Differential Geometry

Lecture Notes

Dr. Silvio Fanzon

23 Dec 2024

Academic Year 2024/25

Department of Mathematics

University of Hull

Table of contents

Welcome **5**
Revision Guide . . . 5
Digital Notes . . . 5
Readings . . . 5
Visualization . . . 6

1. Curves **7**
1.1. Parametrized curves . . . 11
1.2. Parametrizing Cartesian curves . . . 11
1.3. Smooth curves . . . 17
1.4. Tangent vectors . . . 20
1.5. Length of curves . . . 24
1.6. Arc-length . . . 33
1.7. Scalar product in $\mathbb{R}^n$ . . . 37
1.8. Speed of a curve . . . 41
1.9. Reparametrization . . . 42
1.10. Unit-speed reparametrization . . . 47
1.11. Closed curves . . . 58

2. Curvature and Torsion **67**
2.1. Curvature . . . 67
2.2. Vector product in $\mathbb{R}^3$ . . . 74
2.3. Curvature formula in $\mathbb{R}^3$ . . . 83
2.4. Signed curvature of plane curves . . . 85
2.5. Space curves . . . 87
2.6. Torsion . . . 95
2.7. Frenet-Serret equations . . . 103
2.8. Fundamental Theorem of Space Curves . . . 106
2.9. Applications of Frenet-Serret . . . 109
2.10. Proof: Curvature and torsion formulas . . . 117
2.11. Proof: Fundamental Theorem of Space Curves . . . 120

3. Topology **129**
3.1. Closed sets . . . 133
3.2. Comparing topologies . . . 135
3.3. Convergence . . . 138
3.4. Metric spaces . . . 141
3.5. Interior, closure and boundary . . . 145

3.6. Density 151
3.7. Hausdorff spaces 154
3.8. Continuity 158
3.9. Subspace topology 166
3.10. Topological basis 168
3.11. Product topology 172
3.12. Connectedness 173
3.13. Intermediate Value Theorem 180
3.14. Path-connectedness 183

4. Surfaces **188**
4.1. Preliminaries 189
4.1.1. Linear algebra 190
4.1.2. Topology of $\mathbb{R}^n$ 196
4.1.3. Smooth functions 197
4.1.4. Diffeomorphisms 200
4.2. Surfaces 203
4.3. Regular Surfaces 210
4.4. Reparametrizations 219
4.5. Transition maps 223
4.6. Functions between surfaces 226
4.7. Tangent plane 232
4.8. Unit normal and orientability 242
4.9. Differential of smooth functions 247
4.10. Examples of Surfaces 254
4.10.1. Level surfaces 255
4.10.2. Quadrics 258
4.10.3. Ruled surfaces 263
4.10.4. Surfaces of Revolution 266
4.11. First fundamental form 268
4.11.1. Length of curves 274
4.11.2. Isometries 275
4.11.3. Angles on surfaces 285
4.11.4. Angles between curves 287
4.11.5. Conformal maps 289
4.11.6. Conformal parametrizations 296
4.11.7. Areas 300
4.11.8. Equiareal Maps 304
4.11.9. Equiareal parametrizations 309
4.11.10. Summary 312
4.12. Second fudamental form 314
4.12.1. Gauss and Weingarten maps 317
4.12.2. Matrix of Weingarten map 325
4.13. Curvatures 330
4.13.1. Gaussian and mean curvature 330

4.13.2. Principal curvatures 334
4.13.3. Normal and geodesic curvatures 341
4.14. Local shape of a surface 354
4.15. Conclusion: FTS and Theorema Egregium 366

License **375**
Reuse 375
Citation 375

References **376**

Appendices **377**

A. Plots with Python **377**
A.1. Curves in Python 377
A.1.1. Curves in 2D 377
A.1.2. Implicit curves 2D 384
A.1.3. Curves in 3D 388
A.1.4. Interactive plots 396
A.2. Surfaces in Python 401
A.2.1. Plots with Matplotlib 401
A.2.2. Plots with Plotly 409

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

Revision Guide

A Revision Guide to prepare for the Exam is available at

silviofanzon.com/2024-Differential-Geometry-Revision

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

Other good readings are the books:

- do Carmo [3], 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 [9, 10].

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 [4, 5].

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

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

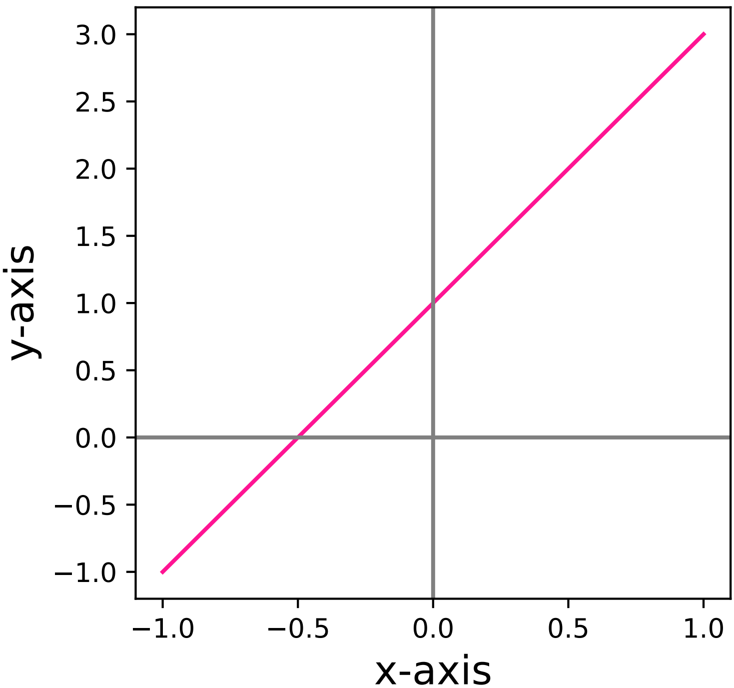


Figure 1.1.: Plotting straight line $y = 2x + 1$

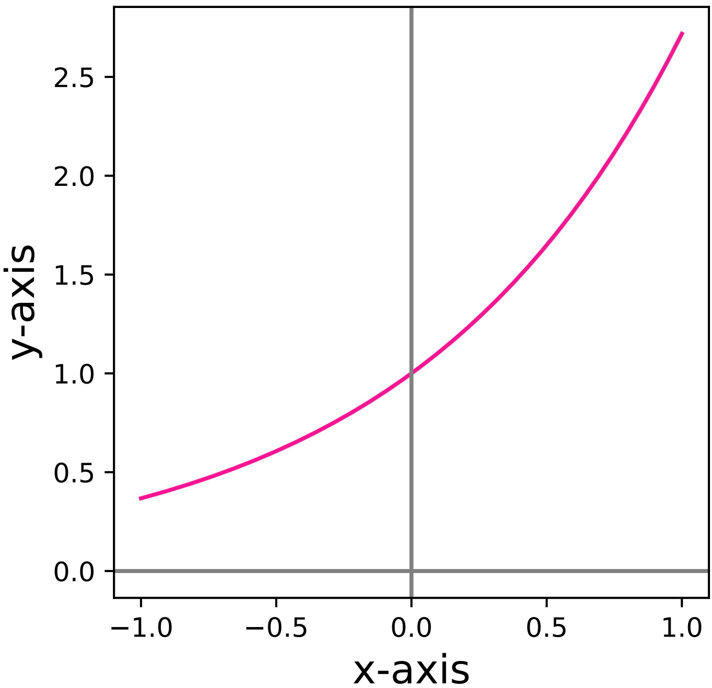


Figure 1.2.: Plot of hyperbole $y = e^x$

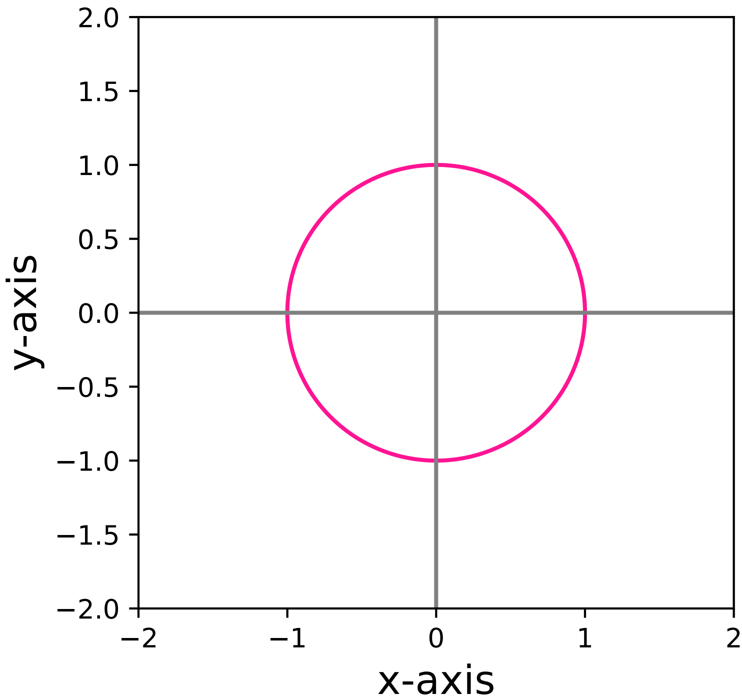


Figure 1.3.: Plot of unit circle of equation $x^2 + y^2 = 1$

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.

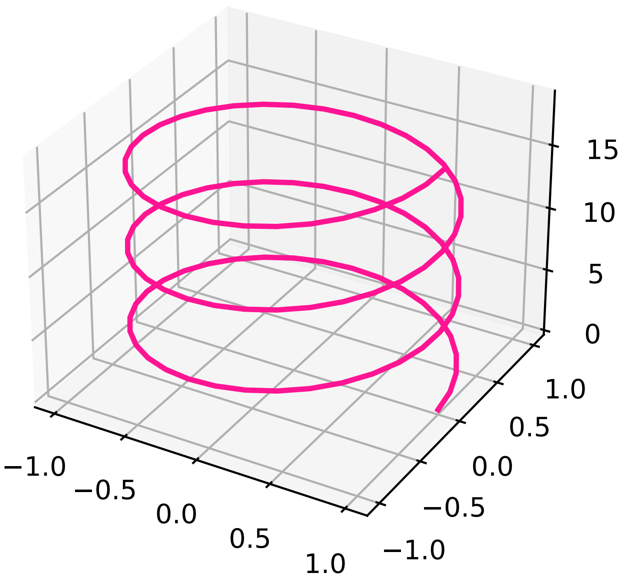


Figure 1.4.: Plot of a 3D Helix

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

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

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.

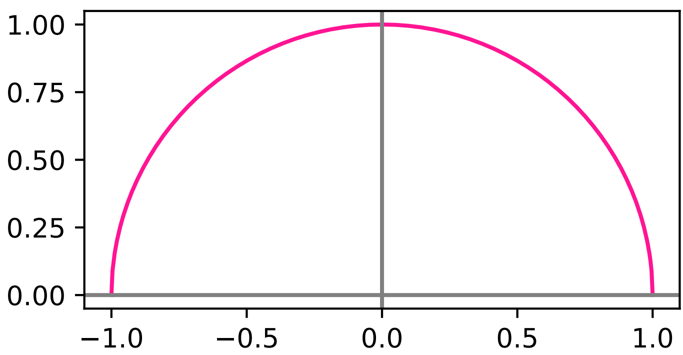


Figure 1.5.: Upper semi-circle

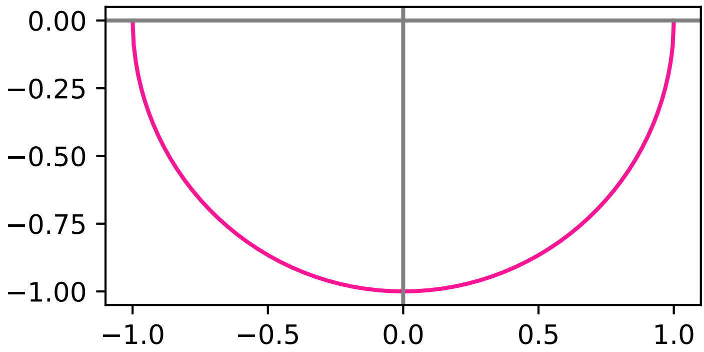


Figure 1.6.: Lower semi-circle

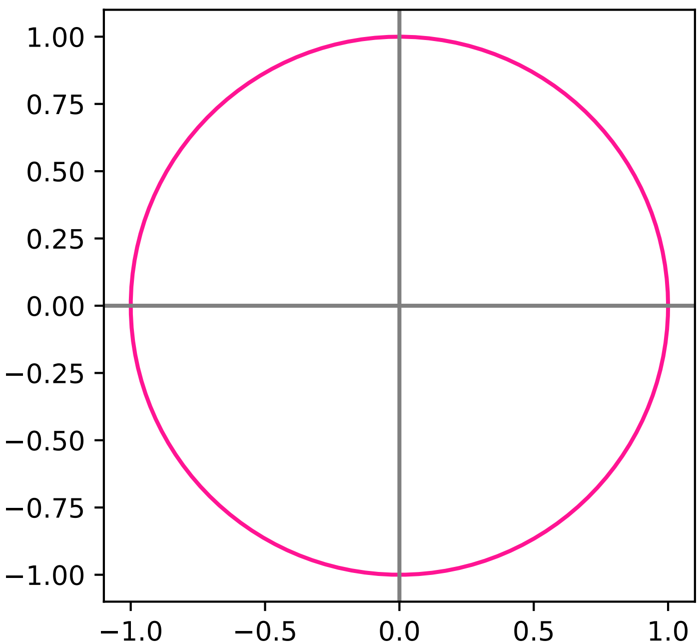


Figure 1.7.: Lower semi-circle

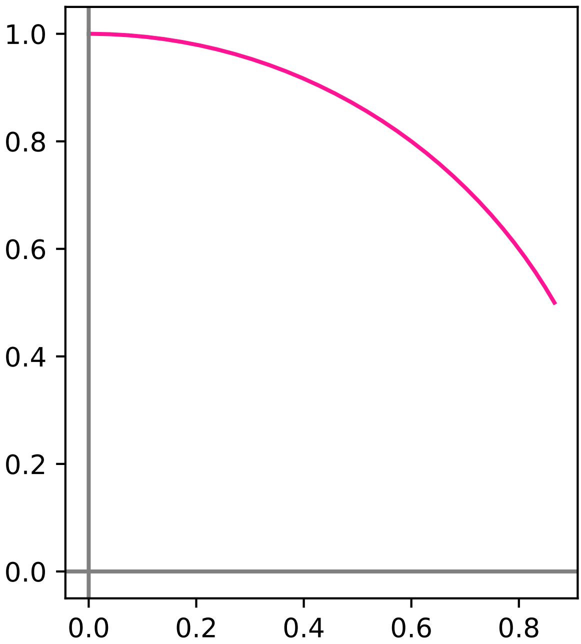


Figure 1.8.: Plotting a portion of C

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

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

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.

We will work under the following assumption.

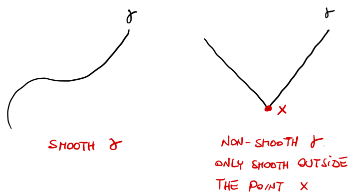


Figure 1.9.: Example of smooth and non-smooth curves

Assumption

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

Example 1.12

The circle

$$\gamma(t) = (\cos(t), \sin(t))$$

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

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

For example the derivative of γ at the point $(0, 1)$ is given by

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

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

1.4. Tangent vectors

Looking at Figure 1.10, it seems like the vector

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

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

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

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

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

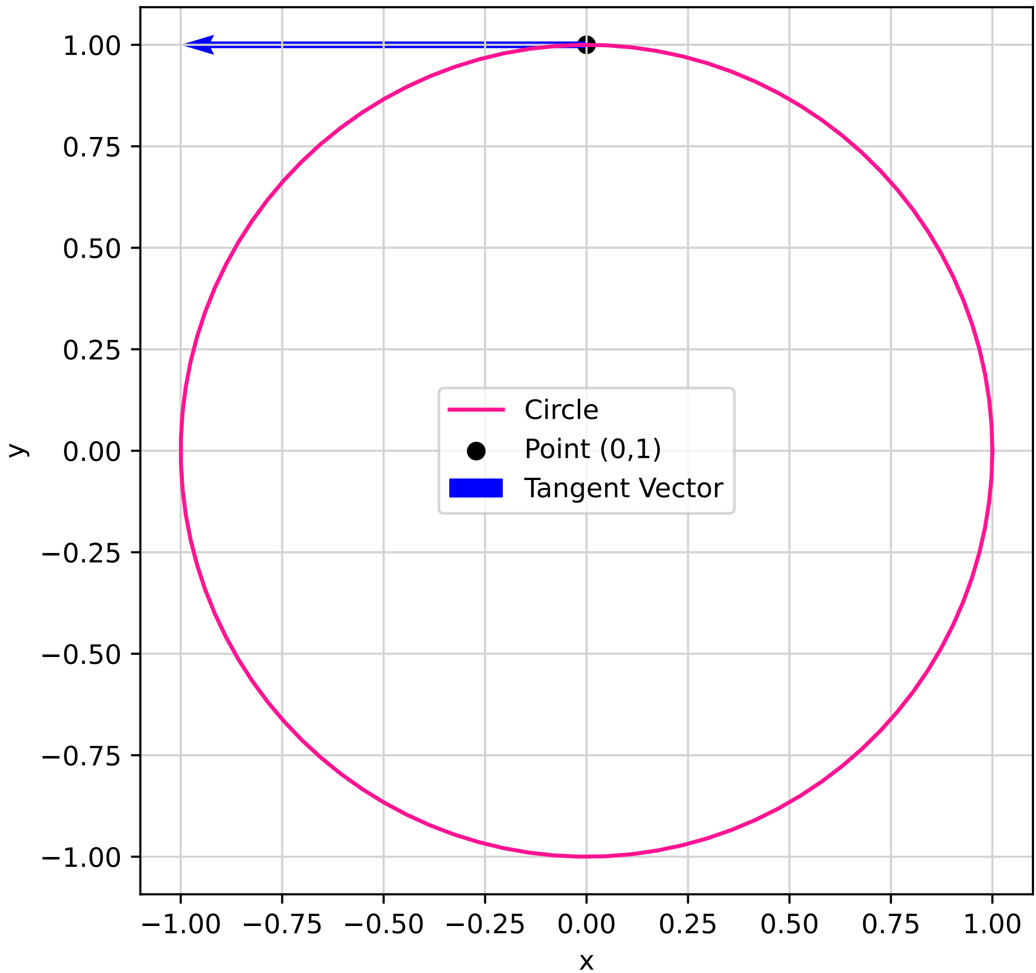


Figure 1.10.: Plot of Circle and Tangent Vector at $(0, 1)$

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.

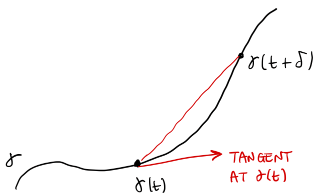


Figure 1.11.: Approximating the tangent vector

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.

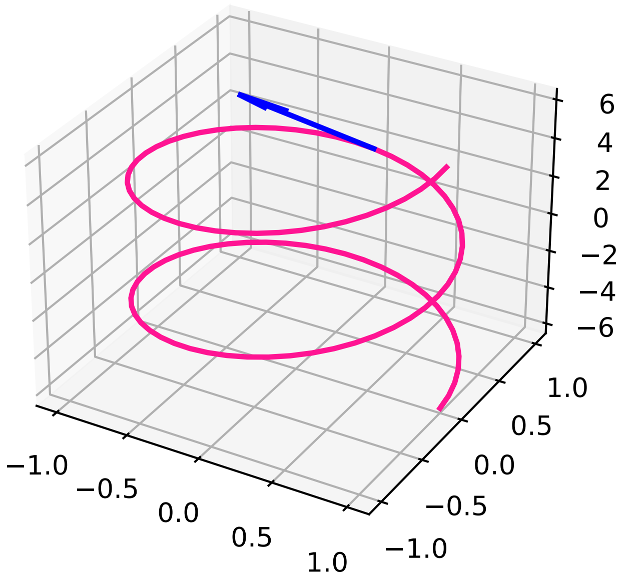


Figure 1.12.: Plot of Helix with tangent vector

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

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

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.

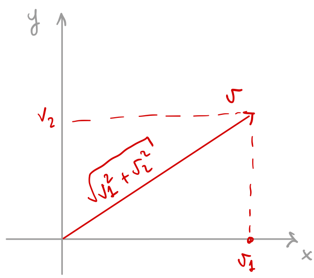


Figure 1.14.: Interpretation of $\|v\|$ in $\mathbb{R}^2$

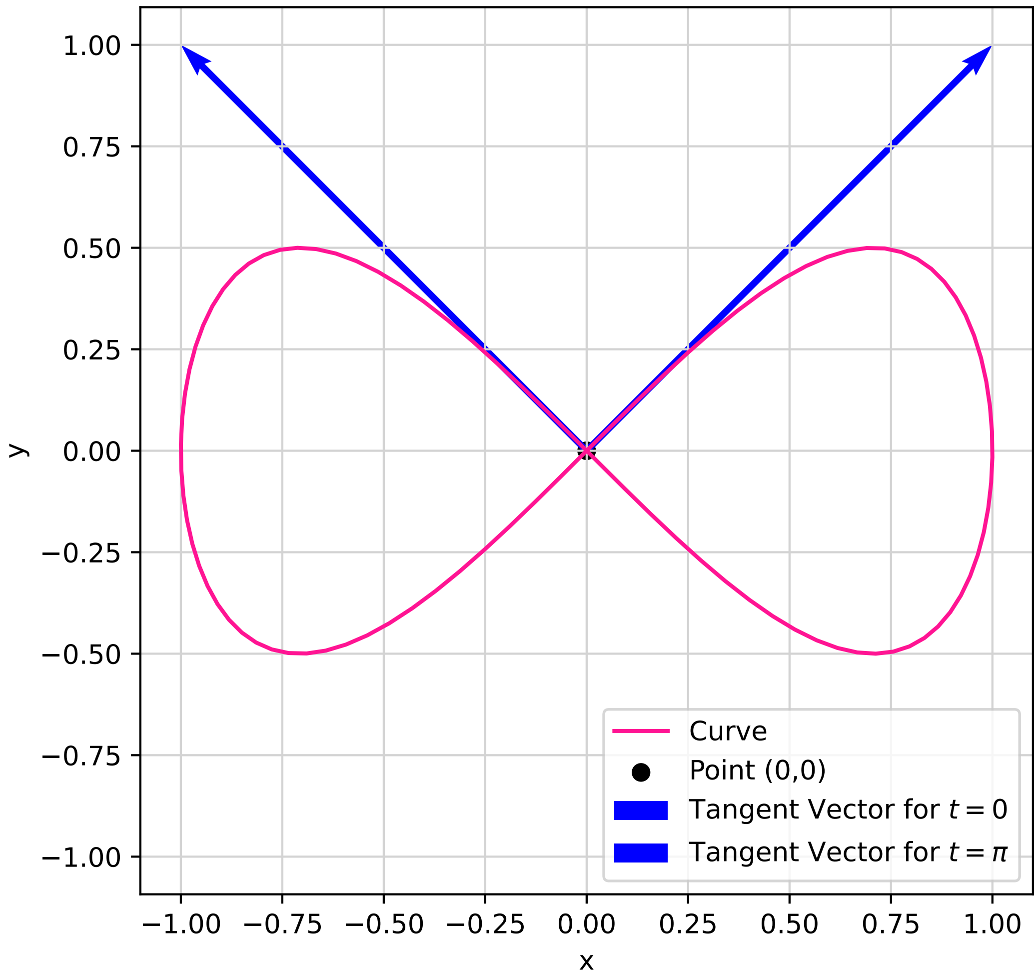


Figure 1.13.: The Lemniscate curve

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.

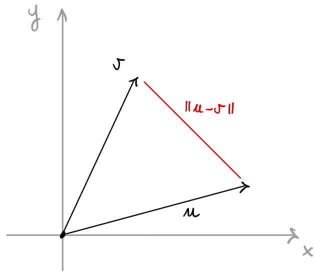


Figure 1.15.: Interpretation of $\|\mathbf{u} - \mathbf{v}\|$ in $\mathbb{R}^2$

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

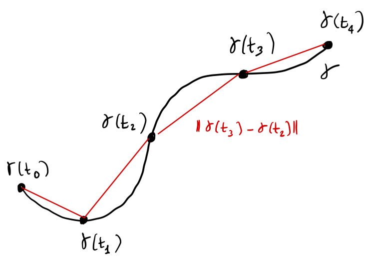


Figure 1.16.: Approximating the length of $\boldsymbol{\gamma}$

In formulae, this means choosing some time instants

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

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

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

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.

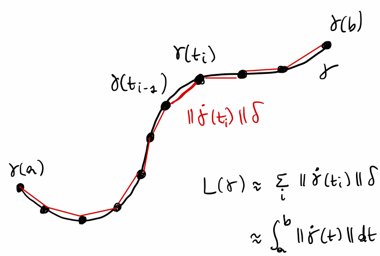


Figure 1.17.: Approximating $L(\boldsymbol{\gamma})$ via $\dot{\boldsymbol{\gamma}}$

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

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

for some constant $C \ge 0$. Therefore

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

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

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

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

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

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

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

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

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

By standard properties of integral we also have

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

so that (1.15) implies

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

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

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

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

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

We can therefore further estimate (1.16) and obtain

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

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

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

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

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

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

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

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

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

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 the Circle

Question. Compute the length of the circle of radius R

$$\boldsymbol{\gamma}(t) = (x_0 + R\cos(t), y_0 + R\sin(t), 0)\,.$$

Solution. We compute

$$\begin{aligned}
\dot{\boldsymbol{\gamma}}(t) &= (-R\sin(t), R\cos(t), 0) \\
\|\dot{\boldsymbol{\gamma}}(t)\| &= \sqrt{\dot{\gamma}_1^2(t) + \dot{\gamma}_2^2(t)} = R \\
L(\boldsymbol{\gamma}) &= \int_0^{2\pi} \|\dot{\boldsymbol{\gamma}}(t)\| \, dt = \int_0^{2\pi} R \, dt = 2\pi R\,.
\end{aligned}$$

Example 1.23: Length of the Helix

Question. Compute the length of the Helix

$$\gamma(t) = (R\cos(t), R\sin(t), Ht) , \quad t \in (0, 2\pi) .$$

Solution. We compute

$$\dot{\gamma}(t) = (-R\sin(t), R\cos(t), H) \qquad \|\dot{\gamma}(t)\| = \sqrt{R^2 + H^2}$$
$$L(\gamma) = \int_0^{2\pi} \|\dot{\gamma}(u)\| \, du = 2\pi\sqrt{R^2 + H^2}$$

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 smooth 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 of a curve

The **arc-length** along $\gamma : (a, b) \to \mathbb{R}^3$ from t_0 to t is

$$s : (a, b) \to \mathbb{R} , \qquad s(t) = \int_{t_0}^t \|\dot{\gamma}(u)\| du .$$

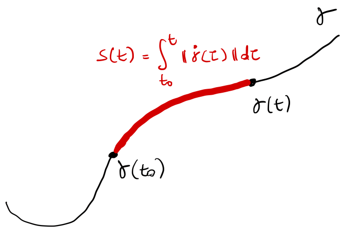


Figure 1.18.: Arc-length of $\boldsymbol{\gamma}$ starting at $\boldsymbol{\gamma}(t_0)$

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: Arc-length of Circle

Question. Consider the circle of radius R, parametrized by

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t), 0) \,.$$

Compute the arc-length function of $\boldsymbol{\gamma}$.
Solution. We have

$$\dot{\boldsymbol{\gamma}}(t) = (-R\sin(t), R\cos(t), 0) \,, \quad \|\dot{\boldsymbol{\gamma}}(t)\| = R \,.$$

Therefore, for any fixed t_0, 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: Arc-length of Logarithmic Spiral

Question. Compute the arc-length of

$$\boldsymbol{\gamma}(t) = (e^{kt}\cos(t), e^{kt}\sin(t), 0) \,.$$

Solution. The arc-length starting from t_0 is

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) &= e^{kt}(k\cos(t) - \sin(t), k\sin(t) + \cos(t), 0) \\ \|\dot{\boldsymbol{\gamma}}(t)\|^2 &= (k^2+1)e^{2kt} \\ s(t) &= \int_{t_0}^{t} \|\dot{\boldsymbol{\gamma}}(\tau)\| \, d\tau = \frac{\sqrt{k^2+1}}{k}(e^{kt} - e^{kt_0})\,. \end{aligned}$$

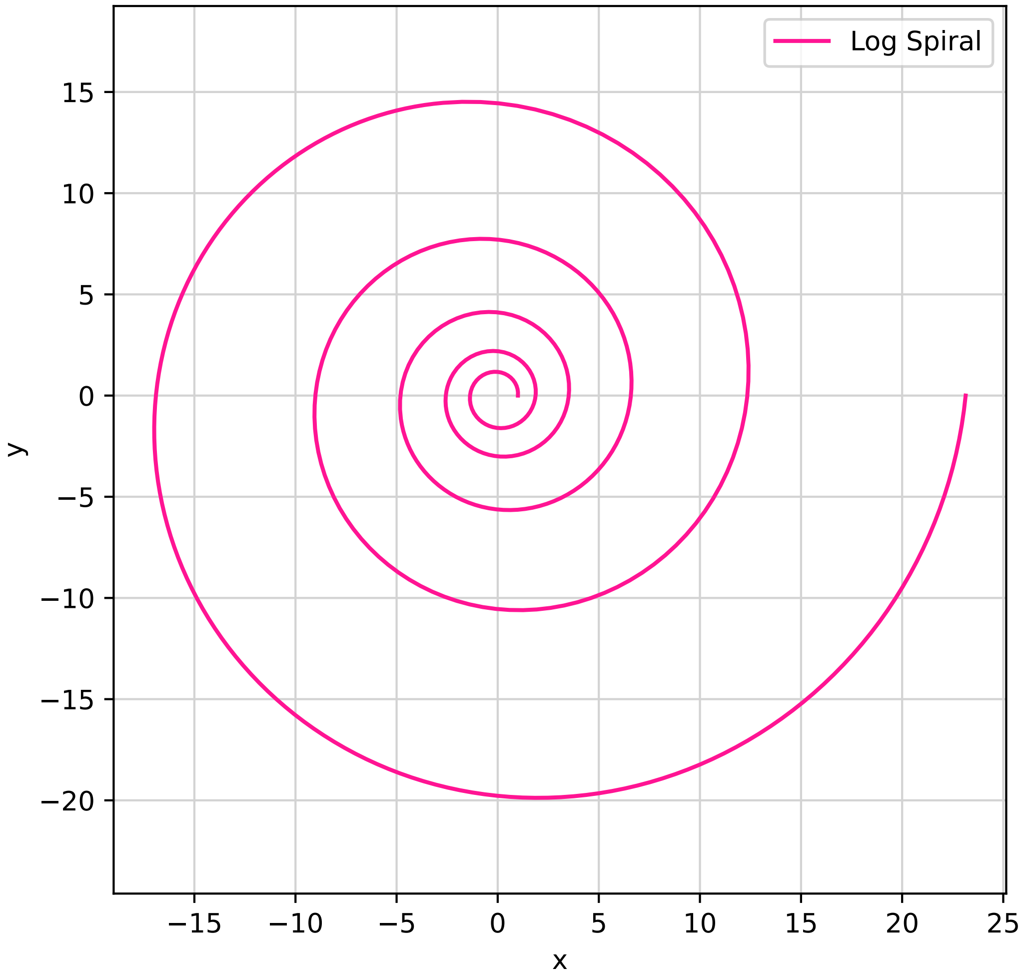


Figure 1.19.: Plot of Logarithmic Spiral with $k = 0.1$

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

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

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

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

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

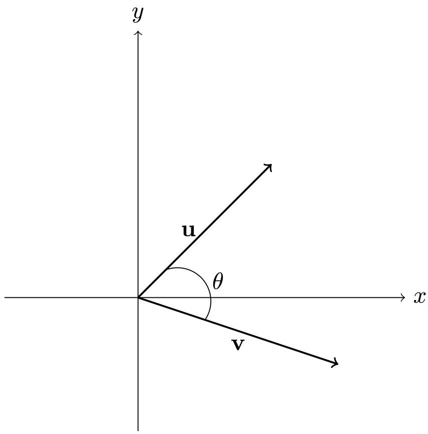


Figure 1.20.: Vectors u and v in $\mathbb{R}^2$ forming angle θ

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 **u** and **v** are parallel.

Definition 1.30: Orthogonal vectors

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

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

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

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

Remark 1.38

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

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

Definition 1.39: Unit-speed curve

A curve $\boldsymbol{\gamma} \,:\, (a,b) \to \mathbb{R}^3$ is **unit-speed** if

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

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

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

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

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

and therefore

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

We can differentiate the LHS of (1.18) to get

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

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

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

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.

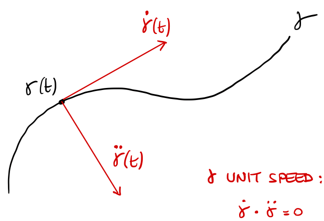


Figure 1.21.: If $\boldsymbol{\gamma}$ is unit-speed then $\dot{\boldsymbol{\gamma}}$ and $\ddot{\boldsymbol{\gamma}}$ are orthogonal

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 $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$. A **reparametrization** of $\boldsymbol{\gamma}$ is a curve $\tilde{\boldsymbol{\gamma}} : (\tilde{a}, \tilde{b}) \to \mathbb{R}^3$ such that

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

for $\phi : (\tilde{a}, \tilde{b}) \to (a, b)$ diffeomorphism. 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 (**?@eq-reparametrization**). This concept is sketched in Figure 1.22.

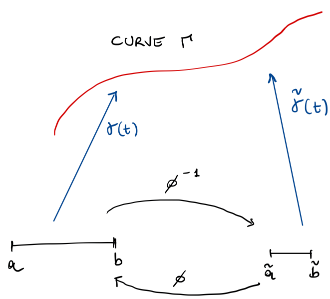


Figure 1.22.: Sketch of 1D shape Γ parametrized by $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$. We can switch parametrization by means of the diffeomorphism ϕ.

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

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

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,

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

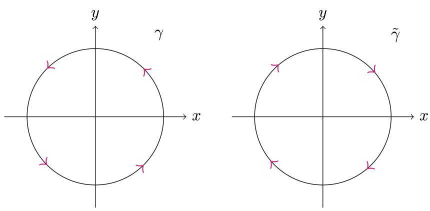


Figure 1.23.: Unit circle with usual parametrization $\boldsymbol{\gamma}$, and with reversed orientation $\tilde{\boldsymbol{\gamma}}$

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

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

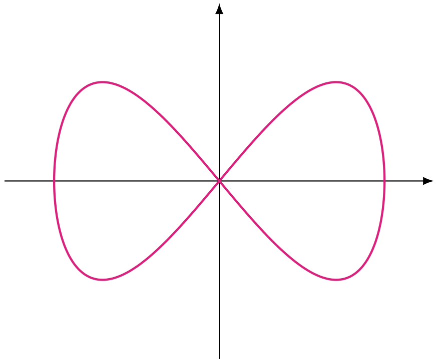


Figure 1.24.: Lemniscate curve

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 curve

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. $\boldsymbol{\gamma}$ 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

$$\frac{d}{dt} \left(\psi(\phi(t))\right) = 1 . \tag{1.21}$$

We can differentiate the LHS by chain rule

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

From (1.21) 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

$$\dot\phi(t) = 3t^2 \qquad \Longrightarrow \qquad \dot\phi(0) = 0 \,,$$

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 $t > 0$.

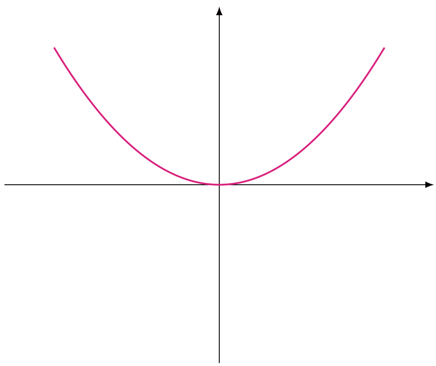


Figure 1.25.: Parabola Γ

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

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) \neq 0 \quad \iff \quad \dot{\boldsymbol{\gamma}}(\phi(t)) \neq 0 \,. \tag{1.22}$$

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

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

From (1.22) 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 $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^n$. 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, that is,

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

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 Theorem 1.57 below.

Theorem 1.56: Existence of unit-speed reparametrization

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

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

$$\dot{s}(t) = \|\dot{\boldsymbol{\gamma}}(t)\| \tag{1.23}$$

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

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

or in other words

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

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 (1.23). Taking the absolute value of the above yields

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \left\|\dot{\tilde{\boldsymbol{\gamma}}}(s(t))\right\| \|\dot{\boldsymbol{\gamma}}(t)\|\,. \tag{1.24}$$

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

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

Therefore we can divide (1.24) 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

$$\|\dot{\boldsymbol{\gamma}}(\phi(t))\|\;|\dot{\phi}(t)| = 1\,, \quad \forall\, t \in (\tilde{a}, \tilde{b})\,. \tag{1.25}$$

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.

Theorem 1.57: Characterization of unit-speed reparametrizations

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 a reparametrization of $\boldsymbol{\gamma}$, that is,

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

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

1. If $\tilde{\boldsymbol{\gamma}}$ is unit-speed, there exists $c \in \mathbb{R}$ such that

$$\phi(t) = \pm s(t) + c\,, \quad \forall\, t \in (a, b)\,. \tag{1.26}$$

2. If ϕ is given by (1.26), 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

$$\dot{s}(t) = |\dot{\phi}(t)| \quad \forall\, t \in (a,b) \tag{1.27}$$

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 (1.27) 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 $\tilde{\gamma}$ be reparametrization of γ

$$\gamma(t) = \tilde{\gamma}(\phi(t)) \,, \quad \forall \, t \in (a,b) \tag{1.28}$$

with

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

for some $c \in \mathbb{R}$. Differentiating (1.28) we get

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

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

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

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

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

As ϕ is invertible, the above is equivalent to

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

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

Definition 1.58: Arc-length reparametrization

Let γ be regular. The **arc-length reparametrization** of γ is

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

with s^{-1} inverse of the arc-length function of γ.

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 $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^n$ be a regular curve and $s : (a, b) \to (\tilde{a}, \tilde{b})$ 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 Circle

Question. The circle of radius $R > 0$ is

$$\boldsymbol{\gamma}(t) = (x_0 + R\cos(t), y_0 + \sin(t), 0).$$

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

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

The inverse is $t(s) = s/R$. The arc-length reparametrization of $\boldsymbol{\gamma}$ is

$$\tilde{\boldsymbol{\gamma}}(s) = \boldsymbol{\gamma}(t(s)) = \left(x_0 + R\cos\left(\frac{s}{R}\right), y_0 + \sin\left(\frac{s}{R}\right), 0\right).$$

Example 1.60: Reparametrization by arc-length

Question. Consider the curve

$$\boldsymbol{\gamma}(t) = (5\cos(t), 5\sin(t), 12t)\,.$$

Prove that $\boldsymbol{\gamma}$ is regular, and reparametrize it by arc-length.
Solution. $\boldsymbol{\gamma}$ is regular because

$$\dot{\boldsymbol{\gamma}}(t) = (-5\sin(t), 5\cos(t), 12)\,, \qquad \|\dot{\boldsymbol{\gamma}}(t)\| = 13 \neq 0$$

The arc-length of $\boldsymbol{\gamma}$ starting from $t_0 = 0$, and its inverse, are

$$s(t) = \int_0^t \|\dot{\boldsymbol{\gamma}}(u)\|\, du = 13t\,, \qquad t(s) = \frac{s}{13}\,.$$

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

$$\tilde{\boldsymbol{\gamma}}(s) = \boldsymbol{\gamma}(t(s)) = \left(5\cos\left(\frac{s}{13}\right), 5\sin\left(\frac{s}{13}\right), \frac{12}{13}s\right)\,.$$

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 1.61: Twisted cubic

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

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

Therefore

$$\dot{\boldsymbol{\gamma}}(t) = (1, 2t, 3t^2) \neq \mathbf{0}\,,$$

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

$$s(t) = \int_{t_0}^t \sqrt{1 + 4\tau^2 + 9\tau^4}\, d\tau\,. \tag{1.29}$$

Since $\boldsymbol{\gamma}$ is regular, by Theorem 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 (1.29) 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.

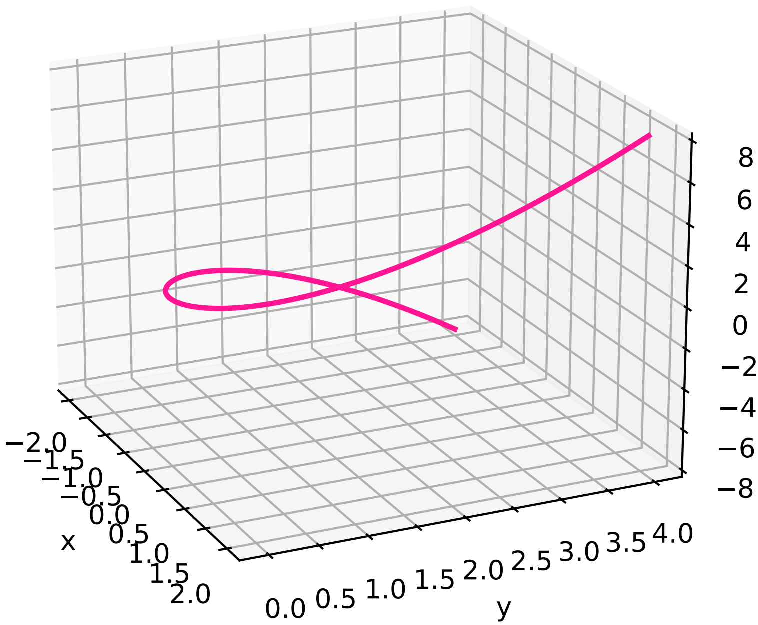


Figure 1.26.: Plot of Twisted Cubic for t between -2 and 2

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 1.62

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

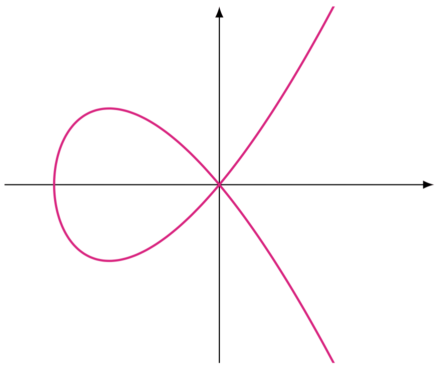


Figure 1.27.: Plot of curve $\boldsymbol{\gamma}(t) = (t^2 - 1, t^3 - 1)$ for $t \in [-2, 2]$

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

Definition 1.63: 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 1.64: 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 1.65

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
   $$\boldsymbol{\gamma}(a) = \boldsymbol{\gamma}(b), \quad \frac{d^k \boldsymbol{\gamma}}{dt^k}(a) = \frac{d^k \boldsymbol{\gamma}}{dt^k}(b) \tag{1.30}$$
   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 (1.30) 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 (1.30) 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.

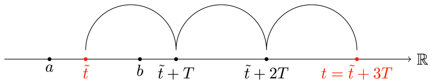


Figure 1.28.: The points $t \in \mathbb{R}$ and $\tilde{t} \in [a, b]$ from Point 3 in Remark 1.65. In this skecth $t = \tilde{t} + 3T$, with $T = b - a$.

Definition 1.66

Let γ be a closed curve. The **period** of γ is the smallest $T > 0$ such that γ is T-periodic, that is

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

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

Proposition 1.67

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

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

admits positive minumum

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

Proof

We make 2 observations about the set S:

- Since γ is closed, we have that γ is T-periodic for some $T \neq 0$. By Remark 1.65 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. γ 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

$$T_1 > T_2 > \dots > T_n > \dots > 0 .$$

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

$$P \in S \quad \iff \quad P > 0 .$$

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 1.65 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

$$\boldsymbol{\gamma}(t) = \boldsymbol{\gamma}(t_n) , \quad \forall\, n \in \mathbb{N} . \tag{1.31}$$

Also notice that

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

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

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

Using the continuity of $\boldsymbol{\gamma}$, we can pass to the limit in (1.31) 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.

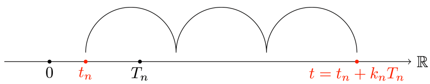


Figure 1.29.: For each $t \in \mathbb{R}$ there exist unique $k_n \in \mathbb{Z}$ and $\tilde{t}_n \in [0, T_n]$ such that $t = \tilde{t} + k_n T_n$. In this skecth $k_n = 3$.

Example 1.68: 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 1.62
$$\boldsymbol{\gamma}(t) := (t^2 - 1, t^3 - t) \,, \quad t \in \mathbb{R} \,.$$
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

$$p = \gamma(-1) = \gamma(1) .$$

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

Definition 1.69: Self-intersecting curve

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

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

Remark 1.70

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 $\gamma : \mathbb{R} \to \mathbb{R}^n$ and let T be its period. Then by Point 6 in Remark 1.65 we have

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

Therefore every point $\gamma(a)$ 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 1.71

Let us go back to the curve of Example 1.62, that is,

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

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

$$p = \gamma(-1) = \gamma(1) .$$

Example 1.72: 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 $(0, 0)$.

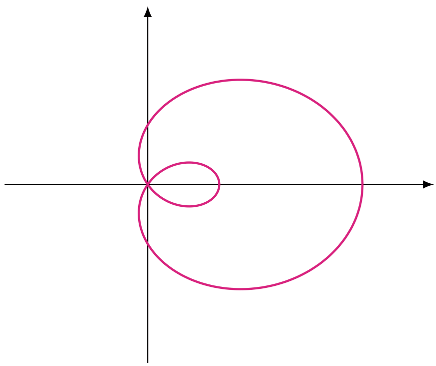


Figure 1.30.: Limaçon curve

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 $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^2$ 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

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

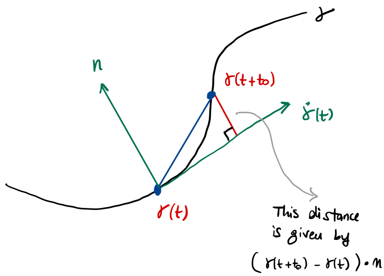


Figure 2.1.: Amount that $\boldsymbol{\gamma}$ deviates from tangent is $[\boldsymbol{\gamma}(t+t_0) - \boldsymbol{\gamma}(t)] \cdot \mathbf{n}(t)$

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

The **curvature** of a unit-speed curve $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ 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

Question. Compute the curvature of the circle of radius $R > 0$

$$\boldsymbol{\gamma}(t) = \left(x_0 + R\cos\left(\frac{t}{R}\right), y_0 + \sin\left(\frac{t}{R}\right), 0 \right) \,.$$

Solution. First, check that $\boldsymbol{\gamma}$ is unit-speed:

$$\dot{\boldsymbol{\gamma}}(t) = \left(-\sin\left(\frac{t}{R}\right), \cos\left(\frac{t}{R}\right), 0 \right) \,, \qquad \|\dot{\boldsymbol{\gamma}}(t)\| = 1$$

Now, compute second derivative and curvature

$$\begin{aligned} \ddot{\boldsymbol{\gamma}}(t) &= \left(-\frac{1}{R}\cos\left(\frac{t}{R}\right), -\frac{1}{R}\sin\left(\frac{t}{R}\right), 0 \right) \,, \\ \kappa(t) &= \|\ddot{\boldsymbol{\gamma}}(t)\| = \frac{1}{R} \,. \end{aligned}$$

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 $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi$ and $\phi \colon (a,b) \to (\tilde{a},\tilde{b})$. Let $\tilde{\kappa} \colon (\tilde{a},\tilde{b}) \to \mathbb{R}$ be the curvature of $\tilde{\boldsymbol{\gamma}}$. The **curvature** of $\boldsymbol{\gamma}$ is

$$\kappa(t) = \tilde{\kappa}(\phi(t)) \,.$$

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: Computing curvature of regular $\boldsymbol{\gamma}$

1. Compute the arc-length $s(t)$ of $\boldsymbol{\gamma}$ and its inverse $t(s)$.
2. Compute the arc-length reparametrization
$$\tilde{\boldsymbol{\gamma}}(s) = \boldsymbol{\gamma}(t(s)) \, .$$
3. Compute the curvature of $\tilde{\boldsymbol{\gamma}}$
$$\tilde{\kappa}(s) = \left\| \ddot{\tilde{\boldsymbol{\gamma}}}(s) \right\| \, .$$
4. The curvature of $\boldsymbol{\gamma}$ 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**.

Definition 2.6: Hyperbolic functions

The **hyperbolic functions** are defined by:

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

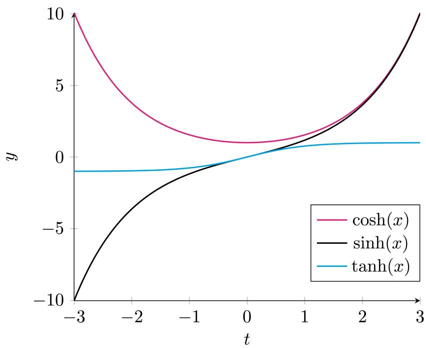


Figure 2.2.: Plot of cosh, sinh, tanh.

Theorem 2.7: Properties of Hyperbolic Functions

Identities:

$$\cosh^2(t) - \sinh^2(t) = 1 \qquad \operatorname{sech}^2(t) + \tanh^2(t) = 1$$

Derivatives:

$$\sinh(t)' = \cosh(t) \qquad \cosh(t)' = \sinh(t)$$
$$\tanh(t)' = \operatorname{sech}^2(t) \qquad \operatorname{sech}(t)' = -\operatorname{sech}(t)\tanh(t)$$

Integrals:

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

Definition 2.8

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 $\boldsymbol{\gamma}$.

Example 2.9: Curvature of the Catenary

Question. Consider the Catenary curve

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

1. Prove that $\boldsymbol{\gamma}$ is regular.
2. Compute the arc-length reparametrization of $\boldsymbol{\gamma}$.
3. Compute the curvature of $\tilde{\boldsymbol{\gamma}}$.
4. Compute the curvature of $\boldsymbol{\gamma}$.

Solution.

1. $\boldsymbol{\gamma}$ is regular because

$$\dot{\boldsymbol{\gamma}}(t) = (1, \sinh(t))$$
$$\|\dot{\boldsymbol{\gamma}}\| = \sqrt{1 + \sinh^2(t)} = \cosh(t) \geq 1$$

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

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

where we used that $\sinh(0) = 0$. Moreover,

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

Substitute $y = e^t$ to obtain

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

Notice that

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

by definition of absolute value. Therefore,

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

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

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

3. Compute the curvature of $\tilde{\boldsymbol{\gamma}}$

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

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

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

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

The discussion in this section follows [3]. We start by defining **orientation** for a vector space.

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

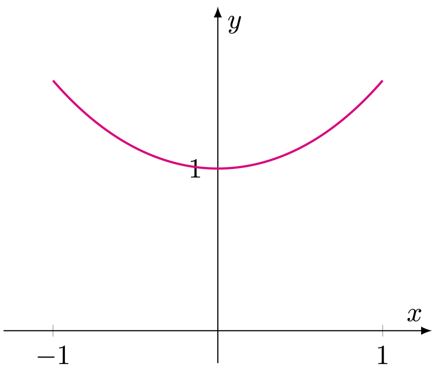


Figure 2.4.: Plot of the catenary curve $\boldsymbol{\gamma}(t) = (t, \cosh(t))$.

Definition 2.10: 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

$$\det P > 0$$

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 2.11: Orientation

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

Definition 2.12: Positive orientation

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

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

where we set

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

Then:

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

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

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

Example 2.13

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 2.14: 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 2.15

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 2.16

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 2.17: 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 2.16 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 2.16 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 2.18: 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 $\mathbf{u}, \mathbf{v}$

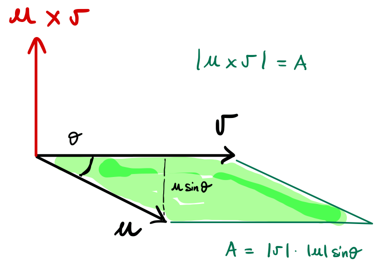


Figure 2.5.: For $\mathbf{u}, \mathbf{v}$ linearly independent, $\mathbf{u}\times\mathbf{v}$ is orthogonal to the plane generated by $\mathbf{u}, \mathbf{v}$. Moreover $\|\mathbf{u}\times\mathbf{v}\|$ is the area of the parallelogram with sides $\mathbf{u}, \mathbf{v}$, and $(\mathbf{u}, \mathbf{v},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

- $\|\mathbf{u}\times\mathbf{v}\|$ is the area of the parallelogram with sides $\mathbf{u}, \mathbf{v}$
- The triple $(\mathbf{u}, \mathbf{v}, \mathbf{u}\times\mathbf{v})$ 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 $\mathbb{R}^3$.

Theorem 2.19

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.

Theorem 2.20

Let $\boldsymbol{\gamma}, \boldsymbol{\eta} \,:\, (a, b) \to \mathbb{R}^3$. Then, the curve $\boldsymbol{\gamma}\times\boldsymbol{\eta}$ 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 $\tilde{\boldsymbol{\gamma}}$.

Theorem 2.21: Curvature formula

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be regular. The curvature of $\boldsymbol{\gamma}$ is

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

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 [7].

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

Example 2.22: Curvature of the Straight Line

Question. Consider the straight line

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

for some $\mathbf{a}, \mathbf{v} \in \mathbb{R}^3$ fixed, with $\mathbf{v} \neq \mathbf{0}$.

1. Prove that $\boldsymbol{\gamma}$ is regular.
2. Compute the curvature of $\boldsymbol{\gamma}$.

Solution.

1. $\boldsymbol{\gamma}$ is regular because

$$\dot{\boldsymbol{\gamma}}(t) = \mathbf{v} \neq \mathbf{0} .$$

2. We have $\|\dot{\boldsymbol{\gamma}}\| = \|\mathbf{v}\|$ and $\ddot{\boldsymbol{\gamma}} = \mathbf{v}$. Thus,

$$\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} = \mathbf{v} \times \mathbf{0} = \mathbf{0}\,,$$

and the curvature is

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

Example 2.23: Curvature of the Helix

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

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

1. Prove that $\boldsymbol{\gamma}$ is regular.
2. Compute the curvature of $\boldsymbol{\gamma}$.

Solution.

1. $\boldsymbol{\gamma}$ is regular because

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) &= (-R\sin(t), R\cos(t), H) \\ \|\dot{\boldsymbol{\gamma}}(t)\| &= \sqrt{R^2 + H^2} \geq R > 0 \end{aligned}$$

2. Compute the curvature using the formula:

$$\begin{aligned} \ddot{\boldsymbol{\gamma}}(t) &= (-R\cos(t), -R\sin(t), 0) \\ \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} &= \left(RH\sin(t), -RH\cos(t), R^2\right) \\ \|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| &= R\sqrt{R^2 + H^2} \\ \kappa(t) &= \frac{\|\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t)\|}{\|\dot{\boldsymbol{\gamma}}(t)\|^3} = \frac{R}{R^2 + H^2} \end{aligned}$$

Remark 2.24

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.

Example 2.25: Calculation of curvature

Question. Define the curve

$$\boldsymbol{\gamma}(t) = \left(\frac{8}{5}\cos(t), 1 - 2\sin(t), \frac{6}{5}\cos(t) \right) .$$

1. Prove that $\boldsymbol{\gamma}$ is regular.
2. Compute the curvature of $\boldsymbol{\gamma}$.

Solution.

1. $\boldsymbol{\gamma}$ is regular because

$$\dot{\boldsymbol{\gamma}} = \left(-\frac{8}{5}\sin(t), -2\cos(t), -\frac{6}{5}\sin(t) \right) , \qquad \|\dot{\boldsymbol{\gamma}}\| = 2 \neq 0 \,.$$

2. Compute the curvature using the formula:

$$\ddot{\boldsymbol{\gamma}} = \left(-\frac{8}{5}\cos(t), 2\sin(t), -\frac{6}{5}\cos(t) \right) \qquad \|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| = 4$$
$$\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} = \left(-\frac{12}{5}, 0, \frac{16}{5} \right) \qquad \kappa = \frac{1}{2} \,.$$

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 2.26

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 2.27

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 2.28

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 2.29

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 [7].

For curves which are not unit-speed, we define the signed curvature as the signed curvature of the unit-speed reparametrization.

Definition 2.30

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 2.31: 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 [7].

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 2.32: Different curves, same curvature

Question Let $\boldsymbol{\gamma}$ be a circle

$$\boldsymbol{\gamma}(t) = (2\cos(t), 2\sin(t), 0)\,,$$

and $\boldsymbol{\eta}$ be a helix of radius $S > 0$ and rise $H > 0$

$$\boldsymbol{\eta}(t) = (S\cos(t), S\sin(t), Ht)\,.$$

Find S and H such that $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ have the same curvature.
Solution. Curvatures of $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ were already computed:

$$\kappa^{\boldsymbol{\gamma}} = \frac{1}{2}\,, \quad \kappa^{\boldsymbol{\eta}} = \frac{S}{S^2 + H^2}\,.$$

Imposing 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 $\kappa^{\boldsymbol{\gamma}} = \kappa^{\boldsymbol{\eta}}$.

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 2.33: Principal normal vector

Let $\boldsymbol{\gamma} \colon (a, b) \to \mathbb{R}^3$ be unit-speed, with $\kappa \neq 0$. The **principal normal vector** to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is

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

Remark 2.34

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

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

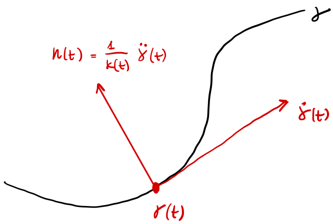


Figure 2.6.: Principal normal vector $\mathbf{n}(t)$ to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$.

Question 2.35

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

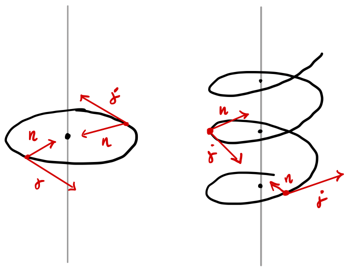


Figure 2.7.: Left: Principal normal to a circle. Note that $\mathbf{n}$ always points towards the origin $\mathbf{0}$. Right: Principal normal to a helix. Note that $\mathbf{n}$ points towards the z-axis, but never towards the same point.

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 2.36: Binormal vector

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be unit-speed, with $\kappa \neq 0$. The **binormal vector** to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is

$$\mathbf{b}(t) = \dot{\boldsymbol{\gamma}}(t) \times \mathbf{n}(t) .$$

To each unit-speed curve $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ with non-vanishing curvature, we can associate a triple of vectors, known as the Frenet frame.

Definition 2.37: Frenet frame

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be unit-speed, with $\kappa \neq 0$. The **Frenet frame** of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is the triple

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

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

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

The Frenet frame is a positive orthonormal basis of $\mathbb{R}^3$, in the following sense.

