}}(t) \cdot \ddot{\boldsymbol{\gamma}}(t) + \dot{\tilde{\boldsymbol{\gamma}}}(t) \cdot \dot{\boldsymbol{\gamma}}(t) \\ &= \mathbf{N}(\boldsymbol{\gamma}(t)) \cdot \ddot{\boldsymbol{\gamma}}(t) + d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) \end{aligned}$$

where in the last equation we used (4.20) and (4.21). Hence

$$-d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) = \mathbf{N}(\boldsymbol{\gamma}(t)) \cdot \ddot{\boldsymbol{\gamma}}(t) . \tag{4.22}$$

By definition of Weingarten and Gauss map we get

$$\mathcal{W}_{\mathbf{p},\mathcal{S}}(\dot{\boldsymbol{\gamma}}(t)) = -d_{\mathbf{p}}\mathcal{G}(\dot{\boldsymbol{\gamma}}(t)) = -d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) . \tag{4.23}$$

Therefore, using (4.22) and (4.23), we infer

$$\begin{aligned} II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}(t), \dot{\boldsymbol{\gamma}}(t)) &= \mathcal{W}_{\mathbf{p},\mathcal{S}}(\dot{\boldsymbol{\gamma}}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) \\ &= -d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) \\ &= \mathbf{N}(\boldsymbol{\gamma}(t)) \cdot \ddot{\boldsymbol{\gamma}}(t) \\ &= \kappa_n , \end{aligned}$$

where in the last equality we used (4.17).
Part 2. Let $\boldsymbol{\sigma}$ be a chart at $\mathbf{p}$ and

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)) .$$

Differentiating the above expression we get

$$\dot{\boldsymbol{\gamma}}(t) = \dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v .$$

By definition of du and dv, see Definition 4.105, we have

$$du(\dot{\boldsymbol{\gamma}}(t)) = \dot{u}(t) , \quad dv(\dot{\boldsymbol{\gamma}}(t)) = \dot{v}(t) .$$

Therefore, using Part 1 and Theorem 4.160, we obtain

$$\begin{aligned} \kappa_n &= II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}(t), \dot{\boldsymbol{\gamma}}(t)) \\ &= L du(\dot{\boldsymbol{\gamma}}(t))^2 + 2M du(\dot{\boldsymbol{\gamma}}(t)) dv(\dot{\boldsymbol{\gamma}}(t)) + N dv(\dot{\boldsymbol{\gamma}}(t))^2 \\ &= L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2 . \end{aligned}$$

Example 4.185: Curves on the sphere

Consider the chart for the sphere

$$\boldsymbol{\sigma}(u,v) = (\cos(u)\sin(v), \sin(u)\sin(v), \cos(v))$$

Show that

$$\kappa_n(t) = 1$$

for all unit speed curves on the sphere.

Solution. We have computed in Example 4.178 that the second fundamental form of $\boldsymbol{\sigma}$ is

$$\mathscr{F}_2 = \sin^2(v)du^2 + dv^2$$

Let $\boldsymbol{\gamma}$ be a unit speed curve on the sphere, that is,

$$\boldsymbol{\gamma}(t) = \sigma(u(t), v(t)) \,. \tag{4.24}$$

By Theorem 4.184 the normal curvature of $\boldsymbol{\gamma}$ is

$$\kappa_n = \sin^2(v)\dot{u}^2 + \dot{v}^2 \,.$$

Differentiating (4.24) we get

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) &= \frac{d}{dt}(\cos(u(t))\sin(v(t)), \sin(u(t))\sin(v(t)), \cos(v(t))) \\ &= (-\dot{u}\sin(u)\sin(v) + \dot{v}\cos(u)\cos(v), \dot{u}\cos(u)\sin(v) + \\ &\quad\ \dot{v}\sin(u)\cos(v), -\dot{v}\sin(v)) \end{aligned}$$

so that

$$\|\dot{\boldsymbol{\gamma}}(t)\|^2 = \sin^2(v)\dot{u}^2 + \dot{v}^2 \,.$$

Since $\boldsymbol{\gamma}$ is unit speed, we also get

$$\|\dot{\boldsymbol{\gamma}}\|^2 = 1 \,,$$

showing that

$$\kappa_n = \sin^2(v)\dot{u}^2 + \dot{v}^2 = 1 \,,$$

as required.

The normal curvature κ_n is related to the principal curvatures κ_1 and κ_2.

Theorem 4.186: Euler's Theorem

Let $\mathcal{S}$ be a regular surface and denote by κ_1, κ_2 the principal curvatures with principal vectors $\mathbf{t}_1, \mathbf{t}_2$. Let $\boldsymbol{\gamma}$ be a unit speed curve on $\mathcal{S}$. The normal curvature of $\boldsymbol{\gamma}$ is given by

$$\kappa_n = \kappa_1 \cos^2(\theta) + \kappa_2 \sin^2(\theta) \,,$$

where θ is the angle between $\dot{\boldsymbol{\gamma}}$ and $\mathbf{t}_1$.

Proof

Let $\boldsymbol{\gamma}$ be a unit speed curve on $\mathcal{S}$ and set

$$\mathbf{p} := \boldsymbol{\gamma}(t) .$$

By Theorem 4.171 the principal vectors $\{\mathbf{t}_1, \mathbf{t}_2\}$ form an orthonormal basis of $T_{\mathbf{p}}\mathcal{S}$. Since by definition

$$\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}}\mathcal{S} ,$$

there exist scalars $\lambda, \mu \in \mathbb{R}$ such that

$$\dot{\boldsymbol{\gamma}}(t) = \lambda \mathbf{t}_1 + \mu \mathbf{t}_2 .$$

As $\boldsymbol{\gamma}$ is unit speed and $\mathbf{t}_1, \mathbf{t}_2$ orthonormal, we infer

$$1 = \|\dot{\boldsymbol{\gamma}}(t)\|^2 = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = \lambda^2 + \mu^2 .$$

Therefore there exists $\theta \in [0, 2\pi]$ such that

$$\lambda = \cos(\theta), \quad \mu = \sin(\theta) .$$

Hence

$$\dot{\boldsymbol{\gamma}}(t) = \cos(\theta)\mathbf{t}_1 + \sin(\theta)\mathbf{t}_2 . \tag{4.25}$$

In particular, we can take the scalar product of (4.25) with $\mathbf{t}_1$ to get

$$\cos(\theta) = \lambda = \dot{\boldsymbol{\gamma}}(t) \cdot \mathbf{t}_1 .$$

Since $\dot{\boldsymbol{\gamma}}$ and $\mathbf{t}_1$ are unit vectors, from the above equation we conclude that θ is the angle between $\dot{\boldsymbol{\gamma}}$ and $\mathbf{t}_1$. In addition, recall that