Definition 2.38: 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

$$\|v_i\| = 1\,, \quad v_i \cdot v_j = 0\,, \quad \text{for } i \neq j\,.$$

Theorem 2.39: Frenet frame is orthonormal basis

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

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

is a positive orthonomal basis of $\mathbb{R}^3$ for each $t \in (a,b)$.

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 2.16, it follows that

$$\mathbf{b} \cdot \dot{\boldsymbol{\gamma}} = 0\,, \quad \mathbf{b} \cdot \mathbf{n} = 0\,.$$

By the calculation in Remark 2.17 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 2.17 Point 6, we also know that

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

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

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 2.40

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 2.41: Frenet frame of reparametrization

Let $\boldsymbol{\gamma}$ be unit-speed, 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 1.4 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:

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

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

Example 2.42: Frenet frame of Helix

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

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

1. Compute the arc-length reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$.
2. Compute the Frenet frame of $\tilde{\boldsymbol{\gamma}}$.

Solution.

1. The arc-length of $\boldsymbol{\gamma}$ starting at $t_0 = 0$, and its inverse, are

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) &= (-R\sin(t), R\cos(t), H) \\ \|\dot{\boldsymbol{\gamma}}\| &= \rho, \qquad \rho := \sqrt{R^2 + H^2} \\ s(t) &= \int_0^t \|\dot{\boldsymbol{\gamma}}(u)\| \, du = \rho t , \qquad t(s) = \frac{s}{\rho} . \end{aligned}$$

The arc-length reparametrization $\tilde{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$ is

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

2. Compute the tangent vector to $\tilde{\boldsymbol{\gamma}}$ and its derivative

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

The curvature of $\tilde{\boldsymbol{\gamma}}$ is

$$\tilde{\kappa}(s) = \|\ddot{\tilde{\boldsymbol{\gamma}}}(s)\| = \|\dot{\tilde{\mathbf{t}}}(s)\| = \frac{R}{R^2 + H^2} .$$

The principal normal vector and binormal are

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

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 2.41.

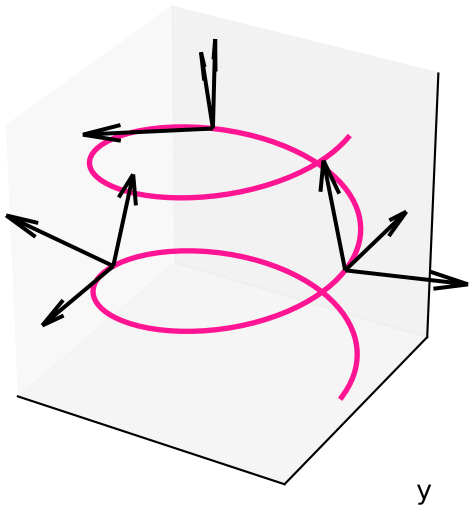


Figure 2.8.: Frenet frame of the Helix with R = 1 and H = 1.

It is of course possible to derive formulas to compute the Frenet frame of a regular curve. These are obtained by using the arc-length reparametrization. We give the formulas without proof.

Theorem 2.43: General Frenet frame formulas

The Frenet frame of a regular curve $\boldsymbol{\gamma}$ is

$$\mathbf{t} = \frac{\dot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}}\|}, \quad \mathbf{b} = \frac{\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}, \quad \mathbf{n} = \mathbf{b} \times \mathbf{t} = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \times \dot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| \, \|\dot{\boldsymbol{\gamma}}\|}.$$

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 2.44

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be unit-speed, 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 2.45: Torsion of unit-speed curve with $\kappa \neq 0$

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be unit-speed, with $\kappa \neq 0$. The **torsion** of $\boldsymbol{\gamma}$ is the unique scalar $\tau(t)$ such that

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

In particular,

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

Warning

We defined the torsion **only** for space curves $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ 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 2.46: Torsion of regular curve with $\kappa \neq 0$

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be a regular curve with $\kappa \neq 0$. Let $\tilde{\boldsymbol{\gamma}}$ be a unit-speed reparametrization of $\boldsymbol{\gamma}$ with $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}} \circ \phi$ and $\phi : (a,b) \to (\tilde{a},\tilde{b})$. Let $\tilde{\tau} : (\tilde{a},\tilde{b}) \to \mathbb{R}$ be the torsion of $\tilde{\boldsymbol{\gamma}}$. The **torsion** of $\boldsymbol{\gamma}$ is

$$\tau(t) = \tilde{\tau}(\phi(t))\,.$$

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 2.47: τ is invariant for unit-speed reparametrization

Consider the setting of Definition 2.47. 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 2.41 and infer that the Frenet frames are linked by the formulas

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

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 $\dot{\xi} \equiv \pm 1$. The torsions of $\tilde{\boldsymbol{\gamma}}$ and $\dot{\hat{\boldsymbol{\gamma}}}$ are computed by

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

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

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

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.

Theorem 2.48: Torsion formula

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be regular, with $\kappa \neq 0$. The torsion of $\boldsymbol{\gamma}$ is

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

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 [7].

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

Example 2.49: Torsion of the Helix with formula

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

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

1. Prove that $\boldsymbol{\gamma}$ is regular with non-vanishing curvature.
2. Compute the torsion of $\boldsymbol{\gamma}$.

Solution.

1. $\boldsymbol{\gamma}$ is regular with non-vanishing curvature, since

$$\|\dot{\boldsymbol{\gamma}}(t)\| = \sqrt{R^2 + H^2} \geq R > 0\,, \qquad \kappa = \frac{R}{R^2 + H^2} > 0\,.$$

2. We compute the torsion using the formula:

$$\begin{aligned}
\dot{\boldsymbol{\gamma}}(t) &= (-R\sin(t), R\cos(t), H) \\
\ddot{\boldsymbol{\gamma}}(t) &= (-R\cos(t), -R\sin(t), 0) \\
\dddot{\boldsymbol{\gamma}}(t) &= (R\sin(t), -R\cos(t), 0) \\
\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} &= \left(RH\sin(t), -RH\cos(t), R^2\right) \\
\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| &= R\sqrt{R^2 + H^2} \\
(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} &= R^2 H \\
\tau(t) &= \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} = \frac{H}{R^2 + H^2}
\end{aligned}$$

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

Example 2.50: Curvature and Torsion of the 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

$$\kappa = \frac{R}{R^2 + H^2}\,, \qquad \tau = \frac{H}{R^2 + H^2}\,.$$

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

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

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.

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 2.9.

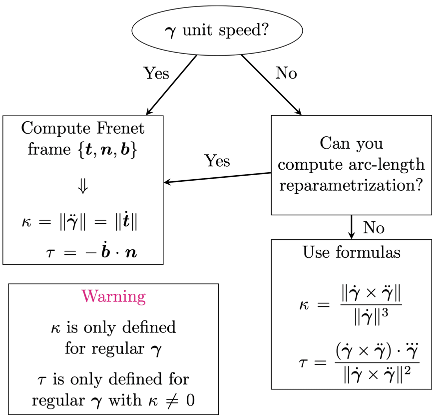


Figure 2.9.: How to compute κ and τ for regular curve $\boldsymbol{\gamma}$.

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 provide an example where we compute curvature and torsion by making use of the Frenet frame.

Example 2.51: 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 2.9. We ask the first question:

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

We have already computed in Example 1.42 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 1.42. 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 1.42, the Frenet frame and curvature of $\tilde{\boldsymbol{\gamma}}$ are

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

we are left to compute the torsion using formula

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

Indeed, we have

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

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 2.49.

Example 2.52: Calculation of torsion

Question. Compute the torsion of the curve

$$\boldsymbol{\gamma}(t) = \left(\frac{8}{5}\cos(t), 1 - 2\sin(t), \frac{6}{5}\cos(t)\right).$$

Solution. Resuming calculations from Example 2.25,

$$\begin{aligned} \dddot{\boldsymbol{\gamma}} &= \left(\frac{8}{5}\sin(t), 2\cos(t), \frac{6}{5}\sin(t)\right) \\ (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} &= \frac{96}{25}\sin(t) - \frac{96}{25}\sin(t) = 0 \\ \tau(t) &= \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} = 0 \end{aligned}$$

Example 2.53: Twisted cubic

Question. Let $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^3$ be the *twisted cubic*

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

1. Is $\boldsymbol{\gamma}$ regular/unit-speed? Justify your answer.
2. Compute the curvature and torsion of $\boldsymbol{\gamma}$.
3. Compute the Frenet frame of $\boldsymbol{\gamma}$.

Solution.

1. $\boldsymbol{\gamma}$ is regular, but not-unit speed, because

$$\begin{aligned} &\dot{\boldsymbol{\gamma}}(t) = (1, 2t, 3t^2) \\ &\|\dot{\boldsymbol{\gamma}}(t)\| = \sqrt{1 + 4t^2 + 9t^4} \geq 1 \qquad \|\dot{\boldsymbol{\gamma}}(1)\| = \sqrt{14} \neq 1 \end{aligned}$$

2. Compute the following quantities

$$\begin{aligned} \ddot{\boldsymbol{\gamma}} &= (0, 2, 6t) & \|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| &= 2\sqrt{1 + 9t^2 + 9t^4} \\ \dddot{\boldsymbol{\gamma}} &= (0, 0, 6) & (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} &= 12 \\ \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} &= (6t^2, -6t, 2) & & \end{aligned}$$

Compute curvature and torsion using the formulas:

$$\kappa(t) = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} = \frac{2\sqrt{1+9t^2+9t^4}}{(1+4t^2+9t^4)^{3/2}}$$

$$\tau(t) = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} = \frac{3}{1+9t^2+9t^4}\,.$$

3. By the Frenet frame formulas and the above calculations,

$$\mathbf{t} = \frac{\dot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}}\|} = \frac{1}{\sqrt{1+4t^2+9t^4}}\,(1, 2t, 3t^2)$$

$$\mathbf{b} = \frac{\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|} = \frac{1}{\sqrt{1+9t^2+9t^4}}(3t^2, -3t, 1)$$

$$\mathbf{n} = \mathbf{b} \times \mathbf{t} = \frac{(-9t^3 - 2t, 1 - 9t^4, 6t^3 + 3t)}{\sqrt{1+9t^2+9t^4}\,\sqrt{1+4t^2+9t^4}}$$

2.7. Frenet-Serret equations

For unit-speed curves $\boldsymbol{\gamma} \colon (a, b) \to \mathbb{R}^3$ 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 2.54: Frenet frame is right-handed

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

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

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 2.55: Frenet-Serret equations

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

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

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 2.44. Now, recall that in Proposition 2.54 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 2.56: Vectorial form of Frenet-Serret equations

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

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

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

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

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

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

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 $\mathbf{\Gamma}$ 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 2.57: 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 $\mathbf{p} \in \mathbb{R}^3$, and $R \in \mathrm{SO}(3)$ **rotation matrix**,

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

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 2.58: Fundamental Theorem of Space Curves

Let $\kappa, \tau \ : \ (a,b) \to \mathbb{R}$ be smooth, 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

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

There exists a rigid motion $M \colon \mathbb{R}^3 \to \mathbb{R}^3$ 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 1.31, which said that signed curvature characterizes 2D curves.

The proof of Theorem 2.58 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 2.58.

Example 2.59: Application of FTSC

Question. Consider the curve

$$\boldsymbol{\gamma}(t) = \left(\sqrt{3}\,t - \sin(t), \sqrt{3}\sin(t) + t, 2\cos(t)\right).$$

1. Calculate the curvature and torsion of $\boldsymbol{\gamma}$.
2. The helix of radius R and rise H is parametrized by

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

Recall that $\boldsymbol{\eta}$ has curvature and torsion

$$\kappa^{\boldsymbol{\eta}} = \frac{R}{R^2 + H^2}, \qquad \tau^{\boldsymbol{\eta}} = \frac{H}{R^2 + H^2}.$$

Prove that there exist a rigid motion $M \colon \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$\boldsymbol{\gamma}(t) = M(\boldsymbol{\eta}(t)), \quad \forall\, t \in \mathbb{R}. \tag{2.29}$$

Solution.

1. Compute curvature and torsion with the formulas

$$\begin{aligned}
\dot{\boldsymbol{\gamma}}(t) &= \left(\sqrt{3} - \cos(t), \sqrt{3}\cos(t) + 1, -2\sin(t)\right) \\
\ddot{\boldsymbol{\gamma}}(t) &= \left(\sin(t), -\sqrt{3}\sin(t), -2\cos(t)\right) \\
\dddot{\boldsymbol{\gamma}}(t) &= \left(\cos(t), -\sqrt{3}\cos(t), 2\sin(t)\right) \\
\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t) &= \left(-2\left(\sqrt{3} + \cos(t)\right), 2\left(\sqrt{3}\cos(t) - 1\right), -4\sin(t)\right) \\
\|\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t)\|^2 &= 32 \\
\|\dot{\boldsymbol{\gamma}}(t)\|^2 &= 8 \\
(\dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t)) \cdot \dddot{\boldsymbol{\gamma}}(t) &= -8 \\
\kappa(t) &= \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} = \frac{\sqrt{32}}{8^{\frac{3}{2}}} = \frac{1}{4} \\
\tau(t) &= \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} = \frac{-8}{32} = -\frac{1}{4}\,.
\end{aligned}$$

2. Equating $\kappa = \kappa^{\boldsymbol{\eta}}$ and $\tau = \tau^{\boldsymbol{\eta}}$, we obtain

$$\frac{R}{R^2 + H^2} = \frac{1}{4}\,, \qquad \frac{H}{R^2 + H^2} = -\frac{1}{4}$$

Rearranging both equalities we get

$$R^2 + H^2 = 4R\,, \qquad R^2 + H^2 = -4H\,,$$

from which we find the relation $R = -H$. Substituting into $R^2 + H^2 = -4H$, we get

$$H = -2\,, \quad R = -H = 2\,.$$

For these values of R and H we have $\kappa = \kappa^{\boldsymbol{\eta}}$ and $\tau = \tau^{\boldsymbol{\eta}}$. By the FTSC, there exists a rigid motion $M : \mathbb{R}^3 \to \mathbb{R}^3$ satisfying (2.29).

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 2.60: 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 2.10.

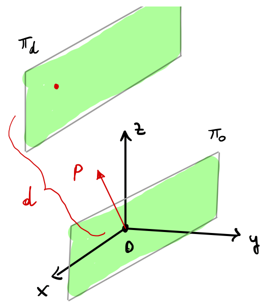


Figure 2.10.: The plane $\boldsymbol{\pi}_0$ is the set of points of $\mathbb{R}^3$ orthogonal to $\mathbf{P}$. The plane $\boldsymbol{\pi}_d$ is obtained by translating $\boldsymbol{\pi}_0$ by a quantity d in direction $\mathbf{P}$.

Theorem 2.61: Curves contained in a plane - Part I

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be regular with $\kappa \neq 0$. They are equivalent:

1. The torsion of $\boldsymbol{\gamma}$ satisfies
$$\tau(t) = 0\,, \quad \forall\, t \in (a,b)\,.$$
2. $\boldsymbol{\gamma}$ is contained in a plane: 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 2.11. 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

$$\boldsymbol{\gamma}(t) \cdot \mathbf{b} = d, \quad \forall\, t \in (a,b). \tag{2.30}$$

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 (2.31) we conclude that $\lambda(t)$ is constant and non-zero. Thus

$$\mathbf{b}(t) = \hat{\lambda}\mathbf{P} , \quad \forall\, t \in (a, b) , \tag{2.31}$$

for some $\hat{\lambda} \neq 0$. Differentiating (2.31) 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.

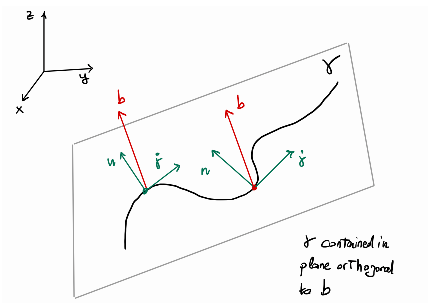


Figure 2.11.: If $\mathbf{b}$ is constant, then $\boldsymbol{\gamma}$ lies in the plane spanned by $\dot{\boldsymbol{\gamma}}$ and $\mathbf{n}$. Note that $\mathbf{b}$ is the unit normal to such plane.

As a corollary of Step 1 in the proof of Theorem 2.61 we obtain the following statement.

Theorem 2.62: Curves contained in a plane - Part II

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be regular, with $\kappa \neq 0$ and $\tau = 0$. Then, the binormal $\mathbf{b}$ is a constant vector, and $\boldsymbol{\gamma}$ is contained in the plane of equation

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

Proof

Following the proof of Step 1 of Theorem 2.61, we get to the conclusion (2.30) that $\boldsymbol{\gamma}$ satisfies

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

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

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

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 2.63: A planar curve

Question. Consider the curve

$$\boldsymbol{\gamma}(t) = (t, 2t, t^4) \,, \quad t > 0 \,.$$

1. Prove that $\boldsymbol{\gamma}$ is regular.
2. Compute the curvature and torsion of $\boldsymbol{\gamma}$.
3. Prove that $\boldsymbol{\gamma}$ is contained in a plane. Compute the equation of such plane.

Solution.

1. $\boldsymbol{\gamma}$ is regualar because $\dot{\boldsymbol{\gamma}}(t) = (1, 2, 4t^3) \neq \mathbf{0}$.
2. Compute the following quantities

$$\begin{aligned}
\|\dot{\boldsymbol{\gamma}}\| &= \sqrt{5 + 16t^4} & \qquad \dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} &= 12\,(2t^2, -t^2, 0) \\
\ddot{\boldsymbol{\gamma}} &= 12\,(0, 0, t^2) & \|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| &= 12\sqrt{5}\, t^2 \\
\dddot{\boldsymbol{\gamma}} &= 24\,(0, 0, t) & (\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} &= 0
\end{aligned}$$

Compute curvature and torsion with the formulas

$$\kappa(t) = \frac{\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}}\|^3} = \frac{12\sqrt{5}\,t^2}{\sqrt{5+16t^4}}$$

$$\tau(t) = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|} = 0\,.$$

3. $\boldsymbol{\gamma}$ lies in a plane because $\tau = 0$. The binormal is

$$\mathbf{b} = \frac{\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|} = \frac{1}{\sqrt{5}}(2,-1,0)\,.$$

At $t_0 = 0$ we have $\boldsymbol{\gamma}(0) = \mathbf{0}$. The equation of the plane containing $\boldsymbol{\gamma}$ is then $\mathbf{x} \cdot \mathbf{b} = 0$, which reads

$$\frac{2}{\sqrt{5}}x - \frac{1}{\sqrt{5}}y = 0 \quad \Longrightarrow \quad 2x - y = 0\,.$$

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

Theorem 2.64: Curves contained in a circle

Let $\boldsymbol{\gamma} : (a,b) \to \mathbb{R}^3$ be unit-speed. They are equivalent:

1. $\boldsymbol{\gamma}$ is contained in a circle of radius $R > 0$.
2. There exists $R > 0$ such that

$$\kappa(t) = \frac{1}{R}\,, \quad \tau(t) = 0\,, \quad \forall\, t \in (a,b)\,.$$

Theorem 2.64 is actually a consequence of the Fundamental Theorem of Space Curves Theorem 2.58, 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 2.58, 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 Theorem 2.64, 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 R. Then, up to a rotation and translation, $\boldsymbol{\gamma}$ is parametrized by

$$\boldsymbol{\gamma}(t) = (R\cos(t), R\sin(t), 0) \,.$$

We have already seen that in this case

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

concluding the proof.
Step 2. Suppose that

$$\kappa(t) = \frac{1}{R} \,, \quad \tau(t) = 0 \,, \quad \forall\, t \in (a,b) \,,$$

for some constant $R > 0$. 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

$$\begin{aligned} \dot{\mathbf{t}} &= \kappa \mathbf{n} = \frac{1}{R}\,\mathbf{n} \\ \dot{\mathbf{n}} &= -\kappa \mathbf{t} + \tau \mathbf{b} = -\frac{1}{R}\mathbf{t} \\ \dot{\mathbf{b}} &= -\tau \mathbf{n} = \mathbf{0} \end{aligned}$$

In particular $\dot{\mathbf{b}} = \mathbf{0}$ and so $\mathbf{b}$ is a constant vector. As seen in the proof Thoerem 2.61, this implies that $\boldsymbol{\gamma}$ is contained in a plane $\boldsymbol{\pi}$ orthogonal to $\mathbf{b}$, see Figure 2.11. As c is constant we get

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

where we used that $\dot{\boldsymbol{\gamma}} = \mathbf{t}$ and the second Frenet-Serret equation. Therefore

$$\boldsymbol{\gamma}(t) + R\mathbf{n}(t) = \mathbf{p} \,, \quad t \in (a,b) \,,$$

for some constant point $\mathbf{p} \in \mathbb{R}^3$. In particular

$$\|\boldsymbol{\gamma}(t) - \mathbf{p}\| = \|-R\mathbf{n}(t)\| = R \,,$$

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

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

The intersection of $\mathcal{S}$ with the plane $\boldsymbol{\pi}$ is a circle $\mathcal{C}$ with some radius $r \geq 0$ (note that $\mathcal{C}$ 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

$$\boldsymbol{\gamma} \in \boldsymbol{\pi} \cap \mathcal{S} = \mathcal{C} \,. \tag{2.32}$$

Therefore $\boldsymbol{\gamma}$ parametrizes part of $\mathcal{C}$. 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

$$\kappa = \frac{1}{r}\,, \qquad \tau = 0\,.$$

As by assumption $\kappa = 1/R$, we conclude that $R = r$. Therefore the circle $\mathscr{C}$ has radius R and the thesis follows by (2.32).

Example 2.65

Question. Consider the curve

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

1. Prove that $\boldsymbol{\gamma}$ is unit-speed.
2. Compute Frenet frame, curvature and torsion of $\boldsymbol{\gamma}$.
3. Prove that $\boldsymbol{\gamma}$ is part of a circle.

Note: As seen in the plot in Figure 2.12, $\boldsymbol{\gamma}$ is just a Circle which has been rotated an translated.
Solution.

1. $\boldsymbol{\gamma}$ is unit-speed because

$$\begin{aligned}
\dot{\boldsymbol{\gamma}}(t) &= \left(-\frac{4}{5}\sin(t), -\cos(t), \frac{3}{5}\sin(t)\right) \\
\|\dot{\boldsymbol{\gamma}}(t)\|^2 &= \frac{16}{25}\sin^2(t) + \cos^2(t) + \frac{9}{25}\sin^2(t) = 1
\end{aligned}$$

2. As $\boldsymbol{\gamma}$ is unit-speed, the tangent vector is $\mathbf{t}(t) = \dot{\boldsymbol{\gamma}}(t)$. The curvature, normal, binormal and torsion are

$$\begin{aligned}
\dot{\mathbf{t}}(t) &= \left(-\frac{4}{5}\cos(t), \sin(t), \frac{3}{5}\cos(t)\right) \\
\kappa(t) &= \|\dot{\mathbf{t}}(t)\| = \frac{16}{25}\cos^2(t) + \sin^2(t) + \frac{9}{25}\cos^2(t) = 1 \\
\mathbf{n}(t) &= \frac{1}{\kappa(t)}\ddot{\boldsymbol{\gamma}}(t) = \left(-\frac{4}{5}\cos(t), \sin(t), \frac{3}{5}\cos(t)\right) \\
\mathbf{b}(t) &= \dot{\boldsymbol{\gamma}}(t) \times \mathbf{n}(t) = \left(-\frac{3}{5}, 0, -\frac{4}{5}\right) \\
\dot{\mathbf{b}} &= \mathbf{0} \\
\tau &= -\dot{\mathbf{b}} \cdot \mathbf{n} = 0
\end{aligned}$$

3. The curvature of γ is constant and the torsion is zero. Therefore γ is contained in a circle of radius
$$R = \frac{1}{\kappa} = 1\,.$$

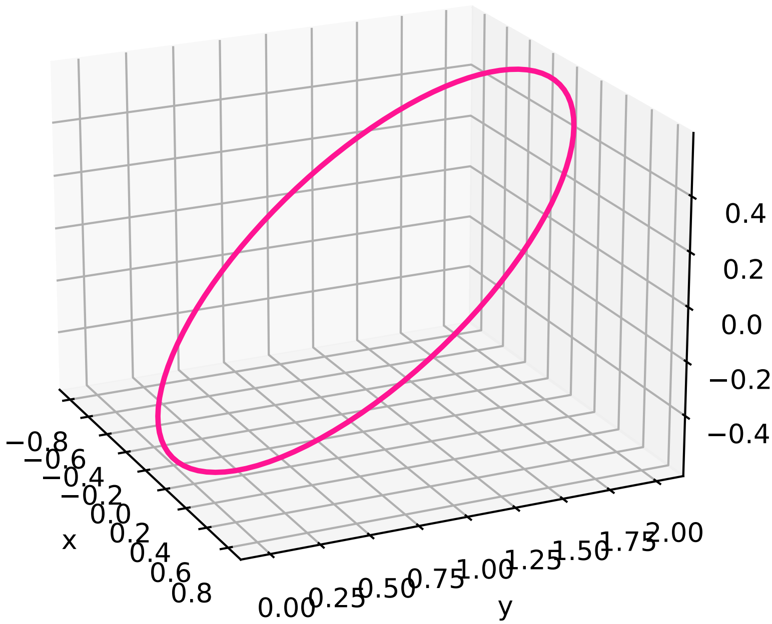


Figure 2.12.: Plot of the curve in Example above

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 Theorem 2.21 and Theorem 2.48. For reader's convenience we recall these two results.

Proposition 2.66: Curvature and torsion formulas

Let $\boldsymbol{\gamma} \colon (a,b) \to \mathbb{R}^3$ be a regular curve. The curvature $\kappa(t)$ of $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma}(t)$ is given by

$$\kappa(t) = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} \,. \tag{2.33}$$

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

$$\tau(t) = \frac{(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}}}{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|^2} \,. \tag{2.34}$$

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

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

The Frenet-Serret equations are

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

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

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

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

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) \times \ddot{\boldsymbol{\gamma}}(t) &= \tilde{\mathbf{t}}\, \dot{s} \times \left[\tilde{\kappa}\, \widetilde{\mathbf{n}}\, \dot{s}^2 + \tilde{\mathbf{t}}\, \ddot{s} \right] \\ &= \dot{s}^3\, \tilde{\kappa}\, \tilde{\mathbf{t}} \times \widetilde{\mathbf{n}} + \dot{s}\, \ddot{s}\, \tilde{\mathbf{t}} \times \tilde{\mathbf{t}} \\ &= \dot{s}^3\, \tilde{\kappa}\, \tilde{\mathbf{b}} \end{aligned} \tag{2.36}$$

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 (2.36) 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 (2.33).
Part 2. Differentiating the second line in (2.35) 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 (2.36), we get

$$\begin{aligned}
(\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}) \cdot \dddot{\boldsymbol{\gamma}} &= \left[\dot{s}^3\tilde{\kappa}\,\tilde{\mathbf{b}}\right] \cdot \left[\ddot{\tilde{\mathbf{t}}}\dot{s}^3 + 3\dot{\tilde{\mathbf{t}}}\dot{s}\ddot{s} + \tilde{\mathbf{t}}\,\dddot{s}\right] \\
&= \dot{s}^6\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \ddot{\tilde{\mathbf{t}}} + 3\dot{s}^4\,\ddot{s}\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \dot{\tilde{\mathbf{t}}} + \dot{s}^3\,\dddot{s}\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \tilde{\mathbf{t}} \qquad (2.37) \\
&= \dot{s}^6\,\tilde{\kappa}\,\tilde{\mathbf{b}} \cdot \ddot{\tilde{\mathbf{t}}}
\end{aligned}$$

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 (2.37) 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 (2.33), 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 (2.34).

2.11. Proof: Fundamental Theorem of Space Curves

In this section we prove the *Fundamental Theorem of Space Curves* Theorem 2.58. For reader's convenience we recall the statement.

Theorem 2.67: 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 [8].

Theorem 2.68: 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 2.69: Orthogonal matrices preserve scalar product and length

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

$$A\mathbf{v} \cdot A\mathbf{w} = \mathbf{v} \cdot \mathbf{w} \tag{2.38}$$

and also

$$\|A\mathbf{v}\| = \|\mathbf{v}\| \tag{2.39}$$

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 2.70: 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

$$A\mathbf{v} \times A\mathbf{w} = \det(A) \left(A^{-1}\right)^T (\mathbf{v} \times \mathbf{w}) \tag{2.40}$$

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

$$R\mathbf{v} \times R\mathbf{w} = R\,(\mathbf{v} \times \mathbf{w}) \tag{2.41}$$

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 (2.40) on the vectors of the orthonormal basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Let us verify (2.40) 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 (2.41) follows from (2.40).

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

Proposition 2.71: 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

$$\tilde{\mathbf{t}} = R\,\mathbf{t}\,, \qquad \tilde{\mathbf{n}} = R\,\mathbf{n}\,, \qquad \tilde{\mathbf{b}} = R\,\mathbf{b}\,, \tag{2.42}$$

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 (2.39), and the assumption of $\boldsymbol{\gamma}$ unit-speed. This concludes the proof that $\tilde{\boldsymbol{\gamma}}$ is unit-speed.
Let us now prove (2.42). The relation $\dot{\tilde{\boldsymbol{\gamma}}} = R\,\dot{\boldsymbol{\gamma}}$ reads

$$\tilde{\mathbf{t}} = R\,\mathbf{t}\,,$$

which gives the first equation in (2.42). Since $\ddot{\tilde{\boldsymbol{\gamma}}} = R\,\ddot{\boldsymbol{\gamma}}$, by (2.39) 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 (2.42). Finally, by definition of binormal and the first two equations in (2.42),

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

where in the third equality we used (2.41). The proof is concluded.

The last Proposition concerns the evolution of orthonormal systems of vectors.

Proposition 2.72: 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

$$\begin{pmatrix} \dot{\mathbf{u}}_1 \\ \dot{\mathbf{u}}_2 \\ \dot{\mathbf{u}}_3 \end{pmatrix} = A \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix} \tag{2.43}$$

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

$$\lambda_{ij}(t) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.44}$$

for all t. Hence the proof is concluded if we show (2.44).
In order to do so, first note that the ODE in (2.43) reads

$$\dot{\mathbf{u}}_i = \sum_{k=1}^{3} a_{ik} \mathbf{u}_k\,, \qquad i = 1,2,3 \tag{2.45}$$

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 (2.45) and the definition of λ_{ij}. The above calculation shows that λ_{ij} solves the ODE

$$\dot{\lambda}_{ij} = \sum_{k=1}^{3} a_{ik}\lambda_{kj} + \sum_{k=1}^{3} a_{jk}\lambda_{ik} \tag{2.46}$$

We claim that δ_{ij} solves (2.46). 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 (2.46). 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 (2.46), by the uniqueness of solutions to ODE systems in Theorem 1.68, we conclude that (2.44) holds for all t. The proof is concluded.

We are finally ready to prove the Fundamental Theorem of Space Curves.

Proof: Proof of Theorem 1.67

Suppose given two smooth functions $\kappa, \tau : (a, b) \to \mathbb{R}$ with $\kappa > 0$.
Part 1. Existence of $\boldsymbol{\gamma}$.
Consider the Frenet-Serret system of ODEs

$$\begin{pmatrix} \dot{\mathbf{t}} \\ \dot{\mathbf{n}} \\ \dot{\mathbf{b}} \end{pmatrix} = A \begin{pmatrix} \mathbf{t} \\ \mathbf{n} \\ \mathbf{b} \end{pmatrix} \tag{2.47}$$

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 1.68 guarantess the existence of a smooth solution $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ such that

$$\mathbf{t}(t_0) = \mathbf{e}_1 \,, \quad \mathbf{n}(t_0) = \mathbf{e}_2 \,, \quad \mathbf{b}(t_0) = \mathbf{e}_3 \tag{2.48}$$

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 2.73 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 (2.48) gives

$$\mathbf{b}(t_0) = \lambda(t_0) \, \mathbf{t}(t_0) \times \mathbf{n}(t_0) \quad \Longrightarrow \quad \mathbf{e}_3 = \lambda(t_0) \, \mathbf{e}_1 \times \mathbf{e}_2$$

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

$$\mathbf{b} = \mathbf{t} \times \mathbf{n} \tag{2.49}$$

Define the curve $\boldsymbol{\gamma} \colon \ (a, b) \to \mathbb{R}^3$ 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 (2.47) 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 (2.49) we deduce that $\mathbf{b}$ is the binormal to $\boldsymbol{\gamma}$
5. Using the third equation in (2.47) 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

$$\tilde{\mathbf{t}}(t_0) = R\,\mathbf{t}(t_0), \quad \tilde{\mathbf{n}}(t_0) = R\,\mathbf{n}(t_0), \quad \tilde{\mathbf{b}}(t_0) = R\,\mathbf{b}(t_0) \tag{2.50}$$

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

$$M(\boldsymbol{\gamma}(t_0)) = \tilde{\boldsymbol{\gamma}}(t_0) \tag{2.51}$$

Define the new curve

$$\hat{\boldsymbol{\gamma}}(t) := M(\boldsymbol{\gamma}(t))$$

By Proposition 1.71 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

$$\hat{\mathbf{t}} = R\,\mathbf{t}, \quad \hat{\mathbf{n}} = R\,\mathbf{n}, \quad \hat{\mathbf{b}} = R\,\mathbf{b} \tag{2.52}$$

From the above relations we deduce that $(\hat{\mathbf{t}}, \hat{\mathbf{n}}, \hat{\mathbf{b}})$ solves the Frenet-Serret equations at (2.47). 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 (2.47). By evaluating (2.52) at t_0, and comparing with (2.50), 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 1.68 to the ODE system (2.47), 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 (2.51) 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 $\mathscr{T}$ a collection of subsets of X. We say that $\mathscr{T}$ is a **topology** on X if the following 3 properties hold:

- (A1) The sets $\emptyset, X$ belong to $\mathscr{T}$,
- (A2) If $\{A_i\}_{i\in I}$ is an arbitrary family of elements of $\mathscr{T}$, then

$$\bigcup_{i\in I} A_i \in \mathscr{T} .$$

- (A3) If $A, B \in \mathscr{T}$ then $A \cap B \in \mathscr{T}$.

Further, we say:

- The pair $(X, \mathscr{T})$ is a **topological space**.
- The elements of X are called **points**.
- The sets in the topology $\mathscr{T}$ are called **open sets**.

Remark 3.2

The intersection property of $\mathscr{T}$, Property (A3) in Definition 3.1, is equivalent to the following:

- (A3') If $A_1, \dots, A_M \in \mathscr{T}$ for some $M \in \mathbb{N}$, then

$$\bigcap_{n=1}^{M} A_n \in \mathscr{T} .$$

The equivalence between (A3) and (A3') can be immediately obtained by induction.

Warning

Notice:

- The union property (A2) of $\mathcal{T}$ holds for an **arbitrary** number of sets, even uncountable!
- The intersection property (A3') of $\mathcal{T}$ holds only for a **finite** number of sets.

There are two main examples of topologies that one should always keep in mind. These are:

- **Trivial topology**: The topology with the smallest possible number of sets.
- **Discrete topology**: The topology with the highest possible number of sets.

Definition 3.3: Trivial topology

Let X be a set. The **trivial topology** on X is the collection of sets

$$\mathcal{T}_{\text{trivial}} := \{\emptyset, X\} .$$

Proof: $\mathcal{T}_{\text{trivial}}$ is a topology on X

To prove $\mathcal{T}_{\text{trivial}}$ is a topology on X, we need to check the axioms:

- (A1) By definition of $\mathcal{T}_{\text{trivial}}$, we have $\emptyset, X \in \mathcal{T}_{\text{trivial}}$.
- (A2) Assume $\{A_i\}_{i\in I}$ is an arbitrary family of elements of $\mathcal{T}_{\text{trivial}}$. There are two possibilities
  - If all the sets A_i are empty, then $\bigcup_i A_i = \emptyset \in \mathcal{T}_{\text{trivial}}$.
  - If $A_i = X$ for at least one index i, then $\bigcup_i A_i = X \in \mathcal{T}_{\text{trivial}}$.

  In both cases, $\bigcup_i A_i \in \mathcal{T}_{\text{trivial}}$, so that (A2) holds.
- (A3) Assume $A, B \in \mathcal{T}_{\text{trivial}}$. We have 3 cases:
  - $A = B = \emptyset$. Then $A \cap B = \emptyset \in \mathcal{T}_{\text{trivial}}$.
  - $A = X$ and $B = \emptyset$. Then $A \cap B = \emptyset \in \mathcal{T}_{\text{trivial}}$.
  - $A = B = X$. Then $A \cap B = X \in \mathcal{T}_{\text{trivial}}$.

  In all the 3 cases we have $A \cap B \in \mathcal{T}_{\text{trivial}}$, so that (A3) holds.

Therefore $\mathcal{T}_{\text{trivial}}$ is a topology on X.

Definition 3.4: Discrete topology

Let X be a set. The **discrete topology** on X is the collection of all subsets of X

$$\mathcal{T}_{\text{discrete}} := \{A \, : \, A \subseteq X\} .$$

Proof: $\mathscr{T}_{\text{discrete}}$ is a topology on X

To prove $\mathscr{T}_{\text{discrete}}$ is a topology on X, we need to check the axioms:

- (A1) We have $\emptyset, X \in \mathscr{T}_{\text{discrete}}$, since $\emptyset$ and X are subsets of X.
- (A2) The arbitrary union of subsets of X is still a subset of X. Therefore $\bigcup A_i \in \mathscr{T}_{\text{discrete}}$ whenever $A_i \in \mathscr{T}_{\text{discrete}}$ for all $i \in I$.
- (A3) The intersection of two subsets of X is still a subset of X. Therefore $A \cap B \in \mathscr{T}_{\text{discrete}}$ whenever $A, B \in \mathscr{T}_{\text{discrete}}$.

Therefore $\mathscr{T}_{\text{discrete}}$ 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.

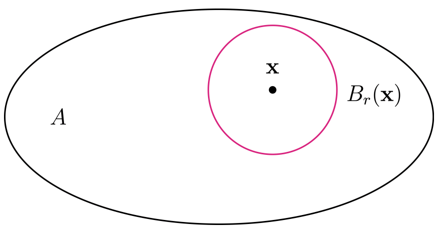


Figure 3.1.: The set $A \subseteq \mathbb{R}^n$ is open if for every $\mathbf{x} \in A$ there exists $r > 0$ such that $B_r(\mathbf{x}) \subseteq A$.

Definition 3.6: Euclidean topology of $\mathbb{R}^n$

The **Euclidean topology** on $\mathbb{R}^n$ is the collection of sets

$$\mathcal{T}_{\text{euclid}} := \{A \, : \, A \subseteq \mathbb{R}^n \, , \, A \text{ is open}\} \, .$$

Proof: $\mathcal{T}_{\text{euclid}}$ is a topology on $\mathbb{R}^n$

To prove $\mathcal{T}_{\text{euclid}}$ is a topology on $\mathbb{R}^n$, we need to check the axioms:

- (A1) We have $\emptyset, \mathbb{R}^n \in \mathcal{T}_{\text{euclid}}$: 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 $A_i \in \mathcal{T}_{\text{euclid}}$ for all $i \in I$. Define the union $A = \bigcup_i A_i$. 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 $B_r(\mathbf{x}) \subseteq A$, so that $A \in \mathcal{T}_{\text{euclid}}$.
- (A3) Let $A, B \in \mathcal{T}_{\text{euclid}}$. 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 $A \cap B \in \mathcal{T}_{\text{euclid}}$.

This proves that $\mathcal{T}_{\text{euclid}}$ 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: $B_r(\mathbf{x})$ is an open set of $\mathcal{T}_{\text{euclid}}$

Let $\mathbb{R}^n$ be equipped with the Euclidean topology $\mathcal{T}_{\text{euclid}}$. Let $r > 0$ and $\mathbf{x} \in \mathbb{R}^n$. Then $B_r(\mathbf{x}) \in \mathcal{T}_{\text{euclid}}$.

Proof

To prove that $B_r(\mathbf{x}) \in \mathcal{T}_{\text{euclid}}$, we need to show 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). Thus, $B_r(\mathbf{x})$ satisfies (3.1), ending the proof.

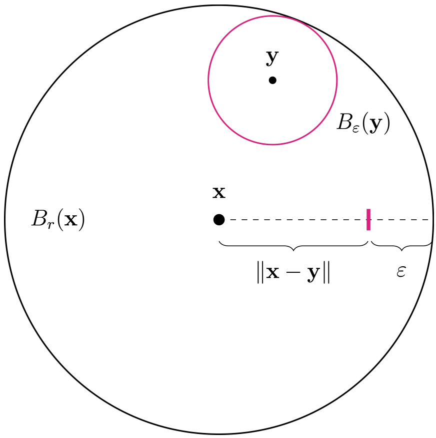


Figure 3.2.: The ball $B_\varepsilon(\mathbf{y})$ is contained in $B_r(\mathbf{x})$ if $\varepsilon := r - \|\mathbf{x} - \mathbf{y}\|$.

3.1. Closed sets

The opposite of open sets are closed sets.

Definition 3.8: Closed set

Let $(X, \mathscr{T})$ be a topological space. A set $C \subseteq X$ is **closed** if

$$C^c \in \mathscr{T}\,,$$

where $C^c := X \smallsetminus C$ 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.
Note: 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:

1. The equivalence between (A1) and (C1) is clear, since
$$\emptyset^c = X \,, \quad X^c = \emptyset \,.$$
2. Suppose C_i is closed for all $i \in I$. Therefore C_i^c is 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.
3. 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

$$\mathscr{C} := \{V(I) \, : \, I \subset \mathbb{K}[x_1, \dots, x_n]\} .$$

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 $V((x^2 + y^2)^3 - 4x^2y^2)$.
Question. Prove that $\mathscr{C}$ satisfies (C1), (C2) and (C3).
Solution. Easy check, left by exercise.

3.2. Comparing topologies

Consider the situation where you have two topologies $\mathscr{T}_1$ and $\mathscr{T}_2$ on the same set X. We would like to have some notions of comparison between $\mathscr{T}_1$ and $\mathscr{T}_2$.

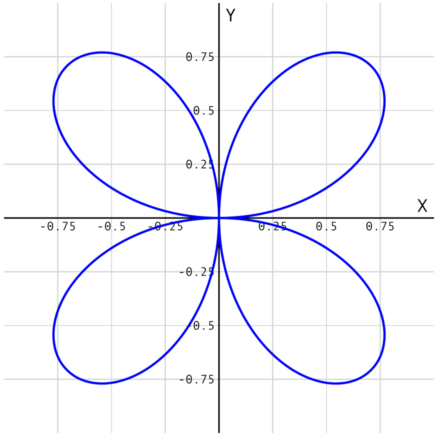


Figure 3.3.: The Quadrifolium is an affine variety with algebraic set $V((x^2 + y^2)^3 - 4x^2y^2)$.

Definition 3.11: Comparing topologies

Let X be a set and let $\mathscr{T}_1$, $\mathscr{T}_2$ be topologies on X.

1. $\mathscr{T}_1$ is **finer** than $\mathscr{T}_2$ if $\mathscr{T}_2 \subseteq \mathscr{T}_1$.
2. $\mathscr{T}_1$ is **strictly finer** than $\mathscr{T}_2$ if $\mathscr{T}_2 \subsetneq \mathscr{T}_1$.
3. $\mathscr{T}_1$ and $\mathscr{T}_2$ are the **same** topology if $\mathscr{T}_1 = \mathscr{T}_2$.

Example 3.12: Comparing $\mathscr{T}_{\text{trivial}}$ and $\mathscr{T}_{\text{discrete}}$

Let X be a set. Then $\mathscr{T}_{\text{trivial}} \subseteq \mathscr{T}_{\text{discrete}}$.

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 $\mathbb{R}$

Question. The **cofinite topology** on $\mathbb{R}$ is the collection of sets

$$\mathscr{T}_{\text{cofinite}} := \{U \subseteq \mathbb{R} \, : \, U^c \text{ is finite, or } U^c = \mathbb{R}\} .$$

1. Prove that $(\mathbb{R}, \mathscr{T}_{\text{cofinite}})$ is a topological space.
2. Prove that $\mathscr{T}_{\text{cofinite}} \subseteq \mathscr{T}_{\text{euclid}}$.
3. Prove that $\mathscr{T}_{\text{cofinite}} \neq \mathscr{T}_{\text{euclid}}$.

Solution. Part 1. Show that the topology properties are satisfied:
(A1) We have $\emptyset \in \mathscr{T}_{\text{cofinite}}$, since $\emptyset^c = \mathbb{R}$. We have $\mathbb{R} \in \mathscr{T}_{\text{cofinite}}$ because $\mathbb{R}^c = \emptyset$ is finite.

(A2) Let $U_i \in \mathcal{T}_{\text{cofinite}}$ for all $i \in I$, and define $U := \bigcup_{i \in I} U_i$. By the De Morgan's laws we have

$$U^c = \left(\cup_{i\in I} U_i\right)^c = \cap_{i\in I} U_i^c \,.$$

We have two cases:

1. There exists $i_0 \in I$ such that $U_{i_0}^c$ is finite. Then
$$U^c = \cap_{i\in I} U_i^c \subset U_{i_0}^c \,,$$
and therefore U^c is finite, showing that $U \in \mathcal{T}_{\text{cofinite}}$.
2. None of the sets U_i^c is finite. Therefore $U_i^c = \mathbb{R}$ for all $i \in I$, from which we deduce
$$U^c = \cap_{i\in I} U_i^c = \mathbb{R} \qquad \Longrightarrow \qquad U \in \mathcal{T}_{\text{cofinite}} \,.$$

In both cases, we have $U \in \mathcal{T}_{\text{cofinite}}$, so that (A2) holds.
(A3) Let $U, V \in \mathcal{T}_{\text{cofinite}}$. Set $A = U \cap V$. Then

$$A^c = U^c \cup V^c \,.$$

We have 2 possibilities:

1. U^c, V^c finite: Then A^c is finite, and $A \in \mathcal{T}_{\text{cofinite}}$.
2. $U^c = \mathbb{R}$ or $V^c = \mathbb{R}$: Then $A^c = \mathbb{R}$, and $A \in \mathcal{T}_{\text{cofinite}}$.

In all cases, we have shown that $A \in \mathcal{T}_{\text{cofinite}}$, so that (A3) holds.
Part 2. Let $U \in \mathcal{T}_{\text{cofinite}}$. We have two cases:

- U^c is finite. Then $U^c = \{x_1, \dots, x_n\}$ for some points $x_i \in \mathbb{R}$. Up to relabeling the points, we can assume that $x_i < x_j$ when $i < j$. Therefore,
$$U = \{x_1, \dots, x_n\}^c = \bigcup_{i=0}^{n} (x_i, x_{i+1}) \,, \qquad x_0 := -\infty \,, \qquad x_{n+1} := \infty \,.$$
The sets (x_i, x_{i+1}) are open in $\mathcal{T}_{\text{euclid}}$, and therefore $U \in \mathcal{T}_{\text{euclid}}$.
- $U^c = \mathbb{R}$. Then $U = \emptyset$, which belongs to $\mathcal{T}_{\text{euclid}}$ by (A1).

In both cases, $U \in \mathcal{T}_{\text{euclid}}$. Therefore $\mathcal{T}_{\text{cofinite}} \subseteq \mathcal{T}_{\text{euclid}}$.
Part 3. consider the interval $U = (0,1)$. Then $U \in \mathcal{T}_{\text{euclid}}$. However U^c is neither $\mathbb{R}$, nor finite. Thus $U \notin \mathcal{T}_{\text{cofinite}}$.

3.3. Convergence

We have generalized the notion of open set to arbitrary sets. Next we generalize the notion of convergence of sequences.

Definition 3.14: Convergent sequence

Let $(X, \mathscr{T})$ be a topological space. Consider a sequence $\{x_n\} \subseteq X$ and a point $x \in X$. We say that x_n converges to x_0 in the topology $\mathscr{T}$, if the following property holds:

$$\begin{aligned} & \forall\, U \in \mathscr{T} \text{ s.t. } x_0 \in U \,, \; \exists\, N = N(U) \in \mathbb{N} \text{ s.t.} \\ & x_n \in U \,, \; \forall\, n \geq N \,. \end{aligned} \tag{3.4}$$

The convergence of x_n to x_0 is denoted by $x_n \to x_0$.

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: Convergent sequences in $\mathscr{T}_{\text{trivial}}$

Let X be equipped with $\mathscr{T}_{\text{trivial}}$. Let $\{x_n\} \subseteq X$, $x_0 \in X$. Then $x_n \to x_0$.

Proof

To show that $x_n \to x_0$ we need to check that (3.4) holds. Let $U \in \mathscr{T}_{\text{trivial}}$ with $x_0 \in U$. We have two cases:

- $U = \emptyset$: There is nothing to prove, since x_0 cannot be in U.
- $U = X$: Take $N = 1$. Since $U = X$, we have $x_n \in U$ for all $n \geq 1$.

Thus (3.4) holds for all the sets $U \in \mathscr{T}_{\text{trivial}}$, showing that $x_n \to x_0$.

Warning

Proposition 1.15 shows the topological limit may **not be unique**!

Proposition 3.16: Convergent sequences in $\mathcal{T}_{\text{discrete}}$

Let X be equipped with $\mathcal{T}_{\text{discrete}}$. Let $\{x_n\} \subseteq X$, $x_0 \in X$. They are equivalent:

1. $x_n \to x_0$ in the topology $\mathcal{T}_{\text{discrete}}$.
2. $\{x_n\}$ is eventually constant: $\exists\, N \in \mathbb{N}$ s.t. $x_n = x_0$, $\forall\, n \geq N$

Proof

Part 1. Assume that $x_n \to x_0$. Let $U = \{x_0\}$. Then $U \in \mathcal{T}_{\text{discrete}}$. 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 $U = \{x_0\}$, we infer $x_n = x_0$ for all $n \geq N$. Hence x_n is eventually constant.
Part 2. Assume that x_n is eventually equal to x_0, that is, there exists $N \in \mathbb{N}$ such that

$$x_n = x_0\,, \quad \forall\, n \geq N\,. \tag{3.5}$$

Let $U \in \mathcal{T}$ be an open set such that $x_0 \in U$. By (3.5) we have that

$$x_n \in U\,, \quad \forall\, n \geq N\,.$$

Since U was arbitrary, we conclude that $x_n \to x_0$.

Before proceeding to examining convergence in the Euclidean topology, let us recall the classical definition of convergence in $\mathbb{R}^n$.

Definition 3.17: Classical convergence in $\mathbb{R}^n$

Let $\{\mathbf{x}_n\} \subseteq \mathbb{R}^n$ and $\mathbf{x}_0 \in \mathbb{R}^n$. We say that $\mathbf{x}_n$ converges $\mathbf{x}_0$ in the classical sense if $\|\mathbf{x}_n - \mathbf{x}_0\| \to 0$, that is,

$$\forall\, \varepsilon > 0,\ \exists\, N \in \mathbb{N}, \quad \text{s.t.} \quad \|\mathbf{x}_n - \mathbf{x}_0\| < \varepsilon\,,\ \forall\, n \geq N\,.$$

Proposition 3.18: Convergent sequences in $\mathcal{T}_{\text{euclid}}$

Let $\mathbb{R}^n$ be equipped with $\mathcal{T}_{\text{euclid}}$. Let $\{x_n\} \subseteq \mathbb{R}^n$, $x_0 \in \mathbb{R}^n$. They are equivalent:

1. $\mathbf{x}_n \to \mathbf{x}_0$ in the topology $\mathcal{T}_{\text{euclid}}$.
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 $\mathcal{T}_{\text{euclid}}$. Fix $\varepsilon > 0$ and define $U := B_\varepsilon(\mathbf{x}_0)$. By Proposition 3.7, we have $U \in \mathcal{T}$. Moreover $\mathbf{x}_0 \in U$. As $\mathbf{x}_n \to \mathbf{x}_0$ with respect to $\mathcal{T}_{\text{euclid}}$, 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 $U \in \mathcal{T}_{\text{euclid}}$ be an arbitrary set 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

$$\|\mathbf{x}_n - \mathbf{x}_0\| < r \,, \quad \forall \, n \geq N \,.$$

By definition of $B_r(\mathbf{x}_0)$, and since $B_r(\mathbf{x}_0) \subseteq U$, the above is equivalent to

$$\mathbf{x}_n \in B_r(\mathbf{x}_0) \subseteq U \,, \quad \forall \, n \geq N \,.$$

As U is arbitrary, we infer $\mathbf{x}_n \to \mathbf{x}_0$ with respect to $\mathcal{T}_{\text{euclid}}$.

Notation

Since classical convergence in $\mathbb{R}^n$ agrees with topological convergence with respect to $\mathcal{T}_{\text{euclid}}$, 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

$$x_n \in U \,, \quad \forall \, n \geq N \,. \tag{3.6}$$

Since $U \in \mathcal{T}_2$, condition (3.6) shows that $x_n \to x_0$ 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 and Metric space

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: $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$
- (M2) Symmetry: $d(x, y) = d(y, x)$
- (M3) Triangle Inequality: $d(x, z) \leq d(x, y) + d(y, z)$

The pair (X, d) is called a **metric space**.

Definition 3.21: Euclidean distance on $\mathbb{R}^n$

The **Euclidean distance** over $\mathbb{R}^n$ is defined by

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{1/2} , \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n .$$

Proposition 3.22

Let d be the Euclidean distance on $\mathbb{R}^n$. Then $(\mathbb{R}^n, d)$ is a metric space.

Proof

It is trivial that d satisfies (M1) and (M2). To show (M3), recall the triangle inequality in $\mathbb{R}^n$:

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| , \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n .$$

Using the above, we obtain

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\| \\ &= \|(\mathbf{x} - \mathbf{z}) + (\mathbf{z} - \mathbf{y})\| \\ &\leq \|\mathbf{x} - \mathbf{z}\| + \|\mathbf{z} - \mathbf{y}\| \\ &= d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}) . \end{aligned}$$

Thus, d satisfies (M3) and $(\mathbb{R}^n, d)$ is a metric space.

Definition 3.23: p-distance on $\mathbb{R}^n$

Let $p \in [1, \infty)$. The ***p*-distance** over $\mathbb{R}^n$ is

$$d_p(\mathbf{x}, \mathbf{y}) := \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}} .$$

Note that d_2 coincides with the Euclidean distance. For $p = \infty$ we set

$$d_\infty(\mathbf{x}, \mathbf{y}) := \max_{i=1\dots,n} |x_i - y_i| .$$

Proposition 3.24

Let d_p be the ***p*-distance** over $\mathbb{R}^n$, with $p \in [1, \infty]$. Then $(\mathbb{R}^n, d_p)$ is a metric space.

Proof

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 $p \in [1, \infty)$. 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

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p , \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n .$$

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 3.25: Topology induced by the metric

Let (X, d) be a metric space. The set $A \subseteq X$ is **open** if it holds

$$\forall \, x \in U \,, \, \exists \, r \in \mathbb{R}, \, r > 0 \quad \text{s.t.} \quad B_r(x) \subseteq U \,,$$

where $B_r(x)$ is the ball centered at x of radius r, defined by

$$B_r(x) = \{ y \in X \; : \; d(x, y) < r \} \,.$$

The topology **induced by the metric** d is the collection of sets

$$\mathcal{T}_d = \{ U \; : \; U \subseteq X, \, U \text{ open} \} \,.$$

The proof that $\mathcal{T}_d$ is a topology on X follows, line by line, the proof that the Euclidean topology $\mathcal{T}_{\text{euclid}}$ is indeed a topology, see proof immediately below Definition 3.6. This is left as an exercise.

Remark 3.26: Topology induced by Euclidean distance

Consider the metric space $(\mathbb{R}^n, d)$ with d the Euclidean distance. Then

$$\mathcal{T}_d = \mathcal{T}_{\text{euclid}} \,,$$

where $\mathcal{T}_{\text{euclid}}$ is the Euclidean topology on $\mathbb{R}^n$.

The proof is trivial, since the metric d is the Euclidean distance.

Example 3.27: Discrete distance

Question. Let X be a set. The **discrete distance** is the function $d : \; X \times X \to \mathbb{R}$ defined by

$$d(x, y) \; := \begin{cases} 0 & \text{if } \; x = y \\ 1 & \text{if } \; x \neq y \end{cases}$$

1. Prove that (X, d) is a metric space.
2. Prove that $\mathcal{T}_d = \mathcal{T}_{\text{discrete}}$.

Solution. See Question 3 in Homework 3.

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 3.28

Let (X, d) be a metric space, $\mathscr{T}_d$ the topology induced by d. Then:

- For all $x \in X$, $r > 0$ we have $B_r(x) \in \mathscr{T}_d$.
- $U \in \mathscr{T}_d$ if and only if $\exists\, I$ family of indices s.t.

$$U = \bigcup_{i \in I} B_{r_i}(x_i)\,, \quad x_i \in X,\ r_i > 0\,.$$

We now define the concept of equivalent metrics.

Definition 3.29: 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 3.30

Let X be a set and d_1, d_2 be metrics on X. Suppose that there exists a constant $\alpha > 0$ such that

$$\frac{1}{\alpha}\, d_2(x,y) \leq d_1(x,y) \leq \alpha\, d_2(x,y)\,, \quad \forall\, x, y \in X\,.$$

Then d_1 and d_2 are equivalent metrics.

The proof of Proposition 3.30 is trivial, and is left as an exercise.

Example 3.31

Let $p > 1$. The metrics d_p and d_∞ on $\mathbb{R}^n$ are equivalent.
Proof Follows from Proposition 3.30 and the estimate

$$d_\infty(\mathbf{x}, \mathbf{y}) \leq d_p(\mathbf{x}, \mathbf{y}) \leq n\, d_\infty(\mathbf{x}, \mathbf{y})\,, \quad \forall \mathbf{x}\,,\ \mathbf{y} \in \mathbb{R}^n\,.$$

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.

We can characterize the convergence of sequences in metric spaces.

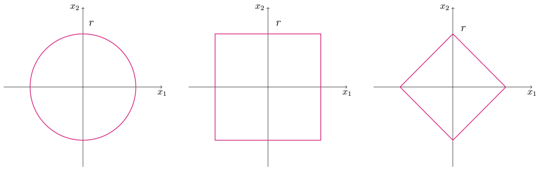


Figure 3.4.: Balls $B_r(0)$ for the metrics d_2, d_∞, d_1 in $\mathbb{R}^2$.

Proposition 3.32: Convergence in metric space

Suppose (X, d) is a metric space and $\mathscr{T}_d$ the topology induced by d. Let $\{x_n\} \subseteq X$ and $x_0 \in X$. They are equivalent:

1. $x_n \to x_0$ with respect to the topology $\mathscr{T}_d$.
2. $d(x_n, x_0) \to 0$ in $\mathbb{R}$.
3. For all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$x_n \in B_r(x_0), \ \forall \, n \geq \mathbb{N}.$$

The proof is similar to the one of Proposition 1.18, 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 3.33: Interior of a set

Let $(X, \mathscr{T})$ be a topological space, $A \subseteq X$. The **interior** of A is

$$\operatorname{Int} A := \bigcup_{\substack{U \subseteq A \\ U \in \mathscr{T}}} U.$$

Remark 3.34

The definition of $\operatorname{Int} A$ is well-posed, since $\emptyset \subseteq A$ and $\emptyset \in \mathscr{T}$. Therefore the union is taken over a non-empty family.

Proposition 3.35: Int A is the largest open set contained in A

Let $(X, \mathscr{T})$ be a topological space, $A \subseteq X$. 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

1. $\operatorname{Int} A$ is open, since it is union of open sets, see property (A2).
2. $\operatorname{Int} A \subseteq A$, since $\operatorname{Int} A$ is union of sets contained in A.
3. Suppose $V \in \mathcal{T}$ and $V \subseteq A$. Therefore
$$V \subseteq \bigcup_{\substack{U \subseteq A \\ U \in \mathcal{T}}} U = \operatorname{Int} A \,.$$
4. Suppose that A is open. Then
$$A \subseteq \bigcup_{\substack{U \subseteq A \\ U \in \mathcal{T}}} U = \operatorname{Int} A \,.$$
As we already know that $\operatorname{Int} A \subseteq A$, we conclude that $A = \operatorname{Int} A$.
Conversely, suppose that $A = \operatorname{Int} A$. Since $\operatorname{Int} A$ is open, then also A is open.

Definition 3.36: Closure of a set

Let $(X, \mathcal{T})$ be a topological space, $A \subseteq X$. The **closure** of A is
$$\overline{A} := \bigcap_{\substack{A \subseteq C \\ C \text{ closed}}} C \,.$$

Remark 3.37

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 3.38: $\overline{A}$ is the smallest closed set containing A

Let $(X, \mathscr{T})$ be a topological space, $A \subseteq X$. 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

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 3.39

Let $(X, \mathscr{T})$ be a topological space, $A \subseteq X$. 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. Since $A \subseteq \overline{A}$, we have
$$A \cap U = A \cap (\overline{A})^c = \emptyset\,.$$
- Assume there exists $U \in \mathcal{T}$ 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 3.40: Boundary of a set

Let $(X, \mathcal{T})$ be a topological space, $A \subseteq X$. The **boundary** of A is
$$\partial A := \overline{A} \smallsetminus \operatorname{Int} A\,.$$

Proposition 3.41

Let $(X, \mathcal{T})$ be a topological space, and $A \subseteq X$. Then ∂A is closed.

Proof

We can write
$$\partial A = \overline{A} \smallsetminus \operatorname{Int} A = \overline{A} \cap (\operatorname{Int} A)^c\,.$$
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 3.42: Set of limit points

Let $(X, \mathcal{T})$ be a topological space, $A \subseteq X$. 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 3.43

Let $(X, \mathcal{T})$ be a topological space, $A \subseteq X$ a set. Let $\{x_n\} \subseteq A$ and $x_0 \in X$ be such that $x_n \to x_0$. Then $x_0 \in \overline{A}$. In particular,

$$L(A) \subseteq \overline{A}\,.$$

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 3.43 is false in general, that is,

   $$\overline{A} \not\subseteq L(A)\,.$$

   We show a counterexample in Example 3.44.

2. The relation

   $$\overline{A} = L(A)\,.$$

   holds in the so-called first countable topological spaces, such as metric spaces, see Proposition 3.45 below.

Example 3.44: Co-countable topology on $\mathbb{R}$

Question. The **co-countable** topology on $\mathbb{R}$ is the collection of sets

$$\mathcal{T}_{\rm cc} := \{A \subseteq \mathbb{R} \,:\; A^c = \mathbb{R} \;\text{ or }\; A^c \;\text{ countable}\,\}\,.$$

1. Prove that $\mathcal{T}_{\rm cc}$ is a topology on $\mathbb{R}$.
2. Prove that a sequence $\{x_n\}$ is convergent in $\mathcal{T}_{\rm cc}$ if and only if it is eventually constant.
3. Define the set $A = (-\infty, 0]$. Prove that $\overline{A} = \mathbb{R}$.

4. Conclude that $\overline{A} \not\subset L(A)$.

Solution.

1. See Question 2 in Homework 3.

2. See Question 2 in Homework 3.

3. Assume C is a closed set such that $A \subseteq C$. Since C is closed, it follows that $C^c \in \mathcal{T}_{cc}$. Therefore $(C^c)^c = C$ is either countable, or equal to $\mathbb{R}$. As $A \subseteq C$, we have that C is uncountable. Therefore, $C = \mathbb{R}$. As C is an arbitrary closed set containing A, we conclude that
$$\overline{A} = \bigcap_{\substack{A \subseteq C \\ C \text{ closed}}} C = \bigcap_{\substack{A \subseteq C \\ C \text{ closed}}} \mathbb{R} = \mathbb{R}.$$

4. By Point 2, convergent sequences are eventually constant. Therefore, if $\{x_n\} \subseteq A$ converges to x_0, we conclude that $x_0 \in A$. This shows
$$L(A) = A = [-\infty, 0).$$
By Point 3, we have $\overline{A} = \mathbb{R}$. We conclude that $\overline{A} \not\subset L(A)$.

In metric spaces we can characterize the interior of a set and the closure in the following way.

Proposition 3.45: Characterization of Int A and $\overline{A}$ in metric space

Let (X, d) be a metric space. Denote by $\mathcal{T}_d$ the topology induced by d. For any $A \subseteq X$, we have

1. $\operatorname{Int} A = \{x \in A \,:\, \exists\, r > 0 \text{ s.t. } B_r(x) \subseteq A\}$,

2. $\overline{A} = L(A) = \{x \in X \text{ s.t. } \exists\, \{x_n\} \subseteq A \text{ s.t. } x_n \to x\}$.

Proof

1. See Question 4 in Homework 3.

2. The inclusion $L(A) \subseteq \overline{A}$ holds by Proposition 3.43. We are left to show that
$$\overline{A} \subseteq L(A).$$
To this end, let $x_0 \in \overline{A}$. For $n \in \mathbb{N}$, consider the ball $B_{1/n}(x_0)$. Since $B_{1/n}(x_0) \in \mathcal{T}_d$ and $x_0 \in B_\varepsilon(x_0)$, we can apply Lemma 3.39 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

$$0 \leq d(x_n, x_0) < \frac{1}{n} \to 0$$

as $n \to \infty$. Thus $x_n \to x_0$, showing that $x_0 \in L(A)$.

Example 3.46

Question. Consider $\mathbb{R}$ equipped with the Euclidean topology. Let $A = [0, 1)$. Prove that:

$$\operatorname{Int} A = (0, 1)\,, \quad \overline{A} = [0, 1]\,, \quad \partial A = \{0, 1\}\,.$$

Note: In particular, this shows

$$\operatorname{Int} A \neq A\,, \quad \overline{A} \neq A\,,$$

so that A is neither open, nor closed.
Solution. See Question 5 in Homework 3.

3.6. Density

Definition 3.47: Density

Let $(X, \mathscr{T})$ be a topological space. We say that a subset $A \subseteq X$ 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 3.48: Characterization of density

Let $(X, \mathscr{T})$ be a topological space, $A \subseteq X$. 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 3.49: $\mathbb{Q}$ is dense in $(\mathbb{R}, \mathcal{T}_{\text{euclid}})$

Question. Consider $\mathbb{R}$ equipped with the Euclidean topology $\mathcal{T}_{\text{euclid}}$.

1. Prove that $\mathbb{Q}$ is dense in $\mathbb{R}$, that is, $\overline{\mathbb{Q}} = \mathbb{R}$.
2. Prove that $\operatorname{Int} Q = \emptyset$.

Note: This shows $\mathbb{Q}$ is neither open, nor closed, since

$$\operatorname{Int} \mathbb{Q} \neq \mathbb{Q} , \quad \overline{\mathbb{Q}} \neq \mathbb{Q} .$$

Solution. To solve the exercise we will need the following well-known analysis result:
Density Thoerem. Let $x \in \mathbb{R}$ and $\varepsilon > 0$. There exists $q \in \mathbb{Q}$ such that

$$|x - q| < \varepsilon .$$

1. We want to prove that $\mathbb{Q}$ is dense in $\mathbb{R}$ according to the topological definition. Therefore, let U be a non-empty open set in $\mathbb{R}$. Let $x \in U$. Since U is open, there exists $r > 0$ such that
$$(x - r, x + r) \subseteq U .$$
By the **Density Theorem**, there exists $q \in \mathbb{R}$ such that
$$|x - q| < r .$$
In particular, this shows $q \in (x - r, x + r)$, so that $q \in U$. Therefore $\mathbb{Q} \cap U \neq \emptyset$, proving that $\mathbb{Q}$ is dense in $\mathbb{R}$. In particular, by Proposition 3.48, we conclude that $\overline{\mathbb{Q}} = \mathbb{R}$.

2. Recall that $(\mathbb{R}, \mathcal{T}_{\text{euclid}})$ is a metric space. Therefore, we can apply Proposition 3.45, and infer that

$$\operatorname{Int} \mathbb{Q} = \{q \in \mathbb{Q} \,:\, \exists\, r > 0 \text{ s.t. } (q-r, q+r) \subseteq \mathbb{Q}\} \tag{3.7}$$

Assume by contradiction that $\operatorname{Int} \mathbb{Q} \neq \emptyset$. Let $q \in \operatorname{Int} \mathbb{Q}$. By (3.7), there exists $r > 0$ such that

$$(q-r, q+r) \subseteq \mathbb{Q}\,.$$

But $(q-r, q+r)$ is an interval of $\mathbb{R}$, and therefore it will contain an irrational number $x \in \mathbb{R} \setminus \mathbb{Q}$. Contradiction. Hence, $\operatorname{Int} \mathbb{Q} = \emptyset$.

Example 3.50: $\mathbb{Z}$ is not dense in $(\mathbb{R}, \mathcal{T}_{\text{euclid}})$

Question. Consider $\mathbb{R}$ equipped with the Euclidean topology $\mathcal{T}_{\text{euclid}}$. Prove that the set of integers $\mathbb{Z}$ is not dense in $\mathbb{R}$, with

$$\overline{\mathbb{Z}} = \mathbb{Z}\,.$$

Solution. The set of integers $\mathbb{Z}$ satisfies

$$\mathbb{Z}^c = \bigcup_{z \in \mathbb{Z}} (z, z+1)\,.$$

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

$$\overline{\mathbb{Z}} = \mathbb{Z}\,.$$

As $\overline{\mathbb{Z}} \neq \mathbb{R}$, by Proposition 3.48, we have that $\mathbb{Z}$ is not dense in $\mathbb{R}$.

If we change topologies, the closure might change.

Example 3.51: $\mathbb{Z}$ is dense in $(\mathbb{R}, \mathcal{T}_{\text{cofinite}})$

Question. Consider $\mathbb{R}$ equipped with the cofinite topology

$$\mathcal{T}_{\text{cofinite}} = \{U \subset \mathbb{R} \,:\, U^c \text{ is finite, or } U^c = \mathbb{R}\}\,.$$

Prove that $\mathbb{Z}$ is dense in $\mathbb{R}$. In particular,

$$\overline{\mathbb{Z}} = \mathbb{R}\,,$$

Solution. Suppose C is a closed set such that $\mathbb{Z} \subseteq C$. By definition of $\mathcal{T}_{\text{cofinite}}$, we have that $(C^c)^c = C$ is either finite, or it coincides with $\mathbb{R}$. Since $\mathbb{Z} \subseteq C$, and $\mathbb{Z}$ is not finite, we conclude $C = \mathbb{R}$. As C is an arbitrary closed set containing $\mathbb{Z}$, we conclude that

$$\overline{\mathbb{Z}} = \bigcap_{\substack{\mathbb{Z} \subseteq C \\ C \text{ closed}}} C = \bigcap_{\substack{\mathbb{Z} \subseteq C \\ C \text{ closed}}} \mathbb{R} = \mathbb{R}\,.$$

In particular, by Proposition 3.48, $\mathbb{Z}$ is dense in $\mathbb{R}$.

3.7. Hausdorff spaces

Hausdorff space are topological spaces in which points can be separated by means of disjoint open sets.

Definition 3.52: Hausdorff space

We say that a topological space $(X, \mathcal{T})$ is **Hausdorff** if for every $x, y \in X$ 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 3.53

Let (X, d) be a metric space, $\mathcal{T}_d$ the topology induced by d. Then $(X, \mathcal{T}_d)$ is a Hausdorff space.

Proof

Let $x, y \in X$ with $x \neq y$. Define

$$U := B_\varepsilon(x)\,, \quad V := B_\varepsilon(y)\,, \quad \varepsilon := \frac{1}{2}\, d(x,y)\,.$$

By Proposition 3.28, 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 3.54: 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 3.55

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 3.53 we know that $(X, \mathcal{T}_d)$ is Hausdorff. Hence $(X, \mathcal{T})$ is Hausdorff.

As a conseuqence of Corollary 3.55 we have that spaces which are not metrizable are not Hausdorff. Let us make a few examples.

Example 3.56: $(X, \mathcal{T}_{\text{trivial}})$ is not Hausdorff

Question. Let X be equipped with the trivial topology $\mathcal{T}_{\text{trivial}}$. Then X is not Hausdorff.
Solution. Assume by contradiction $(X, \mathcal{T}_{\text{trivial}})$ is Hausdorff and let $x, y \in X$ with $x \neq y$. Then, there exist $U, V \in \mathcal{T}_{\text{trivial}}$ such that

$$x \in U , \quad y \in V , \quad U \cap V = \emptyset .$$

In particular $U \neq \emptyset$ and $V \neq \emptyset$. Since $\mathcal{T} = \{\emptyset, X\}$, we conclude that

$$U = V = X \quad \implies \quad U \cap V = X \neq \emptyset .$$

This is a contradiction, and thus $(X, \mathcal{T}_{\text{trivial}})$ is not Hausdorff.

Example 3.57: $(\mathbb{R}, \mathcal{T}_{\text{cofinite}})$ is not Hausdorff

Question. Consider the cofinite topology on $\mathbb{R}$

$$\mathcal{T}_{\text{cofinite}} = \{U \subseteq \mathbb{R} \, : \, U^c \text{ is finite, or } U^c = \mathbb{R}\} .$$

Prove that $(\mathbb{R}, \mathcal{T}_{\text{cofinite}})$ is not Hausdorff.
Solution. Assume by contradiction $(\mathbb{R}, \mathcal{T}_{\text{cofinite}})$ is Hausdorff and let $x, y \in \mathbb{R}$ with $x \neq y$. Then,

there exist $U, V \in \mathcal{T}_{\text{cofinite}}$ such that

$$x \in U\,, \quad y \in V\,, \quad U \cap V = \emptyset\,.$$

Taking the complement of $U \cap V = \emptyset$, we infer

$$\mathbb{R} = (U \cap V)^c = U^c \cup V^c\,. \tag{3.8}$$

There are two possibilities:

1. U^c and V^c are finite. Then $U^c \cup V^c$ is finite, so that (3.8) is a contradiction.
2. Either $U^c = \mathbb{R}$ or $U^c = \mathbb{R}$. If $U^c = \mathbb{R}$, then $U = \emptyset$. This is a contradiction, since $x \in U$. If $V^c = \mathbb{R}$, then $V = \emptyset$. This is a contradiction, since $y \in V$.

Hence $(\mathbb{R}, \mathcal{T}_{\text{cofinite}})$ is not Hausdorff.

Example 3.58: Lower-limit topology on $\mathbb{R}$ is not Hausdorff

Question. The **lower-limit topology** on $\mathbb{R}$ is the collection of sets

$$\mathcal{T}_{\text{LL}} = \{\emptyset, \mathbb{R}\} \cup \{(a, +\infty) \ : \ a \in \mathbb{R}\}\,.$$

1. Prove that $(\mathbb{R}, \mathcal{T}_{\text{LL}})$ is a topological space.
2. Prove that $(\mathbb{R}, \mathcal{T}_{\text{LL}})$ is not Hausdorff.

Solution. Part 1. We show that $(\mathbb{R}, \mathcal{T}_{\text{LL}})$ is a topological space by verifying the axioms:
(A1) By definition $\emptyset, \mathbb{R} \in \mathcal{T}_{\text{LL}}$.
(A2) Let $A_i \in \mathcal{T}_{\text{LL}}$ for all $i \in I$. We have 2 cases:

- If $A_i = \emptyset$ for all i, then $\cup_i A_i = \emptyset \in \mathcal{T}_{\text{LL}}$.
- At least one of the sets A_i is non-empty. As empty-sets do not contribute to the union, we can discard them. Therefore, $A_i = (-\infty, a_i)$ with $a_i \in \mathbb{R} \cup \{\infty\}$. Define:

  $$a := \sup_{i \in I} a_i, \quad A := (-\infty, a).$$

  Then $A \in \mathcal{T}$ and:

  $$A = \cup_{i \in I} A_i.$$

  To prove this, let $x \in A$. Then $x < a$, so there exists $i_0 \in I$ such that $x < a_{i_0}$. Thus, $x \in A_{i_0}$, showing $A \subseteq \cup_{i \in I} A_i$. Conversely, if $x \in \cup_{i \in I} A_i$, then $x \in A_{i_0}$ for some $i_0 \in I$, implying $x < a_{i_0} \leq a$. Thus, $x \in A$, proving $\cup_{i \in I} A_i \subseteq A$.

(A3) Let $A, B \in \mathcal{T}_{\text{LL}}$. We have 3 cases:

- $A = \emptyset$ or $B = \emptyset$. Then $A \cap B = \emptyset \in \mathcal{T}_{\text{LL}}$.
- $A \neq \emptyset$ and $B \neq \emptyset$. Therefore, $A = (-\infty, a)$ and $B = (-\infty, b)$ with $a, b \in \mathbb{R} \cup \{\infty\}$. Define

$$U := A \cap B, \quad z := \min\{a, b\}.$$

Then $U = (-\infty, z) \in \mathcal{T}_{\text{LL}}$.

Thus, $(\mathbb{R}, \mathcal{T}_{\text{LL}})$ is a topological space.
Part 2. To show $(\mathbb{R}, \mathcal{T}_{\text{LL}})$ is not Hausdorff, assume otherwise. Let $x, y \in \mathbb{R}$ with $x \neq y$. Then there exist $U, V \in \mathcal{T}_{\text{LL}}$ such that:

$$x \in U, \quad y \in V, \quad U \cap V = \emptyset.$$

As U, V are non-empty, by definition of $\mathcal{T}_{\text{LL}}$, there exist $a, b \in \mathbb{R} \cup \{\infty\}$ such that $U = (-\infty, a)$ and $V = (-\infty, b)$. Define:

$$z := \min\{a, b\}, \quad Z := U \cap V = (-\infty, z).$$

Hence $Z \neq \emptyset$, contradicting $U \cap V = \emptyset$. Thus, $(\mathbb{R}, \mathcal{T}_{\text{LL}})$ is not Hausdorff.

In Hausdorff spaces the limit of sequences is unique.

Proposition 3.59: Uniqueness of limit in Hausdorff spaces

Let $(X, \mathcal{T})$ be a Hausdorff space. If a sequence $\{x_n\} \subseteq X$ converges, then the limit is unique.

Proof

Let $\{x_n\} \subseteq X$ be convergent, and suppose by contradiction that

$$x_n \to x_0\,, \quad x_n \to y_0\,, \quad x_0 \neq y_0\,.$$

Since X is Hausdorff, there exist $U, V \in \mathcal{T}$ such that

$$x_0 \in U\,, \quad y_0 \in V\,, \quad U \cap V = \emptyset\,.$$

As $x_n \to x_0$ and $U \in \mathcal{T}$ 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 $V \in \mathcal{T}$ 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 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 3.60: Images and Pre-images

Let X, Y be sets and $f : X \to Y$ be a function.

1. Let $U \subseteq X$. The image of U under f is the subset of Y defined by

$$f(U) := \{y \in Y \, : \, \exists \, x \in X \text{ s.t. } y = f(x)\} = \{f(x) \, : \, x \in X\} .$$

2. 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 3.61

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 $g \colon Y \to Z$. 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 3.62: Continuous function

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Let $f \colon X \to Y$ be a function.

1. Let $x_0 \in X$. We say that f is continuous at x_0 if it holds:

$$\forall \, V \in \mathcal{T}_Y \;\text{ s.t. }\; f(x_0) \in V \,, \;\; \exists U \in \mathcal{T}_X \;\text{ s.t. }\; x_0 \in U \,, \;\; f(U) \subseteq V \,.$$

2. 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 3.63

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Let $f \colon X \to Y$ be a function. They are equivalent:

1. f is continuous from $(X, \mathcal{T}_X)$ to $(Y, \mathcal{T}_Y)$.
2. It holds: $f^{-1}(V) \in \mathcal{T}_X$ for all $V \in \mathcal{T}_Y$.

Important

In other words, a function $f \colon X \to Y$ is continuous if and only if the pre-image of open sets in Y are open sets in X.

The proof of Proposition 3.63 is simple, but very tedious. We choose to skip it.

Example 3.64

Question. Let X be a set and $\mathcal{T}_1$, $\mathcal{T}_2$ be topologies on X. Define the identity map

$$\mathrm{Id}_X : (X, \mathcal{T}_1) \to (X, \mathcal{T}_2), \quad \mathrm{Id}_X(x) := x .$$

Prove that they are equivalent:

1. Id_X is continuous from $(X, \mathcal{T}_1)$ to $(X, \mathcal{T}_2)$.
2. $\mathcal{T}_1$ is finer than $\mathcal{T}_2$, that is, $\mathcal{T}_2 \subseteq \mathcal{T}_1$.

Solution. Id_X is continuous if and only if

$$\mathrm{Id}_X^{-1}(V) \in \mathcal{T}_1 , \quad \forall \, V \in \mathcal{T}_2 .$$

But $\mathrm{Id}_X^{-1}(V) = V$, so that the above reads

$$V \in \mathcal{T}_1 , \quad \forall \, V \in \mathcal{T}_2 ,$$

which is equivalent to $\mathcal{T}_2 \subseteq \mathcal{T}_1$.

Let us compare our new definition of contiuity with the classical notion of continuity in $\mathbb{R}^n$. Let us recall the definition of continuous function in $\mathbb{R}^n$.

Definition 3.65: 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 3.66

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 3.67

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 $f : X \to Y$ 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 3.68

Question. 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,
$$\mathcal{T}_Y = \{\emptyset, Y\} \,.$$
Prove that every function $f : X \to Y$ is continuous.
Solution. f is continuous if $f^{-1}(V) \in \mathcal{T}_X$ for all $V \in \mathcal{T}_Y$. 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 3.69

Question. Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Suppose that $\mathcal{T}_Y$ is the discrete topology, that is,
$$\mathcal{T}_Y = \{V \, \text{ s.t. } \, V \subseteq Y\} \,.$$
Let $f : X \to Y$. Prove that they are equivalent:

1. f is continuous from X to Y.
2. $f^{-1}(\{y\}) \in \mathcal{T}_X$ for all $y \in Y$.

Solution. 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 $V =$

$\cup_{y \in V} \{y\}$. 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.

It is useful to introduce the notion of sequential continuity.

Definition 3.70: Sequential continuity

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces, and $f : X \to Y$. We say that f is **sequentially continuous** if the following condition holds:

$$\{x_n\} \subset X \,, \;\; x_n \to x_0 \;\; \text{in} \;\; X \quad \Longrightarrow \quad f(x_n) \to f(x_0) \;\; \text{in} \;\; Y \,.$$

In other words, f is sequentially continuous if it takes convergent sequences into convergent sequences. In any topological space, continuity implies sequential continuity, as proven in the next Proposition.

Proposition 3.71

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Let $f : X \to Y$ be continuous. Then f is sequentially continuous.

Proof

Let $\{x_n\} \subset X$ and suppose that $x_n \to x_0$ in the topology $\mathcal{T}_X$. We need to prove that

$$f(x_n) \to f(x_0), \;\; \text{in} \;\; Y \,.$$

To this end, 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

1. The reverse implication of Proposition 3.71 is false:

$$\text{sequential continuity} \quad \not\Longrightarrow \quad \text{continuity}$$

   A counterexample is given in Example 3.73 below.

2. Continuity is equivalent to sequential continuity if the topologies on X and Y are first countable. This is the case for metrizable topologies, see Proposition 3.72 below.

Proposition 3.72

Let (X, d_X) and (Y, d_Y) be metric spaces. They are equivalent:

1. f is continuous.
2. f is sequentially continuous.

Proof

Part 1. We have already proven that continuity implies sequential continuity in any topological space.
Part 2. Assume f is sequentially continuous. 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 3.73: Sequential continuity does not imply continuity

Question. Consider the co-countable and discrete topologies on $\mathbb{R}$

$$\mathcal{T}_{\text{cc}} = \{A \subseteq \mathbb{R} \, : \, A^c = \mathbb{R} \text{ or } A^c \text{ countable}\}$$
$$\mathcal{T}_{\text{discrete}} = \{A \subseteq \mathbb{R}\}$$

Consider the identity function

$$f \colon (\mathbb{R}, \mathcal{T}_{\text{cc}}) \to (\mathbb{R}, \mathcal{T}_{\text{discrete}}) \,, \quad f(x) := x \,.$$

Prove that

1. f is not continuous.
2. f is sequentially continuous.

Hint: You can use the following fact: Sequences in $(\mathbb{R}, \mathcal{T}_{\text{cc}})$ *and* $(\mathbb{R}, \mathcal{T}_{\text{discrete}})$ *converge if and only if they are eventually constant.*

Solution.

1. We have $\{x\} \in \mathcal{T}_{\text{discrete}}$. However,

$$f^{-1}(\{x\}) = \{x\} \notin \mathcal{T}_{\text{cc}} \,,$$

since $\{x\}^c = \mathbb{R} \smallsetminus \{x\}$ is neither equal to $\mathbb{R}$, nor countable. Therefore f is not continuous.

2. Assume that $\{x_n\}$ is convergent in $\mathcal{T}_{\text{cc}}$. By the Hint, we have that $\{x_n\}$ is eventually constant. Again by the Hint, we infer that $\{x_n\}$ is convergent in $\mathcal{T}_{\text{discrete}}$. Since $f(x_n) = x_n$, we conclude that f is sequentially continuous.

Let us make an observation on continuity of compositions.

Proposition 3.74: Continuity of compositions

Let $(X, \mathcal{T}_X), (Y, \mathcal{T}_Y), (Z, \mathcal{T}_Z)$ be topological spaces. Assume $f \colon X \to Y$ and $g \colon Y \to Z$ are continuous. Then $(g \circ f) \colon X \to Z$ is continuous.

Proof

Let $C \in \mathcal{T}_Z$. As g is continuous, we have that

$$g^{-1}(C) \in \mathcal{T}_Y \,.$$

Since f is continuous, we also have

$$f^{-1}(g^{-1}(C)) \in \mathcal{T}_X \,.$$

Therefore

$$(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C)) \in \mathcal{T}_X \,,$$

so that $g \circ f$ is continuous.

We conclude the section by introducing homeomorphisms.

Definition 3.75: Homeomorphism

Let $(X, \mathcal{T}_X), (Y, \mathcal{T}_Y)$ be topological space. A function $f \colon X \to Y$ is called an **homeomorphism** if they hold:

1. f is continuous.
2. f admits continuous inverse $f^{-1} \colon Y \to X$.

Therefore f is a homeomorphism if it is continuous and admits a continuous inverse. Homeomorphisms are used to study similarities between topological spaces: When 2 topological spaces are homeomorphic, they can be essentially considered to be the same space.

3.9. Subspace topology

Any subset Y in a topological space X naturally inherits a topological structure. Such structure is called **subspace topology**.

Definition 3.76: Subspace topology

Let $(X, \mathcal{T})$ be a topological space and $Y \subseteq X$ a subset. Define the family of sets

$$\begin{aligned} \mathcal{S} &:= \{A \subseteq Y \, : \, \exists \, U \in \mathcal{T} \text{ s.t. } A = U \cap Y\} \\ &= \{U \cap Y , \; U \in \mathcal{T}\} \,. \end{aligned}$$

The family $\mathcal{S}$ is called **subspace topology** on Y induced by the inclusion $Y \subseteq X$.

Proof: Well-posedness of Definition 3.76

We have to show that $(Y, \mathcal{S})$ is a topological space:

- (A1) $\emptyset \in \mathcal{S}$ since $\emptyset = \emptyset \cap Y$ and $\emptyset \in \mathcal{T}$. Similarly, we have $Y \in \mathcal{S}$, since $Y = X \cap Y$ and $X \in \mathcal{T}$.
- (A2) Let $A_i \in \mathcal{S}$ for $i \in I$. By definition there exist $U_i \in \mathcal{T}$ such that $A_i = U_i \cap Y$ for all $i \in I$. Therefore
$$\cup_{i \in I} A_i = \cup_{i \in I}(U_i \cap Y) = (\cup_{i \in I} U_i) \cap Y \,.$$
The above proves that $\cup_{i \in I} A_i \in \mathcal{S}$, since $\cup_{i \in I} U_i \in \mathcal{T}$.
- (A3) Let $A_1, A_2 \in \mathcal{S}$. By definition there exist $U_1, U_2 \in \mathcal{T}$ such that $A_1 = U_1 \cap Y$ and $A_2 = U_2 \cap Y$. 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, the subspace topology coincides with the original topology, as see in the next Proposition.

Proposition 3.77

Let $(X, \mathcal{T})$ be a topological space and $Y \in \mathcal{T}$. Let $A \subseteq Y$. 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, $A \subseteq Y \subseteq X$. In general we could have

$$A \in \mathcal{S} \quad \text{and} \quad A \notin \mathcal{T}\,.$$

Example. Let $X = \mathbb{R}$ with $\mathcal{T}_{\text{euclid}}$. Consider the subset $Y = [0, 2)$, and equip Y with the subspace topology $\mathcal{S}$. Let $A = [0, 1)$. Then $A \notin \mathcal{T}_{\text{euclid}}$ but $A \in \mathcal{S}$, since

$$A = (-1, 1) \cap Y\,, \qquad (-1, 1) \in \mathcal{T}_{\text{euclid}}\,.$$

Example 3.78

Question. Let $X = \mathbb{R}$ be equipped with $\mathcal{T}_{\text{euclid}}$. Let $\mathcal{S}$ be the subspace topology on $\mathbb{Z}$. Prove that

$$\mathcal{S} = \mathcal{T}_{\text{discrete}}\,.$$

Solution. To prove that $\mathcal{S} = \mathcal{T}_{\text{discrete}}$, we need to show that all the subsets of $\mathbb{Z}$ are open in $\mathcal{S}$.

1. Let $z \in \mathbb{Z}$ be arbitrary. Notice that
$$\{z\} = (z - 1, z + 1) \cap \mathbb{Z}$$
and $(z - 1, z + 1) \in \mathcal{T}_{\text{euclid}}$. Thus $\{z\} \in \mathcal{S}$.
2. Let now $A \subseteq \mathbb{Z}$ be an arbitrary subset. Trivially,
$$A = \cup_{z \in A} \{z\}\,.$$
As $\{z\} \in \mathcal{S}$, we infer that $A \in \mathcal{S}$ by (A2).

3.10. Topological basis

We have seen that in metric spaces every open set is union of open balls, see Propostion 3.28. We can then regard the open balls as the 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 3.79: 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 $\{B_i\}_{i \in I} \subseteq \mathcal{B}$, with I family of indices, such

that
$$U = \bigcup_{i \in I} B_i \,. \tag{3.12}$$

Example 3.80

Question. Prove the following statements.

1. Let $(X, \mathscr{T})$ be a topological space. Then $\mathscr{B} := \mathscr{T}$ is a basis for $\mathscr{T}$.
2. Let (X, d) be a metric space with topology $\mathscr{T}_d$ induced by the metric. Then
$$\mathscr{B} := \{B_r(x) \, : \, x \in X \, , \; r > 0\}$$
is a basis for $\mathscr{T}_d$.
3. Let X be equipped with $\mathscr{T}_{\text{discrete}}$. Then
$$\mathscr{B} := \{\{x\} \, : \, x \in X\}$$
is a basis for $\mathscr{T}_{\text{discrete}}$.

Solution.

1. This is true because one can just take $B = U$ in (3.12).
2. This is true by Propostion 3.28.
3. This is true because for any $U \in \mathscr{T}$ we have
$$U = \bigcup_{x \in U} \{x\} \,.$$

Example 3.81

Question. Consider $\mathbb{R}$ equipped with the Euclidean topology $\mathscr{T}_{\text{euclid}}$. Which of the following collection of sets are basis for $\mathscr{T}_{\text{euclid}}$? Motivate your answer.

1. $\mathscr{B}_1 = \{(a, b) \, : \, a, b \in \mathbb{R}\}$.
2. $\mathscr{B}_2 = \{(a, b) \, : \, a, b \in \mathbb{Q}\}$.
3. $\mathscr{B}_3 = \{(a, b) \, : \, a, b \in \mathbb{Z}\}$.

Solution.

1. $\mathscr{B}_1$ is a basis for $\mathscr{T}_{\text{euclid}}$, for the following reason. Let d be the Euclidean distance on $\mathbb{R}$,

and $\mathscr{T}_d$ the topology induced by d. We know that $\mathscr{T}_d = \mathscr{T}_{\text{euclid}}$, therefore

$$\mathscr{B} := \{B_r(x) \, : \, x \in \mathbb{R} \, , \, r > 0\}$$

is a basis for $\mathscr{T}_{\text{euclid}}$ by Propostion 3.28. Note that balls in $\mathbb{R}$ are just open intervals

$$B_r(x) = (x - r, x + r) \, .$$

Hence $\mathscr{B}_1 = \mathscr{B}$, so that $\mathscr{B}_1$ is a basis for $\mathscr{T}_{\text{euclid}}$.

2. $\mathscr{B}_2$ is a basis for $\mathscr{T}_{\text{euclid}}$. This is because any open interval (a, b) with $a, b \in \mathbb{R}$ can be written as

$$\bigcup_{q,r \in \mathbb{Q}, \, a<q, \, s<b} (q, s) = (a, b) \, .$$

Therefore, since $\mathscr{B}_1$ is a basis for $\mathscr{T}_{\text{euclid}}$, we conclude that also $\mathscr{B}_2$ is a basis for $\mathscr{T}_{\text{euclid}}$.

3. $\mathscr{B}_3$ is not a basis for $\mathscr{T}_{\text{euclid}}$. Indeed, consider $U = (0, 1/2)$, which is open in $\mathscr{T}_{\text{euclid}}$. It is clear that U cannot be obtained as the union of intervals (q, s) with $q, s \in \mathbb{Z}$.

Proposition 3.82

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 3.82 are sufficient for generating a topology.

Proposition 3.83

Let X be a set, and $\mathscr{B}$ a collection of subsets of X satisfying (B1)-(B2). Define

$$\mathscr{T} := \left\{ U \subseteq X \, : \, U = \bigcup_{i \in I} B_i \, , \;\; B_i \in \mathscr{B} \right\} .$$

Then $\mathscr{T}$ is a topology on X, with basis given by $\mathscr{B}$.

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

$$U_i = \bigcup_{k \in K_i} B_k^i$$

for some family of indices K_i and $B_k^i \in \mathscr{B}$. Therefore

$$U := \bigcup_{i \in I} U_i = \bigcup_{i \in I, \, k \in K_i} B_k^i \, ,$$

showing that $U \in \mathscr{T}$.

- (A3) Suppose that $U_1, U_2 \in \mathscr{T}$. Then

$$U_1 = \bigcup_{i \in I_1} B_i^1 \, , \quad U_2 = \bigcup_{i \in I_2} B_i^2$$

for $B_i^1, B_i^2 \in \mathscr{B}$. From the above we have

$$U_1 \cap U_2 = \bigcup_{i \in I_1, \, k \in I_2} B_i^1 \cap B_k^2 \, .$$

From property (B2) we have that for each pair of indices (i,k) the set $B_i^1 \cap B_k^2$ is the union of sets in $\mathscr{B}$. 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 $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ we would like to equip the cartesian product

$$X \times Y = \{(x, y) \,:\, x \in X\,,\ y \in Y\}$$

with a topology. We proceed as follows.

Proposition 3.84

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. Define the family $\mathcal{B}$ of subsets of $X \times Y$ as

$$\mathcal{B} := \{U \times V \,:\, U \in \mathcal{T}_X\,,\ V \in \mathcal{T}_Y\} \subseteq \mathcal{T}_X \times \mathcal{T}_Y\,.$$

Then $\mathcal{B}$ satisfies properties (B1) and (B2) from Proposition 3.82. In particular,

$$\mathcal{T}_{X\times Y} := \left\{ Z \,:\, Z = \bigcup_{i\in I} U_i \times V_i,\ U_i \times V_i \in \mathcal{B} \right\} \tag{3.13}$$

is a topology on $X \times Y$.

Proof

The proof that $\mathcal{B}$ satisfies (B1)-(B2) is an easy check, and is left as an exercise. As $\mathcal{B}$ satisfies (B1)-(B2), by Proposition 3.83 we know that $\mathcal{T}_{X\times Y}$ is a topology on $X \times Y$.

Definition 3.85: Product topology

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. The **product topology** on $X \times Y$ is the collection of sets $\mathcal{T}_{X\times Y}$ at (3.13).

Example 3.86

Let $\mathbb{R}$ be equipped with the (one dimensional) Euclidean topology. The product topology on $\mathbb{R}\times\mathbb{R}$ coincides with the topology on $\mathbb{R}^2$ equipped with the (two dimensional) Euclidean topology.

Definition 3.87: Projection maps

Given two sets X, Y we define the **projection maps** as

$$\pi_X : X \times Y \to X\,, \quad \pi_X(x, y) := x\,,$$
$$\pi_Y : X \times Y \to Y\,, \quad \pi_Y(x, y) := y\,.$$

Proposition 3.88: Projections π_X, π_Y are continuous for $\mathscr{T}_{X\times Y}$

Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be topological spaces and equip $X \times Y$ with the product topology $\mathscr{T}_{X\times Y}$. Then π_X and π_Y are continuous.

Proof

Let $U \in \mathscr{T}_X$. Then

$$\pi_X^{-1}(U) = U \times Y .$$

We have that $U \times Y \in \mathscr{T}_{X\times Y}$ since $U \in \mathscr{T}_X$ and $Y \in \mathscr{T}_Y$. Therefore π_X is continuous. The proof that π_Y is continuous is similar, and is left as an exercise.

The following proposition gives a useful criterion to check whether a map into $X\times Y$ is continuous.

Proposition 3.89

Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be topological spaces and equip $X \times Y$ with the product topology $\mathscr{T}_{X\times Y}$. Let $(Z, \mathscr{T}_Z)$ 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 $(X, \mathscr{T})$ is a topological space. By property (A1) we have that

$$\emptyset \,,\ X \in \mathscr{T}$$

Therefore

$$\emptyset^c = X \,, \quad X^c = \emptyset$$

are closed. It follows that $\emptyset$ and X are both open and closed.

Definition 3.90: Connected space

Let $(X, \mathcal{T})$ be a topological space. We say that:

1. X is **connected** if the only subsets of X which are both open and closed are $\emptyset$ and X.
2. 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 3.91: Proper subset

Let X be a set. A subset $A \subseteq X$ is **proper** if $A \neq \emptyset$ and $A \neq X$.

Proposition 3.92: 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 $A := U$, $B := U^c$. 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

$$A^c = X \smallsetminus A = B ,$$

showing that A^c is closed, and hence A is open. Therefore A is proper, open and closed, showing that X is disconnected.

In the following we will use Point 2 and Point 3 in Proposition 3.92 as equivalent definitions of disconnected topological space.

Example 3.93

Question. Consider the set $X = \{0, 1\}$ with the subspace topology induced by the inclusion $X \subseteq \mathbb{R}$, where $\mathbb{R}$ is equipped with the Euclidean topology $\mathcal{T}_{\text{euclid}}$. Prove that X is disconnected.
Solution. Note that

$$X = \{0\} \cup \{1\} , \quad \{0\} \cap \{1\} = \emptyset .$$

The set $\{0\}$ is open for the subspace topology, since

$$\{0\} = X \cap (-1, 1) , \quad (-1, 1) \in \mathcal{T}_{\text{euclid}} .$$

Similarly, also $\{1\}$ is open for the subspace topology, since

$$\{1\} = X \cap (0, 2) , \quad (0, 2) \in \mathcal{T}_{\text{euclid}} .$$

Since $\{0\}$ and $\{1\}$ are proper subsets of X, we conclude that X is disconnected.

Example 3.94

Question. Let $\mathbb{R}$ be equipped with $\mathcal{T}_{\text{euclid}}$, and let $p \in \mathbb{R}$. Prove that the set $X = \mathbb{R} \smallsetminus \{p\}$ is disconnected.
Solution. Define the sets

$$A = (-\infty, p) , \quad B = (p, \infty) .$$

A and B are proper subsets of X. Moreover

$$X = A \cup B , \quad A \cap B = \emptyset .$$

Finally, A, B are open for the subspace topology on X, since they are open in $(\mathbb{R}, \mathcal{T}_{\text{euclid}})$. Therefore X is disconnected.

As it is easy to imagine, the conclusion of 3.94 is false if the dimension is $n \geq 2$. In order to prove the statement rigorously, we need a technical lemma.

Lemma 3.95

Let $(X, \mathscr{T})$ 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:

$$A = (A \cap U) \cup (A \cap U^c)$$
$$\emptyset = (A \cap U) \cap (A \cap U^c)$$

Now, suppose by contradiction $A \nsubseteq U$. 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. Thus $A \subseteq U$.

Example 3.96

Question. Let $n \geq 2$, and $A \subseteq \mathbb{R}^n$ be open and connected. Let $\mathbf{p} \in A$. Prove that $X = A \smallsetminus \{\mathbf{p}\}$ is connected.
Solution. Assume that

$$X = U \cup V,$$

with U, V disjoint and open in X. If we show that U, V are not proper, we conclude that X is connected. In order to prove it, start by noting that $X = A \smallsetminus \{\mathbf{p}\}$ is open, since A is open. As the sets U, V are open for the subspace topology on X, and X is open in $\mathbb{R}^n$, we conclude that U, V are open in $\mathbb{R}^n$. As U, V are also closed for the subspace topology, we conclude that they are closed in $\mathbb{R}^n$. As A is open, and $\mathbf{p} \in A$, there exists $r > 0$ such that $B_r(\mathbf{p}) \subseteq A$. Since $X = A \smallsetminus \{\mathbf{p}\}$, we have

$$B_r(\mathbf{p}) \smallsetminus \{\mathbf{p}\} \subseteq X.$$

As $X = U \cup V$, we have

$$(B_r(\mathbf{p}) \smallsetminus \{\mathbf{p}\}) \cap U \neq \emptyset \quad \text{or} \quad (B_r(\mathbf{p}) \smallsetminus \{\mathbf{p}\}) \cap V \neq \emptyset.$$

Without loss of generality, assume that $(B_r(\mathbf{p}) \smallsetminus \{\mathbf{p}\}) \cap U \neq \emptyset$ (the argument is similar in the other case). Since $B_r(\mathbf{p}) \smallsetminus \{\mathbf{p}\}$ is open, and U is open and closed, by Lemma 3.95 we conclude that

$$B_r(\mathbf{p}) \smallsetminus \{\mathbf{p}\} \subseteq U \quad \Longrightarrow \quad B_r(\mathbf{p}) \subseteq U' := U \cup \{\mathbf{p}\}.$$

Let $\mathbf{q} \in U'$. We have two cases:

- $\mathbf{q} \neq \mathbf{p}$: Then $\mathbf{q} \in U$. As U is open, there exists $\varepsilon > 0$ such that

$$B_\varepsilon(\mathbf{q}) \subseteq U \subseteq U' .$$

- $\mathbf{q} = \mathbf{p}$: We have shown that $B_r(\mathbf{p}) \subseteq U'$.

This shows U' is open in $\mathbb{R}^n$. In conclusion, we have

$$X = U \cup V \quad \Longrightarrow \quad A = U' \cup V ,$$

with U', V disjoint and open in A (since they are open in $\mathbb{R}^n$). As $\mathbf{p} \in U'$, we conclude that $U' \neq \emptyset$. By connectedness of A, we must have $V = \emptyset$. This implies $U = X$. Therefore U and V are not proper, implying that X is connected.

The next theorem shows that connectedness is preserved by continuous maps.

Theorem 3.97

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

$$A = \emptyset \quad \text{or} \quad B = \emptyset \tag{3.15}$$

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

$$f^{-1}(\widetilde{A}) = \emptyset \quad \text{or} \quad f^{-1}(\widetilde{B}) = \emptyset \,.$$

The above implies

$$\widetilde{A} \cap f(X) = \emptyset \quad \text{or} \quad \widetilde{B} \cap f(X) = \emptyset \,.$$

Recalling (3.16), we obtain (3.15), ending the proof.

An immediate corollary of Theorem 3.97 is that connectedness is a topological invariant, e.g., connectedness is preserved by homeomorphisms.

Theorem 3.98: Connectedness is topological invariant

Let $(X, \mathcal{T}_X)$, $(Y, \mathcal{T}_Y)$ be homeomorhic topological spaces. Then

$$X \text{ is connected} \quad \iff \quad Y \text{ is connected}$$

The proof follows immediately by Theorem 3.97, and is left to the reader as an exercise.

Example 3.99

Question. Let $n \geq 2$. Prove that $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}$.
Solution. Suppose by contradiction that there exists a homeomorphism $f \colon \mathbb{R}^n \to \mathbb{R}$. Define $p = f(\mathbf{0})$ and the restriction

$$g \colon \; (\mathbb{R}^n \smallsetminus \{\mathbf{0}\}) \to (\mathbb{R} \smallsetminus \{p\}) \,, \quad g(x) = f(x) \,.$$

Note that g is a homeomorphism, being restriction of a homeomorphism. By Example 3.96, we have that $\mathbb{R}^n \smallsetminus \{\mathbf{0}\}$ is connected. Hence, by Theorem 3.98, we infer that $\mathbb{R} \smallsetminus \{p\}$ is connected. This is a contradiction, since $\mathbb{R} \smallsetminus \{p\}$ is disconnected, as shown in Example 3.94.

A stronger version of the statement in Example 3.99 holds.

Theorem 3.100: Topological invariance of dimension

Let $n \neq m$. Then $\mathbb{R}^n$ is not homeomorphic to $\mathbb{R}^m$.

Unfortunately, the argument in Example 3.99 cannot be extended to prove Theorem 3.100. 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 3.100 is called **homology**, which requires a serious effort to construct/define.

Let us give another example of spaces which are not homeomorphic.

Example 3.101

Question. Define the one dimensional unit circle

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

Prove that $\mathbb{S}^1$ and $[0, 1]$ are not homeomorphic.
Solution. Suppose by contradiction that there exists a homeomorphism

$$f \colon [0, 1] \to \mathbb{S}^1 .$$

The restriction of f to $[0, 1] \smallsetminus \{\frac{1}{2}\}$ defines a homeomorphism

$$g \, : \, \left([0, 1] \smallsetminus \left\{ \frac{1}{2} \right\} \right) \to \left(\mathbb{S}^1 \smallsetminus \{\mathbf{p}\} \right) , \quad \mathbf{p} := f\left(\frac{1}{2}\right) .$$

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 a homeomorphism, 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 3.97, we conclude that $\mathbb{S}^1 \smallsetminus \{\mathbf{p}\}$ is connected. Contradiction.

3.13. Intermediate Value Theorem

Another consequence of Theorem 3.97 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 3.102: 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 3.103: Intervals are connected

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 3.104: 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 3.97 we know that $f(X)$ is connected in $\mathbb{R}$. By Theorem 3.103 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

$$f(a) \leq c \leq f(b)$$

we conclude that $c \in f(X)$, since $f(X)$ is an interval. Hence there exists $\xi \in X$ such that $f(\xi) = c$.

There is an alternative proof to the fact that intervals are connected. It makes use of the classical Intermediate Value Theorem in $\mathbb{R}$. It is an interesting exercise.

Example 3.105: Intervals are connected - Alternative proof

Question. Prove the following statements.

1. Let $(X, \mathcal{T})$ be a disconnected topological space. Prove that there exists a function $f : X \to \{0, 1\}$ which is continuous and surjective.
2. Consider $\mathbb{R}$ equipped with the Euclidean topology. Let $I \subseteq \mathbb{R}$ be an interval. Use point (1), and the Intermediate Value Theorem in $\mathbb{R}$ (see statement below), to show that I is connected.

Intermediate Value Theorem in $\mathbb{R}$: Suppose that $f : [a, b] \to \mathbb{R}$ is continuous, and $f(a) < f(b)$. Let $c \in \mathbb{R}$ be such that $f(a) \leq c \leq f(b)$. Then, there exists $\xi \in [a, b]$ such that $f(\xi) = c$.
Solution. Part 1. Since X is disconnected, there exist $A, B \in \mathcal{T}$ proper and such that

$$X = A \cup B, \quad A \cap B = \emptyset .$$

Define $f : X \to \{0, 1\}$ by

$$f(x) = \begin{cases} 0 & \text{if } x \in A \\ 1 & \text{if } x \in B \end{cases}$$

Since A and B are non-empty, it follows that f is surjective. Moreover f is continuous: Indeed suppose $U \subseteq \mathbb{R}$ is open. We have 4 cases:

- $0, 1 \notin U$. Then $f^{-1}(U) = \emptyset \in \mathcal{T}$.
- $0 \in U$, $1 \notin U$. Then $f^{-1}(U) = A \in \mathcal{T}$.
- $0 \notin U$, $1 \in U$. Then $f^{-1}(U) = B \in \mathcal{T}$.
- $0, 1 \in U$. Then $f^{-1}(U) = X \in \mathcal{T}$.

Then $f^{-1}(U) \in \mathcal{T}$ for all $U \subseteq \mathbb{R}$ open, showing that f is continuous.
Part 2. Let $I \subseteq \mathbb{R}$ be an interval. Suppose by contradiction I is disconnected. By Point (1), there exists a map $f : I \to \{0, 1\}$ which is continuous and surjective. As f is surjective, there exist $a, b \in I$ such that

$$f(a) = 0 , \quad f(b) = 1 .$$

Since f is continuous, and $f(a) = 0 < 1 = f(b)$, by the *Intermediate Value Theorem in* $\mathbb{R}$, there exists $\xi \in [a, b]$ such that $f(\xi) = 1/2$. As I is an interval, $a, b \in I$, and $a \leq \xi \leq b$, it follows that $\xi \in I$. This is a contradiction, since f maps I into $\{0, 1\}$, and $f(\xi) = 1/2 \notin \{0, 1\}$. Therefore I is connected.

3.14. Path-connectedness

Definition 3.106: 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

$$\alpha : [a,b] \to X \quad \text{s.t.} \quad \alpha(a) = x \,, \quad \alpha(b) = y \,.$$

It turns out that path-connectedness implies connectedness.

Theorem 3.107: Path-connectedness implies connectedness

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 $\alpha(0) = x$ and $\alpha(1) = y$. 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.

Example 3.108

Question. Let $A \subseteq \mathbb{R}^n$ be convex. Show that A is path-connected, and hence connected.
Solution. 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 $\alpha(0) = x, \alpha(1) = y$.

Example 3.109: Spaces of matrices

Let $\mathbb{R}^{2\times 2}$ denote the space of real 2×2 matrices. Assume $\mathbb{R}^{2\times 2}$ has the euclidean topology obtained by identifying it with $\mathbb{R}^4$.

1. Consider the set of orthogonal matrices

$$\mathrm{O}(2) = \{A \in \mathbb{R}^{2\times 2} \ : \ A^T A = I\}.$$

   Prove that $\mathrm{O}(2)$ is disconnected.

2. Consider the set of rotations

$$\mathrm{SO}(2) = \{A \in \mathbb{R}^{2\times 2} \ : \ A^T A = I, \ \det(A) = 1\}.$$

   Prove that $\mathrm{SO}(2)$ is path-connected, and hence connected.

Solution. Let $A \in \mathrm{O}(2)$, and denote its entries by a, b, c, d. By direct calculation, the condition $A^T A = I$ is equivalent to

$$a^2 + b^2 = 1, \qquad b^2 + c^2 = 1, \qquad ac + bd = 0.$$

From the first condition, we get that $a = \cos(t)$ and $b = \sin(t)$, for a suitable $t \in [0, 2\pi)$. From the second and third conditions, we get $c = \pm\sin(t)$ and $d = \mp\cos(t)$. We decompose $\mathrm{O}(2)$ as

$$\begin{aligned} &\mathrm{O}(2) = A \cup B, \\ &A = \mathrm{SO}(2) = \left\{ \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}, t \in [0, 2\pi) \right\} \\ &B = \left\{ \begin{pmatrix} \cos(t) & \sin(t) \\ \sin(t) & -\cos(t) \end{pmatrix}, t \in [0, 2\pi) \right\}. \end{aligned}$$

1. The determinant function $\det : \mathrm{O}(2) \to \mathbb{R}$ is continuous. If $M \in A$, we have $\det(M) = 1$. If instead $M \in B$, we have $\det(M) = -1$. Moreover,

$$\det{}^{-1}(\{1\}) = A, \qquad \det{}^{-1}(\{0\}) = B.$$

   As det is continuous, and $\{0\}, \{1\}$ closed, we conclude that A and B are closed. Therefore, A and B are closed, proper and disjoint. Since $\mathrm{O}(2) = A \cup B$, we conclude that $\mathrm{O}(2)$ is disconnected.

2. Define the function $\psi : [0, 2\pi) \to \mathrm{SO}(2)$ by

$$\psi(t) = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} .$$

Clearly, ψ is continuous. Let $R, Q \in \mathrm{SO}(2)$. Then R is determined by an angle t_1, while Q by an angle t_2. Up to swapping R and Q, we can assume $t_1 < t_2$. Define the function $f : [0, 1] \to \mathrm{SO}(2)$ by

$$f(\lambda) = \psi(t_1(1 - \lambda) + t_2\lambda) .$$

Then, f is continuous and

$$f(0) = \psi(t_1) = R, \quad f(1) = \psi(t_2) = Q .$$

Thus $\mathrm{SO}(2)$ is path-connected.

Warning

In general connectedness does not imply path-connectedness, as seen in Proposition 3.110.

Proposition 3.110: Topologist curve

Consider $\mathbb{R}^2$ with the Euclidean topology, and define the sets

$$A := \left\{ \left(t, \sin\left(\frac{1}{t}\right) \right) \, : \, t > 0 \right\}$$
$$B := \{(0, t) \, : \, t \in [-1, 1]\} , \qquad X := A \cup B .$$

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 $\alpha : [0, 1] \to X$ 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 $\boldsymbol{\gamma} : (0, \infty) \to \mathbb{R}^2$ 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 3.97 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 Theorem 3.98 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 3.95 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 3.95 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 3.95 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.

We conclude with the observation that connectedness and path-connectedness are equivalent for **open** sets of $\mathbb{R}^n$.

Theorem 3.111

Let $A \subseteq \mathbb{R}^n$ 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. We conclude with an interesting example concerning spaces of matrices.

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

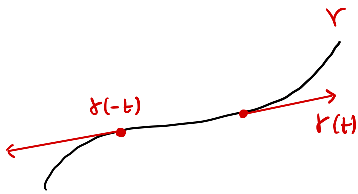


Figure 4.1.: Sketch of a curve $\boldsymbol{\gamma}$.

Surfaces are 2D objects in $\mathbb{R}^3$. There are two directions in which one can move on a surface.

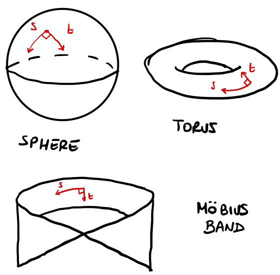


Figure 4.2.: Sketch of a surfaces: Sphere, Torus, Möbius band.

Question 4.1

How to dercribe a surface mathematically?

A curve $\Gamma \subseteq \mathbb{R}^3$ can be described with one function $\boldsymbol{\gamma} \colon (a,b) \to \Gamma$. The idea is that Γ looks locally like $\mathbb{R}$.

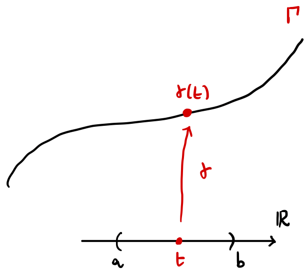


Figure 4.3.: A curve Γ can be described by a function $\boldsymbol{\gamma} \colon (a,b) \to \Gamma$.

How do we represent a surface? Suppose given a function $\boldsymbol{\sigma} \colon U \to \mathbb{R}^3$, 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** $\boldsymbol{\sigma}_i \colon U_i \to \mathcal{S}$, with $U_i \subseteq \mathbb{R}^2$ open. The charts $\boldsymbol{\sigma}_i$ represent $\mathcal{S}$ if they cover the whole surface:

$$\mathcal{S} = \bigcup_i \boldsymbol{\sigma}_i(U_i)\,.$$

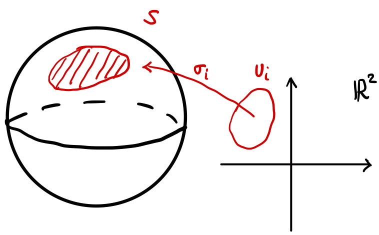


Figure 4.4.: A surface $\mathcal{S}$ can be described by a family of charts $\boldsymbol{\sigma}_i \colon U_i \to \mathcal{S}$ with $U_i \subseteq \mathbb{R}^2$ open set.

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
$$\mathbf{v} = \lambda_1 \mathbf{v}_1 + \ldots + \lambda_n \mathbf{v}_n .$$
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 $B : V \times V \to \mathbb{R}$ we define the matrix

$$M := \left(B(\mathbf{v}_i, \mathbf{v}_j)\right)_{i,j=1}^{n} \in \mathbb{R}^{n \times n} .$$

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 $L \colon V \to V$ 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 $L : V \to V$ be a linear map and A the associated matrix with respect to any basis of V. Then

$$L(\mathbf{v}) = A\mathbf{x}\,, \quad \forall\, \mathbf{v} \in V\,,$$

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

$$P := (\mathbf{v}_1 | \dots | \mathbf{v}_n)$$

$$D := \operatorname{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Remark 4.20

The correspondence between Theorem 4.13 and Theorem 4.19 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 $L \colon V \to V$ 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 $\mathbb{R}^n$

Definition 4.21: Topology of $\mathbb{R}^n$

The Euclidean norm on $\mathbb{R}^n$ is denoted by

$$\|\mathbf{x}\| := \sqrt{\sum_{i=1}^{n} x_i^2} \,, \quad \mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n \,.$$

Define the Euclidean distance $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.