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) = \kappa_1 \mathbf{t}_1 , \quad \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) = \kappa_2 \mathbf{t}_2 ,$$

and $\mathbf{t}_1$, $\mathbf{t}_2$ are orthonormal. Thus

$$\begin{aligned}
II_{\mathbf{p}}(\mathbf{t}_1, \mathbf{t}_1) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) \cdot \mathbf{t}_1 = \kappa_1 \|\mathbf{t}_1\|^2 = \kappa_1 \\
II_{\mathbf{p}}(\mathbf{t}_1, \mathbf{t}_2) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) \cdot \mathbf{t}_2 = \kappa_1 \mathbf{t}_1 \cdot \mathbf{t}_2 = 0 \\
II_{\mathbf{p}}(\mathbf{t}_2, \mathbf{t}_1) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) \cdot \mathbf{t}_1 = \kappa_2 \mathbf{t}_2 \cdot \mathbf{t}_1 = 0 \\
II_{\mathbf{p}}(\mathbf{t}_2, \mathbf{t}_2) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) \cdot \mathbf{t}_2 = \kappa_2 \|\mathbf{t}_2\|^2 = \kappa_2
\end{aligned}$$

By Theorem 4.184, equation (4.25), and bilinearity of $II_{\mathbf{p}}$, we get

$$\begin{aligned}
\kappa_n &= II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}}) \\
&= \cos^2(\theta)\, II_{\mathbf{p}}(\mathbf{t}_1, \mathbf{t}_1) + \cos(\theta)\sin(\theta)\, II_{\mathbf{p}}(\mathbf{t}_1, \mathbf{t}_2) \\
&\quad + \sin(\theta)\cos(\theta)\, II_{\mathbf{p}}(\mathbf{t}_2, \mathbf{t}_1) + \sin^2(\theta)\, II_{\mathbf{p}}(\mathbf{t}_2, \mathbf{t}_2) \\
&= \cos^2(\theta)\kappa_1 + \sin^2(\theta)\kappa_2
\end{aligned}$$

ending the proof.

As an immediate corollary of the Euler's Theorem we get the next statement.

Corollary 4.187

Let $\mathcal{S}$ be a regular surface and κ_1, κ_2 its principal curvatures at $\mathbf{p}$ with principal vectors $\mathbf{t}_1, \mathbf{t}_2$. Then:

- κ_1 and κ_2 are the minimum and maximum values of κ_n, for all unit speed curves on $\mathcal{S}$ passing through $\mathbf{p}$.
- The directions of lowest and highest curvature on $\mathcal{S}$ are given by $\mathbf{t}_1$ and $\mathbf{t}_2$.

In Example 4.185 we have shown with a direct argument that

$$\kappa_n = 1$$

for all unit speed curves on the sphere. Thanks to Euler's Theorem we can obtain an immediate proof of this fact.

Example 4.188: Curves on the sphere

Let us consider again the chart for the sphere

$$\boldsymbol{\sigma}(u, v) = (\cos(u)\sin(v), \sin(u)\sin(v), \cos(v))$$

as seen in Example 4.185. By Example 4.178, the principal curvatures of $\boldsymbol{\sigma}$ are

$$\kappa_1 = \kappa_2 = 1\,.$$

By Euler's Theorem, for any curve $\boldsymbol{\gamma}$ on the sphere we have

$$\kappa_n = \kappa_1 \cos^2(\theta) + \kappa_2 \sin^2(\theta) = 1\,.$$

4.14 Local shape of a surface

The principal curvatures κ_1 and κ_2 determine the maximum and minimum curvature of a surface $\mathcal{S}$, see Corollary 4.187. Hence we can study the local shape of $\mathcal{S}$ in function of κ_1 and κ_2.

Theorem 4.189: Local structure of surfaces

Let $\mathcal{S}$ be a regular surface and $\mathbf{p} \in \mathcal{S}$. In the vicinity of $\mathbf{p}$ the surface $\mathcal{S}$ is approximated by the quadric surface of equation

$$z = \frac{1}{2}\left(x^2\kappa_1(\mathbf{p}) + y^2\kappa_2(\mathbf{p})\right)\,, \tag{4.26}$$

where $\kappa_1(\mathbf{p}), \kappa_2(\mathbf{p})$ are the principal curvatures of $\mathcal{S}$ at $\mathbf{p}$.

Proof

By Theorem 4.171 the principal vectors $\{\mathbf{t}_1, \mathbf{t}_2\}$ are an orthonormal basis of $T_{\mathbf{p}}\mathcal{S}$. Therefore the standard unit normal $\mathbf{N}$ at $\mathbf{p}$ is orthogonal to both $\mathbf{t}_1$ and $\mathbf{t}_2$. Up to rotations and translations, we can assume WLOG that $\mathbf{p} = \mathbf{0}$ and

$$\mathbf{t}_1 = (1,0,0)\,, \quad \mathbf{t}_2 = (0,1,0)\,, \quad \mathbf{N} = (0,0,1)\,. \tag{4.27}$$

Let $\boldsymbol{\sigma}$ be a chart for $\mathcal{S}$ at $\mathbf{p}$. Up to reparametrizing, we can assume that

$$\boldsymbol{\sigma}(0,0) = \mathbf{p} = \mathbf{0}\,.$$

As $\mathbf{N} = (0,0,1)$, it follows that $T_{\mathbf{p}}\mathcal{S}$ is the xy-plane

$$T_{\mathbf{p}}\mathcal{S} = \mathbb{R}^2 = \{(x,y,0) \,:\, x,y \in \mathbb{R}\}\,.$$

Since $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis for $T_{\mathbf{p}}\mathcal{S}$, we have that for each $(x,y) \in \mathbb{R}^2$ there exist $(s,t) \in \mathbb{R}^2$ such that

$$(x,y,0) = s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v\,, \tag{4.28}$$

where $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are evaluated at $(0,0)$. The Taylor approximation of $\boldsymbol{\sigma}$ at $(0,0)$ is

$$\begin{aligned}
\boldsymbol{\sigma}(s,t) &= \boldsymbol{\sigma}(0,0) + s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v \\
&\qquad + \frac{1}{2}\left(s^2\boldsymbol{\sigma}_{uu} + 2st\boldsymbol{\sigma}_{uv} + t^2\boldsymbol{\sigma}_{vv}\right) + R\,, \\
&= (x,y,0) + \frac{1}{2}\left(s^2\boldsymbol{\sigma}_{uu} + 2st\boldsymbol{\sigma}_{uv} + t^2\boldsymbol{\sigma}_{vv}\right) + R
\end{aligned}$$

where R is a remainder and the derivatives of $\boldsymbol{\sigma}$ are evaluated at $(0,0)$. Hence, if x, y are small (and thus s, t are small), we have that

$$\boldsymbol{\sigma}(s,t) \approx (x,y,z)$$

where

$$\begin{aligned}
z &:= \frac{1}{2}\left(s^2\boldsymbol{\sigma}_{uu} + 2st\boldsymbol{\sigma}_{uv} + t^2\boldsymbol{\sigma}_{vv}\right) \cdot \mathbf{N} \\
&= \frac{1}{2}\left(Ls^2 + 2Mst + Nt^2\right)\,,
\end{aligned}$$

with L, M, N coefficients of the second fundamental form of $\boldsymbol{\sigma}$ at $(0,0)$. Set

$$\mathbf{v} := s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v\,.$$

By Theorem 4.160 we have

$$Ls^2 + 2Mst + Nt^2 = II_{\mathbf{p}}(\mathbf{v},\mathbf{v}) = \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) \cdot \mathbf{v}\,.$$