1. The pair $(\mathbb{R}^n, d)$ is a metric space.
2. The topology induced by the metric d is called the Euclidean topology, denoted by $\mathscr{T}$.
3. 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 $\varepsilon > 0$ centered at $\mathbf{x}$. We write $U \in \mathscr{T}$, with $\mathscr{T}$ the Euclidean topology in $\mathbb{R}^n$.
4. A set $V \subseteq \mathbb{R}^n$ is **closed** if $V^c := \mathbb{R}^n \setminus U$ is open.

Example 4.22

- The n-dimensional unit sphere
$$\mathbb{S}^n = \{\mathbf{x} \in \mathbb{R}^{n+1} \ : \ \|\mathbf{x}\| = 1\}$$
is closed in $\mathbb{R}^{n+1}$. Indeed, define $f \colon \mathbb{R}^n \to \mathbb{R}$ by
$$f(\mathbf{x}) = \|\mathbf{x}\| .$$
Then f is continuous and
$$\mathbb{S}^n = f^{-1}(\{1\}) .$$
Since $\{1\}$ is closed in $\mathbb{R}$, and f is continuous, we conclude that $\mathbb{S}^n$ is closed.
- The n-dimensional unit cube
$$C \ := \{\mathbf{x} \in \mathbb{R}^n \ : \ |x_1| + \ldots + |x_n| < 1\}$$
is open in $\mathbb{R}^n$. Indeed, define $f \colon \mathbb{R}^n \to \mathbb{R}$ by
$$f(\mathbf{x}) = |x_1| + \ldots + |x_n| .$$
Then f is continuous and
$$C = f^{-1}((-\infty, 1)) .$$
Since $(-\infty, 1)$ is open in $\mathbb{R}$, and f is continuous, we conclude that C is open.
- 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 4.23: Subspace Topology

Let $A \subseteq \mathbb{R}^n$. The **subspace topology** on A is the family
$$\mathcal{T}_A \ := \{U \subseteq A \ : \ \exists \ W \in \mathcal{T} \ \text{ s.t. } \ U = A \cap W\} .$$
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 4.24: Continuous Function

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open. We say that f is **continuous** at $\mathbf{x} \in U$ if $\forall\, \varepsilon > 0$, $\exists\, \delta > 0$ such that

$$\|\mathbf{x} - \mathbf{y}\| < \delta \quad \Longrightarrow \quad \|f(\mathbf{x}) - f(\mathbf{y})\| < \varepsilon \,.$$

f is continuous in U if it is continuous for all $\mathbf{x} \in U$.

The above "classical' ' definition of continuity is equivalent to the topological one, in the following sense:

Theorem 4.25: Continuity: Topological definition

Let $f : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$, 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 4.26: Homeomorphism

Let $f : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$ with U, V open. We say that f is a **homeomorphism** if:

1. f is continuous;
2. f admits continuous inverse $f^{-1} : V \to U$.

Definition 4.27: 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 $d_{\mathbf{x}} f : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$d_{\mathbf{x}} f(\mathbf{h}) = \lim_{\varepsilon \to 0} \frac{f(\mathbf{x} + \varepsilon \mathbf{h}) - f(\mathbf{x})}{\varepsilon} \,,$$

for all $\mathbf{h} \in \mathbb{R}^n$, where the limit is taken in $\mathbb{R}^m$. The linear map $d_{\mathbf{x}} f$ is called the **differential** of f at $\mathbf{x}$.

The idea behind the definition of differentiability is as follows: The function f is differentiable at $\mathbf{x}$ if it can be approximated by the linear map $d_{\mathbf{x}} f$ around the point $\mathbf{x}$.

We denote by $\{\mathbf{e}_i\}_{i=1}^n$ the standard basis of $\mathbb{R}^n$. When f is differentiable, the partial derivatives are defined as follows:

Definition 4.28: Partial Derivative

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, U open, f differentiable. The **partial derivative** of f at $\mathbf{x} \in U$ in direction $\mathbf{e}_i$ is

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) := d_{\mathbf{x}} f(\mathbf{e}_i) = \lim_{\varepsilon \to 0} \frac{f(\mathbf{x} + \varepsilon \mathbf{e}_i) - f(\mathbf{x})}{\varepsilon} \,.$$

Definition 4.29: Jacobian Matrix

Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ be differentiable. The **Jacobian** of f at $\mathbf{x}$ is the $m \times n$ matrix of partial derivatives:

$$Jf(\mathbf{x}) := \left(\frac{\partial f_i}{\partial x_j}(\mathbf{x}) \right)_{i,j} \in \mathbb{R}^{m \times n} \,.$$

If $m = n$ then $Jf \in \mathbb{R}^{n \times n}$ is a square matrix and we can compute its determinant, denoted by $\det(Jf)$.

The differential $d_{\mathbf{x}} f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. As such, it must have a matrix representation with respect to the Euclidean basis. Since the partial derivative is defined as

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) := d_{\mathbf{x}} f(\mathbf{e}_i) \,,$$

we trivially have that $Jf(\mathbf{x})$ is the matrix of $d_{\mathbf{x}} f$ with respect to the standard basis:

Proposition 4.30: Matrix representation of $d_{\mathbf{x}} f$

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be differentiable. The matrix of the linear map $d_{\mathbf{x}} f : \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard basis is given by the Jacobian matrix $Jf(\mathbf{x})$.

Definition 4.31: 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 4.32: Smooth Function

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with U open. 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

For $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ smooth, the partial derivatives are

$$\partial_{x_i} f = f_{x_i} = \frac{\partial f}{\partial x_i}, \qquad \partial_{x_i x_j} f = f_{x_i x_j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$$\partial_{x_i x_j x_k} f = f_{x_i x_j x_k} = \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}$$

For $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ smoothm, the **gradient** is

$$\nabla f(\mathbf{x}) = \left(f_{x_1}(\mathbf{x}), \dots, f_{x_n}(\mathbf{x}) \right) .$$

Note that $\nabla f(\mathbf{x})$ coincides with $J f(\mathbf{x})$.

Example 4.33

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.

4.1.4. Diffeomorphisms

A key definition needed for the study of surfaces is the one of diffeomorphism. In this section we only consider maps from $\mathbb{R}^n$ into $\mathbb{R}^n$.

Definition 4.34: Diffeomorphism

Let $f : U \to V$, with $U, V \subseteq \mathbb{R}^n$ open. We say that f is a **diffeomorphism** between U and V if:

1. f is smooth,

2. f admits smooth inverse $f^{-1} : V \to U$.

Definition 4.35: Local diffeomorphism

$f : \mathbb{R}^n \to \mathbb{R}^n$ is a **local diffeomorphism** at $\mathbf{x}_0 \in \mathbb{R}^n$ if:

1. There exists an open set $U \subseteq \mathbb{R}^n$ such that $\mathbf{x}_0 \in U$,
2. There exists an open set $V \subseteq \mathbb{R}^n$ such that $f(\mathbf{x}_0) \in V$,
3. $f : U \to V$ is a diffeomorphism.

Proposition 4.36

Diffeomorphisms are local diffeomorphisms.

Non-vanishing Jacobian determinant is a necessary condition for being a diffeomorphism, as outlined in the following Proposition.

Proposition 4.37: Necessary condition for being diffeomorphism

Let $f : U \to \mathbb{R}^n$ with $U \subseteq \mathbb{R}^n$ open. Suppose f is a local diffeomorhism at $\mathbf{x}_0 \in U$. Then

$$\det Jf(\mathbf{x}_0) \neq 0 \,. \tag{4.1}$$

Example 4.38

We have already encountered Proposition 4.37 in the scalar case when we were studying curves. Indeed, suppose that

$$\phi : \mathbb{R} \to \mathbb{R}$$

is a local diffeomorphism at $t_0 \in \mathbb{R}$. Then

$$J\phi(t_0) = \dot{\phi}(t_0) \,, \quad \det J\phi(t_0) = \dot{\phi}(t_0) \,,$$

and we recover the already seen result that

$$\dot{\phi}(t_0) \neq 0 \,.$$

The condition at (4.1) is sufficient fot f to be a **local diffeomorphism** at $\mathbf{x}_0$. This is the content of the Inverse Function Theorem.

Theorem 4.39: Inverse Function Theorem

Let $f : U \to \mathbb{R}^n$ with $U \subseteq \mathbb{R}^n$ open, f smooth. Assume

$$\det Jf(\mathbf{x}_0) \neq 0 \,,$$

for some $\mathbf{x}_0 \in U$. Then:

1. There exists an open set $U_0 \subseteq U$ such that $\mathbf{x}_0 \in U_0$,
2. There exists an open set V such that $f(\mathbf{x}_0) \in V$,
3. $f : U_0 \to V$ is a diffeomorphism.

Example 4.40

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

$$\det Jf(x, y) = 0 \quad \iff \quad y = \frac{k\pi}{2} \,, \;\; k \in \mathbb{N} \,.$$

The above condition means that the Jacobian vanishes on each of the lines

$$L_k := \left\{ \left(x, \frac{k\pi}{2} \right) \; : \; x \in \mathbb{R} \right\} .$$

Define the open set U obtained by removing the lines L_k from $\mathbb{R}^n$, that is,

$$U := \mathbb{R}^n \setminus \bigcup_{k=1}^{\infty} L_k \,.$$

In particular, we have

$$\det Jf(x, y) \neq 0 \,, \quad \forall \, (x, y) \in U \,.$$

By the Inverse Function Theorem 4.39, f is a local diffeomorphism at each point $(x, y) \in U$.

Warning

Condition (4.1) is **not sufficient** for f to be a global diffeomorphism, in the following sense: There exist differentiable functions $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ such that:

1. $\det Jf(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in U$,
2. f is not a diffeomorphism between U and $f(U)$.

We will show this in the next Example.

Example 4.41: A local diffeomorphism which is not global

Question. Define the function $f : \mathbb{R}^2 \to \mathbb{R}^2$

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

Prove f is a local diffeomorphism but not a diffeomorphism.
Solution. f is a local diffeomorphism at each point $(x, y) \in \mathbb{R}^2$ by the Inverse Function Theorem, since

$$\begin{aligned} Jf(x, y) &= e^x \begin{pmatrix} \cos(y) & \sin(y) \\ -\sin(y) & \cos(y) \end{pmatrix} \\ \det Jf(x, y) &= e^{2x} \neq 0 . \end{aligned}$$

However, f is not invertible because it is not injective, since

$$f(x, y) = f(x, y + 2n\pi) , \quad \forall (x, y) \in \mathbb{R}^2, n \in \mathbb{N} .$$

Hence, f cannot be a diffeomorphism of $\mathbb{R}^2$ into $\mathbb{R}^2$.

4.2. Surfaces

We give the main definition of surface in $\mathbb{R}^3$.

Definition 4.42: 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

1. $\mathbf{p} \in \boldsymbol{\sigma}(U)$,
2. $\boldsymbol{\sigma}(U)$ is open in $\mathcal{S}$,
3. $\boldsymbol{\sigma}$ is a homeomorphism between U and $\boldsymbol{\sigma}(U)$.

$\boldsymbol{\sigma}$ is called a **surface chart** at $\mathbf{p}$.

A visual interpretation of the definition of surface is given in Figure 4.5.

Remark 4.43

1. $\mathcal{S}$ is a topological space with the subspace topology induced by the inclusion $\mathcal{S} \subseteq \mathbb{R}^3$. This means that a subset $V \subseteq \mathcal{S}$ is open in $\mathcal{S}$, if there exists an open set $W \subseteq \mathbb{R}^3$ such that
$$V = W \cap \mathcal{S}\,.$$
2. $\mathcal{S}$ is required to be **connected** with respect to the subspace topology.
3. 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.
4. 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 there exists and open set $W \subseteq \mathbb{R}^3$ such that
$$\boldsymbol{\sigma}(U) = W \cap \mathcal{S}\,.$$
5. The homeomorphism condition is saying that the surface patch
$$\boldsymbol{\sigma}(U) \subseteq \mathcal{S}$$
can be continuously deformed into the open set
$$U \subseteq \mathbb{R}^2\,.$$

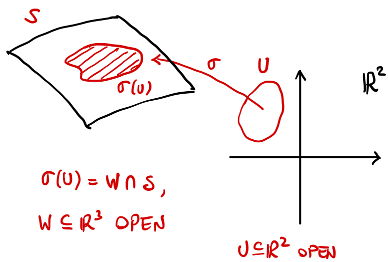


Figure 4.5.: Sketch of the surface $\mathcal{S}$ and chart $\boldsymbol{\sigma} : U \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S}$. The set $U \subseteq \mathbb{R}^2$ is open in $\mathbb{R}^2$ and $\boldsymbol{\sigma}(U)$ is open in $\mathcal{S}$. This means there exists W open in $\mathbb{R}^3$ such that $\boldsymbol{\sigma}(U) = \mathcal{S} \cap W$.

Notation

1. Points in U will be denoted with the pair (u, v).
2. 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.,

$$\boldsymbol{\sigma}_{uu} := \frac{\partial^2 \boldsymbol{\sigma}}{\partial u^2}, \qquad \boldsymbol{\sigma}_{uv} := \frac{\partial^2 \boldsymbol{\sigma}}{\partial u \partial v},$$
$$\boldsymbol{\sigma}_{vu} := \frac{\partial^2 \boldsymbol{\sigma}}{\partial v \partial u}, \qquad \boldsymbol{\sigma}_{vv} := \frac{\partial^2 \boldsymbol{\sigma}}{\partial v^2},$$

3. Components of $\boldsymbol{\sigma}$ will be denoted by

$$\boldsymbol{\sigma} = (\sigma^1, \sigma^2, \sigma^3) = (x, y, z).$$

An atlas of a surface is a collection of charts which cover the whole surface:

Definition 4.44: Atlas of a surface

Let $\mathcal{S}$ be a surface. Assume given a collection of charts

$$\mathcal{A} = \{\boldsymbol{\sigma}_i\}_{i \in I}, \qquad \boldsymbol{\sigma}_i : U_i \to \boldsymbol{\sigma}(U_i) \subseteq \mathcal{S}.$$

The family $\mathcal{A}$ is an **atlas** of $\mathcal{S}$ if

$$\mathcal{S} = \bigcup_{i \in I} \boldsymbol{\sigma}_i(U_i).$$

Example 4.45: 2D Plane in $\mathbb{R}^3$

Planes in $\mathbb{R}^3$ are surfaces with atlas made by one chart. To prove it, note that a plane $\boldsymbol{\pi} \subseteq \mathbb{R}^3$ is described by the equation

$$\boldsymbol{\pi} = \{\mathbf{x} \in \mathbb{R}^3 \ : \ \mathbf{x} \cdot \mathbf{w} = \lambda\},$$

for some $\mathbf{w} \in \mathbb{R}^3$ and $\lambda \in \mathbb{R}$. Let

- $\mathbf{p}, \mathbf{q} \in \mathbb{R}^3$ be orthonormal, and orthogonal to $\mathbf{w}$.
- $\mathbf{a} \in \boldsymbol{\pi}$ be any point in the plane.

This construction is represented in Figure 4.6. Let $\mathbf{x} \in \boldsymbol{\pi}$. Then $\mathbf{x} - \mathbf{a}$ satisfies

$$(\mathbf{x} - \mathbf{a}) \cdot \mathbf{w} = 0.$$

Hence, the vector $\mathbf{x} - \mathbf{a}$ is orthogonal to $\mathbf{w}$, meaning it can be written as linear combination of the vectors $\mathbf{p}$ and $\mathbf{q}$:

$$\mathbf{x} - \mathbf{a} = u\mathbf{p} + v\mathbf{q} \,,$$

for some coefficients $u, v \in \mathbb{R}$. Therefore the plane $\boldsymbol{\pi}$ can be equivalently represented as

$$\boldsymbol{\pi} = \{\mathbf{a} + u\mathbf{p} + v\mathbf{q} \;:\; u, v \in \mathbb{R}\} \,.$$

The above suggests to define the chart $\boldsymbol{\sigma} : \mathbb{R}^2 \to \boldsymbol{\pi}$ by

$$\boldsymbol{\sigma}(u, v) := \mathbf{a} + u\mathbf{p} + v\mathbf{q} \,.$$

Then $\boldsymbol{\sigma}$ is a chart for $\boldsymbol{\pi}$, and

$$\mathscr{A} = \{\boldsymbol{\sigma}\}$$

is an atlas, implying that $\boldsymbol{\pi}$ is a surface.
Proof. Check that $\boldsymbol{\sigma}$ is a chart:

- $\boldsymbol{\sigma}$ is smooth.
- $\mathbb{R}^2$ is obviously open.
- $\boldsymbol{\sigma}(\mathbb{R}^2)$ is open in $\boldsymbol{\pi}$ for the subspace topology, since $\boldsymbol{\sigma}(\mathbb{R}^2) = \boldsymbol{\pi}$, and $\boldsymbol{\pi}$ is open in the subspace topology.
- Suppose $\mathbf{x} = \boldsymbol{\sigma}(u, v)$. Then

  $$(\mathbf{x} - \mathbf{a}) \cdot \mathbf{p} = u \,, \quad (\mathbf{x} - \mathbf{a}) \cdot \mathbf{q} = v \,,$$

  given that $\mathbf{p}$ and $\mathbf{q}$ are orthonormal.
- The above shows that the inverse of $\boldsymbol{\sigma}$ is $\boldsymbol{\sigma}^{-1} : \boldsymbol{\pi} \to \mathbb{R}^2$ given by

  $$\boldsymbol{\sigma}^{-1}(\mathbf{x}) = ((\mathbf{x} - \mathbf{a}) \cdot \mathbf{p}, (\mathbf{x} - \mathbf{a}) \cdot \mathbf{q}) \,.$$

  Clearly, $\boldsymbol{\sigma}^{-1}$ is continuous.
- Thus, $\boldsymbol{\sigma}$ is a homeomorphism between $\mathbb{R}^2$ and $\boldsymbol{\pi}$.
- Therefore $\boldsymbol{\sigma}$ is a chart for $\boldsymbol{\pi}$. Since

Notice that

$$\boldsymbol{\sigma}(\mathbb{R}^2) = \boldsymbol{\pi} \,,$$

and therefore $\mathscr{A} = \{\boldsymbol{\sigma}\}$ is an atlas for $\boldsymbol{\pi}$, showing that $\boldsymbol{\pi}$ is a surface.

Example 4.46: Unit cylinder

Consider the infinite unit cylinder

$$\mathscr{S} = \{(x, y, z) \in \mathbb{R}^3 \;:\; x^2 + y^2 = 1\} \,.$$

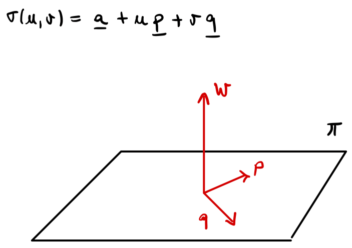


Figure 4.6.: A plane $\boldsymbol{\pi}$ is a surface with atlas containing a single chart $\boldsymbol{\sigma} : \mathbb{R}^2 \to \boldsymbol{\pi}$.

Define the map

$$\boldsymbol{\sigma} : \mathbb{R}^2 \to \mathbb{R}^3 , \quad \boldsymbol{\sigma}(u, v) := (\cos(u), \sin(u), v) .$$

Setting $V := [0, 2\pi) \times \mathbb{R}$, we notice that

$$\boldsymbol{\sigma}(V) = \mathcal{S} .$$

Moreover $\boldsymbol{\sigma} : V \to \mathcal{S}$ is clearly bijective, with inverse

$$\boldsymbol{\sigma}^{-1}(x, y, z) = (\theta, z) ,$$

with θ the angle formed by the vector $\mathbf{p} = (x, y)$ with the x-axis. However, V is not open in $\mathbb{R}^2$, and therefore $\boldsymbol{\sigma}$ cannot be a chart. To overcome this issue, let us cover V with two open sets: For example,

$$U_1 := \left(0, \frac{3\pi}{2}\right) \times \mathbb{R} , \quad \left(\pi, \frac{5\pi}{2}\right) \times \mathbb{R} ,$$

so that

$$V = U_1 \cup U_2 ,$$

with U_1 and U_2 open. We can now define two charts

$$\boldsymbol{\sigma}_1 : U_1 \to \mathcal{S} , \quad \boldsymbol{\sigma}_2 : U_2 \to \mathcal{S} ,$$

by restricting $\boldsymbol{\sigma}$:

$$\boldsymbol{\sigma}_1 := \boldsymbol{\sigma}|_{U_1} , \quad \boldsymbol{\sigma}_2 := \boldsymbol{\sigma}|_{U_2} .$$

The images of the two charts $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are shown in Figure 4.7. We have that $\mathcal{S}$ is a surface with the atlas

$$\mathcal{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\} .$$

Check:

- $\boldsymbol{\sigma}_i$ is smooth, since $\boldsymbol{\sigma}$ 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 clearly invertible from U_i to $\boldsymbol{\sigma}_i(U_i)$, and the inverse is continuous.
- Thus, $\boldsymbol{\sigma}_i$ is a homeomorphism between U_i and $\boldsymbol{\sigma}(U_i)$.
- $\mathcal{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$ is an atlas for $\mathcal{S}$, since

$$\mathcal{S} = \boldsymbol{\sigma}_1(U_1) \cup \boldsymbol{\sigma}_2(U_2) .$$

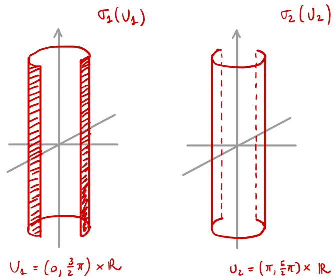


Figure 4.7.: Unit cylinder $\mathcal{S}$ is a surface with atlas $\mathcal{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$. Depicted are the images $\boldsymbol{\sigma}_1(U_1)$ and $\boldsymbol{\sigma}_2(U_2)$.

Example 4.47: 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

$$\Gamma_f := \{(u, v, f(u, v)) : (u, v) \in U\} .$$

Γ_f is a surface with atlas given by

$$\mathcal{A} = \{\boldsymbol{\sigma}\}$$

where $\boldsymbol{\sigma} : U \to \Gamma_f$ is

$$\boldsymbol{\sigma}(u, v) := (u, v, f(u, v)) .$$

Proof. 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 $\boldsymbol{\sigma}^{-1} : \Gamma_f \to U$ defined as

$$\boldsymbol{\sigma}^{-1}(u, v, f(u, v)) := (u, v) .$$

  Clearly $\boldsymbol{\sigma}^{-1}$ 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 4.48: Circular cone

Consider the circular cone

$$\mathcal{S} := \{(x,y,z) \in \mathbb{R}^3 \,:\, x^2 + y^2 = z^2\}\,.$$

Then $\mathcal{S}$ is not a surface. This is, essentially, a consequence of the fact that

$$\mathcal{S} \setminus \{\mathbf{0}\}$$

is a disconnected set, see Figure 4.8.
To see that $\mathcal{S}$ is not a surface, suppose there exists an atlas $\{\boldsymbol{\sigma}_i\}$ of $\mathcal{S}$

$$\boldsymbol{\sigma}_i : U_i \to \boldsymbol{\sigma}_i(U_i) \subseteq \mathcal{S}\,.$$

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

$$\boldsymbol{\sigma}(U) \cap B_\delta(\mathbf{0}) \subseteq \boldsymbol{\sigma}(U) \cap W = \boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0))\,.$$

The ball $B_\delta(\mathbf{0})$ intersects both $\mathcal{S}^-$ and $\mathcal{S}^+$, with

$$\mathcal{S}^- := \mathcal{S} \cap \{z < 0\}\,, \quad \mathcal{S}^+ := \mathcal{S} \cap \{z > 0\}\,.$$

Therefore $\boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0))$ intersects both $\mathcal{S}^-$ and $\mathcal{S}^+$. This implies that the set

$$V := \boldsymbol{\sigma}(B_\varepsilon(\mathbf{x}_0)) \setminus \{\mathbf{0}\}$$

is disconnected, with disconnection given by

$$V = (V \cap \mathcal{S}^-) \cup (V \cap \mathcal{S}^+) .$$

However, V is homeomorphic to

$$B_\varepsilon(\mathbf{x}_0) \setminus \{\mathbf{x}_0\} ,$$

which is instead connected. We have obtained a contradiction, and therefore $\mathcal{S}$ is not a surface.

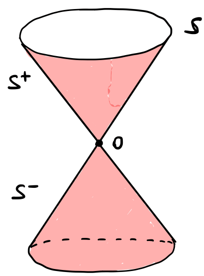


Figure 4.8.: The circular cone is not a surface. This is because $\mathcal{S} \setminus \{\mathbf{0}\}$ is disconnected.

4.3. Regular Surfaces

We have defined a regular curve to be a map $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^n$ such that

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

Regularity allowed us to reparametrize by arc-length and define the Frenet frame, curvature and torsion. We then proved that curvature and torsion completely characterize $\boldsymbol{\gamma}$, up to rigid motions.

We want to do something similar for surfaces: We look for a condition that eventually will allow us to define the tangent plane to the surface. Specifically, we require that the partial derivatives $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ of a chart $\boldsymbol{\sigma}$ are linearly independent. In this case $\boldsymbol{\sigma}$ is called a regular chart. In details:

Definition 4.49: Regular Chart

Let $U \subseteq \mathbb{R}^2$ be open. A map $\boldsymbol{\sigma} = \boldsymbol{\sigma}(u,v) : U \to \mathbb{R}^3$ is 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 $(u,v) \in U$.

We are now ready to define regular surfaces.

Definition 4.50: Regular surface

Let $\mathcal{S}$ be a surface. We say that:

- $\mathcal{A}$ is a **regular atlas** if any $\boldsymbol{\sigma}$ in $\mathcal{A}$ is regular.
- $\mathcal{S}$ is a **regular surface** if it admits a regular atlas.

Before making some examples, we highlight give some equivalent methods for checking the regularity condition.

Theorem 4.51: Characterization of regular charts

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ with $U \subseteq \mathbb{R}^2$ open. They are equivalent:

1. $\boldsymbol{\sigma}$ is a regular chart.
2. $d_{\mathbf{x}}\boldsymbol{\sigma} : \mathbb{R}^2 \to \mathbb{R}^3$ is injective for all $\mathbf{x} \in U$.
3. The Jacobian matrix $J\boldsymbol{\sigma}$ has rank 2 for all $(u,v) \in U$.
4. $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq 0$ for all $(u,v) \in U$.

Proof

Part 1. Equivalence of Point 1 and Point 4.
By the properties of vector product, we have that

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq 0 \qquad \forall\, (u,v) \in U$$

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 $d_{\mathbf{x}}\boldsymbol{\sigma} : \mathbb{R}^2 \to \mathbb{R}^3$ 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.

Example 4.52: 2D Plane in $\mathbb{R}^3$

Question. Let $\mathbf{a}, \mathbf{p}, \mathbf{q} \in \mathbb{R}^3$, with $\mathbf{p}$ and $\mathbf{q}$ orthonormal. The plane

$$\boldsymbol{\pi} = \{\mathbf{a} + u\mathbf{p} + v\mathbf{q} \ : \ u, v \in \mathbb{R}\}$$

is a surface with atlas $\mathscr{A} = \{\boldsymbol{\sigma}\}$, where

$$\boldsymbol{\sigma} : \mathbb{R}^2 \to \boldsymbol{\pi}, \quad \boldsymbol{\sigma}(u, v) := \mathbf{a} + u\mathbf{p} + v\mathbf{q}.$$

Prove that $\boldsymbol{\pi}$ is a regular surface.
Solution. We have $\boldsymbol{\sigma}_u = \mathbf{p}, \boldsymbol{\sigma}_v = \mathbf{q}$. Since $\mathbf{p}$ and $\mathbf{q}$ are orthonormal, we conclude that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent and $\boldsymbol{\sigma}$ is regular. $\boldsymbol{\pi}$ is a regular surface because $\boldsymbol{\sigma}$ is a regular chart.

Example 4.53: Unit cylinder

Question. 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 atlas $\mathscr{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$, with

$$\boldsymbol{\sigma}(u, v) = (\cos(u), \sin(u), v), \qquad \boldsymbol{\sigma}_1 = \boldsymbol{\sigma}|_{U_1}, \quad \boldsymbol{\sigma}_2 = \boldsymbol{\sigma}|_{U_2},$$

$$U_1 = \left(0, \frac{3\pi}{2}\right) \times \mathbb{R}, \qquad U_2 = \left(\pi, \frac{5\pi}{2}\right) \times \mathbb{R}.$$

Prove that $\mathcal{S}$ is a regular surface.
Solution. The map $\boldsymbol{\sigma}$ is regular because

$$\boldsymbol{\sigma}_u = (-\sin(u), \cos(u), 0), \quad \boldsymbol{\sigma}_v = (0, 0, 1),$$

are linearly independent, since the last components of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are 0 and 1. Therefore, also $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are regular charts, being restrictions of $\boldsymbol{\sigma}$. Thus, $\mathscr{A}$ is a regular atlas and $\mathcal{S}$ a regular surface.

The infinite cylinder can also be parametrized using a single chart, as shown in the next Example.

Example 4.54: Unit cylinder: Single chart atlas

Consider again infinite unit cylinder

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 = 1\} .$$

Define the open set

$$U := \mathbb{R}^2 \smallsetminus \{(0, 0)\} ,$$

and the map $\boldsymbol{\sigma} : U \to \mathcal{S}$ by

$$\boldsymbol{\sigma}(u, v) = \left(\frac{u}{\sqrt{u^2 + v^2}}, \frac{v}{\sqrt{u^2 + v^2}}, \log\left(\sqrt{u^2 + v^2}\right) \right) .$$

Then $\boldsymbol{\sigma}$ is regular and $\mathcal{A} = \{\boldsymbol{\sigma}\}$ is a regular atlas for $\mathcal{S}$.
Proof: Left as an exercise.

Example 4.55: Graph of a function

Question. Let $f : U \to \mathbb{R}$ be smooth, $U \subseteq \mathbb{R}^2$ open. Define

$$\Gamma_f = \{(u, v, f(u, v)) \ : \ (u, v) \in U\} ,$$

the graph of f. Then Γ_f is surface with atlas $\mathcal{A} = \{\boldsymbol{\sigma}\}$, where

$$\boldsymbol{\sigma} : U \to \Gamma_f , \quad \boldsymbol{\sigma}(u, v) := (u, v, f(u, v)) .$$

Prove that Γ_f is a regular surface.
Solution. The Jacobian matrix of $\boldsymbol{\sigma}$ is

$$J\boldsymbol{\sigma}(u, v) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ f_u & f_v \end{pmatrix} .$$

$J\boldsymbol{\sigma}$ has rank 2, because the first minor is the 2×2 identity matrix. Therefore, $\boldsymbol{\sigma}$ is regular. This implies $\mathcal{A}$ is a regular atlas, and $\mathcal{S}$ is a regular surface.

We now want to consider the sphere

$$\mathbb{S}^2 := \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 + z^2 = 1\} .$$

In order to prove that $\mathbb{S}^2$ is a regular surface, we need to introduce spherical coordinates.

Definition 4.56: Spherical coordinates

The **spherical coordinates** of $\mathbf{p} = (x, y, z) \neq \mathbf{0}$ are

$$
\begin{aligned}
x &= \rho \cos(\theta) \cos(\varphi) \\
y &= \rho \sin(\theta) \cos(\varphi) \\
z &= \rho \sin(\varphi)
\end{aligned}
$$

where

$$
\rho := \sqrt{x^2 + y^2 + z^2} \,, \quad \theta \in [-\pi, \pi] \,, \quad \varphi \in \left[-\frac{\pi}{2}, \frac{\pi}{2} \right] \,,
$$

with the angles θ and φ as in Figure 4.9.
Check: It is clear that $z = \rho \sin(\varphi)$. To compute x and y, we note that the segment joining $\mathbf{0}$ to $\mathbf{q}$ has length

$$
L = \rho \cos(\varphi) \,.
$$

Therefore we get

$$
\begin{aligned}
x &= L \cos(\theta) = \rho \cos(\theta) \cos(\varphi) \\
y &= L \sin(\theta) = \rho \sin(\theta) \cos(\varphi)
\end{aligned}
$$

concluding.

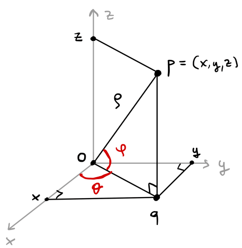


Figure 4.9.: Spherical coordinates in $\mathbb{R}^3$.

Example 4.57: Unit sphere in spherical coordinates

Consider the unit sphere in $\mathbb{R}^3$

$$
\mathbb{S}^2 := \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 + z^2 = 1\} \,.
$$

Spherical coordinates allow us to define an atlas on $\mathbb{S}^2$. In details, define the set

$$
U := \left\{ (\theta, \varphi) \in \mathbb{R}^2 \ : \ \theta \in (-\pi, \pi), \, \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \right\} \,,
$$

and the map $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ by

$$\boldsymbol{\sigma}(\theta, \varphi) := (\cos(\theta)\cos(\varphi), \sin(\theta)\cos(\varphi), \sin(\varphi)) .$$

In order to name some of the parallels and meridians on $\mathbb{S}^2$, let us identify $\mathbb{S}^2$ with the Earth. With reference to Figure 4.10, we make the following definitions:

- The *Equator Line* corresponds to the angle $\varphi = 0$, that is,

$$\text{Equator Line } = \mathbb{S}^2 \cap \{z = 0\} .$$

- The *Greenwich meridian* corresponds to the angle $\theta = 0$. Hence:

$$\text{Greenwich } = \left\{ (\cos(\varphi), 0, \sin(\varphi)) \,, \ \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \right\} .$$

- The *Date Line* is the meridian opposite to the Greenwich one. This corresponds to $\theta = \pi$, and is parametrized by:

$$\text{Date Line } = \left\{ (-\cos(\varphi), 0, \sin(\varphi)) \,, \ \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \right\} .$$

- The *North Pole* and *South Pole* have coordinates

$$N = (0, 0, 1) \,, \quad S = (0, 0, -1) .$$

- The *Northern Hemisphere* is the *top-half* of $\mathbb{S}^2$, that is,

$$\text{Northern Hemisphere } = \mathbb{S}^2 \cap \{z \geq 0\} .$$

- The *Southern Hemisphere* is the *bottom-half* of $\mathbb{S}^2$, that is,

$$\text{Southern Hemisphere } = \mathbb{S}^2 \cap \{z \leq 0\} .$$

Notice that the angles

$$\theta = \pi \,, \quad \varphi = \pm \frac{\pi}{2}$$

are excluded in the definition of U. Therefore the parametrization $\boldsymbol{\sigma}$ misses the Date Line, as well as the North and South Poles, see the left picture in Figure 4.11. In formulas:

$$\begin{aligned} \boldsymbol{\sigma}(U) &= \mathbb{S}^2 \setminus \{\text{Date Line, North Pole, South Pole}\} \\ &= \mathbb{S}^2 \setminus \{(x, 0, z) \in \mathbb{R}^3 \ : \ x \leq 0\} . \end{aligned}$$

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

$$\tilde{\boldsymbol{\sigma}} := (-\cos(\theta)\cos(\varphi), -\sin(\varphi), -\sin(\theta)\cos(\varphi)) .$$

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, see the right picture in Figure 4.11. Thus,

$$\tilde{\boldsymbol{\sigma}}(U) = \mathbb{S}^2 \setminus \{(x, y, 0) \in \mathbb{R}^3 \ : \ x \geq 0\}.$$

In particular, we have shown that

$$\mathbb{S}^2 = \boldsymbol{\sigma}(U) \cup \tilde{\boldsymbol{\sigma}}(U).$$

Question. Show that

$$\mathscr{A} := \{\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}\}$$

is a regular atlas for $\mathbb{S}^2$.

Solution. Check that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are charts:

- $\boldsymbol{\sigma}$ is smooth.
- U is open in $\mathbb{R}^2$.
- Moreover
$$\boldsymbol{\sigma}(U) = \mathbb{S}^2 \setminus \{(x, 0, z) \in \mathbb{R}^3 \ : \ x \leq 0\}.$$
This is clearly an open set in $\mathbb{S}^2$.
- The spherical coordinates on the sphere are invertible. Therefore $\boldsymbol{\sigma}$ is invertible, with continuous inverse.
- Thus, $\boldsymbol{\sigma}$ is a homeomorphism from U into $\boldsymbol{\sigma}(U)$.
- This shows $\boldsymbol{\sigma}$ is a chart of $\mathbb{S}^2$.
- Since $\tilde{\boldsymbol{\sigma}}$ is obtained from $\boldsymbol{\sigma}$ by composing two rotations, we conclude that also $\tilde{\boldsymbol{\sigma}}$ is a chart.

Show that $\boldsymbol{\sigma}$ is a regular chart:

$$\begin{aligned}
\boldsymbol{\sigma}_\theta &= (-\sin(\theta)\cos(\varphi), \cos(\theta)\cos(\varphi), 0) \\
\boldsymbol{\sigma}_\varphi &= (-\cos(\theta)\sin(\varphi), -\sin(\theta)\sin(\varphi), \cos(\varphi)).
\end{aligned}$$

Since $(\theta, \varphi) \in U$, we have $\varphi \in (-\pi/2, \pi/2)$. Therefore, the last component of $\boldsymbol{\sigma}_\varphi$ is non-zero, i.e.,

$$\cos(\varphi) \neq 0, \quad \forall \, \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Since the last component of $\boldsymbol{\sigma}_\theta$ is 0, we conclude that $\boldsymbol{\sigma}_\theta$ and $\boldsymbol{\sigma}_\varphi$ are linearly independent for all $(\theta, \varphi) \in U$. Therefore $\boldsymbol{\sigma}$ is regular. Alternatively, we could have computed:

$$\boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\varphi = (\cos(\theta)\cos^2(\varphi), \sin(\theta)\cos^2(\varphi), \cos(\varphi)\sin(\varphi)),$$

from which

$$\left\| \boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\varphi \right\| = |\cos(\varphi)|.$$

Since $(\theta, \varphi) \in U$, we have $\varphi \in (-\pi/2, \pi/2)$, and so

$$\left\| \boldsymbol{\sigma}_\theta \times \boldsymbol{\sigma}_\varphi \right\| = \cos(\varphi) \neq 0.$$

Thus $\boldsymbol{\sigma}_\theta$ and $\boldsymbol{\sigma}_\varphi$ are linearly independent, and $\boldsymbol{\sigma}$ is regular.
Since $\tilde{\boldsymbol{\sigma}}$ is obtained from $\boldsymbol{\sigma}$ by applying two rotations, it follows that $\tilde{\boldsymbol{\sigma}}$ is regular. Therefore

$$\mathscr{A} = \{\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}\}$$

is a regular atlas for $\mathbb{S}^2$.

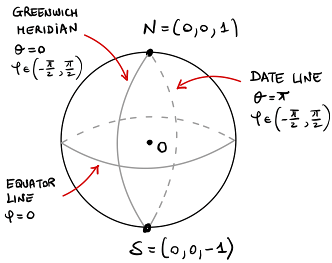


Figure 4.10.: Equator Line, Greenwich Meridian, Date Line, North and South Poles on the sphere.

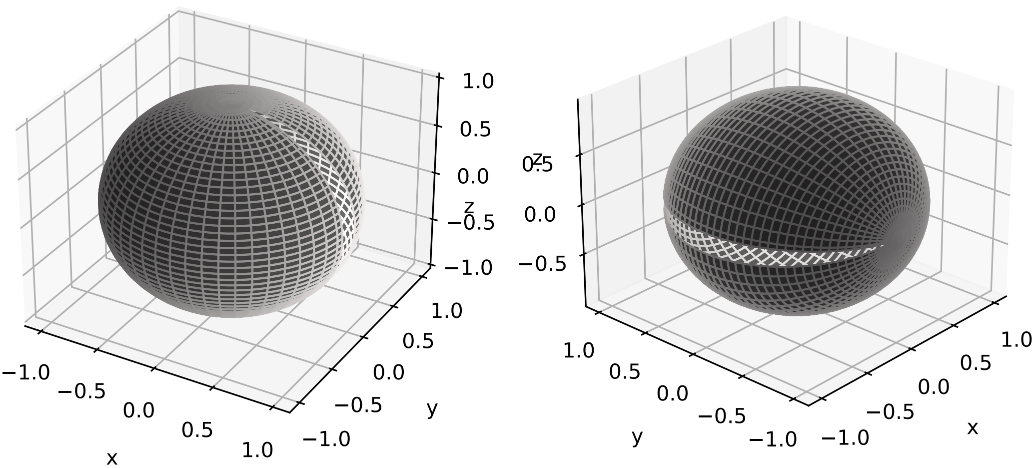


Figure 4.11.: Image of the charts of the sphere from the above example.

In alternative, the sphere can be parametrized in Cartesian coordinates.

Example 4.58: Unit sphere in Cartesian coordinates

Question. Define the following collection of charts on the sphere $\mathbb{S}^2$

$$\mathscr{A} = \{\boldsymbol{\sigma}_i\}_{i=1}^6 ,$$

where $\boldsymbol{\sigma}_i$ is 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

$$\begin{aligned}
\boldsymbol{\sigma}_1(u,v) &= \left(u, v, \sqrt{1-u^2-v^2}\right)\\
\boldsymbol{\sigma}_2(u,v) &= \left(u, v, -\sqrt{1-u^2-v^2}\right)\\
\boldsymbol{\sigma}_3(u,v) &= \left(u, \sqrt{1-u^2-v^2}, v\right)\\
\boldsymbol{\sigma}_4(u,v) &= \left(u, -\sqrt{1-u^2-v^2}, v\right)\\
\boldsymbol{\sigma}_5(u,v) &= \left(\sqrt{1-u^2-v^2}, u, v\right)\\
\boldsymbol{\sigma}_6(u,v) &= \left(-\sqrt{1-u^2-v^2}, u, v, \right)
\end{aligned}$$

Prove that $\mathscr{A}$ is a regular atlas.
Solution. Let us check that $\mathbb{S}^2$ is a surface:

- $\boldsymbol{\sigma}_1$ is smooth, since in U we have $u^2 + v^2 < 1$.
- U is open, being the open ball of radius 1 in $\mathbb{R}^2$.
- $\boldsymbol{\sigma}_1(U)$ is clearly open in $\mathbb{S}^2$: This is because $\boldsymbol{\sigma}_1(U)$ coincides with the Northern Hemisphere, with the Equator Line removed.
- The inverse of $\boldsymbol{\sigma}_1$ is given by $\boldsymbol{\sigma}^{-1} : \boldsymbol{\sigma}_1(U) \to U$ defined by

$$\boldsymbol{\sigma}^{-1}(u, v, \sqrt{1-u^2-v^2}) := (u,v) .$$

- $\boldsymbol{\sigma}^{-1}$ is continuous, and thus $\boldsymbol{\sigma}_1$ is a homeomorphism of U with $\boldsymbol{\sigma}_1(U)$.
- With similar arguments, we can see that all the maps $\boldsymbol{\sigma}_i$ are charts.
- Note that $\boldsymbol{\sigma}_1$ charts the Northern Hemisphere (excluding the Equator), while $\boldsymbol{\sigma}_2$ charts the Southern Hemisphere (excluding the Equator). Thus,

$$\boldsymbol{\sigma}_1(U) \cup \boldsymbol{\sigma}_2(U) = \mathbb{S}^2 \smallsetminus \{z = 0\} .$$

By including the other 4 charts $\boldsymbol{\sigma}_3, \boldsymbol{\sigma}_4, \boldsymbol{\sigma}_5, \boldsymbol{\sigma}_6$, we can cover the whole sphere, that is,

$$\mathbb{S}^2 = \bigcup_{i=1}^{6} \boldsymbol{\sigma}_i(U) .$$

This shows that $\mathscr{A} = \{\boldsymbol{\sigma}_i\}_{i=1}^6$ is an atlas for $\mathbb{S}^2$.

Let us now check that $\mathbb{S}^2$ is a regular surface:

- The first chart $\boldsymbol{\sigma}_1$ has derivatives

$$(\boldsymbol{\sigma}_1)_u = (1, 0, f_u) \,, \quad (\boldsymbol{\sigma}_1)_v = (0, 1, f_v) \,,$$

where f_u, f_v are the partial derivatives of

$$f(u, v) := \sqrt{1 - u^2 - v^2} \,.$$

Therefore, the Jacobian matrix of $\boldsymbol{\sigma}_1$ is

$$J\boldsymbol{\sigma}_1(u, v) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ f_u & f_v \end{pmatrix} .$$

The first minor of $J\boldsymbol{\sigma}_1$ is the identity matrix, and therefore $J\boldsymbol{\sigma}$ has rank 2, showing that $(\boldsymbol{\sigma}_1)_u$ and $(\boldsymbol{\sigma}_2)_v$ are linearly independent. Hence $\boldsymbol{\sigma}_1$ is regular.
- Clearly, $J\boldsymbol{\sigma}_i$ has rank 2 for each of the charts $\boldsymbol{\sigma}_i$. Therefore $\boldsymbol{\sigma}_i$ is regular.
- We conclude that $\mathscr{A}$ is a regular atlas, making $\mathbb{S}^2$ a regular surface.

Let us conclude the section with the example of a non-regular surface.

Example 4.59: A non-regular chart

Question. Prove that the following chart is not regular

$$\boldsymbol{\sigma}(u, v) = (u, v^2, v^3) \,.$$

Solution. We have

$$\boldsymbol{\sigma}_v = (0, 2v, 3v^2) \,, \qquad \boldsymbol{\sigma}_v(u, 0) = (0, 0, 0) \,.$$

$\boldsymbol{\sigma}$ is not regular because $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly dependent along the line $L = \{(u, 0) \, : \, u \in \mathbb{R}\}$.

Looking at Figure Figure 4.12, it is clear that $\mathcal{S}$ is not regular, since $\mathcal{S}$ has a cusp along the line $\boldsymbol{\sigma}(L)$.

4.4. Reparametrizations

We have already considered reparametrizations when we studied curves. In a similar way, one can reparametrize surface charts.

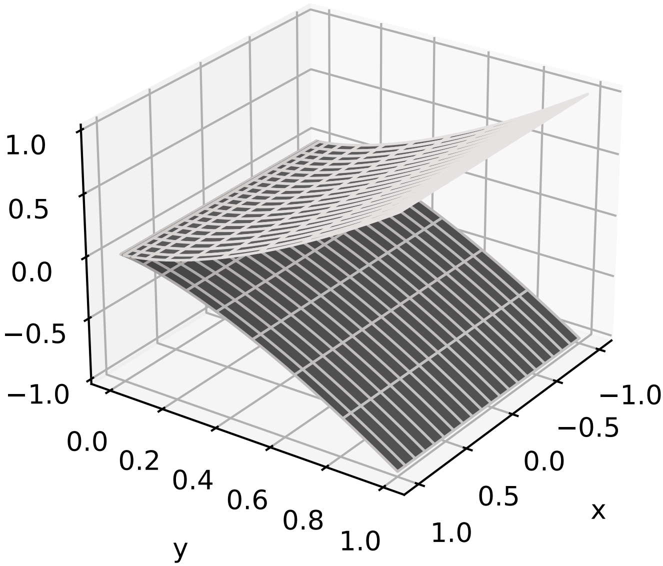


Figure 4.12.: Example of a non-regular surface.

Definition 4.60: Reparametrization

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

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi .$$

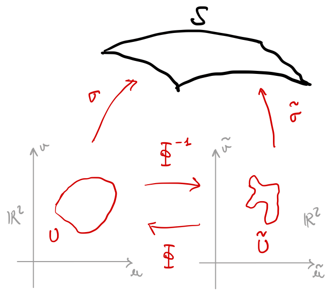


Figure 4.13.: Schematic illustration of surface chart $\boldsymbol{\sigma}$ and reparametrization $\tilde{\boldsymbol{\sigma}}$.

We will show that reparametrizations of regular charts are regular. To prove this, first we need to recall the chain rule for vector valued functions of several variables.

Remark 4.61: 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 carrying out the matrix multiplication on the right hand sinde of 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 expressions are quite cumbersome. This motivates the introduction of more compact notations 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 by:

$$
J\Phi = \begin{pmatrix} \Phi^1_{\tilde{u}} & \Phi^1_{\tilde{v}} \\ \Phi^2_{\tilde{u}} & \Phi^2_{\tilde{v}} \end{pmatrix} \quad \rightsquigarrow \quad \begin{pmatrix} \frac{\partial u}{\partial \tilde{u}} & \frac{\partial u}{\partial \tilde{v}} \\ \frac{\partial v}{\partial \tilde{u}} & \frac{\partial v}{\partial \tilde{v}} \end{pmatrix} .
$$

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.

Theorem 4.62: Reparametrizations of regular charts are regular

Let $U, \widetilde{U} \subseteq \mathbb{R}^2$ be open and $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular. Suppose given a diffeomorphism $\Phi : \widetilde{U} \to U$. The reparametrization

$$
\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3 , \qquad \tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi
$$

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

$$
\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \neq 0 . \tag{4.2}
$$

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

$$\begin{aligned}
\tilde{\boldsymbol{\sigma}}_{\tilde{u}} \times \tilde{\boldsymbol{\sigma}}_{\tilde{v}} &= \left(\boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{u}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{u}} \right) \times \left(\boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{v}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{v}} \right) \\
&= \frac{\partial u}{\partial \tilde{u}} \frac{\partial u}{\partial \tilde{v}} \left(\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_u \right) + \frac{\partial u}{\partial \tilde{u}} \frac{\partial v}{\partial \tilde{v}} \left(\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \right) \\
&+ \frac{\partial v}{\partial \tilde{u}} \frac{\partial u}{\partial \tilde{v}} \left(\boldsymbol{\sigma}_v \times \boldsymbol{\sigma}_u \right) + \frac{\partial v}{\partial \tilde{u}} \frac{\partial v}{\partial \tilde{v}} \left(\boldsymbol{\sigma}_v \times \boldsymbol{\sigma}_v \right) \\
&= \left(\frac{\partial u}{\partial \tilde{u}} \frac{\partial v}{\partial \tilde{v}} - \frac{\partial v}{\partial \tilde{u}} \frac{\partial u}{\partial \tilde{v}} \right) \left(\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \right) \\
&= \det \begin{pmatrix} \frac{\partial u}{\partial \tilde{u}} & \frac{\partial u}{\partial \tilde{v}} \\ \frac{\partial v}{\partial \tilde{u}} & \frac{\partial v}{\partial \tilde{v}} \end{pmatrix} \left(\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \right) \\
&= \det J\Phi \, \left(\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \right) .
\end{aligned}$$

Since Φ is a diffeomorphism, we have that

$$\det J\Phi \neq 0 \,,$$

from which we conclude (4.2).

4.5. Transition maps

Suppose that a surface $\mathcal{S}$ has atlas given by $\mathscr{A} = \{\boldsymbol{\sigma}_i\}_{i \in I}$. By definition of atlas, it holds that

$$\mathcal{S} = \bigcup_{i \in I} \boldsymbol{\sigma}_i(U_i) \,.$$

As the images $\boldsymbol{\sigma}_i(U_i)$ are open in $\mathcal{S}$, and cover the whole surface, in general it will happen that two (or more) images will overlap, i.e.,

$$I := \boldsymbol{\sigma}_i(U_i) \cap \boldsymbol{\sigma}_j(U_j) \neq \emptyset \,,$$

for some $i \neq j$. It is natural to ask whether the charts $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}_j$ are reparametrizations of each other on the overlapping region I, see Figure 4.14. This is indeed the case, see Theorem 4.65 below. Such reparametrization is called a *transition map*.

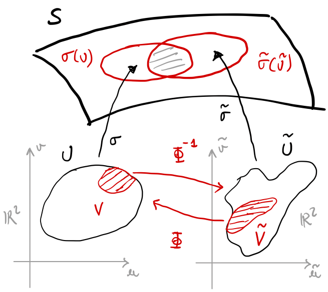


Figure 4.14.: If the two regular charts $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have overlapping image, then they are reparametrization of each other, through a transition map $\Phi = \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}}$.

Definition 4.63: Transition map

Let $\mathcal{S}$ be a regular surface, $\boldsymbol{\sigma} : U \to \mathcal{S}$, $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathcal{S}$ regular charts. Suppose the images of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ overlap

$$I := \boldsymbol{\sigma}(U) \cap \tilde{\boldsymbol{\sigma}}(\widetilde{U}) \neq \emptyset .$$

I is open in $\mathcal{S}$, being intersection of open sets. Define

$$V := \boldsymbol{\sigma}^{-1}(I) \subseteq U , \quad \widetilde{V} := \tilde{\boldsymbol{\sigma}}^{-1}(I) \subseteq \widetilde{U} .$$

V and $\widetilde{V}$ are open, by continuity of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$. Moreover, as $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are homeomorphisms, we have $\boldsymbol{\sigma}(V) = \tilde{\boldsymbol{\sigma}}(\widetilde{V}) = I$. Therefore, they are well defined the restriction homeomorphisms

$$\boldsymbol{\sigma}|_V : V \to I , \quad \tilde{\boldsymbol{\sigma}}|_{\widetilde{V}} : \widetilde{V} \to I .$$

The **transition map** from $\boldsymbol{\sigma}$ to $\tilde{\boldsymbol{\sigma}}$ is the homeomorphism

$$\Phi : \widetilde{V} \to V , \quad \Phi := \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}} .$$

The following theorem states that the transition maps between regular charts are diffeomorphisms. The proof is somewhat technical and relies on the Implicit Function Theorem. A similar argument will be used for Lemma 4.79 in Section 4.7. We have chosen to omit the proof here, but interested readers can refer to page 117 of [7] for details.

Theorem 4.64

Transition maps between regular charts are diffeomorphisms.

The immediate consequence of Theorem 4.64 is that transition maps are reparametrizations. To fix notations, let us state this fact precisely.

Theorem 4.65: Transition maps are reparametrizations

Let $\mathcal{S}$ be a regular surface, $\boldsymbol{\sigma} : U \to \mathcal{S}$, $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathcal{S}$ regular charts, with $I := \boldsymbol{\sigma}(U) \cap \tilde{\boldsymbol{\sigma}}(\widetilde{U}) \neq \emptyset$. Define the transition map

$$\Phi = \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}} : \widetilde{V} \to V, \quad V = \boldsymbol{\sigma}^{-1}(I), \quad \widetilde{V} = \tilde{\boldsymbol{\sigma}}^{-1}(I).$$

Then $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are reparametrization of each other, with

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi, \qquad \boldsymbol{\sigma} = \tilde{\boldsymbol{\sigma}} \circ \Phi^{-1}.$$

Example 4.66: Reparametrization of $\mathbb{S}^2$

In Example 4.57 and Example 4.58 we gave two different regular parametrizations of the sphere $\mathbb{S}^2$:

1. Spherical coordinates: The sphere, excluding the Date Line and the Poles, is charted by
$$\boldsymbol{\sigma}(\theta, \varphi) = (\cos(\theta)\cos(\varphi), \sin(\theta)\cos(\varphi), \sin(\varphi)),$$
defined over the set
$$U = \left\{ (\theta, \varphi) \in \mathbb{R}^2 \,:\, \theta \in (-\pi, \pi),\; \varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2} \right) \right\}.$$

2. Cartesian coordinates: The Northen Hemisphere is charted by
$$\tilde{\boldsymbol{\sigma}}(u, v) = (u, v, \sqrt{1 - u^2 - v^2}),$$
defined over the set
$$\widetilde{U} = \{(u, v) \in \mathbb{R}^2 \,:\, u^2 + v^2 < 1\}.$$

The intersection of the images

$$I = \boldsymbol{\sigma}(U) \cap \tilde{\boldsymbol{\sigma}}(\widetilde{U})$$

is non-empty. Indeed, the two charts overlap across the Northern Hemisphere, excluding the Date Line and North Pole. Define the open sets

$$V := \boldsymbol{\sigma}^{-1}(I), \quad \widetilde{V} := \tilde{\boldsymbol{\sigma}}^{-1}(I),$$

and the transition map

$$\Phi : \widetilde{V} \to V, \quad \Phi := \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}}.$$

Since $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are regular, Theorem 4.65 guarantees that Φ is a reparametrization map. Therefore $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are reparametrization of each other, with

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi.$$

Conclusion: the two parametrizations $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ of $\mathbb{S}^2$ are interchangeable!

Important

Theorem 4.65 demonstrates that there is no single preferred way to parametrize a surface: When two regular charts overlap, they are reparametrizations of each other in the overlapping region. This observation has a significant consequence:
It allows us to define a property of any regular surface by using charts, as long as we ensure that the definition is independent of reparametrization and, therefore, of the specific chart chosen.

4.6. Functions between surfaces

We aim to define the concept of a smooth function

$$f \colon \mathcal{S}_1 \to \mathcal{S}_2 \,,$$

where $\mathcal{S}_1$ and $\mathcal{S}_2$ are regular surfaces. Up to this point, we only know how to define smooth functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. The idea is to use surface charts to extend this definition of smoothness to functions between surfaces, see Figure 4.15.

Definition 4.67: Smooth functions between surfaces

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be regular surfaces and $f \colon \mathcal{S}_1 \to \mathcal{S}_2$ a map.

1. f is *smooth at* $\mathbf{p} \in \mathcal{S}_1$, if there exist charts

$$\boldsymbol{\sigma}_i \colon U_i \to \mathcal{S}_i \text{ such that } \mathbf{p} \in \boldsymbol{\sigma}_1(U_1) \,, \; f(\mathbf{p}) \in \boldsymbol{\sigma}_2(U_2) \,,$$

   and that the following map is smooth

$$\Psi \colon U_1 \to U_2 \,, \quad \Psi = \boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1 \,.$$

2. f is *smooth*, if it is smooth for each $\mathbf{p} \in \mathcal{S}_1$.

Remark 4.68

1. Definition 4.67 makes sense because $\boldsymbol{\sigma}_2^{-1}$ exists.
2. The map $\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1$ is only defined for the points $\mathbf{x} \in U_1$ such that

$$f(\boldsymbol{\sigma}_1(\mathbf{x})) \in \boldsymbol{\sigma}_2(U_2) \,.$$

3. The function $\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1$ maps from $\mathbb{R}^2$ into $\mathbb{R}^2$, therefore smoothness is intended in the classical sense.
4. Definition 4.67 is well-posed: Smoothness of f does not depend on the specific choice of charts $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$.

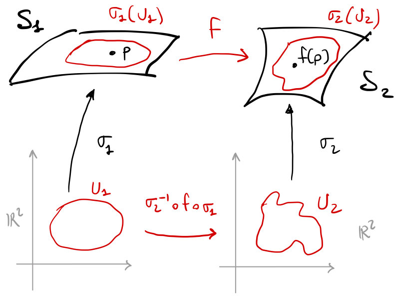


Figure 4.15.: The function $f : \mathcal{S}_1 \to \mathcal{S}_2$ is smooth at $\mathbf{p}$, if the vector valued function $\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1$ is smooth.

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 4.64 there exist open sets

$$V_i \subseteq U_i , \qquad \widetilde{V}_i \subseteq \widetilde{U}_i ,$$

and reparametrization maps

$$\Phi_i : \widetilde{V}_i \to V_i , \qquad \tilde{\boldsymbol{\sigma}}_i = \boldsymbol{\sigma}_i \circ \Phi_i .$$

Hence

$$\begin{aligned}
\tilde{\boldsymbol{\sigma}}_2^{-1} \circ f \circ \tilde{\boldsymbol{\sigma}}_1 &= \tilde{\boldsymbol{\sigma}}_2^{-1} \circ (\boldsymbol{\sigma}_2 \circ \boldsymbol{\sigma}_2^{-1}) \circ f \circ (\boldsymbol{\sigma}_1 \circ \boldsymbol{\sigma}_1^{-1}) \circ \tilde{\boldsymbol{\sigma}}_1 \\
&= (\tilde{\boldsymbol{\sigma}}_2^{-1} \circ \boldsymbol{\sigma}_2) \circ (\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1) \circ (\boldsymbol{\sigma}_1^{-1} \circ \tilde{\boldsymbol{\sigma}}_1) \\
&= \Phi_2^{-1} \circ (\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1) \circ \Phi_1 .
\end{aligned}$$

Since Φ_1, Φ_2^{-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 $f : \mathcal{S}_1 \to \mathcal{S}_2$ and $g : \mathcal{S}_2 \to \mathcal{S}_3$ are smooth maps between surfaces, then the composition

$$(g \circ f) : \mathcal{S}_1 \to \mathcal{S}_3$$

is smooth.

Proof

Fix $\mathbf{p} \in \mathcal{S}_1$ and choose charts

$$\boldsymbol{\sigma}_i : U_i \to \mathcal{S}_i$$

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, by definition the maps

$$\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1, \qquad \boldsymbol{\sigma}_3^{-1} \circ g \circ \boldsymbol{\sigma}_2,$$

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.

The inverse function of a chart is a differentiable.

Proposition 4.70: Inverse of a regular chart is smooth

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular. Then $\boldsymbol{\sigma}^{-1} : \boldsymbol{\sigma}(U) \to U$ is smooth.

Proof

First of all, note that:

- $\boldsymbol{\sigma}^{-1}$ exists, as $\boldsymbol{\sigma}$ is required to be a homeomorphism;
- $\boldsymbol{\sigma}(U)$ can be regarded as a surface, being an open subset of the surface $\mathcal{S}$.

Let $\mathbf{p} \in \boldsymbol{\sigma}(U)$ and $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathcal{S}$ be a regular chart at $\mathbf{p}$, that is,

$$p \in \tilde{\boldsymbol{\sigma}}(\widetilde{U}).$$

In order to prove that $\boldsymbol{\sigma}^{-1} : \boldsymbol{\sigma}(U) \to \mathbb{R}^2$ is a differentiable map, we need to check that the map

$$\boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}}$$

is differentiable (where it is defined). To this end, define the intersection

$$I := \boldsymbol{\sigma}(U) \cap \tilde{\boldsymbol{\sigma}}(\widetilde{U}) .$$

Clearly $I \neq \emptyset$, since $\mathbf{p} \in I$. We can then define the open sets

$$V = \boldsymbol{\sigma}^{-1}(I) , \quad \widetilde{V} = \tilde{\boldsymbol{\sigma}}^{-1}(I) ,$$

and the transition map

$$\Phi : \widetilde{V} \to V , \quad \Phi := \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}} .$$

By Theorem 4.64, the map Φ is differentiable. As $\Phi = \boldsymbol{\sigma}^{-1} \circ \tilde{\boldsymbol{\sigma}}$, the proof is concluded.

The following Theorem gives a very useful sufficient condition to check differentiability.

Theorem 4.71

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be regular surfaces. Assume:

1. $V \subseteq \mathbb{R}^3$ is open, with $\mathcal{S}_1 \subseteq V$,
2. $f : V \to \mathbb{R}^3$ is differentiable, with $f(\mathcal{S}_1) \subseteq \mathcal{S}_2$.

The restriction $f|_{\mathcal{S}_1} : \mathcal{S}_1 \to \mathcal{S}_2$ is a smooth map.

Proof

Let $\mathbf{p} \in \mathcal{S}_1$ and charts $\boldsymbol{\sigma}_1 : U_1 \to \mathcal{S}_1, \boldsymbol{\sigma}_2 : U_2 \to \mathcal{S}_2$, with

$$p \in \boldsymbol{\sigma}_1(U_1) , \qquad f(\mathbf{p}) \in \boldsymbol{\sigma}_2(U_2) .$$

The map

$$\boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}_1 : U_1 \to U_2$$

is differentiable because composition of differentiable functions: $\boldsymbol{\sigma}_2^{-1}$ is differentiable by Proposition 4.70; f is differentiable by assumption; $\boldsymbol{\sigma}_1$ is differentiable by definition of chart.

Example 4.72

Let $\mathcal{S}$ be a regular surface.

1. Assume $\mathcal{S}$ is symmetric relative to the $\{z = 0\}$ plane, that is,

$$(x, y, z) \in \mathcal{S} \quad \iff \quad (x, y, -z) \in \mathcal{S} .$$

   Then the map $f : \mathcal{S} \to \mathcal{S}$, which takes $p \in S$ into its symmetrical point, is differentiable.

This is because f is the restriction to $\mathcal{S}$ of the map

$$f : \mathbb{R}^3 \to \mathbb{R}^3, \qquad f(x,y,z) = (x,y,-z),$$

which is clearly differentiable.

2. Let $\pi : \mathcal{S} \to \mathbb{R}^2$ be the map which takes each $\mathbf{p} \in \mathcal{S}$ into its orthogonal projection over

$$\mathbb{R}^2 = \{(x,y,0) : x,y \in \mathbb{R}\}.$$

π is differentiable because restriction of the differentiable map

$$\pi : \mathbb{R}^3 \to \mathbb{R}^3, \quad \pi(x,y,z) = (x,y,0).$$

3. Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$f(x,y,z) = (xa, yb, zc),$$

where a, b, and c are non-zero real numbers. Clearly, f is differentiable. Therefore, the restriction $f|_{\mathbb{S}^2}$ is a differentiable map from the Sphere

$$\mathbb{S}^2 = \left\{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\right\}$$

into the Ellipsoid

$$\mathbb{E} = \left\{(x,y,z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\right\},$$

because $f(\mathbb{S}^2) \subseteq \mathbb{E}$.

Definition 4.73: Diffeomorphism of surfaces

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be regular surfaces.

1. $f : \mathcal{S}_1 \to \mathcal{S}_2$ is a **diffeomorphism**, if f is smooth and admits smooth inverse.
2. $\mathcal{S}_1$, $\mathcal{S}_2$ are **diffeomorphic** if there exists $f : \mathcal{S}_1 \to \mathcal{S}_2$ diffeomorphism.

The key ideas around diffeomorphisms are:

1. Two diffeomorphic surfaces are essentially the same.

   It is easy to check that being diffeomorphic is an equivalence relation on the set of regular surfaces. Therefore, two diffeomorphic surfaces can be identified.

2. Two diffeomorphic surfaces have essentially the same charts, as shown in the next Proposition.

Proposition 4.74: Image of charts under diffeomorphisms

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces, $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ diffeomorphism. If $\boldsymbol{\sigma} : U \to \mathcal{S}$ is a regular chart for $\mathcal{S}$ at $\mathbf{p}$, then

$$\tilde{\boldsymbol{\sigma}} : U \to \widetilde{\mathcal{S}}\,, \qquad \tilde{\boldsymbol{\sigma}} := f \circ \boldsymbol{\sigma}\,,$$

is a regular chart for $\widetilde{\mathcal{S}}$ 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

$$\Phi : U \to U_2\,, \qquad \Phi := \boldsymbol{\sigma}_2^{-1} \circ f \circ \boldsymbol{\sigma}\,,$$

is a diffeomorphism. Therfore

$$(f \circ \boldsymbol{\sigma})(u, v) = \boldsymbol{\sigma}_2\,(\Phi(u, v))$$

with Φ diffeomorphism, meaning that $f \circ \boldsymbol{\sigma}$ is a reparametrization of $\boldsymbol{\sigma}_2$. Since $\boldsymbol{\sigma}_2$ is regular, by Theorem 4.62 we deduce that $f \circ \boldsymbol{\sigma}$ is regular.

We conclude with the definition of local diffeomorphism between surfaces.

Definition 4.75: Local diffeomorphism

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be regular surfaces, and $f : \mathcal{S}_1 \to \mathcal{S}_2$ smooth.

1. f is a **local diffeomorphism** at $\mathbf{p} \in \mathcal{S}_1$ if:
   - There exists An open set $V \subseteq \mathcal{S}_1$ with $\mathbf{p} \in V$;
   - $f(V) \subseteq \mathcal{S}_2$ is open;
   - $f : V \to f(V)$ is smooth between surfaces.
2. f is a **local diffeomorphism** in $\mathcal{S}_1$, if it is a local diffeomorphism at each $\mathbf{p} \in \mathcal{S}_1$.
3. $\mathcal{S}_1$ is **locally diffeomorphic** to $\mathcal{S}_2$, if for all $\mathbf{p} \in \mathcal{S}_1$ there exists f local diffeomorphism at $\mathbf{p}$.

Two remarks:

1. The above definition is well-posed, since open subsets of surfaces are themselves surfaces.
2. Being locally diffeomorphic is **not** an equivalence relation: $\mathcal{S}_1$ **locally diffeomorphic** to $\mathcal{S}_2$ does not imply that $\mathcal{S}_2$ is **locally diffeomorphic** to $\mathcal{S}_1$.

4.7. Tangent plane

The tangent vector to a curve $\boldsymbol{\gamma} : \mathbb{R} \to \mathbb{R}^3$ at the point $\boldsymbol{\gamma}(t)$ is just $\dot{\boldsymbol{\gamma}}(t)$, the derivative of the curve at t. Tangent vectors to a surface $\mathcal{S}$ can be defined as the tangent vectors of curves $\boldsymbol{\gamma} : \mathbb{R} \to \mathcal{S}$ with values in $\mathcal{S}$, see Figure 4.16.

To simplify statements, we make the following assumption.

Assumption 4.76

From now on, all the surfaces will be regular and all the charts will be regular.

Definition 4.77: Tangent vectors and tangent plane

Let $\mathcal{S}$ be a surface and $\mathbf{p} \in \mathcal{S}$.

1. $\mathbf{v} \in \mathbb{R}^3$ is a **tangent vector** to $\mathcal{S}$ at $\mathbf{p}$, if there exists a smooth curve $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ such that
$$\boldsymbol{\gamma}(-\varepsilon, \varepsilon) \subseteq \mathcal{S}\,, \quad \boldsymbol{\gamma}(0) = \mathbf{p}\,, \quad \mathbf{v} = \dot{\boldsymbol{\gamma}}(0)\,.$$

2. The **tangent plane** of $\mathcal{S}$ at $\mathbf{p}$ is the set
$$T_{\mathbf{p}}\mathcal{S} := \{\mathbf{v} \in \mathbb{R}^3 \ : \ \mathbf{v} \text{ tangent vector of } \mathcal{S} \text{ at } \mathbf{p}\}\,.$$

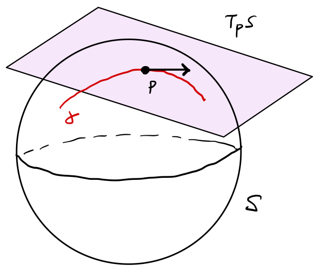


Figure 4.16.: Tangent plane $T_{\mathbf{p}}\mathcal{S}$ of surface $\mathcal{S}$ at the point $\mathbf{p}$. A tangent vector $\mathbf{v}$ has to satisfy $\mathbf{v} = \dot{\boldsymbol{\gamma}}(0)$, for some smooth curve $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}$ such that $\boldsymbol{\gamma}(0) = \mathbf{p}$.

Let us start with the most basic example: We want to compute the tangent plane to an open set in $\mathbb{R}^2$.

Example 4.78

Let $U \subseteq \mathbb{R}^2$ be open and $\mathbf{p} \in U$. Then
$$T_{\mathbf{p}}U = \mathbb{R}^2\,.$$

Proof. Suppose first that $\mathbf{v} \in T_{\mathbf{p}}U$. By definition of tangent vector, there exists a smooth curve

$$\gamma : (-\varepsilon, \varepsilon) \to U$$

such that

$$\boldsymbol{\gamma}(0) = \mathbf{p} \qquad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v} .$$

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 $B_\varepsilon(\mathbf{p}) \subseteq U$. 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

$$T_{\mathbf{p}}U = \mathbb{R}^2$$

for any open set $U \subseteq \mathbb{R}^2$. In general, if $\mathcal{S}$ is a regular surface, then $T_{\mathbf{p}}\mathcal{S}$ is a vector space isomorphic to $\mathbb{R}^2$, in symbols

$$T_{\mathbf{p}}\mathcal{S} \simeq \mathbb{R}^2 .$$

This means that there exists a map

$$\Phi : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}^2$$

which is an isomorphism of vector spaces, i.e., Φ is invertible and linear:

$$\Phi(\lambda\mathbf{v} + \mu\mathbf{w}) = \lambda\Phi(\lambda) + \mu\Phi(\mathbf{w}) ,$$

for all $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ and $\lambda, \mu \in \mathbb{R}$.

To prove this result, we need a Lemma concerning curves with values on surfaces: The lemma says that when $\mathcal{S}$ is regular, all the smooth curves $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ with values in $\mathcal{S}$, are of the form

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)) , \quad \forall\, t \in (-\varepsilon, \varepsilon) ,$$

for a pair of smooth functions $u, v : (-\varepsilon, \varepsilon) \to \mathbb{R}$.

Lemma 4.79: Curves with values on surfaces

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart and $\mathcal{S} := \boldsymbol{\sigma}(U)$. Let $\mathbf{p} \in \mathcal{S}$ and $(u_0, v_0) = \boldsymbol{\sigma}^{-1}(\mathbf{p})$. Assume $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ is a smooth curve such that

$$\boldsymbol{\gamma}(-\varepsilon, \varepsilon) \subseteq \mathcal{S}\,, \quad \boldsymbol{\gamma}(0) = \mathbf{p}\,.$$

There exist smooth functions $u, v : (-\varepsilon, \varepsilon) \to \mathbb{R}$ such that

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))\,, \ \forall\, t \in (-\varepsilon, \varepsilon)\,, \quad u(0) = u_0\,, \ v(0) = v_0\,.$$

Proof

To visualize the geometric ideas of this part of the proof, see Figure 4.17. Let $\mathbf{p}$ in $\mathcal{S} = \boldsymbol{\sigma}(U)$, and assume given a smooth curve $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ such that $\boldsymbol{\gamma}(0) = \mathbf{p}$ and

$$\boldsymbol{\gamma}(t) \in \mathcal{S}\,, \quad \forall\, t \in (-\varepsilon, \varepsilon)\,.$$

Denote the coordinates of the chart $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ by

$$\boldsymbol{\sigma}(u, v) = (f(u, v), g(u, v), h(u, v))\,.$$

The Jacobian of $\boldsymbol{\sigma}$ is

$$J\boldsymbol{\sigma} = \begin{pmatrix} f_u & f_v \\ g_u & g_v \\ h_u & h_v \end{pmatrix}.$$

Since $\boldsymbol{\sigma}$ is regular, by definition $J\boldsymbol{\sigma}$ 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

$$F : U \subseteq \mathbb{R}^2 \to \mathbb{R}^2\,, \qquad F(u, v) := (f(u, v), g(u, v))\,.$$

The Jacobian of F is

$$JF = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix},$$

which is invertible at (u_0, v_0) by assumption. Hence, by the Inverse Function Theorem 4.39, there exist

- $U_0 \subseteq U$ open set, with $(u_0, v_0) \in U_0$,
- $V \subseteq \mathbb{R}^2$ open set, with $F(u_0, v_0) \in V$,

such that

$$F : U_0 \to V$$

is a diffeomorphism. In particular, the inverse function

$$F^{-1} : V \to U_0$$

is smooth. Define the projection map

$$\pi : \mathbb{R}^3 \to \mathbb{R}^2 , \qquad \pi(x, y, z) = (x, y) ,$$

and notice that, by construction,

$$F = \pi \circ \boldsymbol{\sigma} .$$

The composition

$$\pi \circ \boldsymbol{\gamma} : (\varepsilon, \varepsilon) \to \mathbb{R}^2$$

is smooth, and such that

$$\begin{aligned}
(\pi \circ \boldsymbol{\gamma})(0) & = \pi(\boldsymbol{\gamma}(0)) \\
& = \pi(\mathbf{p}) \\
& = \pi(\boldsymbol{\sigma}(u_0, v_0)) \\
& = F(u_0, v_0) .
\end{aligned}$$

Since $F(u_0, v_0) \in V$, with V open in $\mathbb{R}^2$, and since $\pi \circ \boldsymbol{\gamma}$ is continuous, there exists $\varepsilon_0 \in (0, \varepsilon]$ such that

$$(\pi \circ \boldsymbol{\gamma})(t) \in V , \qquad \forall\, t \in (-\varepsilon_0, \varepsilon_0) .$$

Since F^{-1} maps V into U_0, it is well-defined the composition

$$\boldsymbol{\eta} : (-\varepsilon_0, \varepsilon_0) \to U_0 , \qquad \boldsymbol{\eta} := F^{-1} \circ \pi \circ \boldsymbol{\gamma} .$$

Notice that $\boldsymbol{\eta}$ is smooth, since F^{-1}, π and $\boldsymbol{\gamma}$ are smooth. In particular, the componenents of $\boldsymbol{\eta}$ are two smooth functions

$$u, v : (-\varepsilon_0, \varepsilon_0) \to \mathbb{R} ,$$

such that

$$(u(t), v(t)) = (F^{-1} \circ \pi \circ \boldsymbol{\gamma})(t) , \qquad \forall\, t \in (-\varepsilon_0, \varepsilon_0) . \tag{4.3}$$

We are now ready to conclude:

- Recalling that $F = \pi \circ \boldsymbol{\sigma}$, and that $\boldsymbol{\sigma}$ is invertible, we infer that

  $$F = \pi \circ \boldsymbol{\sigma} \quad \Longrightarrow \quad F \circ \boldsymbol{\sigma}^{-1} = \pi \quad \Longrightarrow \quad \boldsymbol{\sigma}^{-1} = F^{-1} \circ \pi .$$

  Hence, we can substitute $F^{-1} \circ \pi = \boldsymbol{\sigma}^{-1}$ in (4.3), and obtain

  $$(u(t), v(t)) = (F^{-1} \circ \pi \circ \boldsymbol{\gamma})(t) = (\boldsymbol{\sigma}^{-1} \circ \boldsymbol{\gamma})(t) .$$

  Applying $\boldsymbol{\sigma}$ to both sides gives the desired equation

  $$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)) , \qquad \forall\, t \in (-\varepsilon_0, \varepsilon_0) .$$

- We have computed that
$$(\pi \circ \boldsymbol{\gamma})(0) = F(u_0, v_0)\,.$$
In particular, substituting $t = 0$ in (4.3) gives,
$$\begin{aligned}(u(0), v(0)) &= (F^{-1} \circ \pi \circ \boldsymbol{\gamma})(0) \\ &= F^{-1}((\pi \circ \boldsymbol{\gamma})(0)) \\ &= F^{-1}(F(u_0, v_0)) \\ &= (u_0, v_0)\,,\end{aligned}$$
showing that
$$u(0) = u_0\,, \qquad v(0) = v_0\,.$$

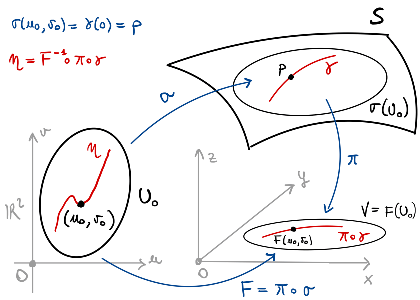


Figure 4.17.: Image associated with the Proof for Lemma 4.79. The smooth coordinates $\boldsymbol{\eta}(t) = (u(t), v(t))$ are constructed by setting $\boldsymbol{\eta} := F^{-1} \circ \pi \circ \boldsymbol{\gamma}$.

We are now ready to characterize $T_{\mathbf{p}}\mathcal{S}$ when $\mathcal{S}$ is a regular surface.

Theorem 4.80: Characterization of Tangent Plane

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart and $\mathcal{S} := \boldsymbol{\sigma}(U)$. Let $\mathbf{p} \in \mathcal{S}$. 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 $(u, v) = \boldsymbol{\sigma}^{-1}(\mathbf{p})$.

Proof

Let $\boldsymbol{\sigma} : U \to \mathcal{S}$ be a regular chart.

- First, suppose $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. By definition of tangent plane, there exists a smooth curve

$\boldsymbol{\gamma} \colon (-\varepsilon, \varepsilon) \to \mathcal{S}$ such that
$$\boldsymbol{\gamma}(0) = \mathbf{p}\,, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v}\,.$$
By Lemma 4.79, there exist $\varepsilon_0 \in (0, \varepsilon]$ and smooth functions
$$u, v \colon (-\varepsilon_0, \varepsilon_0) \to \mathbb{R}\,, \quad u(0) = u_0\,, \quad v(0) = v_0\,,$$
such that
$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))\,, \quad \forall\, t \in (-\varepsilon_0, \varepsilon_0)\,.$$
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
$$\mathbf{v} \in \operatorname{span}\{\boldsymbol{\sigma}_u(u_0, v_0), \boldsymbol{\sigma}_v(u_0, v_0)\}\,.$$

- Conversely, 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
$$\mathbf{v} = \lambda \boldsymbol{\sigma}_u(u_0, v_0) + \mu \boldsymbol{\sigma}_v(u_0, v_0)\,.$$
The map $s \colon \mathbb{R} \to \mathbb{R}^2$ defined by
$$s(t) := (u_0 + \lambda t, v_0 + \mu t)\,.$$
is continuous, being s is the line through (u_0, v_0) in the direction (λ, μ). Moreover,
$$s(0) = (u_0, v_0) \in U\,.$$
Since U is open in $\mathbb{R}^2$, there exists $R > 0$ such that
$$B_R(s(0)) = B_R((u_0, v_0)) \subseteq U\,.$$
In particualar, by continuity of s, there exists $\varepsilon > 0$ such that
$$|t - 0| < \varepsilon \quad \Longrightarrow \quad |s(t) - s(0)| < R\,,$$

which implies

$$s(t) = (u_0 + \lambda t, v_0 + \mu t) \in U \,, \quad \forall\, t \in (-\varepsilon, \varepsilon)\,.$$

Therefore, it is well-defined the curve

$$\boldsymbol{\gamma} \colon (-\varepsilon, \varepsilon) \to \boldsymbol{\sigma}(U) \subseteq \mathcal{S} \,, \qquad \boldsymbol{\gamma} := \boldsymbol{\sigma} \circ s\,.$$

Write down the definition of $\boldsymbol{\gamma}$ explicitly:

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u_0 + \lambda t, v_0 + \mu t)\,.$$

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.

Remark 4.81

The tangent plane is a vector space isomorphic to $\mathbb{R}^2$, that is,

$$T_{\mathbf{p}}\mathcal{S} \simeq \mathbb{R}^2\,,$$

with isomorphism $\Phi \colon T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}^2$ given by

$$\Phi\,(\lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v) = (\lambda, \mu)\,.$$

Proof

By Theorem 4.80 we have that

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$$

Since $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent, we conclude that $T_{\mathbf{p}}\mathcal{S}$ is a vector space of dimension 2. Therefore $T_{\mathbf{p}}\mathcal{S}$ is canonically isomorphic to $\mathbb{R}^2$ via the map $\Phi \colon T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}^2$ given by

$$\Phi\,(\lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v) = (\lambda, \mu)\,.$$

Remark 4.82

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 $\mathbf{p} \in \mathcal{S}$.

We can easily compute cartesian equations for the tangent plane.

Theorem 4.83: Equation of tangent plane

Let $\boldsymbol{\sigma} : U \to \mathcal{S}$ be regular, $\mathcal{S} = \boldsymbol{\sigma}(U)$. Let $\mathbf{p} \in \mathcal{S}$ and

$$\mathbf{n} := \boldsymbol{\sigma}_u(u,v) \times \boldsymbol{\sigma}_v(u,v) \,, \quad (u,v) := \boldsymbol{\sigma}^{-1}(\mathbf{p}) \,.$$

The equation of the tangent plane $T_{\mathbf{p}}\mathcal{S}$ is given by

$$\mathbf{x} \cdot \mathbf{n} = 0 \,, \quad \forall\, \mathbf{x} \in \mathbb{R}^3 \,.$$

Proof

By Theorem 4.80 we know that

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} \,.$$

By the properties of cross product, the vector $\mathbf{n}$ is orthogonal to both $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Therefore it is orthogonal to $T_{\mathbf{p}}\mathcal{S}$. The equation for $T_{\mathbf{p}}\mathcal{S}$ is then

$$\mathbf{x} \cdot \mathbf{n} = 0 \,, \quad \forall\, \mathbf{x} \in \mathbb{R}^3 \,.$$

Remark 4.84: Equation of affine tangent plane

The equation of the affine tangent plane $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is given by

$$(\mathbf{x} - \mathbf{p}) \cdot \mathbf{n} = 0 \,, \quad \forall\, \mathbf{x} \in \mathbb{R}^3 \,.$$

Proof

The vector $\mathbf{n}$ is orthogonalto $T_{\mathbf{p}}\mathcal{S}$. In particular, the equation for the affine tangent plane $\mathbf{p}+T_{\mathbf{p}}\mathcal{S}$

is

$$\mathbf{x} \cdot \mathbf{n} = k\,, \quad \forall\, \mathbf{x} \in \mathbb{R}^3\,,$$

for some $k \in \mathbb{R}$. To compute k, it is sufficient to evaluate the above equation at any point in $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$. Recalling that $\mathbf{0} \in T_{\mathbf{p}}\mathcal{S}$, we have that $\mathbf{p}$ belongs to $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$. Therefore,

$$k = \mathbf{p} \cdot \mathbf{n}\,.$$

Hence the equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is

$$(\mathbf{x} - \mathbf{p}) \cdot \mathbf{n} = 0\,, \quad \forall\, \mathbf{x} \in \mathbb{R}^3\,,$$

ending the proof.

Example 4.85: Calculation of tangent plane

Question. For $u \in (0, 2\pi)$, $v < 1$, let $\mathcal{S}$ charted by

$$\boldsymbol{\sigma}(u,v) = \left(\sqrt{1-v}\cos(u), \sqrt{1-v}\sin(u), v\right)\,.$$

1. Prove that $\boldsymbol{\sigma}$ charts the paraboloid $x^2 + y^2 - z = 1$.
2. Prove that $\boldsymbol{\sigma}$ is regular and compute $\mathbf{n} = \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$.
3. Give a basis for $T_{\mathbf{p}}\mathcal{S}$ at $\mathbf{p} = (\sqrt{2}/2, \sqrt{2}/2, 0)$.
4. Compute the cartesian equation of $T_{\mathbf{p}}\mathcal{S}$.

Solution.

1. Denote $\boldsymbol{\sigma}(u,v) = (x,y,z)$. We have

$$\begin{aligned} x^2 + y^2 &= \left(\sqrt{1-v}\cos(u)\right)^2 + \left(\sqrt{1-v}\sin(u)\right)^2 \\ &= 1 - v = 1 - z\,. \end{aligned}$$

2. We compute $\mathbf{n} = \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v$ and show that $\boldsymbol{\sigma}$ is regular:

$$\begin{aligned} \boldsymbol{\sigma}_u &= \left(-\sqrt{1-v}\sin(u), \sqrt{1-v}\cos(u), 0\right) \\ \boldsymbol{\sigma}_v &= \left(-\frac{1}{2}(1-v)^{-1/2}\cos(u), -\frac{1}{2}(1-v)^{-1/2}\sin(u), 1\right) \\ \mathbf{n} = \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= \left(\sqrt{1-v}\cos(u), \sqrt{1-v}\sin(u), \frac{1}{2}\right) \neq \mathbf{0} \end{aligned}$$

3. Notice that $\boldsymbol{\sigma}(\pi/4, 0) = \mathbf{p}$. A basis for $T_{\mathbf{p}}\mathcal{S}$ is

$$\boldsymbol{\sigma}_u\left(\frac{\pi}{4}, 0\right) = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 0\right),$$
$$\boldsymbol{\sigma}_v\left(\frac{\pi}{4}, 0\right) = \left(-\frac{\sqrt{2}}{4}, -\frac{\sqrt{2}}{4}, 1\right).$$

4. Using the calculation for $\mathbf{n}$ in Point 2, we find

$$\mathbf{n}\left(\frac{\pi}{4}, 0\right) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, -\frac{1}{2}\right).$$

The equation for $T_{\mathbf{p}}\mathcal{S}$ is $\mathbf{x} \cdot \mathbf{n} = 0$, which reads

$$\sqrt{2}\,x + \sqrt{2}\,y - z = 0\,.$$

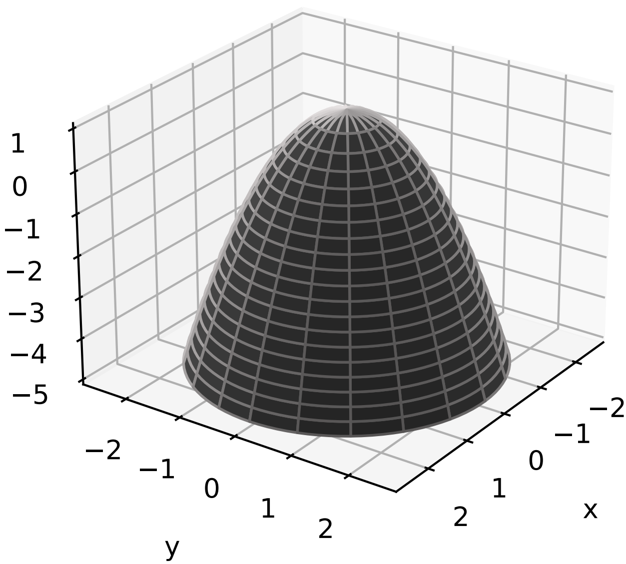


Figure 4.18.: The Paraboloid in the Example.

Remark 4.86: Tangent plane 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].

4.8. Unit normal 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:

$$T_{\mathbf{p}}\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^3 \ : \ \mathbf{x} \cdot \mathbf{N} = 0\}.$$

We will also 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 4.87

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 4.88

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. Thus, $\mathbf{N}_{\boldsymbol{\sigma}}$ 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,

$$\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi \,,$$

with $\Phi : \widetilde{U} \to U$ diffeomorphism. As stated in Thorem 4.62, 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 4.89: 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 4.90: 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
$$\det J\Phi > 0 .$$
2. $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ determine the **opposite orientations** if
$$\det J\Phi < 0 .$$

Example 4.91: Orientation of the plane

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 $\mathbf{a}$ is 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 4.92: Orientable surface

Let $\mathcal{S}$ be a regular surface. Then:

1. Let
$$\mathcal{A} = \{\boldsymbol{\sigma}_i : U_i \to \mathcal{S}\}_{i \in I}$$
be an atlas for $\mathcal{S}$. We say that $\mathcal{A}$ is **oriented**, if the following property holds:
$$\boldsymbol{\sigma}_i(U_i) \cap \boldsymbol{\sigma}_j(U_j) \neq \emptyset \quad \Longrightarrow \quad \det J\Phi > 0\,,$$
where Φ is the transition map between $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}_j$.
2. $\mathcal{S}$ is **orientable** if there exists an oriented atlas $\mathcal{A}$.
3. If an oriented atlas $\mathcal{A}$ is assigned, we say that $\mathcal{S}$ is **oriented** by $\mathcal{A}$.

Warning: Orientability is a global property

Orientability is a global property: To determine if a surface $\mathcal{S}$ is orientable, we need to examine the transition maps for the entire atlas $\mathcal{A}$.

Example 4.93: Möbius band

The classical example of non orientable surface is the Möbius band, see Figure 4.27. We will discuss this example in more details when we introduce ruled srufaces, see Example 4.118.

Example 4.94

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart. Then
$$\mathcal{S}_{\boldsymbol{\sigma}} := \boldsymbol{\sigma}(U)$$
is a regular surface with atlas $\mathcal{A} = \{\boldsymbol{\sigma}\}$. Therefore $\mathcal{S}_{\boldsymbol{\sigma}}$ is orientable.
Check: This is because we have only one chart. Therefore any transition map Φ will be the identity, so that $\det J\Phi = 1 > 0$.

Remark 4.95

Let $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ be regular charts with transition map Φ. We have seen in Remark 4.88 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 4.96: Unit normal of a surface

Let $\mathcal{S}$ be a regular surface. A **unit normal** to $\mathcal{S}$ 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 4.97

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 discussion. For a self-contained proof, we refer the reader to Proposition 4.3.7 in [1].

In view of Propostion 4.97, for an oriented surface there is a natural choice of unit normal, which we call **standard unit normal** of $\mathcal{S}$.

Definition 4.98: 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 $\boldsymbol{\sigma} \in \mathcal{A}$, where

$$\mathbf{N}_{\boldsymbol{\sigma}} \colon U \to \mathbb{R}^3\,, \quad \mathbf{N}_{\boldsymbol{\sigma}} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|}\,,$$

is the standard unit normal of the chart $\boldsymbol{\sigma}$.

Notation

In the following we will often denote by $\mathbf{N}$ both the standard unit normal of $\mathcal{S}$ and of a chart.

Example 4.99: Calculation of $\mathbf{N}$

Question. Compute the standard unit normal to

$$\boldsymbol{\sigma}(u,v) = (e^u, u+v, v)\,, \quad u,v \in \mathbb{R}\,.$$

Solution. The standard unit normal to $\boldsymbol{\sigma}$ is

$$\boldsymbol{\sigma}_u = (e^u, 1, 0)\ ,\ \boldsymbol{\sigma}_v = (0,1,1)\,, \qquad \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = \sqrt{1+2e^{2u}}$$

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (1, -e^u, e^u) \qquad \mathbf{N}_{\boldsymbol{\sigma}} = \frac{(1,-e^u,e^u)}{\sqrt{1+2e^{2u}}}$$

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

$$d_{\mathbf{p}} f : \mathbb{R}^2 \to \mathbb{R}^2$$

which approximates f locally at $\mathbf{p}$. We have seen in Example 4.78 that

$$T_{\mathbf{p}} U = \mathbb{R}^2\,.$$

Therefore we can interpret $d_{\mathbf{p}} f$ as a map between tangent planes:

$$d_{\mathbf{p}} f : T_{\mathbf{p}} U \to T_{\mathbf{p}} U\,.$$

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

$$d_{\mathbf{p}} f : T_{\mathbf{p}} \mathcal{S} \to T_{f(\mathbf{p})} \widetilde{\mathcal{S}}\,.$$

How is the above map defined explicitly? To answer this question, we need a Lemma.

Lemma 4.100

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f \colon \mathcal{S} \to \widetilde{\mathcal{S}}$ a smooth map. Let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$, and

$$\boldsymbol{\gamma} \colon (-\varepsilon, \varepsilon) \to \mathcal{S}$$

be a smooth curve such that

$$\boldsymbol{\gamma}(0) = \mathbf{p}, \quad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v}.$$

Define

$$\tilde{\boldsymbol{\gamma}} := f \circ \boldsymbol{\gamma} \colon (-\varepsilon, \varepsilon) \to \widetilde{\mathcal{S}}.$$

Then $\tilde{\boldsymbol{\gamma}}$ is a smooth curve into $\mathbb{R}^3$ and

$$\dot{\tilde{\boldsymbol{\gamma}}}(0) \in T_{f(\mathbf{p})}\widetilde{\mathcal{S}}.$$

Proof

Note that

$$\tilde{\boldsymbol{\gamma}} = i \circ f \circ \boldsymbol{\gamma},$$

with $i \colon \widetilde{\mathcal{S}} \to \mathbb{R}^3$ inclusion map. Since $i, f, \boldsymbol{\gamma}$ are smooth, we conclude that $\tilde{\boldsymbol{\gamma}} \colon (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ is smooth. Moreover

$$\tilde{\boldsymbol{\gamma}}(0) = f(\boldsymbol{\gamma}(0)) = f(\mathbf{p}).$$

By definition of tangent plane, we conclude that

$$\dot{\tilde{\boldsymbol{\gamma}}}(0) \in T_{f(\mathbf{p})}\widetilde{\mathcal{S}}.$$

Lemma 4.100 justifies the following definition of differential.

Definition 4.101: Differential of smooth function

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f \colon \mathcal{S} \to \widetilde{\mathcal{S}}$ a smooth map. The differential $d_{\mathbf{p}}f$ of f at $\mathbf{p}$ is defined as the map

$$d_{\mathbf{p}}f \colon T_{\mathbf{p}}\mathcal{S} \to T_{f(\mathbf{p})}\widetilde{\mathcal{S}}, \quad d_{\mathbf{p}}f(\mathbf{v}) := (f \circ \boldsymbol{\gamma})'(0),$$

with $\boldsymbol{\gamma} \colon (-\varepsilon, \varepsilon) \to \mathcal{S}$ smooth curve, $\boldsymbol{\gamma}(0) = \mathbf{p}$, $\dot{\boldsymbol{\gamma}}(0) = \mathbf{v}$.

We need to show that Definition 4.100 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 $\boldsymbol{\gamma}$ such that

$$\boldsymbol{\gamma}(0) = \mathbf{p}, \qquad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v}.$$

Therefore, a priori, $d_{\mathbf{p}}f(\mathbf{v})$ could depend on which curve is chosen. This is however, not the case, as shown in the next Proposition. We will also show that the map $df_{\mathbf{p}}$ is linear, and compute its matrix.

Theorem 4.102: Matrix of $d_{\mathbf{p}} f$

Let $\mathcal{S}, \widetilde{\mathcal{S}}$ be regular surfaces, and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ smooth.

1. $d_{\mathbf{p}} f(\mathbf{v})$ depends only on $f, \mathbf{p}, \mathbf{v}$ (and not on $\boldsymbol{\gamma}$).
2. $d_{\mathbf{p}} f$ is linear, that is, for all $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$ and $\lambda, \mu \in \mathbb{R}$
$$d_{\mathbf{p}} f(\lambda \mathbf{v} + \mu \mathbf{w}) = \lambda d_{\mathbf{p}} f(\mathbf{v}) + \mu d_{\mathbf{p}} f(\mathbf{w}) .$$
3. Let $\boldsymbol{\sigma} : U \to \mathcal{S}, \tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \widetilde{\mathcal{S}}$ be regular charts at $\mathbf{p}$, $f(\mathbf{p})$. Let α and β be the components of $\Psi = \tilde{\boldsymbol{\sigma}}^{-1} \circ f \circ \boldsymbol{\sigma}$, so that
$$\tilde{\boldsymbol{\sigma}}(\alpha(u,v), \beta(u,v)) = f(\boldsymbol{\sigma}(u,v)) , \quad \forall \, (u,v) \in U .$$
The matrix of $d_{\mathbf{p}} f$ 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 given by the Jacobian of the map Ψ, that is,
$$J\Psi = \begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix} .$$

Point 3 in the above Theorem says that:

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 $\mathbf{p} \in \mathcal{S}$. 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
$$(u_0, v_0) = \boldsymbol{\sigma}^{-1}(\mathbf{p}) .$$
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

$$\tilde{\boldsymbol{\sigma}}(\alpha(u,v), \beta(u,v)) = f(\boldsymbol{\sigma}(u,v)), \quad \forall\,(u,v) \in U. \tag{4.4}$$

Let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$ and denote by (λ, μ) the components of $\mathbf{v}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$, that is,

$$\mathbf{v} = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v .$$

Define the scalar functions

$$u(t) := u_0 + \lambda t, \quad v(t) := v_0 + \mu t .$$

Since U is open in $\mathbb{R}^2$, and $(u_0, v_0) \in U$, there exists a sufficiently small $\varepsilon > 0$ such that

$$(u(t), v(t)) \in U, \quad \forall\, t \in (-\varepsilon, \varepsilon).$$

In particular, it is well defined the curve

$$\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}, \quad \boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(u(t), v(t)).$$

It is immediate to check that

$$\boldsymbol{\gamma}(0) = \boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}, \quad \dot{\boldsymbol{\gamma}}(0) = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v = \mathbf{v}.$$

By definition of $\boldsymbol{\gamma}$, and 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

$$\begin{aligned}
(f \circ \boldsymbol{\gamma})'(t) &= \tilde{\boldsymbol{\sigma}}_{\tilde{u}} \frac{d}{dt}\left[\alpha(u(t), v(t))\right] + \tilde{\boldsymbol{\sigma}}_{\tilde{v}} \frac{d}{dt}\left[\beta(u(t), v(t))\right] \\
&= \tilde{\boldsymbol{\sigma}}_{\tilde{u}}\left[\alpha_u \dot{u}(t) + \alpha_v \dot{v}(t)\right] + \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\left[\beta_u \dot{u}(t) + \beta_v \dot{v}(t)\right] .
\end{aligned}$$

Noting that

$$\dot{u}(0) = \lambda, \qquad \dot{v}(0) = \mu,$$

we get

$$(f \circ \boldsymbol{\gamma})'(0) = \tilde{\boldsymbol{\sigma}}_{\tilde{u}}\left[\lambda \alpha_u + \mu \alpha_v\right] + \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\left[\lambda \beta_u + \mu \beta_v\right] .$$

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 prove the 3 points stated in the Proposition:

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 $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$ 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 $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$ 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 given by

$$\begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix} = J\Psi\,.$$

Given the above discussion, we have 2 ways of computing the differential of a smooth function

$$f \colon \mathcal{S} \to \widetilde{\mathcal{S}}$$

1. **By using the definition:** 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

$$d_{\mathbf{p}} f(\mathbf{v}) = (f \circ \boldsymbol{\gamma})'(0)\,.$$

Method to construct $\boldsymbol{\gamma}$: Following the proof of Theorem 4.102, one can proceed as follows:

- Find $\boldsymbol{\sigma} \colon U \to \mathcal{S}$ chart of $\mathcal{S}$ at $\mathbf{p}$.
- Compute

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}\,.$$

- For $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$, compute λ, μ such that

$$\mathbf{v} = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v\,.$$

- Compute $(u_0, v_0) = \boldsymbol{\sigma}^{-1}(\mathbf{p})$.

- Define the curve
$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u_0 + \lambda t, v_0 + \mu t)\,.$$
Such curve satisfies
$$\boldsymbol{\gamma}(0) = \mathbf{p}\,, \qquad \dot{\boldsymbol{\gamma}}(0) = \mathbf{v}\,.$$
- Compute
$$d_{\mathbf{p}} f(\mathbf{v}) = (f \circ \boldsymbol{\gamma})'(0)\,.$$

2. **By using the matrix representation:** Let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$
   - Find $\boldsymbol{\sigma} : U \to \mathcal{S}$ chart of $\mathcal{S}$ at $\mathbf{p}$.
   - Find $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \widetilde{\mathcal{S}}$ chart of $\widetilde{\mathcal{S}}$ at $f(\mathbf{p})$.
   - Compute the components α, β of the function
   $$\tilde{\boldsymbol{\sigma}}^{-1} \circ f \circ \boldsymbol{\sigma} : U \to \widetilde{U}\,.$$
   - The matrix of $d_{\mathbf{p}} f$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ on $T_{\mathbf{p}}\mathcal{S}$, and the basis $\{\tilde{\boldsymbol{\sigma}}_u, \tilde{\boldsymbol{\sigma}}_v\}$ on $T_{\mathbf{p}}\widetilde{\mathcal{S}}$, is given by
   $$d_{\mathbf{p}} f = J\Psi = \begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix}.$$
   - Compute λ, μ such that
   $$\mathbf{v} = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v\,.$$
   - Compute
   $$d_{\mathbf{p}} f(\mathbf{v}) = \tilde{\lambda} \tilde{\boldsymbol{\sigma}}_u + \tilde{\mu} \tilde{\boldsymbol{\sigma}}_v$$
   where the coefficients $\tilde{\lambda}$ and $\tilde{\mu}$ are
   $$\begin{pmatrix} \tilde{\lambda} \\ \tilde{\mu} \end{pmatrix} = \begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix}.$$

Let us give example calculations for both methods.

Example 4.103: Computing $d_{\mathbf{p}} f$ using the definition

Question. Consider the plane $\mathcal{S} = \{z = 0\}$, the unit cylinder $\widetilde{\mathcal{S}} = \{x^2 + y^2 = 1\}$, and the map
$$f : S \to \widetilde{\mathcal{S}}\,, \qquad f(x, y, 0) = (\cos x, \sin x, y)\,.$$

1. Compute $T_{\mathbf{p}}\mathcal{S}$.
2. Compute $d_{\mathbf{p}} f$ at $\mathbf{p} = (u_0, v_0, 0)$ and $\mathbf{v} = (\lambda, \mu, 0)$.

Solution.

1. A chart for $\mathcal{S}$ is given by $\boldsymbol{\sigma}(u, v) = (u, v, 0)$. Hence,
$$\boldsymbol{\sigma}_u = (1, 0, 0)\,, \quad \boldsymbol{\sigma}_v = (0, 1, 0)\,,$$
and the tangent space is
$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} = \{(\lambda, \mu, 0) \,:\, \lambda, \mu \in \mathbb{R}\}\,.$$

2. Define the curve $\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}$ by setting

$$\boldsymbol{\gamma}(t) := \boldsymbol{\sigma}(u_0 + t\lambda, v_0 + t\mu) = (u_0 + t\lambda, v_0 + t\mu, 0)\,.$$

Note that $\boldsymbol{\gamma}(0) = \mathbf{p}$ and $\dot{\boldsymbol{\gamma}}(0) = \mathbf{v} = (\lambda, \mu, 0)$. Therefore, the differential is given by

$$\begin{aligned}(f \circ \boldsymbol{\gamma})(t) &= (\cos(u_0 + t\lambda), \sin(u_0 + t\lambda), v_0 + t\mu)\,,\\ (f \circ \boldsymbol{\gamma})'(t) &= (-\lambda \sin(u_0 + t\lambda), \lambda \cos(u_0 + t\lambda), \mu)\,,\\ d_{\mathbf{p}} f(\mathbf{v}) &= (f \circ \boldsymbol{\gamma})'(0) = (-\lambda \sin(u_0), \lambda \cos(u_0), \mu)\,.\end{aligned}$$

Example 4.104: Computing the matrix of $d_{\mathbf{p}} f$

Question. Let $\mathcal{S}$ be the cylinder, and $\widetilde{\mathcal{S}}$ the plane, charted by

$$\boldsymbol{\sigma}(u, v) = (\cos u, \sin u, v)\,, \quad \tilde{\boldsymbol{\sigma}}(u, v) = (u, v, 0)\,,$$

defined on $U = (0, 2\pi) \times \mathbb{R}$ and $\widetilde{U} = \mathbb{R}^2$. Define the map

$$f : \mathcal{S} \to \widetilde{\mathcal{S}}\,, \quad f(x, y, z) = (y, xz, 0)\,.$$

Compute the matrix of $d_{\mathbf{p}} f$ with respect to $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ and $\{\tilde{\boldsymbol{\sigma}}_u, \tilde{\boldsymbol{\sigma}}_v\}$.
Solution. Note that $\tilde{\boldsymbol{\sigma}}^{-1}(u, v, 0) = (u, v)$. Hence,

$$\begin{aligned}\Psi(u, v) &= \tilde{\boldsymbol{\sigma}}^{-1}(f(\boldsymbol{\sigma}(u, v))) = \tilde{\boldsymbol{\sigma}}^{-1}(f(\cos u, \sin u, v))\\ &= \tilde{\boldsymbol{\sigma}}^{-1}(\sin(u), \cos(u)v, 0) = (\sin(u), \cos(u)v)\,.\end{aligned}$$

The components of Ψ are

$$\alpha(u, v) = \sin(u)\,, \quad \beta(u, v) = \cos(u)v\,.$$

The matrix of $d_{\mathbf{p}} f$ is hence

$$J\Psi = \begin{pmatrix} \alpha_u & \alpha_v \\ \beta_u & \beta_v \end{pmatrix} = \begin{pmatrix} \cos(u) & 0 \\ -\sin(u)v & \cos(u) \end{pmatrix}.$$

The differential satisfies the following useful properties.

Proposition 4.105

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

$$I : \mathcal{S} \to \mathcal{S}, \quad I(\mathbf{x}) := \mathbf{x},$$

is the identity map

$$I : T_{\mathbf{p}}(\mathcal{S}) \to T_{\mathbf{p}}(\mathcal{S}), \quad I(\mathbf{v}) := \mathbf{v}.$$

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

$$d_{\mathbf{p}}(g \circ f) = d_{f(\mathbf{p})} g \circ d_{\mathbf{p}} f, \quad \forall \mathbf{p} \in T_{\mathbf{p}} \mathcal{S}_1.$$

3. If $\mathcal{S}_1$, $\mathcal{S}_2$ are regular surfaces and

$$f : \mathcal{S}_1 \to \mathcal{S}_2,$$

is a diffeomorphism, then the differential

$$d_{\mathbf{p}} : T_{\mathbf{p}} \mathcal{S}_1 \to T_{f(\mathbf{p})} \mathcal{S}_2$$

is invertible for all $\mathbf{p} \in \mathcal{S}_1$.

For a proof see Proposition 4.4.5 in [7]. The above Proposition says that the differential of diffeomorphism is invertible. The converse statement is true locally.

Theorem 4.106

Let $\mathcal{S}_1$, $\mathcal{S}_2$ be regular surfaces and $f : \mathcal{S}_1 \to \mathcal{S}_2$ smooth. They are equivalent:

1. f is a local diffeomorphism at $\mathbf{p}$.
2. The differential $d_{\mathbf{p}} f : T_{\mathbf{p}} \mathcal{S}_1 \to T_{f(\mathbf{p})} \mathcal{S}_2$ is invertible at $\mathbf{p}$.

The proof is based on the Inverse Function Theorem 4.39, see Proposition 4.4.6 in [7].

4.10. Examples of Surfaces

In this section we discuss a few families of surfaces:

1. Level surfaces
2. Quadrics

3. Rules surfaces
4. Surfaces of revolution

4.10.1. Level surfaces

Level surfaces are described as the set of zeros of scalar functions.

Definition 4.107: Level surface

Let $f : V \to \mathbb{R}$ be smooth, $V \subseteq \mathbb{R}^3$ open. The **level surface** associated to 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. It can be found in Proposition 3.1.25 of [1]. We decide to omit it.

Theorem 4.108: Regularity of level surfaces

Let $f : V \to \mathbb{R}$ be smooth, with $V \subseteq \mathbb{R}^3$ open. Assume

$$\nabla f(x, y, z) \neq \mathbf{0} , \quad \forall \, (x, y, z) \in V .$$

Then $\mathcal{S}_f$ is a regular surface.

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 4.19, with regular atlas given by $\mathcal{A} = \{\boldsymbol{\sigma}\}$, where

$$\boldsymbol{\sigma} : \ \mathbb{R}^2 \to \mathbb{R}^3 , \quad \boldsymbol{\sigma}(u, v) = (u, v, \sqrt{u^2 + v^2}) .$$

We can also show that $\mathcal{S}^+$ is a regular surface by using Theorem 4.108.

Example 4.109: Circular cone

Question. Prove the circular cone is a regular surface

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 = z^2 \, , \, z > 0\} .$$

Solution. Define the open set $V \subset \mathbb{R}^3$ and $f : V \to \mathbb{R}$ by

$$V = \{(x, y, z) \in \mathbb{R}^3 \ : \ z > 0\} , \quad f(x, y, z) = x^2 + y^2 - z^2 .$$

$\mathcal{S}$ is a regular surface, since $\mathcal{S} = \mathcal{S}_f$ and

$$\nabla f(x, y, z) = (2x, 2y, -2z) \neq \mathbf{0}\,, \quad \forall\,(x, y, z) \in V\,.$$

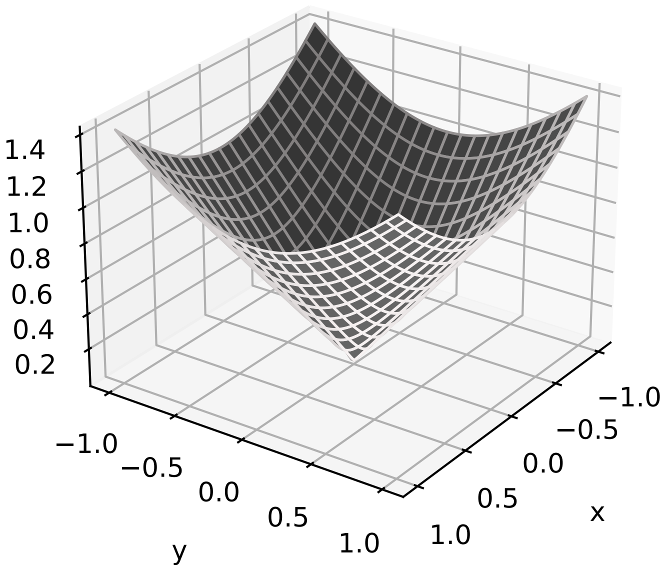


Figure 4.19.: Positive sheet of circular cone.

Let us give a characterization of the tangent plane to level surfaces.

Theorem 4.110: Tangent plane of level surfaces

Let $f : V \to \mathbb{R}$ be smooth, with $V \subseteq \mathbb{R}^3$ open. Assume

$$\nabla f(x, y, z) \neq \mathbf{0}\,, \quad \forall\,(x, y, z) \in V\,.$$

Let $\mathbf{p} \in \mathcal{S}_f$. Then $\nabla f(\mathbf{p}) \perp T_{\mathbf{p}}\mathcal{S}_f$, and

1. The cartesian equation of $T_{\mathbf{p}}\mathcal{S}_f$ is given by
$$\nabla f(\mathbf{p}) \cdot \mathbf{x} = 0\,, \quad \forall\,\mathbf{x} \in \mathbb{R}^3\,,$$
2. The cartesian equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}_f$ is given by
$$\nabla f(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) = 0\,, \quad \forall\,\mathbf{x} \in \mathbb{R}^3\,.$$

Proof

Let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}_f$. By definition of tangent plane, there exists a smooth curve

$$\boldsymbol{\gamma} : (-\varepsilon, \varepsilon) \to \mathcal{S}_f$$

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

$$\nabla f(\mathbf{p}) \perp T_{\mathbf{p}}\mathcal{S}_f.$$

In particular, the equation for $T_{\mathbf{p}}\mathcal{S}_f$ is

$$\nabla f(\mathbf{p}) \cdot \mathbf{x} = 0, \quad \forall \mathbf{x} \in \mathbb{R}^3.$$

Therefore, the equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is given by

$$\nabla f(\mathbf{p}) \cdot \mathbf{x} = k, \quad \forall \mathbf{x} \in \mathbb{R}^3,$$

for some $k \in \mathbb{R}$. Since $\mathbf{p}$ belongs to $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$, we can substitute $\mathbf{p}$ in the above equation to obtain

$$k = \nabla f(\mathbf{p}) \cdot \mathbf{p}.$$

Hence the equation for $\mathbf{p} + T_{\mathbf{p}}\mathcal{S}$ is

$$\nabla f(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) = 0, \quad \forall \mathbf{x} \in \mathbb{R}^3.$$

Example 4.111: Unit cylinder

Question. Consider the unit cylinder $\mathcal{S} = \{x^2 + y^2 = 1\}$.

1. Prove that $\mathcal{S}$ is a regular surface.
2. Find the equation of $T_{\mathbf{p}}\mathcal{S}$ at $\mathbf{p} = \left(\sqrt{2}/2, \sqrt{2}/2, 5\right)$.

Solution.

1. Define the open set $V \subseteq \mathbb{R}^3$ and $f : V \to \mathbb{R}$ by

$$V = \mathbb{R}^3 \setminus \{(0,0,z) \, : \, z \in \mathbb{R}\}, \quad f(x,y,z) := x^2 + y^2 - 1 \,.$$

$\mathcal{S}$ is a regular surface, since $\mathcal{S} = \mathcal{S}_f$ and

$$\nabla f(x,y,z) = (2x, 2y, 0) \neq \mathbf{0} \,, \quad \forall \, (x,y,z) \in V \,.$$

2. Using the expression for ∇f in Point 1, we get

$$\nabla f(\mathbf{p}) = \nabla f \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 5 \right) = (\sqrt{2}, \sqrt{2}, 0) \,.$$

The equation for $T_{\mathbf{p}}\mathcal{S}$ is

$$\nabla f(\mathbf{p}) \cdot \mathbf{x} = 0 \quad \Longleftrightarrow \quad x + y = 0 \,.$$

4.10.2. Quadrics

Quadrics are special 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 4.112

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: $\dfrac{x^2}{p^2} + \dfrac{y^2}{q^2} + \dfrac{z^2}{r^2} = 0$

We refer to Figure 4.20, Figure 4.21, Figure 4.22, Figure 4.23, Figure 4.24, Figure 4.25 for illustrations.

The proof of Theorem 4.112 follows (quite tediously) by diagonalizing the symmetric matrix A, and by studying the eigenvalues, see Theorem 5.5.2 in [7].

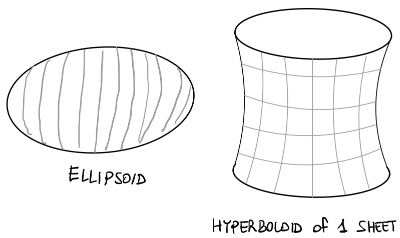


Figure 4.20.: Classification of quadrics: Ellipsoid and Hyperboloid of 1 sheet.

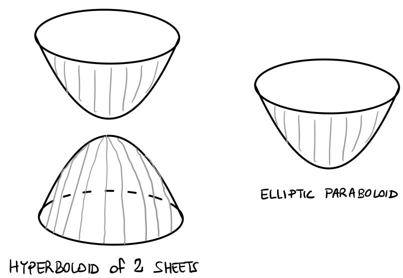


Figure 4.21.: Classification of quadrics: Hyperboloid of 2 sheets and Elliptic Paraboiloid.

Example 4.113

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

$$p = q = r = 1 .$$

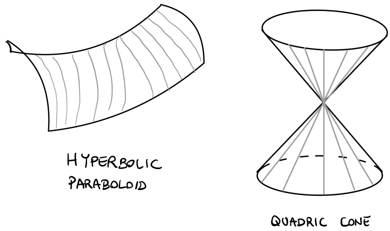


Figure 4.22.: Classification of quadrics: Hyperbolic Paraboloid and Quadric Cone.

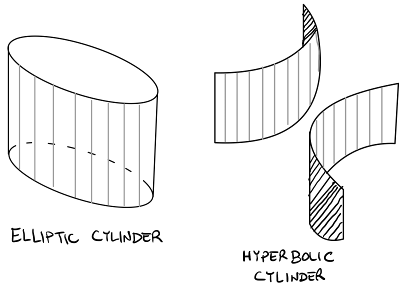


Figure 4.23.: Classification of quadrics: Elliptic Cylinder and Hyperbolic Cylinder.