On the other hand, using (4.27) and (4.28) we get

$$\mathbf{v} = s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v = (x,y,0) = x\mathbf{t}_1 + y\mathbf{t}_2\,.$$

Since the Weingarten map is linear we get

$$\begin{aligned}\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) &= x\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) + y\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2)\\ &= x\kappa_1\mathbf{t}_1 + y\kappa_2\mathbf{t}_2\,,\end{aligned}$$

where we used that $\mathbf{t}_1$ and $\mathbf{t}_2$ are eigenvectors of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ with eigenvalues κ_1 and κ_2. Hence

$$\begin{aligned}\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v})\cdot\mathbf{v} &= x\kappa_1\mathbf{t}_1 + y\kappa_2\mathbf{t}_2\cdot(x\mathbf{t}_1 + y\mathbf{t}_2)\\ &= x^2\kappa_1 + y^2\kappa_2\end{aligned}$$

Therefore

$$\begin{aligned}z &= \frac{1}{2}\left(Ls^2 + 2Mst + Nt^2\right)\\ &= \frac{1}{2}\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v})\cdot\mathbf{v}\\ &= \frac{1}{2}\left(x^2\kappa_1 + y^2\kappa_2\right)\,,\end{aligned}$$

showing that

$$\boldsymbol{\sigma}(t,s) \approx \left(x, y, \frac{1}{2}\left(x^2\kappa_1 + y^2\kappa_2\right)\right)\,.$$

Thanks to Theorem 4.189 we can distinguish between 4 approximating shapes.

Definition 4.190: Local shape types

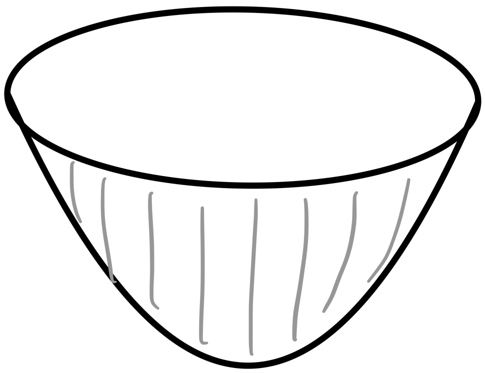
Let $\mathcal{S}$ be a regular surface and denote by $\kappa_1(\mathbf{p})$ and $\kappa_2(\mathbf{p})$ its principal curvatures at $\mathbf{p}$. The point $\mathbf{p}$ is

- **Elliptic** if
$$\kappa_1(\mathbf{p}) > 0\,,\ \kappa_2(\mathbf{p}) > 0 \quad \text{or} \quad \kappa_1(\mathbf{p}) < 0\,,\ \kappa_2(\mathbf{p}) < 0$$
Then (4.26) is the equation of an **elliptic paraboloid**.

- **Hyperbolic** if
$$\kappa_1(\mathbf{p}) < 0 < \kappa_2(\mathbf{p}) \quad \text{or} \quad \kappa_2(\mathbf{p}) < 0 < \kappa_1(\mathbf{p})$$
Then (4.26) is the equation of a **hyperbolic paraboloid**.

- **Parabolic** if
$$\kappa_1(\mathbf{p}) = 0\,,\ \kappa_2(\mathbf{p}) \neq 0 \quad \text{or} \quad \kappa_2(\mathbf{p}) \neq 0,\ \kappa_1(\mathbf{p}) = 0$$
Then (4.26) is the equation of a **parabolic cylinder**.

- **Planar** if
$$\kappa_1(\mathbf{p}) = \kappa_2(\mathbf{p}) = 0$$
Then (4.26) is the equation of a **plane**.

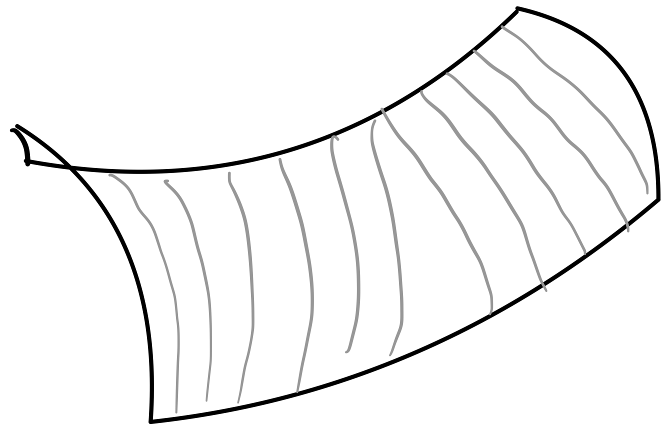


ELLIPTIC

$k_1, k_2 > 0$

OR

$k_1, k_2 < 0$

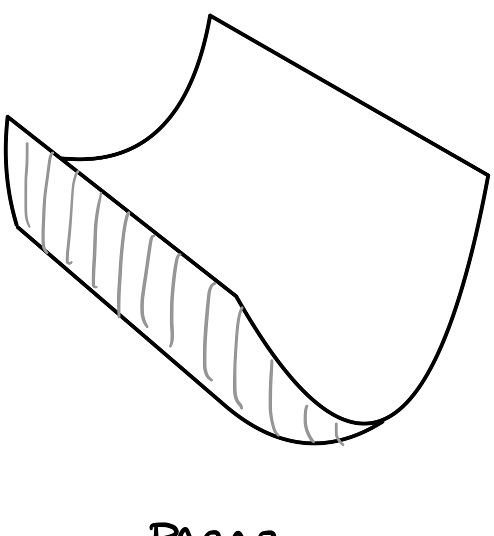


HYPERBOLIC

$k_1 < 0 < k_2$

OR

$k_2 < 0 < k_1$

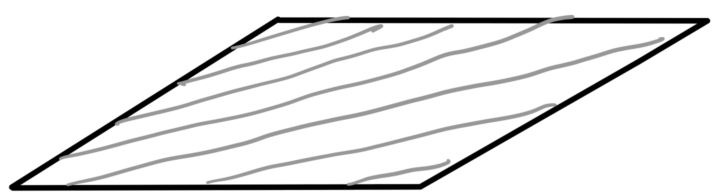


PARABOLIC

$k_1 = 0 \, , \; k_2 \neq 0$

OR

$k_2 = 0 \, , \; k_1 \neq 0$



PLANAR

$k_1 = k_2 = 0$

Example 4.191

Consider the surface chart

$$\boldsymbol{\sigma}(u,v) = \left(u - v, u + v, u^2 + v^2\right) .$$

Show that $\mathbf{p} = \boldsymbol{\sigma}(1,0)$ is an elliptic point. Therefore $\boldsymbol{\sigma}$ is approximated by an *elliptic paraboiloid* in the vicinity of $\mathbf{p}$.

Solution. In Example 4.165 we have shown that the Weingarten matrix of $\boldsymbol{\sigma}$ is

$$\mathscr{W} = \frac{1}{(1+2u^2+2v^2)^{\frac{3}{2}}} \begin{pmatrix} 1+2v^2 & -2uv \\ -2uv & 1+2u^2 \end{pmatrix} .$$

For $u = 1$ and $v = 1$ we obtain

$$\mathscr{W} = \frac{1}{3^{\frac{3}{2}}} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 3^{-\frac{3}{2}} & 0 \\ 0 & 3^{-\frac{1}{2}} \end{pmatrix} .$$

Therefore the principal curvatures at $\mathbf{p}$ are

$$\kappa_1(\mathbf{p}) = 3^{-\frac{3}{2}} , \quad \kappa_2(\mathbf{p}) = 3^{-\frac{1}{2}} .$$

Since $\kappa_1(\mathbf{p}) > 0$ and $\kappa_2(\mathbf{p}) > 0$ we have that $\mathbf{p}$ is an elliptic point.

4.14.1 Umbilical points

Definition 4.192: Umbilical point

Let $\mathcal{S}$ be a regular surface and denote by $\kappa_1(\mathbf{p})$ and $\kappa_2(\mathbf{p})$ its principal curvatures at $\mathbf{p}$. We say that $\mathbf{p}$ is an **umbilic** if

$$\kappa_1(\mathbf{p}) = \kappa_2(\mathbf{p}) .$$

Remark 4.193

Umbilical points might be **planar** or **elliptic**.

Suppose that $\mathbf{p}$ is an umbilic, that is,

$$\kappa_1 = \kappa_2$$

at $\mathbf{p}$. Let κ_n be the normal curvature of a unit speed curve $\boldsymbol{\gamma}$ passing through $\mathbf{p}$. By Theorem 4.186 we have

$$\kappa_n = \kappa_1 \cos^2(\theta) + \kappa_2 \sin^2(\theta) = \kappa_1 .$$

Therefore κ_n does not depend on $\boldsymbol{\gamma}$. Intuitively, this can only happen if in the vicinity of $\mathbf{p}$ the surface looks like a sphere or a plane. Indeed, the following theorem holds.

Theorem 4.194

Let $\mathscr{S}$ be a regular surface such that every point $\mathbf{p} \in \mathscr{S}$ is umbilic. Then $\mathscr{S}$ is an open subset of plane or a sphere.

Proof

By assumption we have

$$\kappa_1(\mathbf{p}) = \kappa_2(\mathbf{p}) = \kappa(\mathbf{p}) , \quad \forall \, \mathbf{p} \in \mathscr{S} . \tag{4.29}$$

Step 1. κ is constant.
By Theorem 4.171 the principal vectors $\{\mathbf{t}_1, \mathbf{t}_2\}$ are an orthonormal basis of $T_{\mathbf{p}}\mathscr{S}$. Hence, for each $\mathbf{v} \in T_{\mathbf{p}}\mathscr{S}$ there exist $\lambda, \mu \in \mathbb{R}$ such that

$$\mathbf{v} = \lambda \mathbf{t}_1 + \mu \mathbf{t}_2 .$$

Using the linearity of $\mathscr{W}_{\mathbf{p},\mathscr{S}}$ and (4.29) we obtain

$$\begin{aligned} \mathscr{W}_{\mathbf{p},\mathscr{S}}(\mathbf{v}) &= \lambda \mathscr{W}_{\mathbf{p},\mathscr{S}}(\mathbf{t}_1) + \mu \mathscr{W}_{\mathbf{p},\mathscr{S}}(\mathbf{t}_2) \\ &= \lambda \kappa \mathbf{t}_1 + \mu \kappa \mathbf{t}_2 \\ &= \kappa \mathbf{v} , \end{aligned}$$

showing that

$$\mathscr{W}_{\mathbf{p},\mathscr{S}}(\mathbf{v}) = \kappa \mathbf{v} , \quad \forall \, \mathbf{v} \in T_{\mathbf{p}}\mathscr{S} . \tag{4.30}$$

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a chart of $\mathscr{S}$. Up to restricting $\boldsymbol{\sigma}$, we can assume that U is connected. By Lemma 4.162 we have

$$\mathscr{W}_{\mathbf{p},\mathscr{S}}(\boldsymbol{\sigma}_u) = -\mathbf{N}_u , \quad \mathscr{W}_{\mathbf{p},\mathscr{S}}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v .$$

On the other hand, by (4.30) we infer

$$\mathscr{W}_{\mathbf{p},\mathscr{S}}(\boldsymbol{\sigma}_u) = \kappa \boldsymbol{\sigma}_u , \quad \mathscr{W}_{\mathbf{p},\mathscr{S}}(\boldsymbol{\sigma}_v) = \kappa \boldsymbol{\sigma}_v ,$$

from which

$$\mathbf{N}_u = -\kappa \boldsymbol{\sigma}_u , \quad \mathbf{N}_v = -\kappa \boldsymbol{\sigma}_v . \tag{4.31}$$

Thus

$$(\kappa \boldsymbol{\sigma}_u)_v = -(\mathbf{N}_u)_v = -(\mathbf{N}_v)_u = (\kappa \boldsymbol{\sigma}_v)_u .$$

Moreover

$$\begin{aligned} (\kappa \boldsymbol{\sigma}_u)_v &= \kappa_v \boldsymbol{\sigma}_u + \kappa \boldsymbol{\sigma}_{uv} \\ (\kappa \boldsymbol{\sigma}_v)_u &= \kappa_u \boldsymbol{\sigma}_v + \kappa \boldsymbol{\sigma}_{uv} , \end{aligned}$$

so that

$$\kappa_v \boldsymbol{\sigma}_u = \kappa_u \boldsymbol{\sigma}_v . \tag{4.32}$$

Recall that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent, being $\mathscr{S}$ regular. Hence the linear combination at (4.32) must be trivial, implying

$$\kappa_u = \kappa_v = 0 .$$

Since U is connected, the above implies that κ is constant.
Step 2. We have the two cases $\kappa = 0$ and $\kappa \neq 0$.