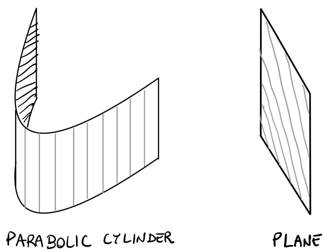


Figure 4.24.: Classification of quadrics: Parabolic Cylinder and Plane.

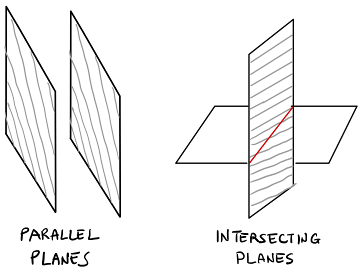


Figure 4.25.: Classification of quadrics: 2 parallel planes and 2 intersecting planes.

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 4.114

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, the surface 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 4.115: Ruled surface

A **ruled surface** is a surface with chart

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\mathbf{a}(u)\,,$$

where $\boldsymbol{\gamma}, \mathbf{a} \colon (a,b) \to \mathbb{R}^3$ are smooth curves, such that

$$\dot{\boldsymbol{\gamma}}(t) \,\text{ and }\, \mathbf{a}(t) \,\text{ are linearly independent for all }\, t \in (a,b).$$

$\boldsymbol{\gamma}$ is the **base curve** and the lines $v \mapsto v\mathbf{a}(u)$ the **rulings**.

Theorem 4.116: Regularity of ruled surfaces

A ruled surface $\mathcal{S}$ is regular if v is sufficiently small.

Proof

A chart for $\mathcal{S}$ is

$$\boldsymbol{\sigma}(u,v) = \boldsymbol{\gamma}(u) + v\mathbf{a}(u)\,,$$

with $\dot{\boldsymbol{\gamma}}$ and $\mathbf{a}$ linerly independent Differentiating, we obtain

$$\boldsymbol{\sigma}_u = \dot{\boldsymbol{\gamma}}(u) + v\dot{\mathbf{a}}(u)\,, \quad \boldsymbol{\sigma}_v = \mathbf{a}(u)\,.$$

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, depending on the choice of rulings. For example, if $\boldsymbol{\gamma}$ is a circle,

$$\boldsymbol{\gamma}(u) = (\cos(u), \sin(u), 0)\,,$$

we can obtain both the unit cylinder, and the Möbius band.

Example 4.117: Unit Cylinder is ruled surface

Question. Prove that the unit cylinder is a ruled surface.
Solution. The unit cylinder $\mathcal{S}$ is charted by

$$\begin{aligned} \boldsymbol{\sigma}(u,v) &= (\cos(u), \sin(u), v) = \boldsymbol{\gamma}(u) + v\mathbf{a}(u) \\ \boldsymbol{\gamma}(u) &= (\cos(u), \sin(u), 0)\,, \quad \mathbf{a} = (0,0,1) \end{aligned}$$

$\mathcal{S}$ is a ruled surface, since the vectors

$$\dot{\boldsymbol{\gamma}} = (-\sin(u), \cos(u), 0)\,, \qquad \mathbf{a} = (0, 0, 1)$$

are orthogonal, and hence linearly independent.

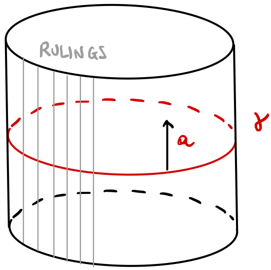


Figure 4.26.: The unit cylinder is a ruled surface with base curve the unit circle, and rulings given by vertical lines.

Example 4.118: Möbius band

Question. The Möbius band is a ruled surface with chart

$$\boldsymbol{\sigma} = \boldsymbol{\gamma}(u) + v\mathbf{a}(u)\,, \qquad u \in (0, 2\pi),\ v \in \left(-\frac{1}{2}, \frac{1}{2}\right)\,,$$

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 half rotation while going around the unit circle $\boldsymbol{\gamma}$. In particular

$$\boldsymbol{\sigma}(u,v) = \left(\left[1 - v\sin\left(\frac{u}{2}\right)\right]\cos(u), \left[1 - v\sin\left(\frac{u}{2}\right)\right]\sin(u), v\cos\left(\frac{u}{2}\right)\right)\,.$$

1. Compute the standard unit normal to $\boldsymbol{\sigma}$.
2. Prove that $\mathcal{S}$ is **non orientable**.

Solution.

1. From the formula for $\boldsymbol{\sigma}$, it is easy to compute that

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = \left(-\cos(u)\cos\left(\frac{u}{2}\right), -\sin(u)\cos\left(\frac{u}{2}\right), -\sin\left(\frac{u}{2}\right)\right)\,.$$

It is also immediate to check that $\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = 1$, and therefore the principal unit normal of $\boldsymbol{\sigma}$ is

$$\mathbf{N}_{\boldsymbol{\sigma}} = \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \, .$$

2. Suppose by contradiction that $\mathcal{S}$ is orientable. This means there exists a globally defined principal unit normal vector

$$\mathbf{N} : \ \mathcal{S} \to \mathbb{R}^3 \, .$$

By definition of principal normal, we have

$$\mathbf{N} \circ \boldsymbol{\sigma} = \mathbf{N}_{\boldsymbol{\sigma}} \, .$$

Consider the point $\mathbf{p} = (1,0,0)$ on $\mathcal{S}$. Notice that, by continuity, $\mathbf{p}$ can be obtained via $\boldsymbol{\sigma}$ through the limits

$$\mathbf{p} = \lim_{u \to 0^+} \boldsymbol{\sigma}(u,0) = \lim_{u \to 2\pi^-} \boldsymbol{\sigma}(u,0) \, .$$

Since $\mathbf{N}$ is continuous, the above implies

$$\mathbf{N}(\mathbf{p}) = \lim_{u \to 0^+} \mathbf{N} \circ \boldsymbol{\sigma}(u,0) = \lim_{u \to 2\pi^-} \mathbf{N} \circ \boldsymbol{\sigma}(u,0) \, . \tag{4.6}$$

However, by direct calculation:

$$\begin{aligned} \lim_{u \to 0^+} \mathbf{N} \circ \boldsymbol{\sigma}(u,0) &= \lim_{u \to 0^+} \mathbf{N}_{\boldsymbol{\sigma}}(u,0) = (-1,0,0) \\ \lim_{u \to 2\pi^-} \mathbf{N} \circ \boldsymbol{\sigma}(u,0) &= \lim_{u \to 2\pi^-} \mathbf{N}_{\boldsymbol{\sigma}}(u,0) = (1,0,0) \end{aligned}$$

This clearly contradicts (4.6). Therefore $\mathbf{N}$ cannot exist, and $\mathcal{S}$ is not orientable.

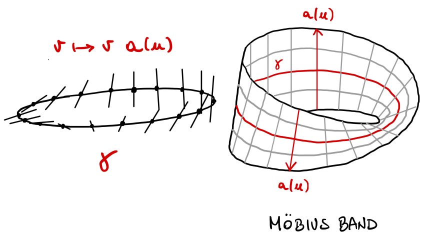


Figure 4.27.: The Möbius band is a ruled surface with base curve $\boldsymbol{\gamma}$ and rulings given by rotating vertical lines.

Example 4.119: A ruled surface

Question. Show that the following surface is ruled

$$S = \left\{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + 10xy + 16x^2 - z = 0\right\} .$$

Solution. We can rearrange

$$x^2 + 10xy + 16x^2 - z = 0 \iff (x + 8y)(x + 2y) = z .$$

Let $u = x + 8y$ and $v = x + 2y$. Therefore $uv = z$ and

$$u - v = 6y \implies y = \frac{u - v}{6} \implies x = u - 8y = \frac{4v - u}{3} .$$

It follows that if $(x, y, z) \in S$ then

$$\begin{aligned} (x, y, z) &= \left(\frac{4v - u}{3}, \frac{u - v}{6}, uv\right) \\ &= \left(-\frac{u}{3}, \frac{u}{6}, 0\right) + v\left(\frac{4}{3}, -\frac{1}{6}, u\right) = \boldsymbol{\gamma}(u) + v\mathbf{a}(u) . \end{aligned}$$

When $u \neq 0$, the vectors

$$\mathbf{a}(u) = \left(\frac{4}{3}, -\frac{1}{6}, u\right) , \quad \dot{\boldsymbol{\gamma}}(u) = \left(-\frac{1}{3}, \frac{1}{6}, 0\right) ,$$

are linearly independent, as the last component of $\dot{\boldsymbol{\gamma}}(u)$ is 0. Also $\mathbf{a}(0)$ and $\dot{\boldsymbol{\gamma}}(0)$ are linearly independent. Thus, $\mathcal{S}$ is a ruled surface.

4.10.4. Surfaces of Revolution

Surfaces of revolution are obtained by rotating a curve about the z-axis.

Definition 4.120: Surface of revolution

Let $\boldsymbol{\gamma} : (a, b) \to \mathbb{R}^3$ be a smooth curve in the (x, z)-plane,

$$\boldsymbol{\gamma}(v) = (f(v), 0, g(v)) , \qquad f > 0 .$$

The surface $\mathcal{S}$ formed by rotating $\boldsymbol{\gamma}$ about the z-axis, called a **surface of revolution**, is charted by $\boldsymbol{\sigma} : U \to \mathbb{R}^3$

$$\boldsymbol{\sigma}(u, v) = (\cos(u) f(v), \sin(u) f(v), g(v)) , \ U = (0, 2\pi) \times (a, b) .$$

Theorem 4.121: Regularity of surfaces of revolution

A surface of revolution is regular if and only if $\boldsymbol{\gamma}$ is regular.

Proof

We have

$$
\begin{aligned}
\boldsymbol{\sigma}_u &= (-\sin(u) f(v), \cos(u) f(v), 0) \\
\boldsymbol{\sigma}_v &= \left(\cos(u) \dot{f}(v), \sin(u) \dot{f}(v), \dot{g}(v)\right) \\
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= \left(\cos(u) f \dot{g}, \sin(u) f \dot{g}, -f \dot{f}\right) , \\
\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| &= f \sqrt{\left(\dot{f}^2 + \dot{g}^2\right)} = f \, \|\dot{\boldsymbol{\gamma}}\| \, .
\end{aligned}
$$

Now, $\boldsymbol{\sigma}$ is regular if and only if $\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v \neq \mathbf{0}$. Since $f > 0$ by definition, we conclude that $\boldsymbol{\sigma}$ is regular if and only if $\dot{\boldsymbol{\gamma}} \neq \mathbf{0}$, that is, $\boldsymbol{\gamma}$ is regular.

Example 4.122: Catenoid is surface of revolution

Question. The Catenoid $\mathcal{S}$ is the surface of revolution formed by rotating the catenary $\boldsymbol{\gamma}(v) = (\cosh(v), 0, v)$ about the z-axis. A chart for $\mathcal{S}$ is given by

$$
\boldsymbol{\sigma}(u, v) = (\cos(u) \cosh(v), \sin(u) \cosh(v), v) \, ,
$$

with $u \in (0, 2\pi)$, $v \in \mathbb{R}$. Prove that $\mathcal{S}$ is a regular surface.
Solution. Note that $f > 0$. $\mathcal{S}$ is regular because $\boldsymbol{\gamma}$ is regular, as

$$
\dot{\boldsymbol{\gamma}} = (\sinh(v), 0, 1) \, , \quad \|\dot{\boldsymbol{\gamma}}\|^2 = 1 + \sinh(v)^2 \geq 1 \, .
$$

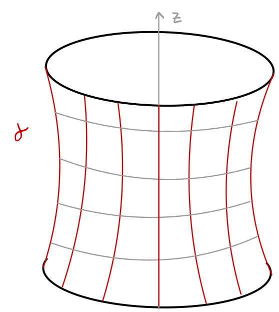


Figure 4.28.: The Catenoid is the surface of revolution obtained by rotating the catenary about the z-axis.

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

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 $\mathbf{q}$, that is, the planar distance between $\mathbf{p}$ and $\mathbf{q}$. 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

$$\boldsymbol{\gamma} : (a,b) \to \mathcal{S}$$

is a smooth curve such that

$$\boldsymbol{\gamma}(a) = \mathbf{p}, \quad \boldsymbol{\gamma}(b) = \mathbf{q} .$$

The distance between $\mathbf{p}$ and $\mathbf{q}$ on $\mathcal{S}$ could be defined as the length of $\boldsymbol{\gamma}$, i.e.,

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

Since $\boldsymbol{\gamma}(t) \in \mathcal{S}$, by definition of tangent plane, we have

$$\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}} S , \quad \mathbf{p} := \boldsymbol{\gamma}(t) .$$

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 4.123: First fundamental form (FFF)

Let $\mathcal{S}$ be a regular surface and $\mathbf{p} \in \mathcal{S}$. The **first fundamental form (FFF)** of $\mathcal{S}$ at $\mathbf{p}$ is the bilinear symmetric map

$$I_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{S} \times T_{\mathbf{p}}\mathcal{S} \to \mathbb{R} , \quad I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) := \mathbf{v} \cdot \mathbf{w} .$$

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 $T_{\mathbf{p}}\mathcal{S}$,
- The first fundamental form of a chart.

Definition 4.124: Coordinate functions on tangent plane

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular, $\mathcal{S} = \boldsymbol{\sigma}(U)$. The **coordinate functions** on $T_{\mathbf{p}}\mathcal{S}$ are the linear maps

$$du, dv : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}, \quad du(\mathbf{v}) := \lambda, \quad dv(\mathbf{v}) := \mu,$$

where $\mathbf{v} = \lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v$, since $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis for $T_{\mathbf{p}}\mathcal{S}$.

Definition 4.125: FFF of a chart

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular, $\mathcal{S} = \boldsymbol{\sigma}(U)$. Define $E, F, G : U \to \mathbb{R}$

$$E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u, \quad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v, \quad G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v.$$

The **FFF** of $\boldsymbol{\sigma}$ is the quadratic form $\mathcal{F}_1 : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$

$$\mathcal{F}_1(\mathbf{v}) = E\, du^2(\mathbf{v}) + 2F\, du(\mathbf{v})\, dv(\mathbf{v}) + G\, dv^2(\mathbf{v}), \quad \forall\, v \in T_{\mathbf{p}}\mathcal{S},$$

for all $\mathbf{p} \in \boldsymbol{\sigma}(U)$, with E, F, G evaluated at $(u, v) = \boldsymbol{\sigma}^{-1}(\mathbf{p})$.

We usually omit the dependence on $\mathbf{v}$ in (**?@eq-fff-chart**), and just write

$$\mathcal{F}_1 = E\, du^2 + 2F\, du\, dv + G\, dv^2.$$

We are now ready to write $I_{\mathbf{p}}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$.

Theorem 4.126: Matrix of FFF

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular, $\mathcal{S} = \boldsymbol{\sigma}(U)$, 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, it holds

$$\mathcal{F}_1(\mathbf{v}) = I_{\mathbf{p}}(\mathbf{v}, \mathbf{v}), \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}.$$

Proof

By Theorem 4.80, 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

$$\mathbf{v} = \lambda_1 \boldsymbol{\sigma}_u + \mu_1 \boldsymbol{\sigma}_v\,, \quad \mathbf{w} = \lambda_2 \boldsymbol{\sigma}_u + \mu_2 \boldsymbol{\sigma}_v\,.$$

Therefore,

$$\begin{aligned}
I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) &= \mathbf{v} \cdot \mathbf{w} \\
&= \lambda_1 \lambda_2 \, \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v + (\lambda_1 \mu_2 + \lambda_2 \mu_1)\, \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v + \mu_1 \mu_2 \, \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v \\
&= E\, du(\mathbf{v}) du(\mathbf{w}) + F\, (du(\mathbf{v})\, dv(\mathbf{w}) + du(\mathbf{w}) dv(\mathbf{v})) \\
&\qquad + G\, dv(\mathbf{v}) dv(\mathbf{w}) \\
&= (du(\mathbf{v}), dv(\mathbf{v})) \begin{pmatrix} E & F \\ F & G \end{pmatrix} (du(\mathbf{w}), dv(\mathbf{w}))^T\,.
\end{aligned}$$

The fact that

$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{v}) = \mathcal{F}_1(\mathbf{v})$$

follows from the first part of the statement and definition of $\mathcal{F}_1$.

Remark 4.127: Linear algebra interpretation

Using linear algebra, Theorem 4.126 has the following clear interpretation: $I_{\mathbf{p}}$ 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 $\mathcal{F}_1$ the 2×2 matrix

$$\mathcal{F}_1 := \begin{pmatrix} E & F \\ F & G \end{pmatrix}.$$

Example 4.128: FFF of Unit cylinder

Question. 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 FFF of $\boldsymbol{\sigma}$ is

$$\mathscr{F}_1 = du^2 + dv^2\,.$$

Solution. We have

$$\begin{aligned} \boldsymbol{\sigma}_u &= (-\sin(u), \cos(u), 0) & F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\ \boldsymbol{\sigma}_v &= (0,0,1) & G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1 \\ E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 & \mathscr{F}_1 &= du^2 + dv^2 \end{aligned}$$

Warning

The first fundamental form $I_{\mathbf{p}}$ depends only on the surface $\mathcal{S}$ and the point $\mathbf{p}$. Instead the local representation of $I_{\mathbf{p}}$

$$\mathscr{F}_1 = E\,du^2 + 2F\,dudv + G\,dv^2\,,$$

depends on the choice of chart $\boldsymbol{\sigma} : U \to \mathbb{R}^3$. The next result explains how $\mathscr{F}_1$ changes when we change chart.

Proposition 4.129: FFF and reparametrizations

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular, and $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3$ a reparametrization, with $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi$ and $\Phi : \widetilde{U} \to U$ diffeomorphism.

1. The matrices $\mathscr{F}_1$ and $\widetilde{\mathscr{F}}_1$ of the FFF of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are related by

$$\widetilde{\mathscr{F}}_1 = (J\Phi)^T\, \mathscr{F}_1\, J\Phi\,, \quad \mathscr{F}_1 = \begin{pmatrix} E & F \\ F & G \end{pmatrix}, \quad \widetilde{\mathscr{F}}_1 = \begin{pmatrix} \widetilde{E} & \widetilde{F} \\ \widetilde{F} & \widetilde{G} \end{pmatrix}.$$

2. The linear maps du, dv and $d\tilde{u}, d\tilde{v}$ are related by

$$\begin{aligned} du &= \frac{\partial u}{\partial \tilde{u}}\, d\tilde{u} + \frac{\partial u}{\partial \tilde{v}}\, d\tilde{v} \\ dv &= \frac{\partial v}{\partial \tilde{u}}\, d\tilde{u} + \frac{\partial v}{\partial \tilde{v}}\, d\tilde{v} \end{aligned}$$

Proof

The pairs $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u\}$ and $\{\tilde{\boldsymbol{\sigma}}_{\tilde{u}}, \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\}$ are both bases for the vector space $T_{\mathbf{p}}\mathcal{S}$. By the chain rule, we have

$$\begin{aligned}\tilde{\boldsymbol{\sigma}}_{\tilde{u}} &= \boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{u}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{u}} \\ \tilde{\boldsymbol{\sigma}}_{\tilde{v}} &= \boldsymbol{\sigma}_u \frac{\partial u}{\partial \tilde{v}} + \boldsymbol{\sigma}_v \frac{\partial v}{\partial \tilde{v}}\end{aligned}$$

The above show that the change of basis matrix between $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_u\}$ and $\{\tilde{\boldsymbol{\sigma}}_{\tilde{u}}, \tilde{\boldsymbol{\sigma}}_{\tilde{v}}\}$ is given exactly by $J\Phi$. Therefore, the two formulas in the Proposition are consequence of change of basis results for bilinear forms and linear maps, respectively.

Example 4.130: FFF of Plane

Question. Let $\mathbf{a}, \mathbf{p}, \mathbf{q} \in \mathbb{R}^3$, with $\mathbf{p}, \mathbf{q}$ orthonormal. The plane in cartesian and polar coordinates is charted by, respectively,

$$\begin{aligned}\boldsymbol{\sigma}(u,v) &= \mathbf{a} + u\mathbf{p} + v\mathbf{q}\,, \quad (u,v) \in \mathbb{R}^2\,, \\ \tilde{\boldsymbol{\sigma}}(\rho,\theta) &= \mathbf{a} + \rho\cos(\theta)\mathbf{p} + \rho\sin(\theta)\mathbf{q}\,, \quad \rho > 0,\ \theta \in (0,2\pi)\,.\end{aligned}$$

1. Show that the FFF of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are

$$\mathcal{F}_1 = du^2 + dv^2\,, \qquad \widetilde{\mathcal{F}}_1 = d\rho^2 + \rho^2 d\theta^2\,.$$

2. Let Φ be the change of variables from polar to cartesian coordinates. Show that

$$\widetilde{\mathcal{F}}_1 = (J\Phi)^T\, \mathcal{F}_1\, J\Phi\,.$$

Solution.

1. Using that $\mathbf{p}$ and $\mathbf{q}$ are orthonormal,

$$\begin{aligned}
\boldsymbol{\sigma}_u &= \mathbf{p}\,, & \tilde{\boldsymbol{\sigma}}_\rho &= \cos(\theta)\mathbf{p} + \sin(\theta)\mathbf{q} \\
\boldsymbol{\sigma}_v &= \mathbf{q} & \tilde{\boldsymbol{\sigma}}_\theta &= -\rho\sin(\theta)\mathbf{p} + \rho\cos(\theta)\mathbf{q} \\
E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u &= 1 & \widetilde{E} = \tilde{\boldsymbol{\sigma}}_\rho \cdot \tilde{\boldsymbol{\sigma}}_\rho &= 1 \\
F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v &= 0 & \widetilde{F} = \tilde{\boldsymbol{\sigma}}_\rho \cdot \tilde{\boldsymbol{\sigma}}_\theta &= 0 \\
G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v &= 1 & \widetilde{G} = \tilde{\boldsymbol{\sigma}}_\theta \cdot \tilde{\boldsymbol{\sigma}}_\theta &= r^2 \\
\mathcal{F}_1 &= du^2 + dv^2 & \widetilde{\mathcal{F}}_1 &= d\rho^2 + \rho^2 d\theta^2
\end{aligned}$$

2. We have $\Phi(\rho,\theta) = (\rho\cos(\theta), \rho\sin(\theta))$. Then

$$
\begin{aligned}
(J\Phi)^T \, \mathscr{F}_1 \, J\Phi &= (J\Phi)^T \, J\Phi \\
&= \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\rho\sin(\theta) & \rho\cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\rho\sin(\theta) \\ \sin(\theta) & \rho\cos(\theta) \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & \rho^2 \end{pmatrix} = \widetilde{\mathscr{F}}_1 \, .
\end{aligned}
$$

Remark 4.131

As seen in Example 4.129, when the plane is charted in cartesian coordinates, the FFF is essentially the Pythagorean Theorem on the plane: In fact, a basis for $T_{\mathbf{p}}\mathcal{S}$ is

$$
T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} = \{\lambda\mathbf{p} + \mu\mathbf{q} \; : \; \lambda, \mu \in \mathbb{R}\} \, .
$$

Therefore, for

$$
\mathbf{v} \in T_{\mathbf{p}}\mathcal{S} \, , \qquad \mathbf{v} = \lambda\mathbf{p} + \mu\mathbf{q} \, ,
$$

we have

$$
\|\mathbf{v}\|^2 = \mathscr{F}_1(\mathbf{v}) = du^2(\mathbf{v}) + dv^2(\mathbf{v}) = \lambda^2 + \mu^2 \, .
$$

Hence, the square of the length of the vector $\mathbf{v}$, which has coordinates λ, μ in the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$, is equal to $\lambda^2 + \mu^2$.

Remark 4.132

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, angles and areas are the same on the two surfaces.

Example 4.133

Question. Find the FFF of the surface chart

$$
\boldsymbol{\sigma}(u,v) = \left(u - v, u + v, u^2 + v^2\right) \, .
$$

Solution. We compute

$$
\begin{aligned}
\boldsymbol{\sigma}_u &= (1, 1, 2u) & F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 4uv \\
\boldsymbol{\sigma}_v &= (-1, 1, 2v) & G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 2\left(1 + 2v^2\right) \\
E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v &= 2\left(1 + 2u^2\right) & \mathscr{F}_1 &= 2 \begin{pmatrix} 1 + 2u^2 & 2uv \\ 2uv & 1 + 2v^2 \end{pmatrix}.
\end{aligned}
$$

4.11.1. Length of curves

The first fundamental form allows to compute the length of curves with values on surfaces.

Proposition 4.134: Length of curves and FFF

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular, $\mathcal{S} = \boldsymbol{\sigma}(U)$. Let $\boldsymbol{\gamma} : (a, b) \to \mathcal{S}$ be a smooth curve. Then

$$
\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)),
$$

for some smooth functions $u, v : (a, b) \to \mathbb{R}$, and

$$
\int_a^b \|\dot{\boldsymbol{\gamma}}(t)\| \, dt = \int_a^b \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2} \, dt,
$$

where $\dot{u}, \dot{v}$ are computed at t, and E, F, G at $(u(t), v(t))$.

Proof

Since $\boldsymbol{\gamma}$ takes values into $\boldsymbol{\sigma}(U)$, by Lemma 4.79 there exist smooth functions u, v such that

$$
\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)), \quad \forall\, t \in (a, b).
$$

By chain rule we have

$$
\dot{\boldsymbol{\gamma}}(t) = \dot{u}(t)\boldsymbol{\sigma}_u(u(t), v(t)) + \dot{v}(t)\boldsymbol{\sigma}_v(u(t), v(t)).
$$

Therefore,

$$
\begin{aligned}
\|\dot{\boldsymbol{\gamma}}(t)\|^2 &= \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} \\
&= (\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \cdot (\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \\
&= E\,\dot{u}^2 + 2F\,\dot{u}\dot{v} + G\,\dot{v}^2.
\end{aligned}
$$

Integrating, we obtain the thesis.

Example 4.135: Curves on the Cone

Question. Consider the cone with chart

$$\boldsymbol{\sigma}(u,v) = (\cos(u)v, \sin(u)v, v)\,, \quad u \in (0,2\pi),\ v > 0\,.$$

1. Compute the first fundamental form of $\boldsymbol{\sigma}$.
2. Compute the length of $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(t,t)$ for $t \in (\pi/2, \pi)$.

Solution.

1. The first fundamental form of $\boldsymbol{\sigma}$ is

$$\begin{aligned} \boldsymbol{\sigma}_u &= (-\sin(u)v, \cos(u)v, 0) & F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\ \boldsymbol{\sigma}_v &= (\cos(u), \sin(u), 1) & G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 2 \\ E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = v^2 & \mathcal{F}_1 &= v^2\, du^2 + 2\, dv^2 \end{aligned}$$

2. $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ with $u(t) = t$ and $v(t) = t$. Then

$$\begin{aligned} \dot{u} &= 1\,,\ \ \dot{v} = 1 & F(u(t),v(t)) &= F(t,t) = 0 \\ E(u(t),v(t)) &= E(t,t) = t^2 & G(u(t),v(t)) &= G(t,t) = 2 \end{aligned}$$

The length of $\boldsymbol{\gamma}$ between $\pi/2$ and π is

$$\int_{\pi/2}^{\pi} \|\dot{\boldsymbol{\gamma}}(t)\|\, dt = \int_{\pi/2}^{\pi} \sqrt{t^2 + 2}\, dt\,.$$

4.11.2. Isometries

Isometries are an important class of maps between surfaces: They are smooth maps which preserve the first fundamental form.

Definition 4.136: Local Isometry and Isometry

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ smooth. We say that:

1. f is a **local isometry**, if for all $\mathbf{p} \in \mathcal{S}$

$$\mathbf{v} \cdot \mathbf{w} = d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{w})\,, \quad \forall\, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}} \mathcal{S}\,.$$

In this case, $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are said to be **locally isometric**.

2. f is an **isometry** if:

- f is a local isometry;
- f is a diffeomorphism of $\mathcal{S}$ into $\widetilde{\mathcal{S}}$.

In this case, $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are said to be **isometric**.

Recall that the first fundamental form of $\mathcal{S}$ is defined by

$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot \mathbf{w}\,, \quad \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}\,.$$

Therefore, condition (**?@eq-loc-iso**) reads

$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) = I_{\mathbf{p}}(d_{\mathbf{p}} f(\mathbf{v}), d_{\mathbf{p}} f(\mathbf{w}))\,, \quad \forall\, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}\,.$$

In this sense, we see that local isometries preserve the first fundamental form.

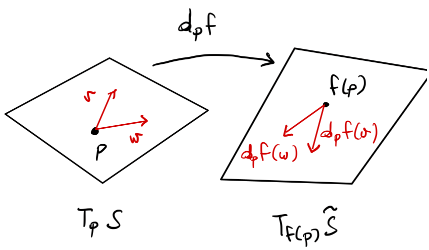


Figure 4.29.: Sketch of a local isometry f between $\mathcal{S}$ and $\widetilde{\mathcal{S}}$. The scalar product between tangent vectors $\mathbf{v}$ and $\mathbf{w}$ is preserved by $d_{\mathbf{p}} f$.

Remark 4.137

A smooth map $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ is a local isometry if and only if

$$\mathbf{v} \cdot \mathbf{v} = d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{v})\,, \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}\,.$$

Proof. The thesis follows immediately 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 are automatically local diffeomorphisms.

Proposition 4.138

Local isometries are local diffeomorphims.

Proof

Let $\mathbf{p} \in \mathcal{S}$. Assume by contradiction that the differential of f

$$d_{\mathbf{p}} f \colon T_{\mathbf{p}} \mathcal{S} \to T_{f(\mathbf{p})} \widetilde{\mathcal{S}} ,$$

is not invertible. As $d_{\mathbf{p}} f$ is linear, this implies $d_{\mathbf{p}} f$ is not injective. Therefore

$$\ker \left(d_{\mathbf{p}} f \right) \neq \{\mathbf{0}\} ,$$

meaning that there exists $\mathbf{v} \in T_{\mathbf{p}} \mathcal{S}$ such that

$$d_{\mathbf{p}} f(\mathbf{v}) = \mathbf{0} , \qquad \mathbf{v} \neq \mathbf{0} .$$

Using that f is a local isometry, we get

$$\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v} = d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{v}) = 0 ,$$

which implies $\mathbf{v} = \mathbf{0}$. This is a contradiction, and therefore $d_{\mathbf{p}} f$ has to be invertible. By Theorem 4.106, we conclude that f is a local diffeomorphism at $\mathbf{p}$. As $\mathbf{p} \in \mathcal{S}$ is arbitrary, we infer that f is a local diffeomorphism of $\mathcal{S}$ into $\widetilde{\mathcal{S}}$.

Remark 4.139

We have just seen that local isometries are also local diffeomorphisms. By definition, isometries are local isometries which are also diffeomorphisms. As such:

1. Isometries are local isometries,
2. Local isometries are not isometries

This is because, in general, local diffeomorphisms are not global diffeomorphisms, see Example 4.145.

Local isometries preserve the length of curves, as shown in the following Proposition.

Theorem 4.140: Local isometries preserve lengths

Let $\mathcal{S}, \widetilde{\mathcal{S}}$ be regular surfaces, $f \colon \mathcal{S} \to \widetilde{\mathcal{S}}$ smooth. Equivalently:

1. f is a local isometry.
2. Let $\boldsymbol{\gamma}$ be a curve on $\mathcal{S}$ and define 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 $\boldsymbol{\gamma} : (a, b) \to \mathcal{S}$ is a smooth curve. Consider the smooth curve

$$\tilde{\boldsymbol{\gamma}} : (a, b) \to \widetilde{\mathcal{S}} , \qquad \tilde{\boldsymbol{\gamma}} := f \circ \boldsymbol{\gamma} .$$

Set $\mathbf{p} := \boldsymbol{\gamma}(t)$, so that

$$\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}}\mathcal{S} .$$

By definition of differential of a function between surfaces, we have

$$d_{\mathbf{p}} f(\dot{\boldsymbol{\gamma}}(t)) = \dot{\tilde{\boldsymbol{\gamma}}}(t) .$$

Using that f is a local isometry gives:

$$\begin{aligned} \|\dot{\tilde{\boldsymbol{\gamma}}}(t)\|^2 & = \dot{\tilde{\boldsymbol{\gamma}}}(t) \cdot \dot{\tilde{\boldsymbol{\gamma}}}(t) \\ & = d_{\mathbf{p}} f(\dot{\boldsymbol{\gamma}}(t)) \cdot d_{\mathbf{p}} f(\dot{\boldsymbol{\gamma}}(t)) \\ & = \dot{\boldsymbol{\gamma}}(t) \cdot \dot{\boldsymbol{\gamma}}(t) \\ & = \|\dot{\boldsymbol{\gamma}}(t)\|^2 \end{aligned}$$

Therefore $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ have the same length:

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

Part 2. We need to prove that f is a local isometry. By Remark 4.137, it is sufficient to show that

$$d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{v}) = \mathbf{v} \cdot \mathbf{v} , \quad \forall \, \mathbf{v} \in T_{\mathbf{p}}(\mathcal{S}) . \tag{4.7}$$

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

$$\tilde{\boldsymbol{\gamma}} : (-\varepsilon, \varepsilon) \to \widetilde{\mathcal{S}} , \qquad \tilde{\boldsymbol{\gamma}} := f \circ \boldsymbol{\gamma} .$$

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 equality is true for each $\varepsilon > 0$, and the functions being 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

$$d_{\mathbf{p}} f(\mathbf{v}) = \dot{\tilde{\boldsymbol{\gamma}}}(0) .$$

Therefore

$$\begin{aligned} d_{\mathbf{p}} f(\mathbf{v}) \cdot d_{\mathbf{p}} f(\mathbf{v}) &= \dot{\tilde{\boldsymbol{\gamma}}}(0) \cdot \dot{\tilde{\boldsymbol{\gamma}}}(0) \\ &= \dot{\boldsymbol{\gamma}}(0) \cdot \dot{\boldsymbol{\gamma}}(0) \\ &= \mathbf{v} \cdot \mathbf{v} . \end{aligned}$$

As $\mathbf{v}$ was arbitrary, we conclude (4.7).

By definition, local isometries preserve the first fundamental form. The next Theorem gives a practical method to check if a map is a local isometry.

Theorem 4.141: Local isometries preserve FFF

Let $\mathcal{S}, \widetilde{\mathcal{S}}$ be regular surfaces, $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ smooth. Equivalently:

1. f is a local isometry.
2. Let $\boldsymbol{\sigma} : U \to \mathcal{S}$ be regular chart of $\mathcal{S}$, and define a chart of $\widetilde{\mathcal{S}}$ as $\tilde{\boldsymbol{\sigma}} : U \to \widetilde{\mathcal{S}}$, with $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$. Then $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same FFF

$$E = \widetilde{E} , \quad F = \widetilde{F} , \quad G = \widetilde{G} .$$

Note: 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

$$\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma} .$$

Since f is a local isometry, then it is also a local diffeomorhims by Proposition 4.138. In particular, Proposition 4.74 ensures that $\tilde{\boldsymbol{\sigma}}$ is a regualar chart of $\widetilde{\mathcal{S}}$ at $f(\mathbf{p})$. Now, recall the statement of Theorem 4.102: 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

$$\begin{aligned} d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) &= 1 \cdot \tilde{\boldsymbol{\sigma}}_u + 0 \cdot \tilde{\boldsymbol{\sigma}}_v = \tilde{\boldsymbol{\sigma}}_u \\ d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) &= 0 \cdot \tilde{\boldsymbol{\sigma}}_u + 1 \cdot \tilde{\boldsymbol{\sigma}}_v = \tilde{\boldsymbol{\sigma}}_v \end{aligned}$$

Using that f is a local isometry gives

$$\begin{aligned} E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) \cdot d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) \\ &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u = \widetilde{E}\,, \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = d_{\mathbf{p}} f(\boldsymbol{\sigma}_u) \cdot d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) \\ &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v = \widetilde{F}\,, \\ G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) \cdot d_{\mathbf{p}} f(\boldsymbol{\sigma}_v) \\ &= \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v = \widetilde{G}\,, \end{aligned}$$

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 Theorem 4.102 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

$$\begin{aligned} \mathbf{v}\cdot\mathbf{v} &= (\lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v)\cdot(\lambda\boldsymbol{\sigma}_u + \mu\boldsymbol{\sigma}_v) \\ &= \lambda^2(\boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_u) + 2\lambda\mu(\boldsymbol{\sigma}_u\cdot\boldsymbol{\sigma}_v) + \mu^2(\boldsymbol{\sigma}_v\cdot\boldsymbol{\sigma}_v) \\ &= \lambda^2(\tilde{\boldsymbol{\sigma}}_u\cdot\tilde{\boldsymbol{\sigma}}_u) + 2\lambda\mu(\tilde{\boldsymbol{\sigma}}_u\cdot\tilde{\boldsymbol{\sigma}}_v) + \mu^2(\tilde{\boldsymbol{\sigma}}_v\cdot\tilde{\boldsymbol{\sigma}}_v) \\ &= (\lambda\tilde{\boldsymbol{\sigma}}_u + \mu\tilde{\boldsymbol{\sigma}}_v)\cdot(\lambda\tilde{\boldsymbol{\sigma}}_u + \mu\tilde{\boldsymbol{\sigma}}_v) \\ &= d_{\mathbf{p}}f(\mathbf{v})\cdot d_{\mathbf{p}}f(\mathbf{v})\,, \end{aligned}$$

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 4.137 we conclude that f is a local isometry.

Remark 4.142

To prove that a smooth map

$$f\colon \mathcal{S} \to \widetilde{\mathcal{S}}$$

is a local isometry, it is sufficient to verify Condition 2 in Theorem 4.141 on one atlas of $\mathcal{S}$. Then, automatically, all other atlases will verify the condition, ensuring that f is a local isometry. To see this, suppose we can verify Condition 2 on the chart $\boldsymbol{\sigma}\colon U \to \mathcal{S}$, that is, suppose we have proven

$$\widetilde{\mathcal{F}}_1 = \mathcal{F}_1\,, \tag{4.8}$$

where $\widetilde{\mathcal{F}}_1$ is the first fundamental form of $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$. Assume that $\hat{\boldsymbol{\sigma}}\colon \hat{U} \to \mathcal{S}$ is another chart of $\mathcal{S}$. By Remark 4.129, the fundamental forms of $\hat{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}$ are related by

$$\hat{\mathcal{F}}_1 = (J\Phi)^T \mathcal{F}_1 J\Phi \tag{4.9}$$

where Φ is the transition map $\Phi\colon \hat{U} \to U$ such that

$$\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi\,.$$

Applying f to $\hat{\boldsymbol{\sigma}}$ gives

$$\boldsymbol{\sigma}^{\dagger} := f \circ \hat{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma} \circ \Phi = \tilde{\boldsymbol{\sigma}} \circ \Phi\,,$$

showing that $\boldsymbol{\sigma}^\dagger$ is a reparametrization of $\tilde{\boldsymbol{\sigma}}$ with transition map Φ. Therefore, by Proposition 4.129, the fundamental forms of $\boldsymbol{\sigma}^\dagger$ and $\tilde{\boldsymbol{\sigma}}$ are related by

$$\mathcal{F}_1^\dagger = (J\Phi)^T \widetilde{\mathcal{F}}_1 J\Phi$$

Recalling (4.8) and (4.9), we get

$$\begin{aligned}\mathcal{F}_1^\dagger &= (J\Phi)^T \widetilde{\mathcal{F}}_1 J\Phi \\ &= (J\Phi)^T \mathcal{F}_1 J\Phi \\ &= \hat{\mathcal{F}}_1 .\end{aligned}$$

We have therefore proven that $\hat{\boldsymbol{\sigma}}$ and $\boldsymbol{\sigma}^\dagger = f \circ \hat{\boldsymbol{\sigma}}$ have the same first fundamental form. Thus $\hat{\boldsymbol{\sigma}}$ satisfies Condition 2 in Theorem 4.141.

Sometimes, we wish to determine if two surfaces $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are locally isometric, but it is not clear how to construct a map

$$f \colon \mathcal{S} \to \widetilde{\mathcal{S}} .$$

As an alternative, we can shift the problem of finding f, to the problem of finding suitable charts $\boldsymbol{\sigma}$ of $\mathcal{S}$, and $\tilde{\boldsymbol{\sigma}}$ of $\widetilde{\mathcal{S}}$, such that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same first fundamntal form. This is detailed in the next Theorem.

Theorem 4.143: Sufficient condition for local isometry

Let $\mathcal{S}, \widetilde{\mathcal{S}}$ be regular surfaces, with charts $\boldsymbol{\sigma} \colon U \to \mathcal{S}$ and $\tilde{\boldsymbol{\sigma}} \colon U \to \widetilde{\mathcal{S}}$. Assume that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same FFF. We have

1. The surfaces $\boldsymbol{\sigma}(U)$ and $\widetilde{\mathcal{S}}$ are locally isometric.
2. A local isometry is given by

$$f \colon \boldsymbol{\sigma}(U) \to \widetilde{\mathcal{S}} , \qquad f = \tilde{\boldsymbol{\sigma}} \circ \boldsymbol{\sigma}^{-1} .$$

Note: $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are defined on the same open set U. Therefore, E, F, G and $\widetilde{E}, \widetilde{F}, \widetilde{G}$ are defined on U, and the equality is intended pointwise.

Proof

Define $f = \tilde{\boldsymbol{\sigma}} \circ \boldsymbol{\sigma}^{-1}$, and notice that

$$f \circ \boldsymbol{\sigma} = \tilde{\boldsymbol{\sigma}} \circ \boldsymbol{\sigma}^{-1} \circ \boldsymbol{\sigma} = \tilde{\boldsymbol{\sigma}} .$$

Therefore, by assumption, $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same first fundamental form. Note that the chart $\boldsymbol{\sigma}$ gives an atlas for the surface $\boldsymbol{\sigma}(U)$. Hence, by Theorem 4.143, we conclude that f is a local isometry between $\boldsymbol{\sigma}(U)$ and $\widetilde{\mathcal{S}}$.

Example 4.144: Plane and Cylinder are locally isometric

Question. Consider the plane $\mathcal{S} = \{x = 0\}$ and the unit cylinder $\widetilde{\mathcal{S}} = \{x^2 + y^2 = 1\}$. Define the function

$$f : \mathcal{S} \to \widetilde{\mathcal{S}}, \qquad f(0, y, z) = (\cos(y), \sin(y), z).$$

Prove that f is a local isometry (you may assume f smooth).
Solution. The plane $\mathcal{S}$ is charted by

$$\boldsymbol{\sigma}(u, v) = (0, u, v), \quad u, v \in \mathbb{R}.$$

We already know that $\boldsymbol{\sigma}$ is regular, with FFF coefficients

$$E = 1, \ F = 0, \ G = 1 \qquad \Longrightarrow \qquad \mathcal{F}_1 = du^2 + dv^2.$$

Define $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$. Therefore,

$$\tilde{\boldsymbol{\sigma}}(u, v) = f(0, u, v) = (\cos(u), \sin(u), v).$$

The FFF of $\tilde{\boldsymbol{\sigma}}$ is

$$\begin{aligned} \tilde{\boldsymbol{\sigma}}_u &= (-\sin(u), \cos(u), 0) & \qquad \widetilde{F} &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v = 0 \\ \tilde{\boldsymbol{\sigma}}_v &= (0, 0, 1) & \widetilde{G} &= \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v = 1 \\ \widetilde{E} &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u = 1 & \widetilde{\mathcal{F}}_1 &= du^2 + dv^2 \end{aligned}$$

Thus, $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same FFF. Since $\mathcal{A} = \{\boldsymbol{\sigma}\}$ is an atlas for $\mathcal{S}$, by Theorem 4.143 we conclude that f is a local isometry of $\mathcal{S}$ into $\widetilde{\mathcal{S}}$.

Example 4.145: Plane and Cylinder are not isometric

Consider again the plane

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \ : \ x = 0\},$$

and the unit cylinder

$$\widetilde{\mathcal{S}} = \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 = 1\}.$$

We have seen in Example 4.144 that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are locally isometric. However, $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are **not isometric**.
Proof. The surfaces $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are nor homeomorphic, and therefore they cannot be isometric (and hence diffeomorphic). We cannot rigorously prove this claim with our current knowledge of topology. However, to give some intuition, here is a sketch of the argument, see Figure 4.30:

- Any simple closed curve $\boldsymbol{\gamma}$ in the plane $\mathcal{S}$ can be shrunk continuously into a point without leaving $\mathcal{S}$. In this case we say that $\mathcal{S}$ is **simply connected**.
- If $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ were to be homeomorphic, then $\widetilde{\mathcal{S}}$ would be simply connected.
- However, a parallel $\boldsymbol{\gamma}$ of the Cylinder $\widetilde{\mathcal{S}}$ cannot be shrunk continuously into a point without leaving the Cylinder. Thus, $\widetilde{\mathcal{S}}$ is not simply connected.
- Hence, the Plane and the Cylinder cannot be homeomorphic.

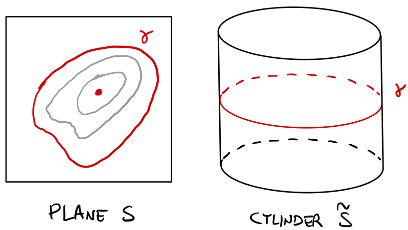


Figure 4.30.: A simple closed curve $\boldsymbol{\gamma}$ in the plane $\mathcal{S}$ can be shrunk continuously into a point without leaving $\mathcal{S}$. A parallel $\boldsymbol{\gamma}$ of the Cylinder $\widetilde{\mathcal{S}}$ cannot be shrunk continuously into a point without leaving $\widetilde{\mathcal{S}}$.

Example 4.146: Plane and Cone are locally isometric

Question. Consider the cone without tip

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \ : \ x^2 + y^2 = z^2 \,, \ z > 0\} \,,$$

and the plane $\widetilde{\mathcal{S}} = \{z = 0\}$.

1. Compute the FFF of the chart of the Cone

$$\boldsymbol{\sigma} : U \to \mathcal{S} \,, \qquad \boldsymbol{\sigma}(\rho, \theta) = (\rho\cos(\theta), \rho\sin(\theta), \rho) \,,$$
$$U = \left\{(\rho, \theta) \in \mathbb{R}^2 \ : \ \rho > 0, \ \theta \in (0, 2\pi)\right\} .$$

2. Compute the FFF of the chart of the plane

$$\tilde{\boldsymbol{\sigma}} : U \to \widetilde{\mathcal{S}} \,, \qquad \tilde{\boldsymbol{\sigma}}(\rho, \theta) = (a\rho\cos(b\theta), a\rho\sin(b\theta), 0) \,,$$

where $a > 0$ and $b \in (0, 1]$ are constants.

3. Prove that $f = \tilde{\boldsymbol{\sigma}} \circ \boldsymbol{\sigma}^{-1}$ is a local isometry between $\mathcal{S}$ and $\widetilde{\mathcal{S}}$, for suitable coefficients a, b.

Solution.

1. As seen in Example 4.135, the coefficients of the FFF of $\boldsymbol{\sigma}$ are

$$E = 2\,, \qquad F = 0\,, \qquad G = \rho^2\,.$$

2. Note that $\tilde{\boldsymbol{\sigma}}$ is well defined for all $(\rho, \theta) \in U$, as

$$\theta \in (0, 2\pi), \quad b \in (0, 1] \quad \Longrightarrow \quad b\theta \in (0, 2\pi)\,.$$

The coefficients of the FFF of $\tilde{\boldsymbol{\sigma}}$ are

$$\tilde{\boldsymbol{\sigma}}_\rho = a\,(\cos(b\theta), \sin(b\theta), 0) \qquad \widetilde{F} = \tilde{\boldsymbol{\sigma}}_\rho \cdot \tilde{\boldsymbol{\sigma}}_\theta = 0$$
$$\tilde{\boldsymbol{\sigma}}_\theta = ab\rho\,(-\sin(b\theta), \cos(b\theta), 0) \qquad \widetilde{G} = \tilde{\boldsymbol{\sigma}}_\theta \cdot \tilde{\boldsymbol{\sigma}}_\theta = a^2 b^2 \rho^2$$
$$\widetilde{E} = \tilde{\boldsymbol{\sigma}}_\rho \cdot \tilde{\boldsymbol{\sigma}}_\rho = a^2$$

3. Imposing that $\widetilde{E} = E$, $\widetilde{F} = F$ and $\widetilde{G} = G$, we obtain

$$a^2 = 2,\ a^2 b^2 = 1 \quad \Longrightarrow \quad a = \sqrt{2}\,,\ b = \frac{1}{\sqrt{2}}\,.$$

Note that $a > 0$ and $0 < b < 1$, showing that a, b are admissible. Hence, for $a = \sqrt{2}$ and $b = 1/\sqrt{2}$, the charts $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ have the same FFF. By Theorem 4.141, we conclude that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are locally isometric, with local isometry given by $f = \tilde{\boldsymbol{\sigma}} \circ \boldsymbol{\sigma}^{-1}$.

4.11.3. Angles on surfaces

We want to define the notion of angle between tangent vectors.

Definition 4.147: 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

$$\theta(\mathbf{v}, \mathbf{w}) \in [0, \pi]\,,$$

such that

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

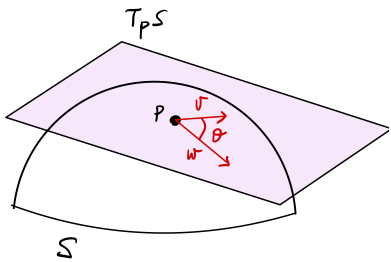


Figure 4.31.: Sketch of angle θ between two vectors $\mathbf{v}, \mathbf{w}$ in $T_{\mathbf{p}}\mathcal{S}$.

The angle between tangent vectors can be computed in terms of local charts.

Proposition 4.148

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 Theorem 4.126 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

$$\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v} = E\lambda^2 + 2F\lambda\mu + G\mu^2$$
$$\|\mathbf{w}\|^2 = \mathbf{w} \cdot \mathbf{w} = E\tilde{\lambda}^2 + 2F\tilde{\lambda}\tilde{\mu} + G\tilde{\mu}^2 .$$

Substituting in (4.10) we conclude.

4.11.4. Angles 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 4.149: Angle between curves

Let $\mathcal{S}$ be a regular surface, and $\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}$ curves on $\mathcal{S}$ intersecting at

$$\boldsymbol{\gamma}(t_0) = \mathbf{p} = \tilde{\boldsymbol{\gamma}}(t_0) .$$

The angle θ between $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ is

$$\cos(\theta) = \frac{\dot{\boldsymbol{\gamma}}(t_0) \cdot \dot{\tilde{\boldsymbol{\gamma}}}(t_0)}{\|\dot{\boldsymbol{\gamma}}(t_0)\| \, \|\dot{\tilde{\boldsymbol{\gamma}}}(t_0)\|} .$$

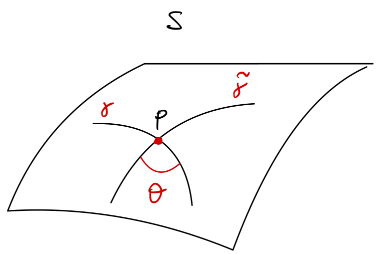


Figure 4.32.: Sketch of angle θ between two curves $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ on $\mathcal{S}$.

Theorem 4.150: Angle between curves and FFF

Let $\mathcal{S}$ be a regular surface, $\boldsymbol{\sigma}$ regular chart at $\mathbf{p}$, and $\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}$ curves on $\mathcal{S}$ intersecting at $\boldsymbol{\gamma}(t_0) = \mathbf{p} = \tilde{\boldsymbol{\gamma}}(t_0)$. There exist smooth functions $u, v, \tilde{u}, \tilde{v}$ such that

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)), \quad \tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}(\tilde{u}(t), \tilde{v}(t)).$$

The angle between $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ is

$$\cos(\theta) = \frac{E\dot{u}\dot{\tilde{u}} + F(\dot{u}\dot{\tilde{v}} + \dot{\tilde{u}}\dot{v}) + G\dot{v}\dot{\tilde{v}}}{(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2}(E\dot{\tilde{u}}^2 + 2F\dot{\tilde{u}}\dot{\tilde{v}} + G\dot{\tilde{v}}^2)^{1/2}},$$

with E, F, G evaluated at $(u(t_0), v(t_0))$, and $\dot{u}, \dot{v}, \dot{\tilde{u}}, \dot{\tilde{v}}$ at t_0.

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 4.79 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,

$$\begin{aligned} \dot{\boldsymbol{\gamma}} \cdot \dot{\tilde{\boldsymbol{\gamma}}} &= (\dot{u}\boldsymbol{\sigma}_u + \dot{v}\boldsymbol{\sigma}_v) \cdot (\dot{\tilde{u}}\boldsymbol{\sigma}_u + \dot{\tilde{v}}\boldsymbol{\sigma}_v) \\ &= E\dot{u}\dot{\tilde{u}} + F(\dot{u}\dot{\tilde{v}} + \dot{\tilde{u}}\dot{v}) + G\dot{v}\dot{\tilde{v}}. \end{aligned}$$

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.

Example 4.151: Calculation of angle between curves

Question. Let S be a surface charted by

$$\boldsymbol{\sigma}(u,v) = (u, v, e^{uv}) \ .$$

1. Calculate the FFF of $\boldsymbol{\sigma}$.
2. Calculate $\cos(\theta)$, where θ is the angle between the two curves

$$\begin{aligned} \boldsymbol{\gamma}(t) &= \boldsymbol{\sigma}(u(t), v(t)), \quad u(t) = t, \ v(t) = t \,, \\ \tilde{\boldsymbol{\gamma}}(t) &= \boldsymbol{\sigma}(\tilde{u}(t), \tilde{v}(t)) \,, \quad \tilde{u}(t) = 1, \ \tilde{v}(t) = t \,. \end{aligned}$$

Solution.

1. The coefficients of the FFF are

$$\begin{aligned} \boldsymbol{\sigma}_u &= (1, 0, e^{uv} v) & F(u,v) &= e^{2uv} uv \\ \boldsymbol{\sigma}_v &= (0, 1, e^{uv} u) & G(u,v) &= 1 + e^{2uv} u^2 \\ E(u,v) &= 1 + e^{2uv} v^2 \end{aligned}$$

2. $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ intersect at $\boldsymbol{\gamma}(1) = \tilde{\boldsymbol{\gamma}}(1) = \boldsymbol{\sigma}(1,1)$. We compute

$$\begin{aligned} \dot{u}(1) &= 1 & E(1,1) &= 1 + e^2 \\ \dot{v}(1) &= 1 & F(1,1) &= e^2 \\ \dot{\tilde{u}}(1) &= 0 & G(1,1) &= 1 + e^2 \\ \dot{\tilde{v}}(1) &= 1 \end{aligned}$$

Therefore, the angle θ satisfies

$$\cos(\theta) = \frac{1 + 2e^2}{\sqrt{2 + 4e^2}\sqrt{1 + e^2}} = \sqrt{\frac{1 + 2e^2}{2 + 2e^2}} \,.$$

4.11.5. 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 4.152: Conformal map

Let $\mathcal{S}, \widetilde{\mathcal{S}}$ be regular surfaces, $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ local diffeomorphism. We say that f is a **conformal map**, if for all $\mathbf{p} \in \mathcal{S}$

$$\theta = \tilde{\theta}, \quad \forall\, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S},$$

- θ is the angle between $\mathbf{v}$ and $\mathbf{w}$,
- $\tilde{\theta}$ is the angle between $d_{\mathbf{p}}f(\mathbf{v})$ and $d_{\mathbf{p}}f(\mathbf{w})$.

In this case, we say that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are **conformal**.

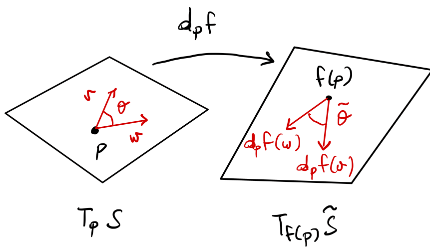


Figure 4.33.: Sketch of conformal map f between $\mathcal{S}$ and $\widetilde{\mathcal{S}}$. The angles between tangent vectors are preserved by $d_{\mathbf{p}}f$.

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 4.153

We have that f is a conformal map if and only if

$$\frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \, \|\mathbf{w}\|} = \frac{\langle \mathbf{v}, \mathbf{w} \rangle_f}{\|\mathbf{v}\|_f \, \|\mathbf{w}\|_f}, \quad \forall\, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}.$$

Proof. Follows immediately by the definition of angle between vectors.

Proposition 4.154

Local isometries are conformal maps.

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 4.159 below. Before giving the example, let us provide a characterization of conformal maps in terms of the first fundamental form.

Theorem 4.155

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ a local diffeomorphism. They are equivalent:

1. f is a conformal map.
2. There exists a function $\lambda : \mathcal{S} \to \mathbb{R}$ such that

$$\langle \mathbf{v}, \mathbf{w} \rangle_f = \lambda(\mathbf{p}) \, \langle \mathbf{v}, \mathbf{w} \rangle \,, \quad \forall \, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S} \,.$$

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

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v} \rangle &= v_1^2 \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 \rangle + 2v_1 v_2 \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle + v_2^2 \langle \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_2 \rangle \\ &= v_1^2 + v_2^2 , \end{aligned}$$

where we used that $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are orthonormal. On the other hand,

$$\begin{aligned} \langle \mathbf{v}, \mathbf{v} \rangle_f &= v_1^2 \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1 \rangle_f + 2v_1 v_2 \langle \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \rangle_f + v_2^2 \langle \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_2 \rangle_f \\ &= v_1^2\, \lambda(\mathbf{p}) + 2v_1 v_2\, \mu(\mathbf{p}) + v_2^2\, \nu(\mathbf{p}) \\ &= \lambda(\mathbf{p})\, (v_1^2 + v_2^2) , \end{aligned}$$

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 4.137 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 $\lambda \colon \mathcal{S} \to \mathbb{R}$ such that

$$\langle \mathbf{v}, \mathbf{w} \rangle_f = \lambda(\mathbf{p})\, \langle \mathbf{v}, \mathbf{w} \rangle , \quad \forall\, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S} .$$

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.

The following result gives a practical necessary and sufficient condition to check if a map is conformal.

Theorem 4.156: Conformal maps and FFF

Let $\mathcal{S}, \widetilde{\mathcal{S}}$ be regular surfaces, $f \colon \mathcal{S} \to \widetilde{\mathcal{S}}$ a local diffeomorphism. Equivalently:

1. f is a conformal map.
2. Let $\boldsymbol{\sigma} : U \to \mathcal{S}$ be regular chart of $\mathcal{S}$, and define a chart of $\widetilde{\mathcal{S}}$ as $\tilde{\boldsymbol{\sigma}} : U \to \widetilde{\mathcal{S}}$, with $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$. Then the FFF of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ satisfy

$$\widetilde{\mathcal{F}}_1 = \lambda(u,v)\mathcal{F}_1 \,, \quad \forall\,(u,v) \in U \,,$$

for some smooth map $\lambda : U \to \mathbb{R}$.