- Assume $\kappa = 0$. By (4.31) we get that
$$\mathbf{N}_u = \mathbf{N}_v = \mathbf{0}\,,$$
which implies $\mathbf{N}$ is constant. Therefore
$$(\mathbf{N} \cdot \boldsymbol{\sigma})_u = \mathbf{N}_u \cdot \boldsymbol{\sigma} + \mathbf{N} \cdot \boldsymbol{\sigma}_u = 0$$
since $\mathbf{N}_u = \mathbf{0}$ and $\mathbf{N} \cdot \boldsymbol{\sigma}_u = 0$ because $\mathbf{N}$ is orthogonal to $T_{\mathbf{p}}\mathcal{S}$. Similarly we get
$$(\mathbf{N} \cdot \boldsymbol{\sigma})_v = 0\,,$$
showing that $\mathbf{N} \cdot \boldsymbol{\sigma}$ is constant. Hence there exists $c \in \mathbb{R}$ such that
$$\mathbf{N} \cdot \boldsymbol{\sigma}(u,v) = c\,, \quad \forall\,(u,v) \in U\,.$$
This shows $\boldsymbol{\sigma}(U)$ is contained in the plane
$$\pi = \{\mathbf{x} \in \mathbb{R}^3 \;:\; \mathbf{N} \cdot \mathbf{x} = c\}\,.$$
- Assume $\kappa \neq 0$. Condition (4.31) implies
$$\mathbf{N} = -\kappa\boldsymbol{\sigma} + \mathbf{a}$$
for some $\mathbf{a} \in \mathbb{R}^3$ constant vector. Thus
$$\left\| \boldsymbol{\sigma} - \frac{1}{\kappa}\mathbf{a} \right\|^2 = \left\| -\frac{1}{\kappa}\mathbf{N} \right\|^2 = \frac{1}{\kappa^2}\,,$$
given that $\|\mathbf{N}\| = 1$. Therefore $\boldsymbol{\sigma}(U)$ is contained in the sphere of center $\mathbf{a}/\kappa$ and radius $1/\kappa$.

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
```

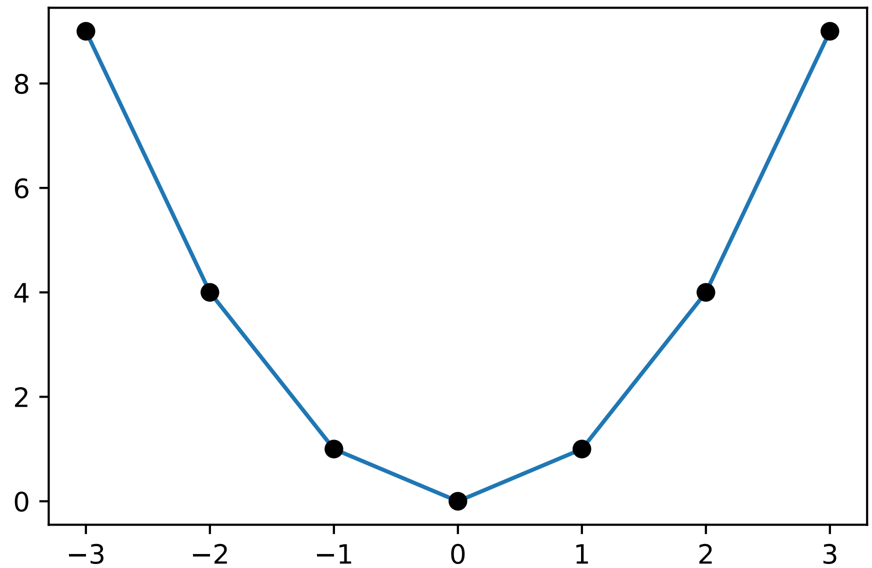
The above imports **numpy** and the module **pyplot** from **matplotlib**, and renames them to `np` and `plt`, respectively. These shorthands are standard in the literature, and they make code much more readable.
The function for plotting 2D graphs is called `plot(x,y)` and is contained in `plt`. As 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.

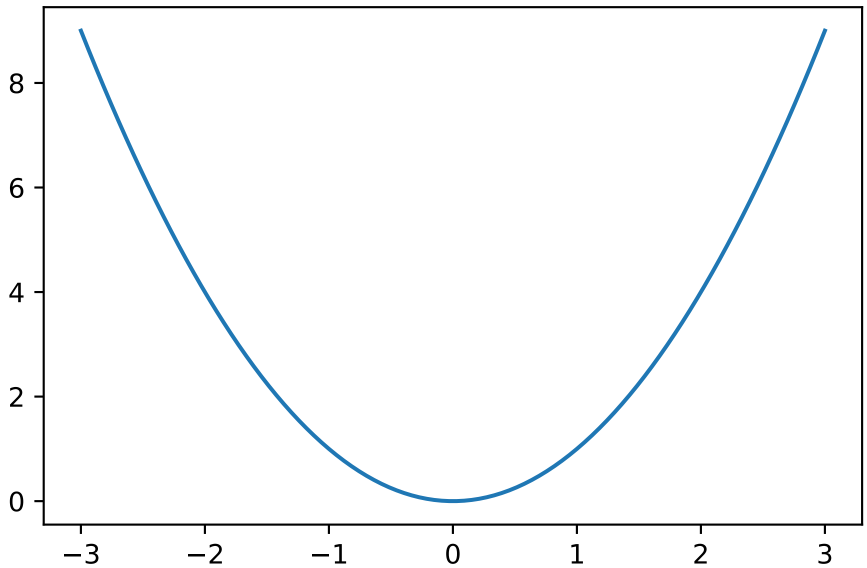
`np.linspace(a,b,n)`. Such call will return a numpy array which contains `n` evenly spaced points, starts at `a`, and ends in `b`. For example `np.linspace(-3,3,7)` returns our original array `t` at 5.1, as shown below

```
# 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,

$$t = [t_1, \dots, t_n] \qquad \Longrightarrow \qquad \sqrt{t} = [\sqrt{t_1}, \dots, \sqrt{t_n}] \,.$$

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:

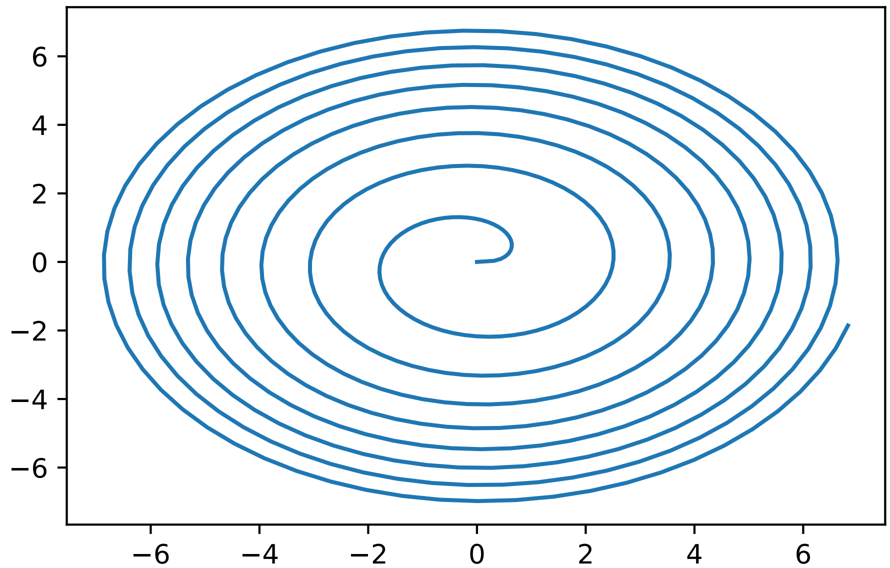


Figure 5.1: Fermat's spiral

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))

# 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()
```

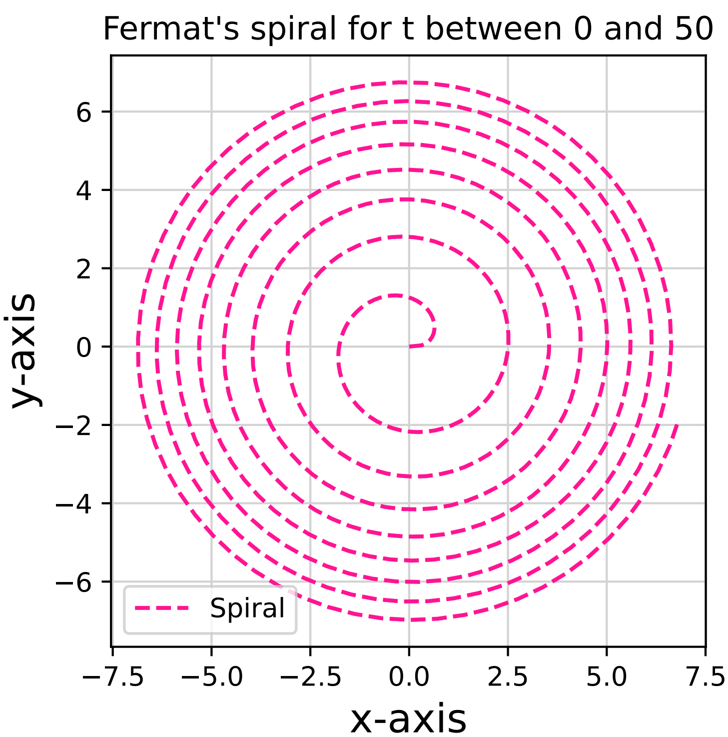


Figure 5.2: Adding a bit of style

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

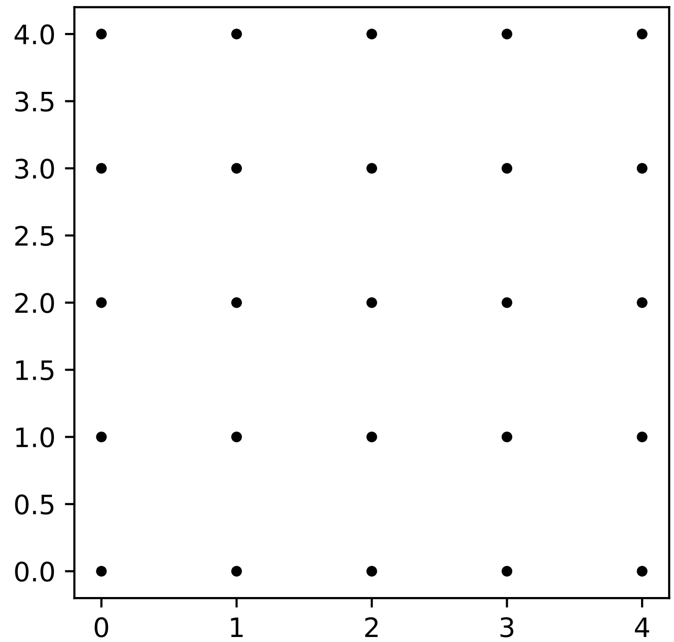


Figure 5.3: The 5 x 5 grid corresponding to the matrix A

To achieve this, first generate x and y coordinates using

```
x = np.linspace(0, 4, 5)
y = np.linspace(0, 4, 5)
```

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()
```

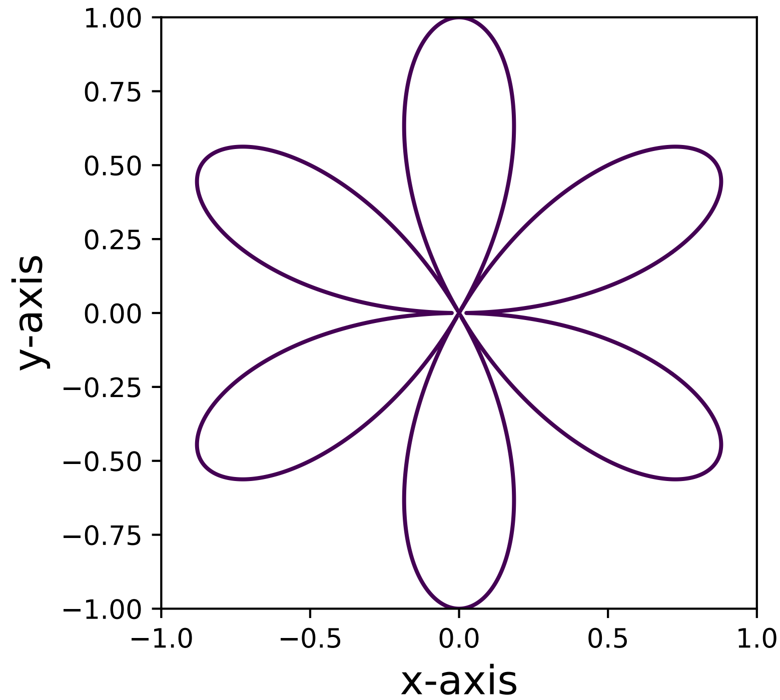


Figure 5.4: Plot of the curve defined by f=0

5.1.3 Curves in 3D

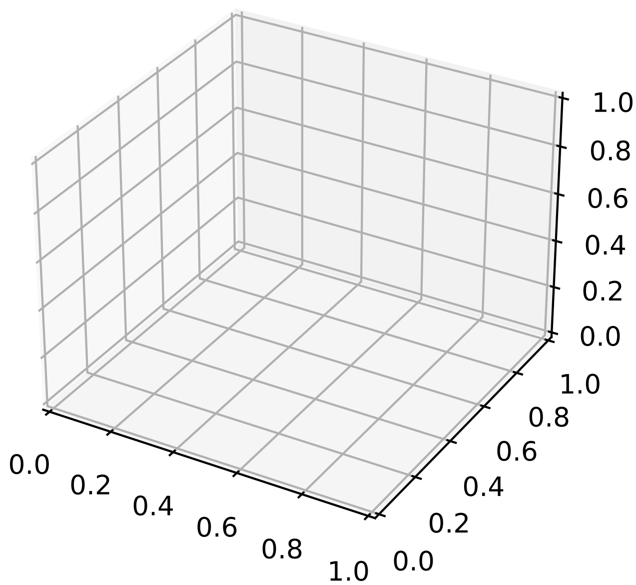
Plotting in 3D with matplotlib requires the `mplot3d` toolkit, see here for documentation. Therefore our first lines will always be

```
# 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()
```