Theorem 4.156 can be easily proven by using Theorem 4.155, and by adapting the argument in the proof of Theorem 4.141. We omit the proof.

Remark 4.157

To prove that a diffeomorphism

$$f : \mathcal{S} \to \widetilde{\mathcal{S}}$$

is a conformal map, it is sufficient to verify Condition 2 in Theorem 4.156 on one atlas of $\mathcal{S}$. Then, automatically, all other atlases will verify the condition, ensuring that f is a conformal map.
Proof. This can be seen by adapting the argument found in Remark 4.142.

The following result gives a sufficient condition to prove that two regular surfaces are conformal.

Theorem 4.158: Sufficient condition for conformality

Let $\mathcal{S}, \widetilde{\mathcal{S}}$ be regular surfaces, with charts $\boldsymbol{\sigma} : U \to \mathcal{S}$ and $\tilde{\boldsymbol{\sigma}} : U \to \widetilde{\mathcal{S}}$. Assume that $\widetilde{\mathcal{F}}_1 = \lambda \mathcal{F}_1$ for some $\lambda : U \to \mathbb{R}$. We have

1. The surfaces $\boldsymbol{\sigma}(U)$ and $\widetilde{\mathcal{S}}$ are conformal.
2. A conformal map is given by

$$f : \boldsymbol{\sigma}(U) \to \widetilde{\mathcal{S}} \,, \qquad f = \tilde{\boldsymbol{\sigma}} \circ \boldsymbol{\sigma}^{-1} \,.$$

The proof follows by adapting the argument in the proof of Theorem 4.143, and by applying Theorem 4.156. This is left as an exercise.

Example 4.159: Stereographic Projection

Question. Consider the unit sphere $\mathbb{S}^2 = \{x^2 + y^2 + z^2 = 1\}$ and define the surface

$$\mathcal{S} = \mathbb{S}^2 \setminus \{N\} \,, \quad N = (0,0,1) \,.$$

Consider the plane $\widetilde{\mathcal{S}} = \{z = 0\}$. The plane $\widetilde{\mathcal{S}}$ slices through the equator of the sphere. Let

$P = (x, y, z)$ be any point on $\mathbb{S}^2$, except the North Pole N. The line joining N to P intersects the plane $\widetilde{\mathcal{S}}$ at the point P', see Figure 4.34. The point P' defines the *Stereographic Projection* map, which is easily computed to be:

$$f : \mathcal{S} \to \widetilde{\mathcal{S}}, \quad f(x, y, z) = \left(\frac{x}{1-z}, \frac{y}{1-z}, 0 \right) .$$

Prove that:

1. f is a conformal map.
2. f is not a local isometry.

Note: In particular, the Sphere and the Plane are conformal.
Solution. It is easy to prove that $f^{-1} = \boldsymbol{\sigma}$, with

$$\boldsymbol{\sigma}(u, v) = \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, 1 - \frac{2}{u^2 + v^2 + 1} \right) .$$

It is straightforward to compute that the FFF of $\boldsymbol{\sigma}$ is

$$\mathcal{F}_1 = \lambda(u, v)(du^2 + dv^2) , \quad \lambda(u, v) = \frac{4}{(u^2 + v^2 + 1)^2} .$$

Let $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$. Since $\boldsymbol{\sigma} = f^{-1}$, we have that

$$\tilde{\boldsymbol{\sigma}}(u, v) = (u, v, 0) .$$

As already computed, the FFF of $\tilde{\boldsymbol{\sigma}}$ is

$$\widetilde{\mathcal{F}}_1 = du^2 + dv^2 .$$

We can now conclude:

1. We have that
$$\widetilde{\mathcal{F}}_1 = \frac{1}{\lambda} \mathcal{F}_1 .$$
Since $\mathcal{A} = \{\boldsymbol{\sigma}\}$ is an atlas for $\mathcal{S}$, by Theorem 4.156 we conclude that f is a conformal map.

2. Since λ is not always equal to 1, we have that
$$\widetilde{\mathcal{F}}_1 \neq \mathcal{F}_1 .$$
Therefore, by Theorem 4.141, we conclude that f cannot be a local isometry.

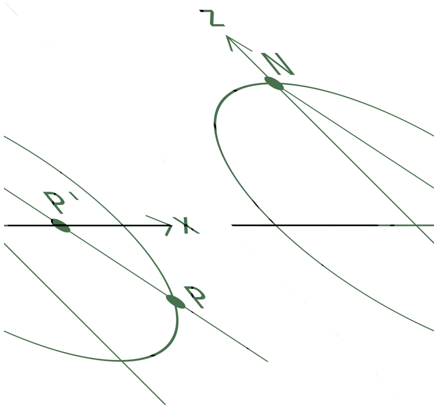


Figure 4.34.: Stereographic projection map from the North pole $N = (0, 0, 1)$. Let $P \in \mathbb{S}^2$. The line through N and P intersects the plane $\{z = 0\}$ at the point P'.

4.11.6. Conformal parametrizations

In the previous section we defined conformal maps between surfaces

$$f : \mathcal{S} \to \widetilde{\mathcal{S}} .$$

It is also useful to define conformal parametrizations.

Definition 4.160: Conformal parametrization

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular. We say that $\boldsymbol{\sigma}$ is a **conformal parametrization** if the FFF of $\boldsymbol{\sigma}$ satisfies

$$\mathcal{F}_1 = \lambda(u, v)(du^2 + dv^2) ,$$

for some smooth function $\lambda : U \to \mathbb{R}$.

The above definition is motivated by the following result, stating that $\mathcal{S}$ admits a conformal parametrization if and only if $\mathcal{S}$ is conformal to a plane.

Theorem 4.161

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular. Let $\boldsymbol{\pi}$ be the plane $\{z = 0\}$. They are equivalent:

1. $\boldsymbol{\sigma}$ is a conformal parametrization.
2. The map $f : \boldsymbol{\pi} \to \boldsymbol{\sigma}(U)$ defined by

$$f(u, v, 0) = \boldsymbol{\sigma}(u, v)$$

is conformal.

Proof

The plane $\boldsymbol{\pi}$ is charted by

$$\tilde{\boldsymbol{\sigma}}(u,v) = (u,v,0)\,, \quad (u,v) \in \mathbb{R}^2$$

The first fundamental form of $\tilde{\boldsymbol{\sigma}}$ is

$$\widetilde{\mathcal{F}}_1 = du^2 + dv^2\,.$$

Consider the map $f : \boldsymbol{\pi} \to \mathcal{S}$ given in Point 2, that is,

$$f(u,v,0) = \boldsymbol{\sigma}(u,v)\,.$$

By definition of $\tilde{\boldsymbol{\sigma}}$,

$$f(\tilde{\boldsymbol{\sigma}}(u,v)) = \boldsymbol{\sigma}(u,v)\,.$$

Theorem 4.156 says that f is a conformal map if and only if there exists $\lambda : U \to \mathbb{R}$ such that

$$\mathcal{F}_1 = \lambda(u,v)\widetilde{\mathcal{F}}_1\,,$$

where $\mathcal{F}_1$ is the first fundamental form of $\boldsymbol{\sigma}$. This happens exactly when $\boldsymbol{\sigma}$ is a conformal parametrization.

As an immediate consequence, we have the following Theorem.

Theorem 4.162: Conformal parametrizations preserve angles

Let $\boldsymbol{\sigma}$ be a conformal parametrization, and $\boldsymbol{\gamma}_1(t), \boldsymbol{\gamma}_2(t)$ be curves in $\mathbb{R}^2$ such that $\dot{\boldsymbol{\gamma}}_1(t_0), \dot{\boldsymbol{\gamma}}_2(t_0)$ make angle θ. Let $\tilde{\boldsymbol{\gamma}}_1 = \boldsymbol{\sigma} \circ \boldsymbol{\gamma}_1$ and $\tilde{\boldsymbol{\gamma}}_2 = \boldsymbol{\sigma} \circ \boldsymbol{\gamma}_2$. Then $\dot{\tilde{\boldsymbol{\gamma}}}_1(t_0), \dot{\tilde{\boldsymbol{\gamma}}}_2(t_0)$ also make angle θ.

It turns out that all regular surfaces admit an atlas formed by conformal charts.

Theorem 4.163

Let $\mathcal{S}$ be a regular surface. There exists an atlas $\mathcal{A}$ such that each chart $\boldsymbol{\sigma} \in \mathcal{A}$, $\boldsymbol{\sigma} : U \to \mathcal{S}$ is conformal, that is, there exists $\lambda : U \to \mathbb{R}$ such that the first fundamental form of $\boldsymbol{\sigma}$ is

$$\mathcal{F}_1 = \lambda(u,v)(du^2 + dv^2)\,.$$

The coordinates (u,v) are called **isothermal**.

The proof of this result is quite technical, see for example the paper [2], which can be read here. The proof in not constructive, and it involeves showing the existence of solutions to a certain PDE. If $\mathcal{S}$ is a minimial surface, existence of isothermal coordinates is easier to prove, see for example Theorem 12.4.1 in [7].

We have already seen examples of conformal parametrization for the Cylinder:

Example 4.164: Unit cylinder

Question. Prove that the following is a conformal parametrization of the unit cylinder

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v), \quad u \in (0, 2\pi), \ v \in \mathbb{R}.$$

Solution. We have already computed that the FFF of $\boldsymbol{\sigma}$ is

$$\mathscr{F}_1 = du^2 + dv^2.$$

Therefore $\boldsymbol{\sigma}$ is a conformal parametrization, with $\lambda = 1$.

An important application of conformal parametrizations is in **cartography**.

Remark 4.165: Drawing World Maps

Suppose given a parametrization of $\mathbb{S}^2$

$$\boldsymbol{\sigma} : U \to \mathbb{S}^2,$$

for some $U \subseteq \mathbb{R}^2$. Set $V := \boldsymbol{\sigma}(U)$. The inverse map

$$\boldsymbol{\sigma}^{-1} : V \subseteq \mathbb{S}^2 \to \mathbb{R}^2$$

is a projection of $\mathbb{S}^2$ into the plane. If $\mathbb{S}^2$ models the Earth, $\boldsymbol{\sigma}^{-1}$ gives a rule to draw a world map! For a map to be useful, we would like it to

- preserve angles and shapes;
- preserve areas.

We know that angles are preserved if $\boldsymbol{\sigma}^{-1}$ is a conformal map, and $\boldsymbol{\sigma}$ a conformal parametrization. Below we discuss two conformal maps of $\mathbb{S}^2$:

- Stereographic Projection,
- Mercator Projection.

In the next sections we will also discuss an equiareal map of of $\mathbb{S}^2$, i.e., a map which preserves areas. This will be called

- Lambert Cylindrical Projection.

We will also see that it is not possible for a map $\boldsymbol{\sigma}$ to both preserve angles and areas. This is because $\boldsymbol{\sigma}$ would be both conformal and equiareal, and hence an isometry of the sphere into the plane. However, isometries between sphere and plane cannot exist.

Example 4.166: Stereographic Projection

Consider the parametrization of the sphere $\mathbb{S}^2$ given by

$$\boldsymbol{\sigma}(u,v) = \left(\frac{2u}{u^2+v^2+1}, \frac{2v}{u^2+v^2+1}, 1 - \frac{2}{u^2+v^2+1} \right) .$$

We have seen in Example 4.159 that $\boldsymbol{\sigma}$ is the inverse of the stereographic projection map. We have also mentioned that the FFF of $\boldsymbol{\sigma}$ is

$$\mathcal{F}_1 = \frac{4}{(u^2+v^2+1)^2}(du^2 + dv^2) .$$

Therefore, $\boldsymbol{\sigma}$ is a conformal parametrization of the sphere.
Note. Let $V = \boldsymbol{\sigma}(U)$. In this case

$$V = \mathbb{S}^2 \setminus \{(1,0,0)\} .$$

In particular, the inverse

$$\boldsymbol{\sigma}^{-1} : V \subseteq \mathbb{S}^2 \to \mathbb{R}^2$$

is a conformal map. This is known as the **Stereographic Projection**, see Figure 4.35 and Figure 4.36.

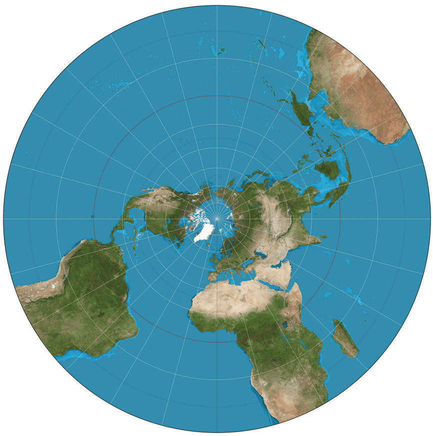


Figure 4.35.: Stereographic projection of the Earth. Image from Wikipedia

Example 4.167: Mercator projection

Question. Prove that the parametrization of $\mathbb{S}^2$ is conformal

$$\boldsymbol{\sigma}(u,v) := (\cos(u)\operatorname{sech}(v), \sin(u)\operatorname{sech}(v), \tanh(v)) .$$

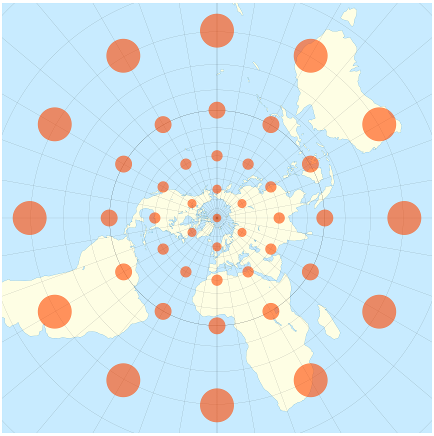


Figure 4.36.: The Stereographic projection is a conformal map: Angles and shapes are preserved. However areas are magnified away from the North pole.

Note. Let $V = \boldsymbol{\sigma}(U)$. In particular, the inverse

$$\boldsymbol{\sigma}^{-1} \colon V \subseteq \mathbb{S}^2 \to \mathbb{R}^2$$

is a conformal map. This is known as the **Mercator Projection**, see Figure 4.37 and Figure 4.38.
Solution. Recall the identities $\operatorname{sech}^2(v) + \tanh^2(v) = 1$ and

$$\operatorname{sech}(v)' = -\operatorname{sech}(v)\tanh(v)\,, \qquad \tanh(v)' = \operatorname{sech}^2(v)\,.$$

The chart $\boldsymbol{\sigma}$ is a conformal parametrization because the FFF is

$$\begin{aligned}
\tilde{\boldsymbol{\sigma}}_u &= \operatorname{sech}(v)\,(-\sin(u), \cos(u), 0) \\
\tilde{\boldsymbol{\sigma}}_v &= \operatorname{sech}(v)\,(-\cos(v)\tanh(v), -\sin(u)\tanh(v), \operatorname{sech}(v)) \\
\widetilde{E} = \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u &= \operatorname{sech}^2(v)(\cos^2(u) + \sin^2(u)) = \operatorname{sech}^2(v) \\
\widetilde{F} = \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v &= 0 \\
\widetilde{G} = \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v &= \operatorname{sech}^2(v)(\tanh^2(v) + \operatorname{sech}^2(v)) = \operatorname{sech}^2(v) \\
\mathcal{F}_1 &= \operatorname{sech}^2(v)\,\left(du^2 + dv^2\right)\,.
\end{aligned}$$

4.11.7. Areas

Suppose given a regular surface $\mathcal{S}$, with chart $\boldsymbol{\sigma} \colon U \to \mathbb{R}^3$. Given $R \subseteq U$, consider the region

$$\boldsymbol{\sigma}(R) \subseteq \mathcal{S}\,.$$

Figure 4.37.: Mercator projection of the Earth. This is the most frequently used map of the Earth. Image from Wikipedia.

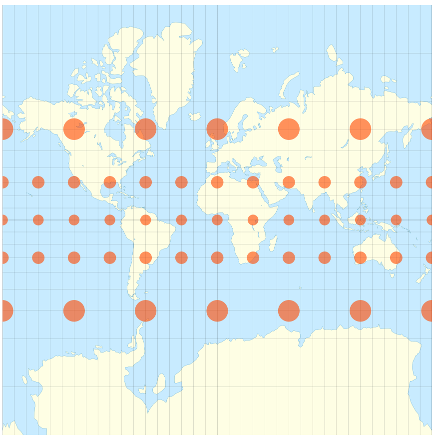


Figure 4.38.: The Mercator Projection is a conformal map: As such it preseves angles and shapes. However areas are distorted away from the Equator. Image from Wikipedia.

If $\boldsymbol{\sigma}(R)$ is sufficiently small, it can be approximated by a small portion of the tangent plane $T_{\mathbf{p}}\mathcal{S}$, namely, by the parallelogram spanned by $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$, see Figure 4.39. By the properties of vector product, the area of such parallelogram is $\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|$. Therefore, the area $A_{\boldsymbol{\sigma}}(R)$ of $\boldsymbol{\sigma}(R)$ is approximated by

$$A_{\boldsymbol{\sigma}}(R) \approx \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| \,.$$

It can be shown that the area of any region $\boldsymbol{\sigma}(R)$ is obtained by integrating $\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|$ on R, see Theorem 4.2.5 in [1]. Specifically,

$$A_{\boldsymbol{\sigma}}(R) = \int_R \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| \, dudv \,. \tag{4.13}$$

The proof of such result is similar to the one we did for showing that the length of a curve is given by

$$L(\boldsymbol{\gamma}) = \int_a^b \|\dot{\boldsymbol{\gamma}}(u)\| \, du \,,$$

where $L(\boldsymbol{\gamma})$ is defined as

$$L(\boldsymbol{\gamma}) = \lim_{\varepsilon \to 0} L(\boldsymbol{\gamma}_\varepsilon) \,,$$

where $L(\boldsymbol{\gamma}_\varepsilon)$ is the length of a piecewise linear approximation $\boldsymbol{\gamma}_\varepsilon$ of $\boldsymbol{\gamma}$. To obtain (4.13), one can approximate $\boldsymbol{\sigma}(R)$ with a piecewise affine mesh $\boldsymbol{\sigma}(R_\varepsilon)$, and then define

$$A_{\boldsymbol{\sigma}}(R) = \lim_{\varepsilon \to 0} A_{\boldsymbol{\sigma}}(R_\varepsilon) \,.$$

Making this argument precise requires a lot of effort. Instead, we just take (4.13) as the definition of area.

Definition 4.168

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart and $R \subseteq U$. The area $A_{\boldsymbol{\sigma}}$ of the region $\boldsymbol{\sigma}(R)$ is

$$A_{\boldsymbol{\sigma}} = \int_R \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| \, dudv \,.$$

Theorem 4.169

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart. Then

$$\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = \sqrt{EG - F^2} \,.$$

Proof

By the properties of vector product, we have

$$\|\mathbf{v} \times \mathbf{w}\|^2 = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 - \mathbf{v} \cdot \mathbf{w} \,, \quad \forall \, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3 \,.$$

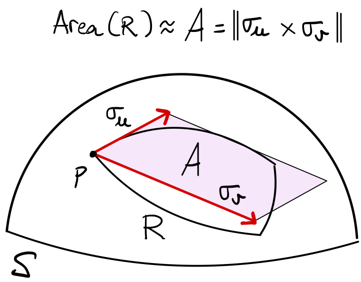


Figure 4.39.: The area $A_{\boldsymbol{\sigma}}$ of a small region $\boldsymbol{\sigma}(R)$ on a surface $\mathcal{S}$ can be approximated by the area of the parallelogram of sides $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Such area is $\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|$.

Applying the above identity to $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ gives

$$\left\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\right\|^2 = \|\boldsymbol{\sigma}_u\|^2\|\boldsymbol{\sigma}_v\|^2 - \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = EG - F^2 \,.$$

Taking the square roots gives the thesis.

It is easy to check that the definition of $A_{\boldsymbol{\sigma}}$ does not depend on the choice of $\boldsymbol{\sigma}$, see Proposition 6.4.3 in [7].

Example 4.170

Question. Let $\mathcal{S}$ be the paraboloid $z = x^2 + y^2$. Compute the area of

$$\mathcal{S} \cap \{z \leq 1\} \,.$$

Solution. The paraboloid $\mathcal{S}$ is charted by

$$\boldsymbol{\sigma}(u, v) = (u, v, u^2 + v^2) \,,$$

for $u, v \in \mathbb{R}$. We compute the first fundamental form of $\boldsymbol{\sigma}$:

$$\begin{aligned} \boldsymbol{\sigma}_u &= (1, 0, 2u) & F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 4uv \\ \boldsymbol{\sigma}_v &= (0, 1, 2v) & G &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 + 4v^2 \\ E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 + 4u^2 & EG - F^2 &= 1 + 4(u^2 + v^2) \end{aligned}$$

The region $\mathcal{S} \cap \{z \leq 1\}$ corresponds to $\boldsymbol{\sigma}(R)$, with

$$R = \{(u, v) \in \mathbb{R}^2 \ : \ u^2 + v^2 \leq 1\} \,.$$

The area of $\boldsymbol{\sigma}(R)$ is then

$$
\begin{aligned}
A_{\boldsymbol{\sigma}}(R) &= \int_R \sqrt{EG - F^2}\, dudv = \int_R \sqrt{1 + 4(u^2 + v^2)}\, dudv \\
&= \int_0^1 \int_0^{2\pi} \sqrt{1 + 4r^2}\, \rho\, d\rho\, d\theta = \frac{\pi}{6}(5^{3/2} - 1)\,,
\end{aligned}
$$

where the integral was computed using polar coordinates

$$
u = \rho\cos(\theta)\,, \quad v = \rho\sin(\theta)\,.
$$

4.11.8. Equiareal Maps

Definition 4.171

Let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ be regular surfaces and $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ a local diffeomorphism. We say that f is an **equiareal map** if it takes any region in $\mathcal{S}$ to a region of the same area in $\widetilde{\mathcal{S}}$. In this case, we say that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are **equiareal**.

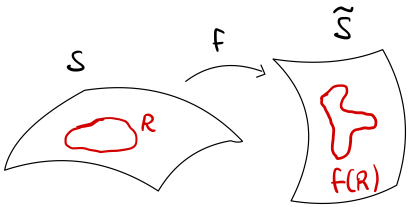


Figure 4.40.: Sketch of equiareal map f between the surfaces $\mathcal{S}$ and $\widetilde{\mathcal{S}}$. The area of the region R is the same as the area of $f(R)$.

Theorem 4.172

Let $\mathcal{S}$, $\widetilde{\mathcal{S}}$ be regular surfaces, $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ a local diffeomorphism. Equivalently:

1. f is an equiareal map.
2. Let $\boldsymbol{\sigma} : U \to \mathcal{S}$ be regular chart of $\mathcal{S}$, and define a chart of $\widetilde{\mathcal{S}}$ as $\tilde{\boldsymbol{\sigma}} : U \to \widetilde{\mathcal{S}}$, with $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$.

Then the FFF of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ satisfy

$$EG - F^2 = \widetilde{E}\widetilde{G} - \widetilde{F}^2 \,.$$

Theorem 4.172 can be proven with arguments similar to the proof of Theorem 4.141. The proof is omitted.

As usual, we give a sufficient condition to prove that two regular surfaces are equiareal.

Theorem 4.173

Let $\mathcal{S}, \widetilde{\mathcal{S}}$ be regular surfaces, with charts $\boldsymbol{\sigma} : U \to \mathcal{S}$ and $\tilde{\boldsymbol{\sigma}} : U \to \widetilde{\mathcal{S}}$. Assume that

$$EG - F^2 = \widetilde{E}\widetilde{G} - \widetilde{F}^2 \,,$$

We have

1. The surfaces $\boldsymbol{\sigma}(U)$ and $\widetilde{\mathcal{S}}$ are equiareal.
2. An equiareal map is given by

$$f : \boldsymbol{\sigma}(U) \to \widetilde{\mathcal{S}} \,, \qquad f = \tilde{\boldsymbol{\sigma}} \circ \boldsymbol{\sigma}^{-1} \,.$$

The proof follows by adapting the argument in the proof of Theorem 4.143, and by applying Theorem 4.172. This is left as an exercise.

Example 4.174: Archimedes map

Question. Consider the surface $\mathcal{S}$ obtained by removing the North and South Poles from the unit sphere

$$\mathcal{S} = \mathbb{S}^2 \setminus \{(0,0,\pm 1)\} \,, \qquad \mathbb{S}^2 = \{x^2 + y^2 + z^2 = 1\} \,.$$

Let $\widetilde{\mathcal{S}} = \{x^2 + y^2 = 1\}$ be the unit cylinder. The *Archimedes map* is

$$f : \mathcal{S} \to \widetilde{\mathcal{S}} \,, \quad f(x,y,z) = \left(\frac{x}{(x^2+y^2)^{1/2}}, \frac{y}{(x^2+y^2)^{1/2}}, z \right) .$$

1. Prove that f is a diffeomorphism.
2. Prove that f is equiareal.

Solution. Note that $f \in \widetilde{\mathcal{S}}$ because

$$\left[\frac{x}{(x^2+y^2)^{1/2}} \right]^2 + \left[\frac{y}{(x^2+y^2)^{1/2}} \right]^2 = 1 \,.$$

Therefore f is well-defined. In order to chart $\mathcal{S}$, introduce

$$
\begin{aligned}
&\boldsymbol{\sigma}(u,v) = (\cos(u)\cos(v), \sin(u)\cos(v), \sin(v)) \\
&U_1 = \left\{(u,v) \in \mathbb{R}^2 \ : \ u \in (-\pi,\pi), \ v \in (-\pi/2, \pi/2)\right\} \\
&U_2 = \left\{(u,v) \in \mathbb{R}^2 \ : \ u \in (0,2\pi), \ v \in (-\pi/2, \pi/2)\right\} \\
&\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}|_{U_1}, \qquad \boldsymbol{\sigma}_2 = \boldsymbol{\sigma}|_{U_2}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\boldsymbol{\sigma}(U_1) = \mathbb{S}^2 \smallsetminus \{\text{Date Line}, (0,0,\pm 1)\}, \\
&\boldsymbol{\sigma}(U_2) = \mathbb{S}^2 \smallsetminus \{\text{Greenwich Meridian}, (0,0,\pm 1)\},
\end{aligned}
$$

Therefore $\mathscr{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$ is an atlas for $\mathcal{S}$. Denote the components of $\boldsymbol{\sigma}$ by

$$
x = \cos(u)\cos(v), \qquad y = \sin(u)\cos(v), \qquad z = \sin(v).
$$

We have that

$$
(x^2+y^2)^{1/2} = |\cos(v)| = \cos(v),
$$

where we used that $\cos(v) > 0$, since $v \in (-\pi/2, \pi/2)$ when $(u,v) \in U_1$ or $(u,v) \in U_2$. Thus, the chart $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ is

$$
\begin{aligned}
\tilde{\boldsymbol{\sigma}}(u,v) &= f(\boldsymbol{\sigma}(u,v)) \\
&= \left(\frac{x}{(x^2+y^2)^{1/2}}, \frac{y}{(x^2+y^2)^{1/2}}, z \right) \\
&= \left(\frac{\cos(u)\cos(v)}{\cos(v)}, \frac{\sin(u)\cos(v)}{\cos(v)}, \sin(v) \right) \\
&= (\cos(u), \sin(u), \sin(v)).
\end{aligned}
$$

It is clear that $\tilde{\boldsymbol{\sigma}}$ charts the part of the unit cylinder between the planes $\{z=-1\}$ and $\{z=1\}$, when $u \in [0,2\pi]$ and $v \in (-\pi/2, \pi/2)$. Therefore, the charts

$$
\widetilde{A} = \{\tilde{\boldsymbol{\sigma}}_1, \tilde{\boldsymbol{\sigma}}_2, \}, \quad \tilde{\boldsymbol{\sigma}}_1, = \tilde{\boldsymbol{\sigma}}|_{U_1}, \quad \tilde{\boldsymbol{\sigma}}_2 = \tilde{\boldsymbol{\sigma}}|_{U_2},
$$

form an atlas for the surface $\widetilde{\mathcal{S}} \cap \{-1 < z < 1\}$.

1. For $i = 1,2$ define the map

$$
\Psi_i \colon U_i \to U_i, \qquad \Psi_i = \tilde{\boldsymbol{\sigma}}^{-1} \circ f \circ \boldsymbol{\sigma}.
$$

Since $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$, we have that Ψ is the identity

$$
\Psi(u,v) = (u,v).
$$

Therefore Ψ is smooth, with smooth inverse. By definition of smooth map between surfaces, this implies that both f and f^{-1} are smooth. Thus, f is a diffeomorphism between $\mathcal{S}$ and $\widetilde{\mathcal{S}} \cap \{-1 < z < 1\}$.

2. We compute the FFF of $\boldsymbol{\sigma}$

$$\begin{aligned}
\boldsymbol{\sigma}_u &= (-\sin(u)\cos(v), \cos(u)\cos(v), 0)\\
\boldsymbol{\sigma}_v &= (-\cos(u)\sin(v), -\sin(u)\sin(v), \cos(v))\\
E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = \cos^2(v)\\
F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0\\
G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1
\end{aligned}$$

The FFF of $\tilde{\boldsymbol{\sigma}}$ is

$$\begin{aligned}
\tilde{\boldsymbol{\sigma}}_u &= (-\sin(u), \cos(u), 0)\\
\tilde{\boldsymbol{\sigma}}_v &= (0, 0, \cos(v))\\
\widetilde{E} &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u = 1\\
\widetilde{F} &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v = 0\\
\widetilde{G} &= \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v = \cos^2(v)
\end{aligned}$$

Therefore, we have that $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ satisfy

$$EG - F^2 = \widetilde{E}\widetilde{G} - \widetilde{F}^2 = \cos^2(v)\,. \tag{4.14}$$

Note that $\boldsymbol{\sigma}_i$ is defined by restricting $\boldsymbol{\sigma}$ to U_i. Hence, $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}_i$ have the same FFF. Similarly, $\tilde{\boldsymbol{\sigma}}_i := f \circ \boldsymbol{\sigma}_i$ is just the restriction to U_i of $\tilde{\boldsymbol{\sigma}}$. Therefore, $\tilde{\boldsymbol{\sigma}}_i$ and $\tilde{\boldsymbol{\sigma}}$ have the same FFF. We conclude that $\tilde{\boldsymbol{\sigma}}_i$ and $\tilde{\boldsymbol{\sigma}}_i = f \circ \boldsymbol{\sigma}_i$ also satisfy (4.14). In particular, since $\mathscr{A} = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\}$ is an Atlas for $\mathscr{S}$, Theorem 4.172 implies that f is an equiareal map.

Derivation of the Archimedes map. The sphere $\mathscr{S}$ is contained inside the cylinder $\widetilde{\mathscr{S}}$, and the two surfaces touch along the circle $x^2 + y^2 = 1$ in the $\{z = 0\}$ plane. Let $P \in \mathbb{S}^2$, except for North or South Pole. Draw the line through P and the z-axis which is parallel to the plane $\{z = 0\}$. This line intersects the cylinder in 2 points. Denote by P' the intersection point which is closest to P, see Figure 4.41. To write the projection map explicitly, denote the coordinates of P and P' by

$$P = (x, y, z)\,, \quad P' = (X, Y, Z)\,.$$

Since the line through P and P' is parallel to the plane $\{z = 0\}$, we have

$$Z = z\,, \quad (X, Y) = \lambda(x, y)$$

for some scalar λ. Using that P' belongs to the cylinder, we get

$$1 = X^2 + Y^2 = \lambda^2(x^2 + y^2)$$

from which we deduce that

$$\lambda = \pm \frac{1}{(x^2 + y^2)^{1/2}}\,.$$

Therefore

$$X = \lambda x = \pm \frac{x}{(x^2 + y^2)^{1/2}}, \quad Y = \lambda y = \pm \frac{y}{(x^2 + y^2)^{1/2}}.$$

Taking the $+$ sign gives the point

$$P' = (X, Y, Z) = \left(\frac{x}{(x^2 + y^2)^{1/2}}, \frac{y}{(x^2 + y^2)^{1/2}}, z \right).$$

This defines the Archimedes map $f : \mathcal{S} \to \widetilde{\mathcal{S}}$

$$f(x, y, z) = \left(\frac{x}{(x^2 + y^2)^{1/2}}, \frac{y}{(x^2 + y^2)^{1/2}}, z \right).$$

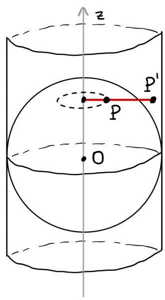


Figure 4.41.: Construction of the Archimedes Map between the Sphere and the Cylinder.

We conclude this section by clarifying the relationship between isometries, conformal and equiareal maps.

Theorem 4.175

Let $f : \mathcal{S} \to \widetilde{\mathcal{S}}$ be a local diffeomorphism between surfaces. They are equivalent:

1. f is a local isometry.
2. f is conformal and equiareal.

Proof

Part 1. Local isometries preserve the FFF. Therefore, f is conformal with $\lambda = 1$. Moreover, also $EG - F^2$ is preserved, implying that f is equiareal.
Part 2. Assume f is conformal and equiareal. Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a chart of $\mathcal{S}$, and define

$\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$. As f is conformal, by Thoerem 4.156 we have that

$$\widetilde{E} = \lambda E\,, \qquad \widetilde{F} = \lambda F\,, \qquad \widetilde{G} = \lambda G\,,$$

for some smooth $\lambda : U \to \mathbb{R}$. As f is equiareal, by Theorem 4.172 we have that

$$EG - F^2 = \widetilde{E}\widetilde{G} - \widetilde{F}^2\,.$$

In particular, we get

$$EG - F^2 = \widetilde{E}\widetilde{G} - \widetilde{F}^2 = (\lambda E)(\lambda F) - (\lambda G)^2 = \lambda^2(EG - F^2)\,.$$

Since $\boldsymbol{\sigma}$ is a regular chart, we have $EG - F^2 > 0$. Therefore, we obtain $\lambda = \pm 1$, which implies

$$\widetilde{E} = \lambda E\,, \quad \lambda = \pm 1\,.$$

Note that

$$E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = \left\|\boldsymbol{\sigma}_u\right\|^2 > 0\,,$$

being $\boldsymbol{\sigma}$ regular (so that $\boldsymbol{\sigma}_u \neq \mathbf{0}$). Similarly, as $\tilde{\boldsymbol{\sigma}}$ is regular, we also have that $\widetilde{E} > 0$. Therefore $\lambda = 1$. In particular, we have shown that $\mathcal{F} = \widetilde{\mathcal{F}}$, implying that f is a local isometry by Thoerem 4.141.

4.11.9. Equiareal parametrizations

In the previous section we defined equiareal maps between surfaces

$$f : \mathcal{S} \to \widetilde{\mathcal{S}}\,.$$

It is also useful to define equiareal parametrizations.

Definition 4.176: Equiareal parametrization

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular. We say that $\boldsymbol{\sigma}$ is an **equiareal parametrization** if the coefficients of the FFF of $\boldsymbol{\sigma}$ satisfy

$$EG - F^2 = 1\,.$$

The above definition is motivated by the following result, stating that $\mathcal{S}$ admits an equiareal parametrization if and only if $\mathcal{S}$ is equiareal to a plane.

Theorem 4.177

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular. Let $\boldsymbol{\pi}$ be the plane $\{z = 0\}$. They are equivalent:

1. $\boldsymbol{\sigma}$ is an equiareal parametrization.

2. The map $f : \boldsymbol{\pi} \to \boldsymbol{\sigma}(U)$ defined by

$$f(u, v, 0) = \boldsymbol{\sigma}(u, v)$$

is equiareal.

Proof

The plane $\boldsymbol{\pi}$ is charted by

$$\tilde{\boldsymbol{\sigma}}(u, v) = (u, v, 0), \quad (u, v) \in \mathbb{R}^2$$

The first fundamental form of $\tilde{\boldsymbol{\sigma}}$ is

$$\widetilde{\mathcal{F}}_1 = du^2 + dv^2 .$$

Therefore

$$\widetilde{E}\widetilde{G} - \widetilde{F}^2 = 1 .$$

Consider the map $f : \boldsymbol{\pi} \to \mathcal{S}$ given in Point 2, that is,

$$f(u, v, 0) = \boldsymbol{\sigma}(u, v) .$$

By definition of $\tilde{\boldsymbol{\sigma}}$,

$$f(\tilde{\boldsymbol{\sigma}}(u, v)) = \boldsymbol{\sigma}(u, v) .$$

Theorem 4.172 says that f is an equiareal map if and only if

$$EG - F^2 = \widetilde{E}\widetilde{G} - \widetilde{F}^2 = 1 .$$

This happens exactly when $\boldsymbol{\sigma}$ is an equiareal parametrization.

As an immediate consequence, we have the following Theorem.

Theorem 4.178

An equiareal parametrization $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ preserves areas, that is,

$$|R| = A_{\boldsymbol{\sigma}}(R) , \quad \forall\, R \subseteq U ,$$

where $|R|$ is the area of R in $\mathbb{R}^2$.

An important application of equiareal parametrization is in cartography: We would like to draw a map of the Earth which preserves areas. One such map is known as the Lambert cylindrical projection.

Example 4.179: Lambert cylindrical projection

Question. Prove that the following parametrization of $\mathbb{S}^2$ is equiareal

$$\boldsymbol{\sigma}(u,v) = \left(\cos(u)\sqrt{1-v^2}, \sin(u)\sqrt{1-v^2}, v\right) .$$

Note: In particular, the inverse

$$\boldsymbol{\sigma}^{-1} : \mathbb{S}^2 \to \mathbb{R}^2$$

is an equiareal map. This is known as the **Lambert cylindrical projection**, see Figure 4.42 and Figure 4.43.

Solution. We compute

$$\begin{aligned}
\boldsymbol{\sigma}_u &= \left(-\sin(u)\sqrt{1-v^2}, \cos(u)\sqrt{1-v^2}, 0\right) \\
\boldsymbol{\sigma}_v &= \left(-\cos(u)v(1-v^2)^{-1/2}, -\sin(u)v(1-v^2)^{-1/2}, 1\right) \\
E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 - v^2 \\
F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\
G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1 + \frac{v^2}{1-v^2}
\end{aligned}$$

and therefore

$$EG - F^2 = (1-v^2)\left(1 + \frac{v^2}{1-v^2}\right) = 1 \,,$$

showing that $\boldsymbol{\sigma}$ is an equiareal parametrization of $\mathbb{S}^2$.

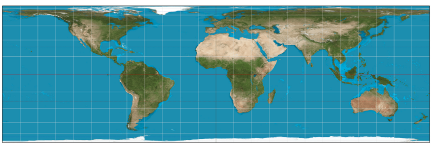


Figure 4.42.: Lambert cylindrical projection of the Earth. Images from Wikipedia.

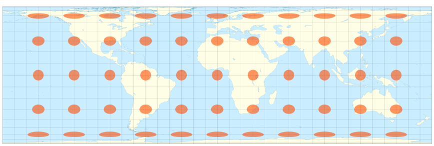


Figure 4.43.: The Lambert cylindrical projection is an equiareal map. As such it preseves areas. However angles and shapes are distorted. Image from Wikipedia.

Remark 4.180: Cartography

We have shown that:

1. The Mercator projection is a conformal map of $\mathbb{S}^2$ into the plane. Such projection preseves angles but distorts areas.

2. The Lambert cylindrical projection is an equiareal map of $\mathbb{S}^2$ into the plane. Such projection preseves areas but distorts angles.

The following question is natural: Is there a projection

$$\boldsymbol{\sigma} : \mathbb{S}^2 \to \mathbb{R}^2$$

which is both conformal and equiareal?
Answer: No. This is because $\boldsymbol{\sigma}$ would be an isometry between $\mathbb{S}^2$ and the plane $\mathbb{R}^2$, see Theorem 4.175. This is impossible because:

- The Theorema Egregium (by Gauss - we will study it) tells us that isometric surfaces must have the same Gaussian curvature;
- However, the plane has zero Gaussian curvature, while the sphere has non-zero Gaussian curvature;
- Therefore the plane and the sphere cannot be isometric.

This means that we cannot draw a map of the Earth which preserves both angles and lengths.

4.11.10. Summary

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ a chart. We have introduced the **first fundamental form** of $\mathcal{S}$ as the restriction of the euclidean scalar product to the tangent space:

$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{w}) = \mathbf{v} \cdot \mathbf{w} \,, \quad \forall\, \mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$$

The first fundamental form of $\boldsymbol{\sigma}$ is

$$\mathcal{F}_1 = Edu^2 + 2Fdudv + Gdv^2$$

where the coefficients are

$$E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u \,, \quad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v \,, \quad G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v \,.$$

Given a chart $\boldsymbol{\sigma}$, we have

$$I_{\mathbf{p}}(\mathbf{v}, \mathbf{v}) = Edu(\mathbf{v})^2 + 2Fdu(\mathbf{v})dv(\mathbf{v}) + Gdv(\mathbf{v})^2$$

Consider two curves $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ on the surface $\boldsymbol{\sigma}(U)$, that is,

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t)) \,, \quad \tilde{\boldsymbol{\gamma}}(t) = \boldsymbol{\sigma}(\tilde{u}(t), \tilde{v}(t)) \,.$$

Moreover, consider a region $R \subseteq U$. The first fundamental form allows to compute:

1. Length of curves:
$$L(\boldsymbol{\gamma}) = \int_a^b \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\, dt$$
2. Angle θ between $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ (at any intersection point):
$$\cos(\theta) = \frac{E\dot{u}\dot{\tilde{u}} + F(\dot{u}\dot{\tilde{v}} + \dot{\tilde{u}}\dot{v}) + G\dot{v}\dot{\tilde{v}}}{(E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2)^{1/2}(E\dot{\tilde{u}}^2 + 2F\dot{\tilde{u}}\dot{\tilde{v}} + G\dot{\tilde{v}}^2)^{1/2}}\,.$$
3. Area $A_{\boldsymbol{\sigma}}(R)$ of the region $\boldsymbol{\sigma}(R)$
$$A_{\boldsymbol{\sigma}}(R) = \int_R \sqrt{EG - F^2}\, dudv\,.$$

We have also introduced maps preserving certain quantities. Specfically, let $\widetilde{\mathcal{S}}$ be another regular surface and
$$f \colon \mathcal{S} \to \widetilde{\mathcal{S}}$$
a local diffeomorphism:

1. f local isometry: Preserves scalar product of tangent vectors, and length of curves
   - Length of $\boldsymbol{\gamma}$ equals that of $f \circ \boldsymbol{\gamma}$ for any $\boldsymbol{\gamma}$.
   - f isometry $\iff$ FFF of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ satisfy
   $$E = \widetilde{E}, \quad F = \widetilde{F}, \quad G = \widetilde{G}$$
2. f conformal map: Preserves the angle between tangent vectors, and between curves
   - Angle between $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\gamma}}$ is the same as the angle between $f \circ \boldsymbol{\gamma}$ and $f \circ \tilde{\boldsymbol{\gamma}}$
   - f conformal $\iff$ FFF of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ satisfy
   $$E = \lambda\widetilde{E}, \quad F = \lambda\widetilde{F}, \quad G = \lambda\widetilde{G}$$
   for some function $\lambda(u, v)$.
3. Equiareal maps: They preserve areas of surface regions
   - Area of $\Omega \subseteq \mathcal{S}$ is the same as the area of $f(\Omega)$.
   - f equiareal $\iff$ FFF of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}} = f \circ \boldsymbol{\sigma}$ satisfy
   $$EG - F^2 = \widetilde{E}\widetilde{G} - \widetilde{F}^2$$

Moreover, they are equivalent:

1. f is a local isometry,
2. f is conformal and equiareal.

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.

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 $\mathcal{S}$ is deviating from the tangent plane $T_{\mathbf{p}}\mathcal{S}$. Recall that

$$T_{\mathbf{p}}\mathcal{S} = \operatorname{span}\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\} \,,$$

where $\boldsymbol{\sigma}$ is a regular chart of $\mathcal{S}$ 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 $T_{\mathbf{p}}\mathcal{S}$. 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

$$\boldsymbol{\sigma}(u_0 + \Delta u, v_0 + \Delta v) \in \mathcal{S}$$

deviates from the tangent plane $T_{\mathbf{p}}\mathcal{S}$. Since $\mathbf{N}$ is orthogonal to $T_{\mathbf{p}}\mathcal{S}$, 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 4.44.

Using Taylor's formula we get

$$\begin{aligned}
\boldsymbol{\sigma}(u_0 + \Delta u, v_0 + \Delta v) = \boldsymbol{\sigma}(u_0, v_0) + \boldsymbol{\sigma}_u(u_0, v_0)\, \Delta u + \boldsymbol{\sigma}_v(u_0, v_0)\, \Delta v \\
+ \frac{1}{2}\Big(\boldsymbol{\sigma}_{uu}(u_0, v_0)(\Delta u)^2 + 2\boldsymbol{\sigma}_{uv}(u_0, v_0)\Delta u \Delta v \\
+ \boldsymbol{\sigma}_{vv}(u_0, v_0)(\Delta v)^2\Big) + R(\Delta u, \Delta v) \,,
\end{aligned}$$

where $R(\Delta u, \Delta v)$ is a remainder such that

$$\lim_{\Delta \to 0} \frac{R(\Delta u, \Delta v)}{\Delta} = 0 \,, \quad \Delta := (\Delta u)^2 + (\Delta v)^2 \,.$$

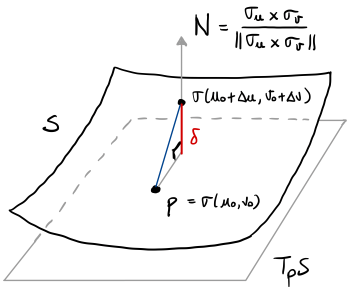


Figure 4.44.: The point $\boldsymbol{\sigma}(u_0 + \Delta u, v_0 + \Delta v)$ on $\mathcal{S}$ deviates from $T_{\mathbf{p}}\mathcal{S}$ by a quantity δ.

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 4.181: Second fundamental form of a chart

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular, $\mathcal{S} = \boldsymbol{\sigma}(U)$. Define $L, M, N : U \to \mathbb{R}$

$$L := \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} , \quad M := \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} , \quad N := \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} ,$$

where $\mathbf{N}$ is the standard unit normal to $\boldsymbol{\sigma}$. The **second fundamental form (SFF)** of $\boldsymbol{\sigma}$ is the quadratic form $\mathcal{F}_2 : T_{\mathbf{p}}\mathcal{S} \to \mathbb{R}$

$$\mathcal{F}_2(\mathbf{v}) = L\, du^2(\mathbf{v}) + 2M\, du(\mathbf{v})\, dv(\mathbf{v}) + N\, dv^2(\mathbf{v}) , \ \forall\, v \in T_{\mathbf{p}}\mathcal{S} ,$$

for all $\mathbf{p} \in \boldsymbol{\sigma}(U)$, with L, M, N evaluated at $(u, v) = \boldsymbol{\sigma}^{-1}(v)$, and du, dv the coordinate functions in Definition 4.124.

Notation

With a little abuse of notation, we also denote by $\mathcal{F}_2$ the 2×2 matrix

$$\mathcal{F}_2 = \begin{pmatrix} L & M \\ M & N \end{pmatrix} .$$

Let us show that a plane and a cylinder have different second fundamental forms.

Example 4.182: SFF of Plane

Question. Let $\mathbf{a}, \mathbf{p}, \mathbf{q} \in \mathbb{R}^3$, with $\mathbf{p}, \mathbf{q}$ orthonormal. The plane is charted by

$$\boldsymbol{\sigma}(u,v) = \mathbf{a} + u\mathbf{p} + v\mathbf{q}\,, \quad (u,v) \in \mathbb{R}^2\,.$$

Prove that the SFF of $\boldsymbol{\sigma}$ is $\mathscr{F}_2 = 0$.
Note: This reflects the intuition that a plane is flat, and therefore there is no *curvature*.
Solution. We have that $\mathscr{F}_2 = 0$, since

$$\begin{gathered}\boldsymbol{\sigma}_u = \mathbf{p}\,, \quad \boldsymbol{\sigma}_v = \mathbf{q}\,, \quad \boldsymbol{\sigma}_{uu} = \boldsymbol{\sigma}_{uv} = \boldsymbol{\sigma}_{vv} = \mathbf{0}\,, \\ L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} = 0\,, \quad M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = 0\,, \quad N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} = 0\,.\end{gathered}$$

Example 4.183: SFF of Unit cylinder

Question. 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 SFF of $\boldsymbol{\sigma}$ is

$$\mathscr{F}_2 = -du^2\,.$$

Note: This reflects the intuition that the cylinder curves only when moving in the u-direction. In such direction we are moving on a circle of radius 1, therefore we expect the curvature to be -1.
Solution. We have

$$\begin{aligned}
\boldsymbol{\sigma}_u &= (-\sin(u), \cos(u), 0) & \mathbf{N} &= \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} \\
\boldsymbol{\sigma}_v &= (0,0,1) & &= (\cos(u), \sin(u), 0) \\
\boldsymbol{\sigma}_{uu} &= (-\cos(u), -\sin(u), 0) & L &= \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} = -1 \\
\boldsymbol{\sigma}_{uv} &= \boldsymbol{\sigma}_{vv} = \mathbf{0} & M &= \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = 0 \\
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= (\cos(u), \sin(u), 0) & N &= \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} = 0 \\
\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| &= 1 & \mathscr{F}_2 &= -du^2
\end{aligned}$$

Remark 4.184

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

$$\mathcal{F}_2 = 0\,, \quad \widetilde{\mathcal{F}}_2 = -du^2\,,$$

respectively.

Remark 4.185: SFF and reparametrizations

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular, and $\tilde{\boldsymbol{\sigma}} : \widetilde{U} \to \mathbb{R}^3$ a reparametrization, with $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} \circ \Phi$ and $\Phi : \widetilde{U} \to U$ diffeomorphism. The matrices $\mathcal{F}_2$ and $\widetilde{\mathcal{F}}_2$ of the SFF of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are related by

$$\widetilde{\mathcal{F}}_2 = \pm (J\Phi)^T \mathcal{F}_2 J\Phi\,, \quad \mathcal{F}_2 = \begin{pmatrix} \tilde{L} & \widetilde{M} \\ \widetilde{M} & \widetilde{N} \end{pmatrix}, \quad \widetilde{\mathcal{F}}_2 \begin{pmatrix} \tilde{L} & \widetilde{M} \\ \widetilde{M} & \widetilde{N} \end{pmatrix},$$

where the formula holds with the plus sign if $\det J\Phi > 0$, and with the minus sign if $\det J\Phi < 0$.

Proof

The formula 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 4.88.

4.12.1. Gauss and Weingarten maps

Another way to quantify how much a surface $\mathcal{S}$ is curving is by examining the behavior of standard unit normal $\mathbf{N}$. If $\mathcal{S}$ 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 $\mathcal{S}$. If $\mathcal{S}$ is a general surface, measuring the variation of $\mathbf{N}$ will tell us how much $\mathcal{S}$ is deviating from being a plane. This is the idea behind the definition of the **Gauss** and **Weingarten** maps.

Remark 4.186

Let $\mathcal{S}$ be oriented and $\mathbf{N} : \mathcal{S} \to \mathbb{R}^3$ be the standard unit normal. In particular $\mathbf{N}$ is a smooth map and

$$\mathbf{N}(\mathbf{p}) \perp T_{\mathbf{p}}\mathcal{S}\,, \quad \|\mathbf{N}(\mathbf{p})\| = 1\,, \quad \forall\, \mathbf{p} \in \mathcal{S}\,.$$

Since $T_{\mathbf{p}}\mathcal{S}$ 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 $\mathbf{N} : \mathcal{S} \to \mathbb{S}^2$.

Definition 4.187: Gauss map

Let $\mathcal{S}$ be an oriented surface with standard unit normal $\mathbf{N}$. The **Gauss map** of $\mathcal{S}$ is

$$\mathscr{G}_{\mathcal{S}} : \mathcal{S} \to \mathbb{S}^2 \,, \quad \mathscr{G}_{\mathcal{S}}(\mathbf{p}) := \mathbf{N}(\mathbf{p}) \,.$$

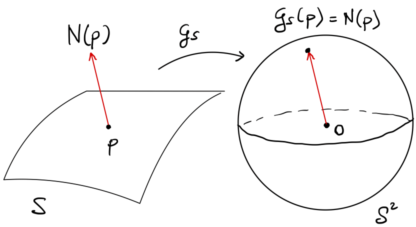


Figure 4.45.: The Gauss map $\mathscr{G}_{\mathcal{S}}$ of $\mathcal{S}$ is defined as $\mathscr{G}_{\mathcal{S}}(\mathbf{p}) := \mathbf{N}(\mathbf{p})$. Note that $\mathscr{G}_{\mathcal{S}}(\mathbf{p}) \in \mathbb{S}^2$.

Remark 4.188

The Gauss map of $\mathcal{S}$ is just the standard unit normal of $\mathcal{S}$. By definition of standard unit normal to $\mathcal{S}$ we obtain that

$$\mathscr{G}_{\mathcal{S}} \circ \boldsymbol{\sigma} = \mathbf{N}$$

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 4.189

1. Suppose $\mathcal{S}$ is the unit sphere $\mathbb{S}^2$. Then $\mathscr{G}_{\mathcal{S}} : \mathcal{S} \to \mathbb{S}^2$ is the identity,

$$\mathscr{G}_{\mathcal{S}}(\mathbf{p}) = \mathbf{p} \,,$$

see Figure 4.46.

2. Let $\mathbf{a}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ with $\mathbf{v}$ and $\mathbf{w}$ linearly independent. Let $\mathcal{S}$ 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 $\mathcal{S}$ is constant:

$$\mathcal{G}_{\mathcal{S}}(\mathbf{p}) = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = \frac{\mathbf{v} \times \mathbf{w}}{\|\mathbf{v} \times \mathbf{w}\|}\,,$$

for all $\mathbf{p} \in \mathcal{S}$, see Figure 4.47.

3. Let $\mathcal{S}$ be the unit cylinder

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v)\,, \quad (u,v) \in (0,2\pi) \times \mathbb{R}\,.$$

We have already compute that the standard unit normal is

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = (\cos(u), \sin(u), 0)\,.$$

Therefore, the Gauss map of $\mathcal{S}$ is

$$\mathcal{G}_{\mathcal{S}}(\mathbf{p}) = (\cos(u_0), \sin(u_0), 0)\,,$$

where (u_0, v_0) is such that $\boldsymbol{\sigma}(u_0, v_0) = \mathbf{p}$. Note that $\mathcal{G}_{\mathcal{S}}$ maps $\mathcal{S}$ into the equator of $\mathbb{S}^2$, see Figure 4.48.

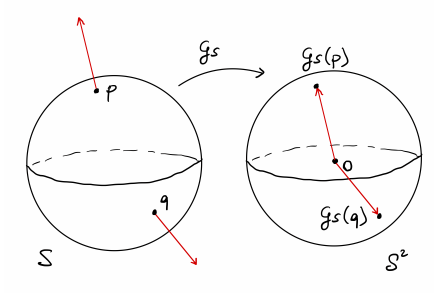


Figure 4.46.: The Gauss map $\mathcal{G}_{\mathcal{S}}$ of a sphere is the identity.

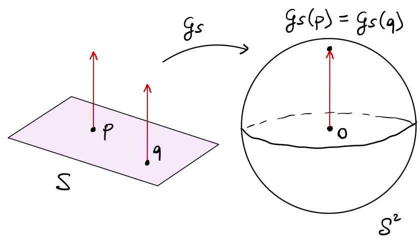


Figure 4.47.: The Gauss map $\mathcal{G}_\mathcal{S}$ of a plane is constant.

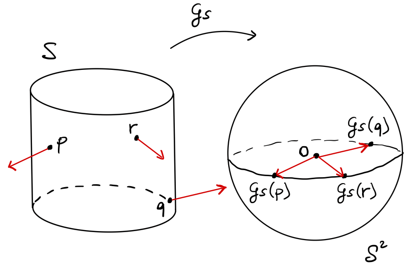


Figure 4.48.: If $\mathcal{S}$ is the unit cylinder, the Gauss map $\mathcal{G}_\mathcal{S}$ maps $\mathcal{S}$ into the equator of $\mathbb{S}^2$.

Remark 4.190

By definition, the Gauss map is a smooth function between surfaces. Therefore the differential of $\mathcal{G}_{\mathcal{S}}$ is well defined, and

$$d_{\mathbf{p}}\mathcal{G}_{\mathcal{S}} \colon T_{\mathbf{p}}\mathcal{S} \to T_{\mathcal{G}_{\mathcal{S}}(\mathbf{p})}\mathbb{S}^2 ,$$

for all $\mathbf{p} \in \mathcal{S}$. We have that

$$T_{\mathcal{G}_{\mathcal{S}}(\mathbf{p})}\mathbb{S}^2 = T_{\mathbf{p}}\mathcal{S} , \tag{4.15}$$

see Figure 4.49. Therefore

$$d_{\mathbf{p}}\mathcal{G}_{\mathcal{S}} \colon T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S} .$$

Proof. The tangent plane $T_{\mathcal{G}_{\mathcal{S}}(\mathbf{p})}\mathbb{S}^2$ passes through the origin and

$$\mathcal{G}(\mathbf{p}) \perp T_{\mathcal{G}_{\mathcal{S}}(\mathbf{p})}\mathbb{S}^2 .$$

By definition $\mathcal{G}(\mathbf{p}) = \mathbf{N}(\mathbf{p})$, and thus

$$\mathbf{N}(\mathbf{p}) \perp T_{\mathcal{G}_{\mathcal{S}}(\mathbf{p})}\mathbb{S}^2 .$$

Since by definition

$$\mathbf{N}(\mathbf{p}) \perp T_{\mathbf{p}}\mathcal{S} ,$$

we infer (4.15).

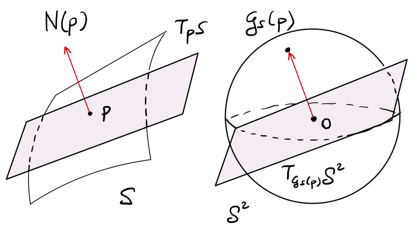


Figure 4.49.: We ca identify $T_{\mathcal{G}_{\mathcal{S}}(\mathbf{p})}\mathbb{S}^2$ with $T_{\mathbf{p}}\mathcal{S}$. This is because $\mathcal{G}(\mathbf{p}) \perp T_{\mathcal{G}_{\mathcal{S}}(\mathbf{p})}\mathbb{S}^2$ and $\mathcal{G}(\mathbf{p}) = \mathbf{N}(\mathbf{p})$.