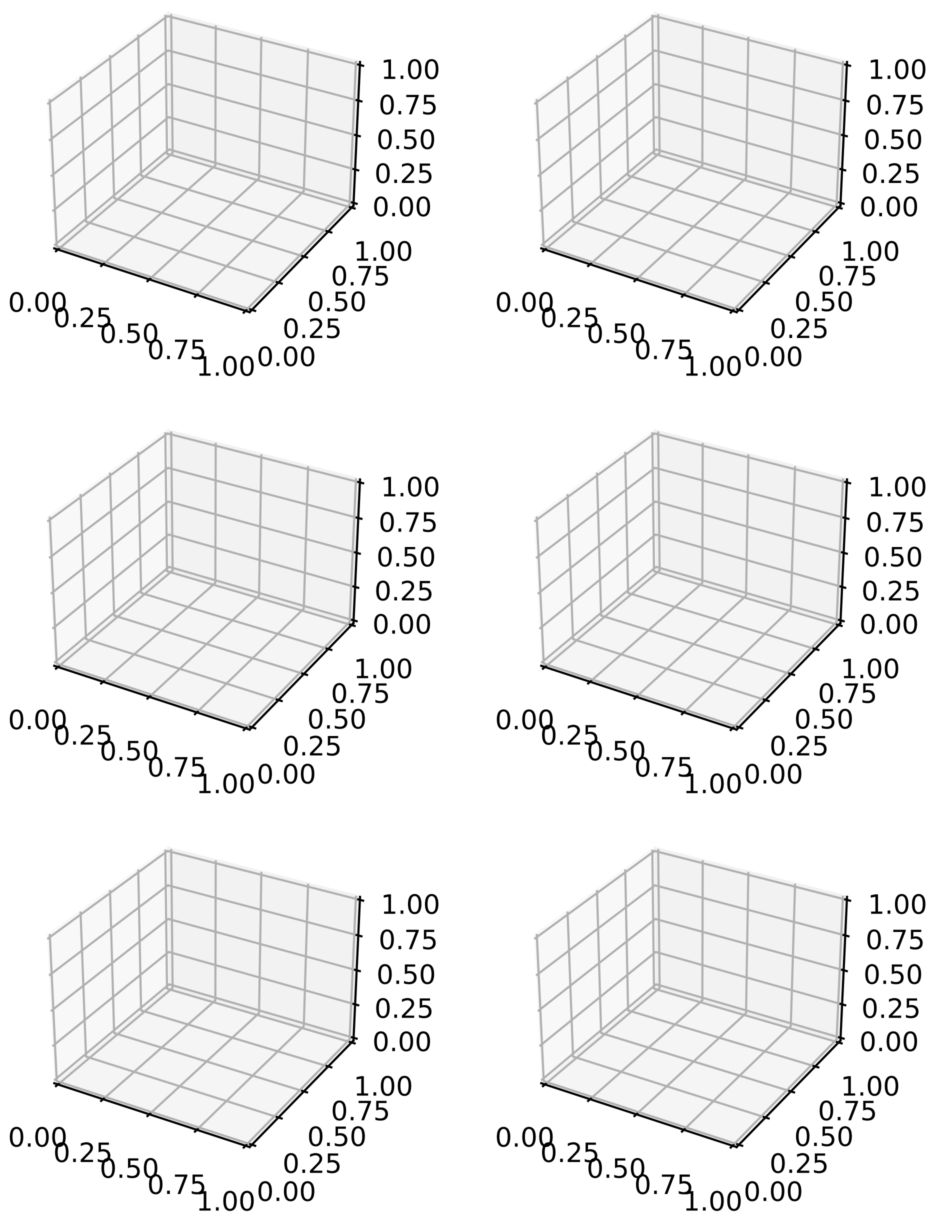
In the above code `fig` is a figure object, while `ax` is an axes object. In practice, the figure object contains the axes objects, and the actual plot information will be contained in axes. If you want multiple plots in the figure container, you should use the command

```
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

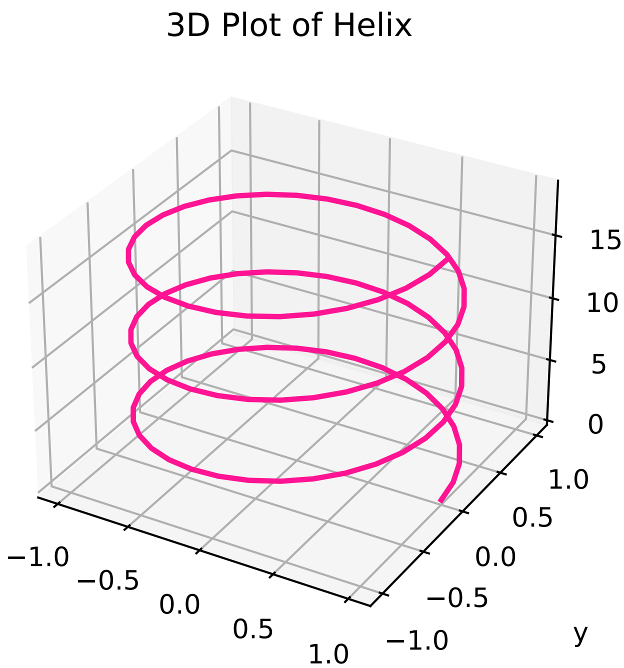
```
# Code to plot 3D curve

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()
```



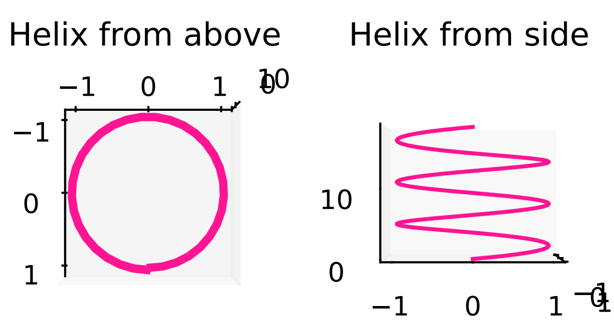
We can also change the viewing angle for a 3D plot store in ax. This is done via

```
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 pre-defined templates descirbed 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
```

```python
# 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()
```

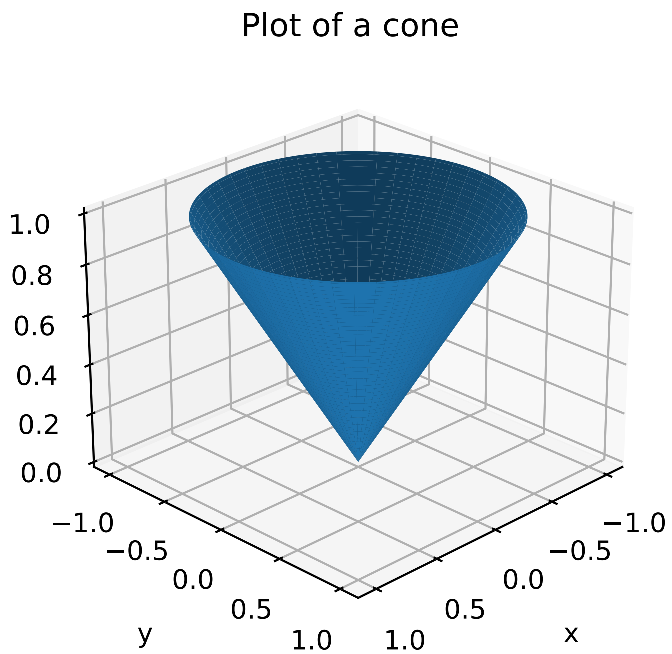
For example let us plot a cone described parametrically by:

$$x = u\cos(v)\,, \quad y = u\sin(v)\,, \quad z = u$$

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:

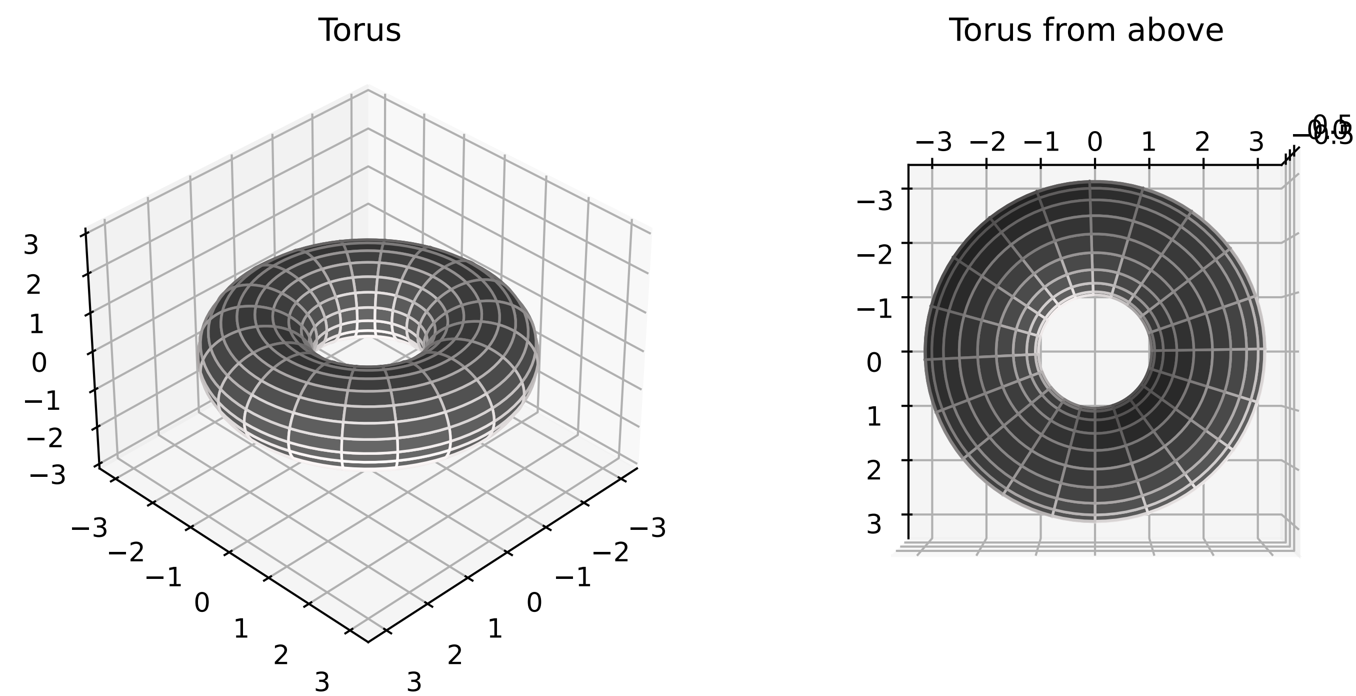
$$
\begin{aligned}
x &= (R + r\cos(u))\cos(v) \\
y &= (R + r\cos(u))\sin(v) \\
z &= r\sin(u)
\end{aligned}
$$

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

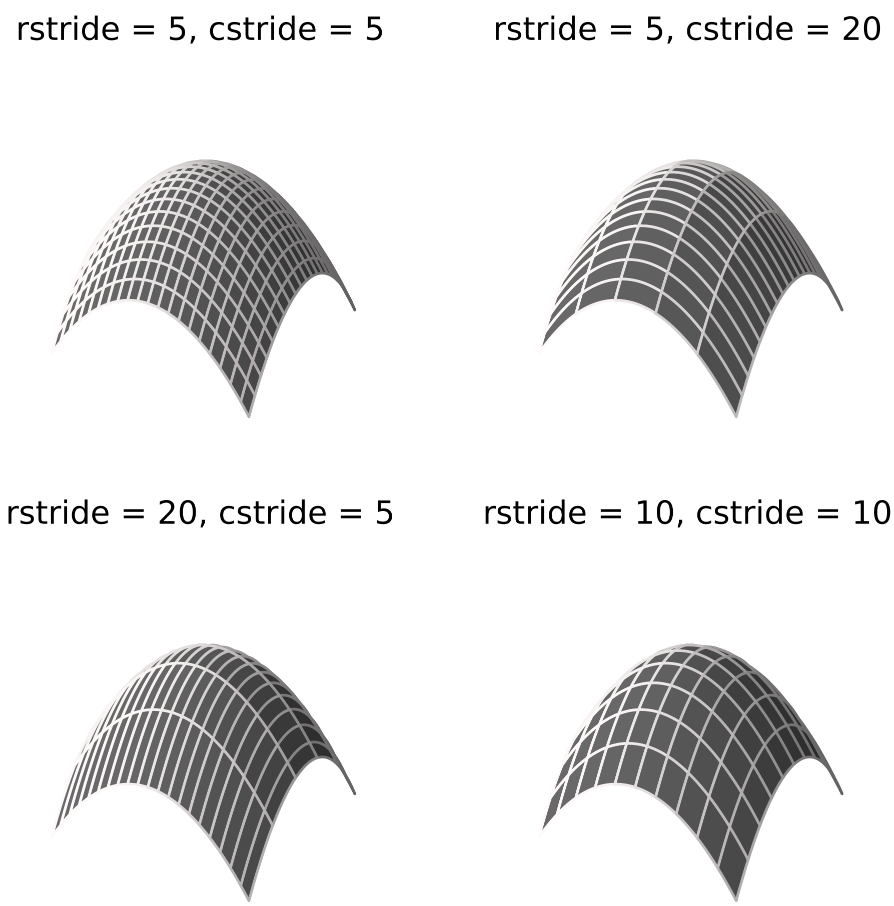
$$
\begin{aligned}
x &= u \\
y &= v \\
z &= -u^2 - v^2
\end{aligned}
$$

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. (2024). *Lecture Notes on Differential Geometry.*
> https://www.silviofanzon.com/2024-Differential-Geometry-Notes/

BibTex citation:

```
@electronic{Fanzon-Diff-Geom-2024,
    author = {Fanzon, Silvio},
    title = {Lecture Notes on Differential Geometry},
    url = {https://www.silviofanzon.com/2024-Differential-Geometry-Notes/},
    year = {2024}}
```

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] W. Walter. *Ordinary differential equations*. Springer-Verlag, 1998.

[8] V. A. Zorich. *Mathematical Analysis I*. Second Edition. Springer, 2015.

[9] V. A. Zorich. *Mathematical Analysis II*. Second Edition. Springer, 2016.