Definition 4.191: Weingarten map

Let $\mathcal{S}$ be an orientable surface with Gauss map $\mathcal{G} \colon \mathcal{S} \to \mathbb{S}^2$. The **Weingarten map** $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ of $\mathcal{S}$ at $\mathbf{p}$ is

$$\mathcal{W}_{\mathbf{p},\mathcal{S}} \colon T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S} , \quad \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) = -d_{\mathbf{p}}\mathcal{G}(\mathbf{v}) .$$

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 4.192

The minus sign in the definition of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ 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 4.193: SFF of a surface

Let $\mathcal{S}$ be an orientable surface with Weingarten map $\mathscr{W}_{\mathbf{p},\mathcal{S}}$. The **SFF** of $\mathcal{S}$ at $\mathbf{p}$ is the bilinear map

$$II_{\mathbf{p}} : T_{\mathbf{p}}\mathcal{S} \times T_{\mathbf{p}}\mathcal{S} \to \mathbb{R} , \quad II_{\mathbf{p}}(\mathbf{v},\mathbf{w}) := \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) \cdot \mathbf{w} .$$

Remark 4.194

The second fudamental form $II_{\mathbf{p}}$ of $\mathcal{S}$ is bilinear.
Check. $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ is linear, being the differential of a smooth map. Hence $II_{\mathbf{p}}$ is bilinear, given that the scalar product is bilinear.

Remark 4.195: 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 the next Theorem, we will prove that

$$A = \mathscr{F}_2 = \begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

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 $\mathscr{F}_2$ of the chart $\boldsymbol{\sigma}$. We prove this statement in the next theorem.

Theorem 4.196: Matrix of the SFF

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be regular, $\mathcal{S} = \boldsymbol{\sigma}(U)$, and $\mathbf{p} \in \boldsymbol{\sigma}(U)$. Then

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 $\mathbf{v}, \mathbf{w} \in T_{\mathbf{p}}\mathcal{S}$.
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 4.196 we use the following two Lemmas.

Lemma 4.197

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 4.198

Let $\mathcal{S}$ be an orientable surface with Weingarten map $\mathscr{W}_{\mathbf{p},\mathcal{S}}$, and $\boldsymbol{\sigma}$ a regular chart at $\mathbf{p}$. Then

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) = -\mathbf{N}_u \,, \quad \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v \,,$$

where $\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v, \mathbf{N}_u, \mathbf{N}_v$ are evaluated at $(u,v) = \boldsymbol{\sigma}^{-1}(\mathbf{p})$.

Proof

Since $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ is defined as $-d_{\mathbf{p}}\mathscr{G}_{\mathcal{S}}$, we can compute $\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u)$ and $\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v)$ 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

$$\tilde{\boldsymbol{\gamma}}(t) := (\mathscr{G}_{\mathcal{S}} \circ \boldsymbol{\gamma})(t) \,.$$

By Remark 4.188

$$\tilde{\boldsymbol{\gamma}}(t) = \mathscr{G}_{\mathcal{S}}(\boldsymbol{\gamma}(t)) = \mathscr{G}_{\mathcal{S}}(\boldsymbol{\sigma}(u_0 + t, v_0)) = \mathbf{N}(u_0 + t, v_0) \,.$$

Thus

$$\dot{\tilde{\boldsymbol{\gamma}}}(t) = \mathbf{N}_u(u_0 + t, v_0) \,, \quad \dot{\tilde{\boldsymbol{\gamma}}}(0) = \mathbf{N}_u(u_0, v_0) \,.$$

By definition of differential, we have

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) = -d_{\mathbf{p}}\mathscr{G}_{\mathcal{S}}(\boldsymbol{\sigma}_u) = -\dot{\tilde{\boldsymbol{\gamma}}}(0) = -\mathbf{N}_u(u_0, v_0) \,,$$

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 4.196

Proof: Proof of Theorem 4.196

By Theorem 4.80 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 4.198 and Lemma 4.197 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 $\mathscr{F}_2$.

4.12.2. Matrix of Weingarten map

The Weingarten map is a linear map

$$\mathscr{W}_{\mathbf{p},\mathcal{S}} : T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}.$$

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 $\mathscr{W}$ which represents $\mathscr{W}_{\mathbf{p},\mathcal{S}}$. It turns out that

$$\mathscr{W} = \mathscr{F}_1^{-1}\mathscr{F}_2 ,$$

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 4.199: Matrix of Weingarten map

Let $\mathcal{S}$ be an orientable surface with Weingarten map $\mathscr{W}_{\mathbf{p},\mathcal{S}}$. Let $\boldsymbol{\sigma}$ be a regular chart at $\mathbf{p}$. The matrix of the Weingarten map with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$ is

$$\mathscr{W} = \mathscr{F}_1^{-1}\mathscr{F}_2 ,$$

where the FFF and SFF are evaluated at $(u,v) = \boldsymbol{\sigma}^{-1}(\mathbf{p})$.

Proof

By Theorem 4.80, we know that $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ is a basis of $T_{\mathbf{p}}\mathcal{S}$. The matrix of the linear map

$$\mathscr{W}_{\mathbf{p},\mathcal{S}} \colon T_{\mathbf{p}}\mathcal{S} \to T_{\mathbf{p}}\mathcal{S}$$

with respect to such basis is given by

$$\mathscr{W} = \begin{pmatrix} a & c \\ b & d \end{pmatrix},$$

where the coefficients $a, b, c, d \in \mathbb{R}$ are such that

$$\begin{aligned} \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) &= a\boldsymbol{\sigma}_u + b\boldsymbol{\sigma}_v \\ \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v) &= c\boldsymbol{\sigma}_u + d\boldsymbol{\sigma}_v . \end{aligned}$$

By Lemma 4.198 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 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 4.197 we have

$$
\mathbf{N}_u \cdot \boldsymbol{\sigma}_u = -L\,, \qquad \mathbf{N}_u \cdot \boldsymbol{\sigma}_v = -M\,,
$$

$$
\mathbf{N}_v \cdot \boldsymbol{\sigma}_u = -M\,, \qquad \mathbf{N}_v \cdot \boldsymbol{\sigma}_v = -N\,.
$$

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

$$
\begin{pmatrix} L & M \\ M & N \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix},
$$

which reads

$$
\mathcal{F}_2 = \mathcal{F}_1 \mathcal{W} = \,.
$$

Now, notice that

$$
\det \mathcal{F}_1 > 0\,.
$$

This is true by Cauchy-Schwarz inequality:

$$
\mathbf{v} \cdot \mathbf{w} \leq \|\mathbf{v}\| \|\mathbf{w}\|\,, \quad \forall\, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3\,,
$$

where the inequality is strict if and only if $\mathbf{v}$ and $\mathbf{w}$ are linearly independent. Since $\mathcal{S}$ 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

$$\begin{aligned}\det(\mathscr{F}_1) &= EG - F^2 \\ &= (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u)(\boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v) - (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)^2 \\ &= \|\boldsymbol{\sigma}_u\|^2 \|\boldsymbol{\sigma}_v\|^2 - (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)^2 > 0 \,.\end{aligned}$$

Alternatively, we could have also noticed that

$$\|\boldsymbol{\sigma}_u\|^2 \|\boldsymbol{\sigma}_v\|^2 - (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)^2 = \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|^2$$

by the properties of vector product. Therefore,

$$\det(\mathscr{F}_1) = EG - F^2 = \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|^2 > 0$$

since $\boldsymbol{\sigma}$ is regular.
In particular the matrix $\mathscr{F}_1$ is invertible, and thus

$$\mathscr{W} = \mathscr{F}_1^{-1} \mathscr{F}_2 \,,$$

concluding the proof.

Remark 4.200: Matrix inverse

A matrix $A \in \mathbb{R}^{2\times 2}$ is invertible if and only if $\det(A) \neq 0$. In such case the inverse A^{-1} is computed via the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} , \qquad \det(A) = ad - bc \,.$$

If the matrix is diagonal, then

$$\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}^{-1} = \begin{pmatrix} 1/\lambda & 0 \\ 0 & 1/\mu \end{pmatrix} .$$

Notation

In the following we denote the matrix of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ by the symbol $\mathscr{W}$.

Example 4.201: Weingarten map of Helicoid

Question. The Helicoid is charted by

$$\boldsymbol{\sigma}(u,v) = (u\cos(v), u\sin(v), \lambda v)\,, \quad u \in \mathbb{R}\,,\ v \in (0,2\pi)\,,$$

with $\lambda > 0$ constant, see Figure 4.50. Compute the matrix of the Weingarten map.
Solution. We compute all the derivatives of $\boldsymbol{\sigma}$

$$\boldsymbol{\sigma}_u = (\cos(v), \sin(v), 0) \qquad \boldsymbol{\sigma}_{uv} = (-\sin(v), \cos(v), 0)$$
$$\boldsymbol{\sigma}_v = (-u\sin(v), u\cos(v), \lambda) \qquad \boldsymbol{\sigma}_{vv} = -u\,(\cos(v), \sin(v), 0)$$
$$\boldsymbol{\sigma}_{uu} = (0,0,0)$$

The FFF and its inverse are

$$E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 \qquad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0$$
$$G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = u^2 + \lambda^2$$
$$\mathcal{F}_1 = \begin{pmatrix} 1 & 0 \\ 0 & u^2+\lambda^2 \end{pmatrix} \qquad \mathcal{F}_1^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{1}{u^2+\lambda^2} \end{pmatrix}.$$

The standard unit normal to $\boldsymbol{\sigma}$ is

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = (\lambda\sin(v), -\lambda\cos(v), u)$$
$$\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = \sqrt{u^2+\lambda^2}$$
$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = \frac{1}{\sqrt{u^2+\lambda^2}}(\lambda\sin(v), -\lambda\cos(v), u)\,.$$

The SFF of $\boldsymbol{\sigma}$ is

$$L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} = 0 \qquad M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = -\frac{\lambda}{\sqrt{u^2+\lambda^2}}$$
$$N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} = 0$$
$$\mathcal{F}_2 = \begin{pmatrix} 0 & -\dfrac{\lambda}{\sqrt{u^2+\lambda^2}} \\ -\dfrac{\lambda}{\sqrt{u^2+\lambda^2}} & 0 \end{pmatrix}.$$

Finally, the matrix of the Weingarten map is

$$\mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2 = \begin{pmatrix} 0 & -\dfrac{\lambda}{(u^2+\lambda^2)^{1/2}} \\ -\dfrac{\lambda}{(u^2+\lambda^2)^{3/2}} & 0 \end{pmatrix}.$$

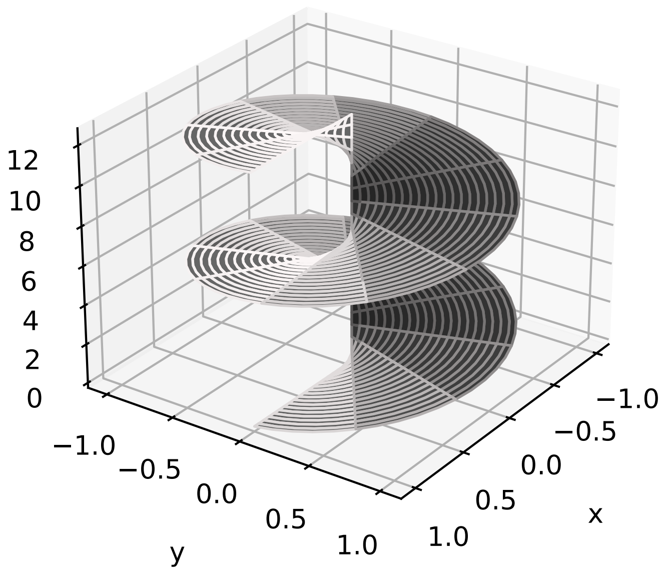


Figure 4.50.: Plot of Helicoid.

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 $\mathscr{S}$ 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 4.202: Gaussian and mean curvature

Let $\mathscr{S}$ be an orientable surface. Let $\mathscr{W}$ be the matrix of the Weingarten map $\mathscr{W}_{\mathbf{p},\mathscr{S}}$ of $\mathscr{S}$ at $\mathbf{p}$. We define:

1. The **Gaussian curvature** of $\mathcal{S}$ at $\mathbf{p}$ is
$$K := \det(\mathcal{W}),$$
2. The **mean curvature** of $\mathcal{S}$ at $\mathbf{p}$ is
$$H := \frac{1}{2}\operatorname{Tr}(\mathcal{W}),$$

Notation: Trace of a matrix

The **trace** of a 2×2 matrix is the sum of the diagonal entries
$$\operatorname{Tr}(A) = a + d, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Remark 4.203

The Gaussian curvature and mean curvature do not depend on the choice of basis of $T_{\mathbf{p}}\mathcal{S}$. Indeed, if $\widetilde{\mathcal{W}}$ 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
$$\det(\mathcal{W}) = \det(\widetilde{\mathcal{W}}), \quad \operatorname{Tr}(\mathcal{W}) = \operatorname{Tr}(\widetilde{\mathcal{W}}).$$

Check. 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
$$\mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2,$$
we can express K and H in terms of the first and second fundamental forms.

Proposition 4.204: Formulas for K and H

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart, and $\mathcal{S} = \boldsymbol{\sigma}(U)$. Then
$$K = \frac{LN - M^2}{EG - F^2}, \quad H = \frac{LG - 2MF - NE}{2(EG - F^2)}.$$

Proof

By Theorem 4.200 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

$$\begin{aligned} H &= \frac{1}{2}\operatorname{Tr}\mathscr{W} \\ &= \frac{1}{2}(w_{11} + w_{22}) \\ &= \frac{LG - 2MF + EN}{2(EG - F^2)} . \end{aligned}$$

Example 4.205: Curvatures of the Plane

Question. Let $\mathbf{a}, \mathbf{p}, \mathbf{q} \in \mathbb{R}^3$, with $\mathbf{p}, \mathbf{q}$ orthonormal. Consider the plane charted by

$$\boldsymbol{\sigma}(u,v) = \mathbf{a} + \mathbf{p}u + \mathbf{q}v \,.$$

1. Compute the matrix of the Weingarten map of $\boldsymbol{\sigma}$.
2. Compute the Gaussian and mean curvatures of the plane.

Solution.

1. From Examples 4.129, 4.181, the FFF and SFF of $\boldsymbol{\sigma}$ are

$$\mathcal{F}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathcal{F}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore the matrix of the Weingarten map is

$$\mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

2. The Gaussian and mean curvatures are

$$K = \det(\mathcal{W}) = 0\,, \qquad H = \frac{1}{2}\,\mathrm{Tr}(\mathcal{W}) = 0\,.$$

Example 4.206: Curvatures of the Unit cylinder

Question. Consider the unit cylinder $\mathcal{S}$ charted by

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v)\,.$$

1. Compute the matrix of the Weingarten map of $\boldsymbol{\sigma}$.
2. Compute the Gaussian and mean curvatures of $\mathcal{S}$.

Solution.

1. From Examples 4.127, 4.185, the FFF and SFF of $\boldsymbol{\sigma}$ are

$$\mathcal{F}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathcal{F}_2 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore the matrix of the Weingarten map is

$$\mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}.$$

2. The Gaussian and mean curvatures are

$$K = \det(\mathcal{W}) = 0\,, \qquad H = \frac{1}{2}\,\mathrm{Tr}(\mathcal{W}) = -\frac{1}{2}\,.$$

4.13.2. Principal curvatures

In order to define the principal curvatures, we need the notions of eigenvalue and eigenvector. For reader's convenience, we recall them in the next Remark.

Remark 4.207: Eigenvalues and eigenvectors

Let V be a two-dimensional vector space, and $L : V \to V$ a linear map. We say that $\lambda \in \mathbb{R}$ is an **eigenvalue** of L with **eigenvector** $\mathbf{v} \in V$ if

$$L(\mathbf{v}) = \lambda \mathbf{v} , \quad \mathbf{v} \neq 0 . \tag{4.16}$$

Let $\{\mathbf{v}_1, \mathbf{v}_2\}$ be a basis of V, and denote by

$$\mathbf{x} = (x_1, x_2) , \quad \mathbf{v} = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 .$$

the coordinates of $\mathbf{v}$ in such basis. Let $A \in \mathbb{R}^{2\times 2}$ be the matrix of L with respect to the basis $\{\mathbf{v}_1, \mathbf{v}_2\}$. Equation (4.16) is equivalent to

$$A\mathbf{x} = \lambda \mathbf{x} ,$$

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

$$A = P^{-1} D P .$$

Theorem 4.208: Eigenvalues of Weingarten map

Let $\mathcal{S}$ be an orientable surface and $\boldsymbol{\sigma}$ a regular chart at $\mathbf{p}$. Let $\mathcal{W}$ be the matrix of the Weingarten map $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$. Then

1. 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

$$\mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) = \kappa_1 \mathbf{t}_1 , \quad \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) = \kappa_2 \mathbf{t}_2 .$$

2. Let $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{R}$ be such that

$$\mathbf{t}_1 = \lambda_1 \boldsymbol{\sigma}_u + \mu_1 \boldsymbol{\sigma}_v , \quad \mathbf{t}_2 = \lambda_2 \boldsymbol{\sigma}_u + \mu_2 \boldsymbol{\sigma}_v .$$

   Denote $\mathbf{x}_1 = (\lambda_1, \mu_1)$ and $\mathbf{x}_2 = (\lambda_2, \mu_2)$. Then κ_1, κ_2 are eingenvalues of $\mathcal{W}$ of eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$

$$\mathcal{W} \mathbf{x}_1 = \kappa_1 \mathbf{x}_1 , \quad \mathcal{W} \mathbf{x}_2 = \kappa_2 \mathbf{x}_2 .$$

In particular, the matrix $\mathscr{W}$ is diagonalizable, with

$$\mathscr{W} = P^{-1}DP, \quad D = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix}, \quad P = \begin{pmatrix} \lambda_1 & \lambda_2 \\ \mu_1 & \mu_2 \end{pmatrix}.$$

Proof

Part 1. 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 $\mathscr{W}$ be the matrix of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ with respect to such basis. By Theorem 4.200 we have

$$\mathscr{W} = \mathscr{F}_1^{-1}\mathscr{F}_2 .$$

Recall that

$$\mathscr{F}_1^{-1} = \frac{1}{EG - F^2}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}.$$

Thus, $\mathscr{F}_1^{-1}$ is symmetric. Since

$$\mathscr{F}_2 = \begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

is symmetric, and the product of symmetric matrices is symmetric, we conclude that $\mathscr{W}$ is symmetric as well. Therefore $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ is self-adjoint, see Remark 4.15. The thesis now follows from the Spectral Theorem, see Theorem 4.13.

Part 2. We have just proven that

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) = \kappa_1 \mathbf{t}_1 , \quad \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) = \kappa_2 \mathbf{t}_2 .$$

As $\mathscr{W}$ is the matrix of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ and $\mathbf{x}_1$, $\mathbf{x}_2$ are the coordinates of $\mathbf{t}_1, \mathbf{t}_2$, we infer

$$\mathscr{W}\mathbf{x}_1 = \kappa_1 \mathbf{x}_1 , \quad \mathscr{W}\mathbf{x}_2 = \kappa_2 \mathbf{x}_2 ,$$

showing that κ_i is eigenvalue of $\mathscr{W}$ with eigenvector $\mathbf{x}_i$. In particular, it follows that $\mathscr{W}$ is diagonal in the basis $\{\mathbf{x}_1, \mathbf{x}_2\}$ of $\mathbb{R}^2$. Therefore $\mathscr{W} = P^{-1}DP$, with D and P as in the statement.

The principal curvatures are the eigenvalues of the matrix of the Weingarten map, and the principal vectors its eigenvectors.

Definition 4.209: Principal curvatures and vectors

Let $\mathcal{S}$ be an orientable surface. Let $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ the Weingarten map of $\mathcal{S}$ at $\mathbf{p}$. We define:

1. The **principal curvatures** of $\mathcal{S}$ at $\mathbf{p}$ are the eigenvalues κ_1, κ_2 of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$.
2. The **principal vectors** corresponding to κ_1 and κ_2 are the eigenvectors $\mathbf{t}_1, \mathbf{t}_2$ of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$.

Theorem 4.208 gives an explicit way to compute the principal curvatures and vectors.

Remark 4.210: Computing principal curvatures and vectors

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart and $\mathcal{S} = \boldsymbol{\sigma}(U)$.

1. Compute the FFF and SFF of $\boldsymbol{\sigma}$, and the matrix of the Weingarten map
$$\mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2 \,.$$
2. Compute the eigenvalues of $\mathcal{W}$, by solving for λ the equation
$$\det(\mathcal{W} - \lambda I) = 0 \,.$$
The two solutions are the principal curvatures κ_1 and κ_2.
3. Find scalars λ, μ which solve the linear system
$$(\mathcal{W} - \kappa_i I)\begin{pmatrix} \lambda \\ \mu \end{pmatrix} = 0 \,.$$
The solution(s) gives the eigenvector(s) of $\mathcal{W}$
$$\mathbf{x}_i = (\lambda, \mu)$$
corresponding to the eigenvalue κ_i.
4. The principal vector(s) associated to κ_i is
$$\mathbf{t}_i = \lambda \boldsymbol{\sigma}_u + \mu \boldsymbol{\sigma}_v$$

Remark 4.211: The case of $\mathcal{W}$ diagonal

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart and $\mathcal{S} = \boldsymbol{\sigma}(U)$. Assume the matrix of the Weingarten map is diagonal
$$\mathcal{W} = \begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} .$$
Then, the eigenvalues of $\mathcal{W}$ 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 of $\mathcal{S}$, 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 4.212: Relationships between curvatures

Let $\mathcal{S}$ be an orientable surface. Then

$$K = \kappa_1 \kappa_2 \,, \quad H = \frac{\kappa_1 + \kappa_2}{2} \,,$$
$$k_i = H \pm \sqrt{H^2 - K} \,.$$

Proof

Part 1. By Theorem 4.208 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

$$\det(P^{-1}) = \frac{1}{\det(P)} \,.$$

Part 2. The trace satisfies

$$\operatorname{Tr}(AB) = \operatorname{Tr}(BA) \,, \quad \forall \, A, B \in \mathbb{R}^{2\times 2} \,.$$

By definition of mean curvature and the above formula we get

$$\begin{aligned} H &= \frac{1}{2} \operatorname{Tr}(\mathscr{W}) \\ &= \frac{1}{2} \operatorname{Tr}\left(P^{-1} D P\right) \\ &= \frac{1}{2} \operatorname{Tr}\left(P P^{-1} D\right) \\ &= \frac{1}{2} \operatorname{Tr}(D) \\ &= \frac{1}{2} (\kappa_1 + \kappa_2) \,. \end{aligned}$$

Part 3. For any matrix $A \in \mathbb{R}^{2\times 2}$, we have

$$\begin{aligned}\det(A - \lambda I) &= \det \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} \\ &= (a-\lambda)(d-\lambda) - bc \\ &= \lambda^2 - (a+d)\lambda + ad - bc \\ &= \lambda^2 - \operatorname{Tr}(A)\lambda + \det(\lambda)\,.\end{aligned}$$

If $A = \mathscr{W}$, we obtain

$$\det(\mathscr{W} - \lambda I) = \lambda^2 - 2H\lambda + K\,.$$

Therefore, the principal curvatures are

$$\kappa_i = H \pm \sqrt{H^2 - K}\,.$$

Example 4.213: Principal curvatures of Unit Cylinder

Question. Consider the unit cylinder charted by

$$\boldsymbol{\sigma}(u,v) = (\cos(u), \sin(u), v)\,.$$

Compute the principal curvature and principal vectors.
Solution. By Example 4.206, 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 the eigenvectors are

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

Therefore, the principal curvatures and principal vectors are

$$\begin{aligned} &\kappa_1 = -1\,, \quad \kappa_2 = 0\,, \\ &\mathbf{t}_1 = \boldsymbol{\sigma}_u = (-\sin(u), \cos(v), 0)\,, \\ &\mathbf{t}_2 = \boldsymbol{\sigma}_v = (0,0,1)\,, \end{aligned}$$

as shown in Figure 4.51.

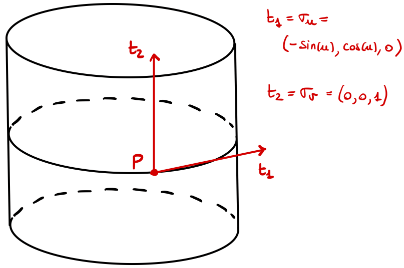


Figure 4.51.: Principal vectors of the unit cylinder.

Example 4.214: Curvatures of Sphere

Question. Consider the chart for the sphere

$$\boldsymbol{\sigma}(u,v) = (\cos(u)\cos(v), \sin(u)\cos(v), \sin(v)),$$

where $u \in (0, 2\pi)$, $v \in (-\pi/2, \pi/2)$. Prove that

$$\mathscr{F}_1 = \mathscr{F}_2 = \begin{pmatrix} \cos^2(v) & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathscr{W} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$
$$K = H = \kappa_1 = \kappa_2 = 1, \quad \mathbf{t}_1 = \boldsymbol{\sigma}_u, \quad \mathbf{t}_2 = \boldsymbol{\sigma}_v.$$

Solution. Compute the FFF of $\boldsymbol{\sigma}$

$$\begin{aligned} \boldsymbol{\sigma}_u &= (-\sin(u)\cos(v), \cos(u)\cos(v), 0) \\ \boldsymbol{\sigma}_v &= (-\cos(u)\sin(v), -\sin(u)\sin(v), \cos(v)) \\ E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = \cos^2(v) \\ F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\ G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1 \\ \mathscr{F}_1 &= \begin{pmatrix} \cos^2(v) & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Moreover

$$\begin{aligned} \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= (\cos(u)\cos^2(v), \sin(u)\cos^2(v), \cos(v)\sin(v)) \\ \|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| &= |\cos(v)| = \cos(v), \end{aligned}$$

where we used that $\cos(v) > 0$ since $v \in (-\pi/2, \pi/2)$. Therefore,

$$\begin{aligned}
\mathbf{N} &= (\cos(u)\cos(v), \sin(u)\cos(v), \sin(v)) \\
\boldsymbol{\sigma}_{uu} &= (-\cos(u)\cos(v), -\sin(u)\cos(v), 0) \\
\boldsymbol{\sigma}_{uv} &= (\sin(u)\sin(v), -\cos(u)\sin(v), 0) \\
\boldsymbol{\sigma}_{vv} &= (-\cos(u)\cos(v), -\sin(u)\cos(v), -\sin(v)) \\
L &= \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} = \cos^2(v) \\
M &= \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = 0 \\
N &= \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} = 1
\end{aligned}$$

Hence, the SFF and matrix of the Weingarten map are

$$\mathcal{F}_2 = \begin{pmatrix} \cos^2(v) & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Since $\mathcal{W}$ is diagonal, the principal curvatures and vectors are

$$\kappa_1 = \kappa_2 = 1\,, \quad \mathbf{t}_1 = \boldsymbol{\sigma}_u\,, \quad \mathbf{t}_2 = \boldsymbol{\sigma}_v\,.$$

Finally, the mean and Gaussian curvatures are

$$H = \frac{\kappa_1 + \kappa_2}{2} = 1\,, \quad K = \kappa_1\kappa_2 = 1\,.$$

Example 4.215: Curvatures of the Torus

Consider a circle $\mathcal{C}$ 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 $\mathcal{C}$ 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

$$\begin{aligned}
\mathcal{F}_1 &= \begin{pmatrix} b^2 & 0 \\ 0 & (a + b\cos(\theta))^2 \end{pmatrix} \\
\mathcal{F}_2 &= \begin{pmatrix} b & 0 \\ 0 & (a + b\cos(\theta))\cos(\theta) \end{pmatrix}.
\end{aligned}$$

Therefore the matrix of the Weingarten map is

$$\mathscr{W} = \mathscr{F}_1^{-1}\mathscr{F}_2 = \begin{pmatrix} \frac{1}{b} & 0 \\ 0 & \frac{\cos(\theta)}{a + b\cos(\theta)} \end{pmatrix}.$$

Since $\mathscr{W}$ 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 $\mathscr{S}$ be a regular surface and consider all the curves $\boldsymbol{\gamma}$ on $\mathscr{S}$ passing through the point $\mathbf{p} \in \mathscr{S}$. The shape of $\mathscr{S}$ at $\mathbf{p}$ influences the curvature of such curves.

Question 4.216

Which curves through $\mathbf{p}$ have greatest or lowest curvature?

We start our analysis with the following Definition.

Definition 4.217: Darboux frame

Let $\mathscr{S}$ be a regular surface, $\boldsymbol{\gamma} : (a,b) \to \mathscr{S}$ a unit-speed curve. The **Darboux frame** of $\boldsymbol{\gamma}$ at t is the triple

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{N}, \mathbf{N} \times \dot{\boldsymbol{\gamma}}\}\,,$$

where $\boldsymbol{\gamma}$ is evaluated at t, and $\mathbf{N}$ is the standard unit normal to $\mathscr{S}$, evaluated at $\mathbf{p} = \boldsymbol{\gamma}(t)$.

Proposition 4.218: Darboux frame is orthonormal basis

Let $\mathcal{S}$ be a regular surface, $\boldsymbol{\gamma} : (a,b) \to \mathcal{S}$ a unit-speed curve. The Darboux frame is an orthornormal basis of $\mathbb{R}^3$ for all $t \in (a,b)$.

Proof

By definition of tangent space, $\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}}\mathcal{S}$ when $\mathbf{p} := \boldsymbol{\gamma}(t)$. As $\mathbf{N}(\boldsymbol{\gamma}(t))$ is normal to $T_{\mathbf{p}}\mathcal{S}$, we get

$$\dot{\boldsymbol{\gamma}} \cdot \mathbf{N}(\boldsymbol{\gamma}(t)) = 0 .$$

As $\boldsymbol{\gamma}$ is unit-speed, we have $\|\dot{\boldsymbol{\gamma}}\| = 1$. Moreover, $\|\mathbf{N}\| = 1$ by definition. Therefore, $\dot{\boldsymbol{\gamma}}$ and $\mathbf{N}$ are orthonormal, and by the properties of vector product:

$$\|\mathbf{N} \times \dot{\boldsymbol{\gamma}}\| = \|\mathbf{N}\| \, \|\dot{\boldsymbol{\gamma}}\| = 1 .$$

Again using the properties of vector product, we have

$$(\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \cdot \mathbf{N} = 0 , \quad (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \cdot \dot{\boldsymbol{\gamma}} = 0 .$$

Therefore $\{\dot{\boldsymbol{\gamma}}, \mathbf{N}, \mathbf{N} \times \dot{\boldsymbol{\gamma}}\}$ is an orthonormal basis of $\mathbb{R}^3$.

Important

In general, the Darboux frame

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{N}, \mathbf{N} \times \dot{\boldsymbol{\gamma}}\}$$

does not coincide with the Frenet frame

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

of $\boldsymbol{\gamma}$. This is because the principal normal to $\boldsymbol{\gamma}$ is

$$\mathbf{n} = \frac{\ddot{\boldsymbol{\gamma}}}{\|\ddot{\boldsymbol{\gamma}}\|} = \frac{\ddot{\boldsymbol{\gamma}}}{\kappa} ,$$

and, in general, $\mathbf{n} \neq \mathbf{N}$.

Proposition 4.219: Coefficients of $\ddot{\boldsymbol{\gamma}}$ in the Darboux frame

Let $\mathcal{S}$ be a regular surface, $\boldsymbol{\gamma} : (a,b) \to \mathcal{S}$ a unit-speed curve. Then

$$\ddot{\boldsymbol{\gamma}} = \kappa_n \mathbf{N} + \kappa_g \, (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) , \tag{4.17}$$

where $\mathbf{N}$ is evaluated at $\mathbf{p} := \boldsymbol{\gamma}(t)$ and κ_n, κ_g are scalars depedent on $\mathbf{p}$. Moreover

$$\kappa_n = \ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} , \quad \kappa_g = \ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) , \tag{4.18}$$

$$\kappa^2 = \kappa_n^2 + \kappa_g^2 \,, \tag{4.19}$$

$$\kappa_n = \kappa \cos(\phi) \,, \quad \kappa_g = \pm \kappa \sin(\phi) \,, \tag{4.20}$$

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 4.218, we know that

$$\{\dot{\boldsymbol{\gamma}}, \mathbf{N}, \mathbf{N} \times \dot{\boldsymbol{\gamma}}\}$$

is an orthornormal basis of $\mathbb{R}^3$. Hence

$$\ddot{\boldsymbol{\gamma}} = a\dot{\boldsymbol{\gamma}} + b\mathbf{N} + c\,(\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \,,$$

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 both $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$, and

$$\dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = \|\dot{\boldsymbol{\gamma}}\|^2 = 1 \,.$$

Therefore, $a = 0$ and

$$\ddot{\boldsymbol{\gamma}} = b\mathbf{N} + c\,(\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \,.$$

Setting $\kappa_n := b$ and $\kappa_g := c$ we conclude (4.17).
Part 2. Taking the scalar product of (4.17) with $\mathbf{N}$ yields

$$\ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} = \kappa_n \|\mathbf{N}\|^2 + \kappa_g\,(\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \cdot \mathbf{N} = \kappa_n \,,$$

where we used that $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ are orthonormal vectors. Similarly, taking the scalar product of (4.17) with $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ yields the second equation in (4.18).
Part 3. By (4.17) we infer

$$\begin{aligned} \|\ddot{\boldsymbol{\gamma}}\|^2 &= \kappa_n^2 \|\mathbf{N}\|^2 + 2\kappa_n\kappa_g \mathbf{N} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) + \kappa_g^2 \|\mathbf{N} \times \dot{\boldsymbol{\gamma}}\|^2 \\ &= \kappa_n^2 + \kappa_g^2 \,, \end{aligned}$$

where we used that $\mathbf{N}$ and $\mathbf{N} \times \dot{\boldsymbol{\gamma}}$ are orthonormal. Since

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

we conclude (4.19).
Part 4. Recalling that

$$\ddot{\boldsymbol{\gamma}} = \kappa \mathbf{n} \,,$$

from the first equation in (4.18) we obtain

$$\begin{aligned} \kappa_n &= \ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} \\ &= \kappa \mathbf{n} \cdot \mathbf{N} \\ &= \kappa \|\mathbf{n}\| \|\mathbf{N}\| \cos(\phi) \\ &= \kappa \cos(\phi) \,, \end{aligned}$$

where we used that $\mathbf{n}$ and $\mathbf{N}$ have unit norm. Hence, the first equation in (4.20) is established. By (4.19) we get

$$\begin{aligned} \kappa_g^2 &= \kappa^2 - \kappa_n^2 \\ &= \kappa^2 - \kappa^2 \cos^2(\phi) \\ &= \kappa^2 (1 - \cos^2(\phi)) \\ &= \kappa^2 \sin^2(\phi) \,, \end{aligned}$$

from which we obtain the second equation in (4.20).

The quantities κ_n and κ_g are the normal and geodesic curvatures of $\boldsymbol{\gamma}$.

Definition 4.220: Normal and geodesic curvatures

Let $\mathcal{S}$ be regular and $\boldsymbol{\gamma} : (a,b) \to \mathcal{S}$ a unit-speed curve. Let $\mathbf{N}$ bet the standard unit normal to $\mathcal{S}$.

1. The **normal curvature** of $\boldsymbol{\gamma}$ is
$$\kappa_n = \ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} \,,$$
2. The **geodesic curvature** of $\boldsymbol{\gamma}$ is
$$\kappa_g = \ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) \,.$$

In particular:

- The normal curvature is the curvature of $\boldsymbol{\gamma}$ *forced* by being on the surface.
- The geodesic curvature is the *residual* curvature.

The normal curvature κ_n can be computed via the second fundamental form, as shown in the theorem below.

Theorem 4.221: Computing κ_n with SFF

Let $\mathcal{S}$ be a regular surface and $\boldsymbol{\gamma} : (a,b) \to \mathcal{S}$ 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 $\mathbf{p} = \boldsymbol{\gamma}(t)$. Then
$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$$
for some smooth functions $u, v : (a,b) \to \mathbb{R}$, and
$$\kappa_n = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2 ,$$
where L, M, N are evaluated at $(u(t), v(t))$, and $\dot{u}, \dot{v}$ at t.

Proof

Part 1. By definition of tangent space, $\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}}\mathcal{S}$ when $\mathbf{p} = \boldsymbol{\gamma}(t)$. By definition of differential, we have
$$d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) = (\mathbf{N} \circ \boldsymbol{\gamma})'(t) . \tag{4.21}$$
Since $\mathbf{N}(\boldsymbol{\gamma}(t))$ is normal to $T_{\mathbf{p}}(\mathcal{S})$ at $\mathbf{p} = \boldsymbol{\gamma}(t)$, and $\dot{\boldsymbol{\gamma}}(t) \in T_{\mathbf{p}}(\mathcal{S})$, we have
$$\mathbf{N}(\boldsymbol{\gamma}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) = 0 .$$
Differentiating the above expression, we get
$$\begin{aligned} 0 &= \frac{d}{dt} [\mathbf{N}(\boldsymbol{\gamma}(t)) \cdot \dot{\boldsymbol{\gamma}}(t)] \\ &= (\mathbf{N} \circ \boldsymbol{\gamma})'(t) \cdot \dot{\boldsymbol{\gamma}}(t) + \mathbf{N}(\boldsymbol{\gamma}(t)) \cdot \ddot{\boldsymbol{\gamma}}(t) \\ &= d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) + \mathbf{N}(\boldsymbol{\gamma}(t)) \cdot \ddot{\boldsymbol{\gamma}}(t) , \end{aligned}$$
where in the last equation we used (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
$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\dot{\boldsymbol{\gamma}}(t)) = -d_{\mathbf{p}}\mathscr{G}(\dot{\boldsymbol{\gamma}}(t)) = -d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) . \tag{4.23}$$
Therefore, using (4.22) and (4.23), we infer
$$\begin{aligned} II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}(t), \dot{\boldsymbol{\gamma}}(t)) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\dot{\boldsymbol{\gamma}}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) \\ &= -d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t)) \cdot \dot{\boldsymbol{\gamma}}(t) \\ &= \mathbf{N}(\boldsymbol{\gamma}(t)) \cdot \ddot{\boldsymbol{\gamma}}(t) = \kappa_n , \end{aligned}$$

where in the last equality we used (4.18).
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 4.124, 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 4.196, we obtain

$$\begin{aligned} \kappa_n &= II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}(t), \dot{\boldsymbol{\gamma}}(t)) \\ &= Ldu(\dot{\boldsymbol{\gamma}}(t))^2 + 2Mdu(\dot{\boldsymbol{\gamma}}(t))dv(\dot{\boldsymbol{\gamma}}(t)) + Ndv(\dot{\boldsymbol{\gamma}}(t))^2 \\ &= L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2 . \end{aligned}$$

Example 4.222: Curves on the sphere

Question. Consider the unit sphere $\mathbb{S}^2$ with chart

$$\boldsymbol{\sigma}(u, v) = (\cos(u)\cos(v), \sin(u)\cos(v), \sin(v)) .$$

Show that, for all unit-speed curves on $\mathbb{S}^2$,

$$\kappa_n(t) = 1 .$$

Solution. Let $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$ be a unit-speed curve on $\mathbb{S}^2$. Differentiating, we get Differentiating, we get

$$\begin{aligned} \dot{\boldsymbol{\gamma}}(t) &= \frac{d}{dt}(\cos(u(t))\cos(v(t)), \sin(u(t))\cos(v(t)), \sin(v(t))) \\ &= (-\dot{u}\sin(u)\cos(v) - \dot{v}\cos(u)\sin(v), \\ &\qquad \dot{u}\cos(u)\cos(v) - \dot{v}\sin(u)\sin(v), \\ &\qquad \dot{v}\cos(v)) \\ \|\dot{\boldsymbol{\gamma}}(t)\|^2 &= \cos^2(v)\dot{u}^2 + \dot{v}^2 . \end{aligned}$$

Since $\boldsymbol{\gamma}$ is unit-speed, we have $\|\dot{\boldsymbol{\gamma}}\| = 1$. Therefore,

$$\cos^2(v)\dot{u}^2 + \dot{v}^2 = 1 .$$

By Example 4.214, the coefficients of the SFF of $\boldsymbol{\sigma}$ are

$$L = \cos^2(v), \quad M = 0, \quad N = 1 .$$

By Theorem 4.220, the normal curvature of $\boldsymbol{\gamma}$ is

$$\kappa_n = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2 = \cos^2(v)\dot{u}^2 + \dot{v}^2 = 1\,.$$

The normal curvature κ_n is related to the principal curvatures κ_1 and κ_2.

Theorem 4.223: Euler's Theorem

Let $\mathcal{S}$ be a regular surface with principal curvatures κ_1, κ_2 and 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 4.208 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

$$\dot{\boldsymbol{\gamma}}(t) = \cos(\theta)\mathbf{t}_1 + \sin(\theta)\mathbf{t}_2\,. \tag{4.24}$$

In particular, we can take the scalar product of (4.24) 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

$$
\begin{aligned}
II_{\mathbf{p}}(\mathbf{t}_1, \mathbf{t}_1) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) \cdot \mathbf{t}_1 = \kappa_1 \left\|\mathbf{t}_1\right\|^2 = \kappa_1 \\
II_{\mathbf{p}}(\mathbf{t}_1, \mathbf{t}_2) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) \cdot \mathbf{t}_2 = \kappa_1 \mathbf{t}_1 \cdot \mathbf{t}_2 = 0 \\
II_{\mathbf{p}}(\mathbf{t}_2, \mathbf{t}_1) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) \cdot \mathbf{t}_1 = \kappa_2 \mathbf{t}_2 \cdot \mathbf{t}_1 = 0 \\
II_{\mathbf{p}}(\mathbf{t}_2, \mathbf{t}_2) &= \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) \cdot \mathbf{t}_2 = \kappa_2 \left\|\mathbf{t}_2\right\|^2 = \kappa_2
\end{aligned}
$$

By Theorem 4.220, equation (4.24), 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 4.224

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 4.222, we have shown by direct calculation that

$$
\kappa_n = 1
$$

for all unit-speed curves on the sphere. Thanks to Euler's Theorem, we can obtain the same result in a quicker way.

Example 4.225: Curves on the sphere (again)

Question. Same question as in Example 4.222.
Solution. By Example 4.214, the principal curvatures of the unit sphere are $\kappa_1 = \kappa_2 = 1$. By

Euler's Theorem, for any unit-speed curve $\boldsymbol{\gamma}$ on the sphere we have

$$\kappa_n = \kappa_1 \cos^2(\theta) + \kappa_2 \sin^2(\theta) = \cos^2(\theta) + \sin^2(\theta) = 1\,.$$

We have only defined normal and geodesic curvatures for unit-speed curves. We now extend the definition to regular curves.

Definition 4.226: κ_n and κ_g for regular $\boldsymbol{\gamma}$

Let $\mathcal{S}$ be regular, and $\boldsymbol{\gamma} : (a,b) \to \mathcal{S}$ a regular curve. Let $\tilde{\boldsymbol{\gamma}}$ be a unit-speed reparametrization of $\boldsymbol{\gamma}$, with

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

Let $\tilde{\kappa}_n$ and $\tilde{\kappa}_g$ be the normal and geodesic curvatures of $\tilde{\boldsymbol{\gamma}}$. The normal and geodesic curvatures of $\boldsymbol{\gamma}$ are

$$\kappa_n(t) = \tilde{\kappa}_n(\phi(t))\,, \qquad \kappa_g(t) = \tilde{\kappa}_g(\phi(t))\,.$$

It is immediate to check that κ_n and κ_g, as defined above, do not depend on the choice of unit-speed reparametrization. Therefore, $\tilde{\boldsymbol{\gamma}}$ can be taken as the arc-length reparametrization of $\boldsymbol{\gamma}$. The next Theorem gives practical formulas to compute κ_n and κ_g.

Theorem 4.227: Formulas for κ_n and κ_g

Let $\mathcal{S}$ be regular, and $\boldsymbol{\gamma} : (a,b) \to \mathcal{S}$ a regular curve.

1. The normal and geodesic curvatures of $\boldsymbol{\gamma}$ are given by

$$\kappa_n = \frac{\ddot{\boldsymbol{\gamma}} \cdot \mathbf{N}}{\|\dot{\boldsymbol{\gamma}}\|^2}\,, \qquad \kappa_g = \frac{\ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}})}{\|\dot{\boldsymbol{\gamma}}\|^3}\,.$$

2. Denote by κ the curvature of $\boldsymbol{\gamma}$. It holds

$$\kappa^2 = \kappa_n^2 + \kappa_g^2\,.$$

3. Let $\boldsymbol{\sigma}$ be a chart for $\mathcal{S}$ at $\mathbf{p} = \boldsymbol{\gamma}(t)$. Then

$$\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(u(t), v(t))$$

for some smooth functions $u, v : (a,b) \to \mathbb{R}$, and

$$\kappa_n = \frac{II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}})}{I_{\mathbf{p}}(\dot{\boldsymbol{\gamma}}, \dot{\boldsymbol{\gamma}})} = \frac{L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2}{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\,, \tag{4.25}$$

with E, F, G, L, M, N evaluated at $(u(t), v(t))$, and $\dot{u}, \dot{v}$ at t.

Proof

Part 1. Denote by s the arc-length function of $\boldsymbol{\gamma}$. In particular, we have

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

Differentiating, we obtain

$$\dot{\boldsymbol{\gamma}}(t) = \dot{\tilde{\boldsymbol{\gamma}}}(s(t))\dot{s}(t) \tag{4.26}$$

$$\ddot{\boldsymbol{\gamma}}(t) = \ddot{\tilde{\boldsymbol{\gamma}}}(s(t))\dot{s}^2(t) + \dot{\tilde{\boldsymbol{\gamma}}}(s(t))\dot{s}(t)\ddot{s}(t) \,. \tag{4.27}$$

Since $\tilde{\boldsymbol{\gamma}}$ is unit-speed, its normal and geodesic curvatures are, by definition

$$\begin{aligned} \tilde{\kappa}_n(s) &= \ddot{\tilde{\boldsymbol{\gamma}}}(s) \cdot \mathbf{N}(\tilde{\boldsymbol{\gamma}}(s)) \\ \tilde{\kappa}_g(s) &= \ddot{\tilde{\boldsymbol{\gamma}}}(s) \cdot \left[\mathbf{N}(\tilde{\boldsymbol{\gamma}}(s)) \times \dot{\tilde{\boldsymbol{\gamma}}}(s)\right] \,. \end{aligned}$$

Taking the scalar product of (4.27) with $\mathbf{N}(\boldsymbol{\gamma}(t))$ gives

$$\begin{aligned} \ddot{\boldsymbol{\gamma}}(t) \cdot \mathbf{N}(\boldsymbol{\gamma}(t)) &= \left[\ddot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot \mathbf{N}(\boldsymbol{\gamma}(t))\right] \dot{s}^2(t) + \\ &\qquad + \left[\dot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot \mathbf{N}(\boldsymbol{\gamma}(t))\right] \dot{s}(t)\ddot{s}(t) \\ &= \left[\ddot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot \mathbf{N}(\boldsymbol{\gamma}(t))\right] \dot{s}^2(t) \,, \end{aligned}$$

where we used that

$$\dot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot \mathbf{N}(\boldsymbol{\gamma}(t)) = \dot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot \mathbf{N}(\tilde{\boldsymbol{\gamma}}(s(t))) = 0 \,,$$

since $\dot{\tilde{\boldsymbol{\gamma}}}(s(t)) \in T_{\mathbf{q}}\mathcal{S}$ when $\mathbf{q} = \tilde{\boldsymbol{\gamma}}(s(t))$, and $\mathbf{N}(\tilde{\boldsymbol{\gamma}}(s(t)))$ is normal to $T_{\mathbf{q}}\mathcal{S}$. Therefore,

$$\begin{aligned} \ddot{\boldsymbol{\gamma}}(t) \cdot \mathbf{N}(\boldsymbol{\gamma}(t)) &= \left[\ddot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot \mathbf{N}(\boldsymbol{\gamma}(t))\right] \dot{s}^2(t) \\ &= \left[\ddot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot \mathbf{N}(\tilde{\boldsymbol{\gamma}}(s(t)))\right] \dot{s}^2(t) \\ &= \tilde{\kappa}_n(s(t)) \left\|\dot{\boldsymbol{\gamma}}(t)\right\|^2 \,, \end{aligned}$$

where we used the definition of $\tilde{\kappa}_n$ and that $\dot{s} = \|\dot{\boldsymbol{\gamma}}\|$. By definition of κ_n, we obtain

$$\kappa_n(t) = \tilde{\kappa}_n(s(t)) = \frac{\ddot{\boldsymbol{\gamma}}(t) \cdot \mathbf{N}(\boldsymbol{\gamma}(t))}{\left\|\dot{\boldsymbol{\gamma}}(t)\right\|^2} \,,$$

as required. Similarly, taking the scalar product of (4.27) with $\mathbf{N}(\boldsymbol{\gamma}(t)) \times \dot{\boldsymbol{\gamma}}(t)$ gives

$$\begin{aligned} \ddot{\boldsymbol{\gamma}}(t) \cdot [\mathbf{N}(\boldsymbol{\gamma}(t)) \times \dot{\boldsymbol{\gamma}}(t)] &= \ddot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot [\mathbf{N}(\boldsymbol{\gamma}(t)) \times \dot{\boldsymbol{\gamma}}(t)]\, \dot{s}^2(t) + \\ &\qquad + \dot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot [\mathbf{N}(\boldsymbol{\gamma}(t)) \times \dot{\boldsymbol{\gamma}}(t)]\, \dot{s}(t)\ddot{s}(t) \\ &= \ddot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot [\mathbf{N}(\boldsymbol{\gamma}(t)) \times \dot{\boldsymbol{\gamma}}(t)]\, \dot{s}^2(t) \,, \end{aligned}$$

where we used (4.26), which implies

$$\begin{aligned} &\dot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot [\mathbf{N}(\boldsymbol{\gamma}(t)) \times \dot{\boldsymbol{\gamma}}(t)]\, \dot{s}(t)\ddot{s}(t) = \\ &\qquad = \dot{\tilde{\boldsymbol{\gamma}}}(s(t)) \cdot \left[\mathbf{N}(\boldsymbol{\gamma}(t)) \times \dot{\tilde{\boldsymbol{\gamma}}}(s(t))\right] \dot{s}^2(t)\ddot{s}(t) = 0 \,, \end{aligned}$$

by the properties of vector product. Therefore,

$$\begin{aligned}\ddot{\boldsymbol{\gamma}}(t)\cdot[\mathbf{N}(\boldsymbol{\gamma}(t))\times\dot{\boldsymbol{\gamma}}(t)] &= \ddot{\tilde{\boldsymbol{\gamma}}}(s(t))\cdot[\mathbf{N}(\boldsymbol{\gamma}(t))\times\dot{\boldsymbol{\gamma}}(t)]\,\dot{s}^2(t)\\ &= \ddot{\tilde{\boldsymbol{\gamma}}}(s(t))\cdot\left[\mathbf{N}(\tilde{\boldsymbol{\gamma}}(s(t)))\times\dot{\tilde{\boldsymbol{\gamma}}}(s(t))\right]\dot{s}^3(t)\\ &= \tilde{\kappa}_g(s(t))\,\|\dot{\boldsymbol{\gamma}}(t)\|^3\,,\end{aligned}$$

where in the last to last equality we used (4.26), while in the last equality we used the definition of $\tilde{\kappa}_g$, and that $\dot{s} = \|\dot{\boldsymbol{\gamma}}\|$. By definition of κ_g, we get

$$\kappa_g(t) = \tilde{\kappa}_g(s(t)) = \frac{\ddot{\boldsymbol{\gamma}}(t)\cdot[\mathbf{N}(\boldsymbol{\gamma}(t))\times\dot{\boldsymbol{\gamma}}(t)]}{\|\dot{\boldsymbol{\gamma}}(t)\|^3}\,,$$

as required.
Part 2. Recall that the curvature of $\boldsymbol{\gamma}$ is defined by

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

where $\tilde{\kappa}$ is the curvature of $\tilde{\boldsymbol{\gamma}}$. By (4.19) we have that

$$\tilde{\kappa}^2(s) = \tilde{\kappa}_n^2(s) + \tilde{\kappa}_g^2(s)\,.$$

Therefore,

$$\begin{aligned}\kappa^2(t) &= \tilde{\kappa}^2(s(t))\\ &= \tilde{\kappa}_n^2(s(t)) + \tilde{\kappa}_g^2(s(t))\\ &= \kappa_n^2(t) + \kappa_g^2(t)\,.\end{aligned}$$

Part 3. Arguing as in the proof of Theorem 4.220, we observe that

$$\dot{\boldsymbol{\gamma}}(t)\cdot\mathbf{N}(\boldsymbol{\gamma}(t)) = 0\,,$$

since $\dot{\boldsymbol{\gamma}}(t)\in T_{\mathbf{p}}\mathcal{S}$ when $\mathbf{p} = \boldsymbol{\gamma}(t)$, and $\mathbf{N}(\boldsymbol{\gamma}(t))$ is orthogonal to $T_{\mathbf{p}}\mathcal{S}$. Differentiating, we get

$$\begin{aligned}0 &= \frac{d}{dt}\left[\dot{\boldsymbol{\gamma}}(t)\cdot\mathbf{N}(\boldsymbol{\gamma}(t))\right]\\ &= \ddot{\boldsymbol{\gamma}}(t)\cdot\mathbf{N}(\boldsymbol{\gamma}(t)) + \dot{\boldsymbol{\gamma}}(t)\cdot(\mathbf{N}\circ\boldsymbol{\gamma})'(t)\\ &= \ddot{\boldsymbol{\gamma}}(t)\cdot\mathbf{N}(\boldsymbol{\gamma}(t)) + \dot{\boldsymbol{\gamma}}(t)\cdot d_{\mathbf{p}}\mathbf{N}(\dot{\boldsymbol{\gamma}}(t))\\ &= \ddot{\boldsymbol{\gamma}}(t)\cdot\mathbf{N}(\boldsymbol{\gamma}(t)) + \dot{\boldsymbol{\gamma}}(t)\cdot d_{\mathbf{p}}\mathscr{G}(\dot{\boldsymbol{\gamma}}(t))\\ &= \ddot{\boldsymbol{\gamma}}(t)\cdot\mathbf{N}(\boldsymbol{\gamma}(t)) - \dot{\boldsymbol{\gamma}}(t)\cdot\mathscr{W}_{\mathbf{p},\mathcal{S}}(\dot{\boldsymbol{\gamma}}(t))\end{aligned}$$

from which we obtain

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\dot{\boldsymbol{\gamma}}(t))\cdot\dot{\boldsymbol{\gamma}}(t) = \ddot{\boldsymbol{\gamma}}(t)\cdot\mathbf{N}(\boldsymbol{\gamma}(t))\,.$$

Therefore, by definition of second fundamental form, we infer

$$II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}}) = \mathscr{W}_{\mathbf{p},\mathcal{S}}(\dot{\boldsymbol{\gamma}}) \cdot \dot{\boldsymbol{\gamma}} = \ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} \,. \tag{4.28}$$

Moreover, by definition of first fundamental form, we have

$$I_{\mathbf{p}}(\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}}) = \dot{\boldsymbol{\gamma}} \cdot \dot{\boldsymbol{\gamma}} = \|\dot{\boldsymbol{\gamma}}\|^2 \,. \tag{4.29}$$

Using (4.28)-(4.29), and the formula for κ_n obtained in Part 1, we get the first equality in (4.25)

$$\kappa_n = \frac{\ddot{\boldsymbol{\gamma}} \cdot \mathbf{N}}{\|\dot{\boldsymbol{\gamma}}\|^2} = \frac{II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}})}{I_{\mathbf{p}}(\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}})} \,.$$

The second equality in (4.25) follows because

$$\dot{\boldsymbol{\gamma}} = \frac{d}{dt}\boldsymbol{\sigma}(u(t),v(t)) = \boldsymbol{\sigma}_u \dot{u} + \boldsymbol{\sigma}_v \dot{v} \,,$$

and therefore

$$\begin{aligned} I_{\mathbf{p}}(\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}}) &= E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2 \\ II_{\mathbf{p}}(\dot{\boldsymbol{\gamma}},\dot{\boldsymbol{\gamma}}) &= L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2 \,. \end{aligned}$$

Example 4.228: Calculation of normal and geodesic curvatures

Question. For $v \neq 0$ and $t \neq 0$, consider the surface chart and curve

$$\boldsymbol{\sigma}(u,v) = \left(u, v, \frac{u}{v}\right), \quad \boldsymbol{\gamma}(t) = \boldsymbol{\sigma}(t^2, t) \,.$$

1. Prove that $\boldsymbol{\sigma}$ is regular.
2. Compute the principal unit normal to $\boldsymbol{\sigma}$.
3. Prove that $\boldsymbol{\gamma}$ is regular.
4. Compute the normal and geodesic curvatures of $\boldsymbol{\gamma}$.
5. Compute κ, the curvature of $\boldsymbol{\gamma}$. Verify that

$$\kappa^2 = \kappa_n^2 + \kappa_g^2 \,.$$

Solution.

1. The chart $\boldsymbol{\sigma}$ is regular because

$$\begin{aligned} \boldsymbol{\sigma}_u &= \left(1, 0, \frac{1}{v}\right), \quad \boldsymbol{\sigma}_v = \left(0, 1, -\frac{u}{v^2}\right) \\ \boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= \left(-\frac{1}{v}, \frac{u}{v^2}, 1\right) \neq \mathbf{0} \end{aligned}$$

2. The principal unit normal is

$$\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = \frac{\left(u^2+v^2+v^4\right)^{1/2}}{v^2}$$

$$\mathbf{N} = \frac{\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v}{\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\|} = \frac{\left(-v,u,v^2\right)}{\left(u^2+v^2+v^4\right)^{1/2}}.$$

3. The curve $\boldsymbol{\gamma}$ is regular because

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

$$\dot{\boldsymbol{\gamma}}(t) = (2t,1,1) \neq \mathbf{0}$$

4. Compute the following quantities

$$\|\dot{\boldsymbol{\gamma}}(t)\| = 2^{1/2}\,(2t^2+1)^{1/2}$$

$$\ddot{\boldsymbol{\gamma}}(t) = (2,0,0)$$

$$\mathbf{N}(t^2,t) = \frac{(-1,t,t)}{\left(2t^2+1\right)^{1/2}}$$

$$\ddot{\boldsymbol{\gamma}} \cdot \mathbf{N} = -\frac{2}{(2t^2+1)^{1/2}}$$

$$\mathbf{N} \times \dot{\boldsymbol{\gamma}} = \left(1+2t^2\right)^{1/2}(0,1,-1)$$

$$\ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}}) = 0$$

The normal and geodesic curvatures are

$$\kappa_n = \frac{\ddot{\boldsymbol{\gamma}} \cdot \mathbf{N}}{\|\dot{\boldsymbol{\gamma}}\|^2} = -\frac{1}{(2t^2+1)^{3/2}},$$

$$\kappa_g = \frac{\ddot{\boldsymbol{\gamma}} \cdot (\mathbf{N} \times \dot{\boldsymbol{\gamma}})}{\|\dot{\boldsymbol{\gamma}}\|^3} = 0\,.$$

5. The curvature of $\boldsymbol{\gamma}$ is

$$\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}} = (0,2,-2)\,, \quad \|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\| = 2^{3/2}$$

$$\kappa = \frac{\|\dot{\boldsymbol{\gamma}} \times \ddot{\boldsymbol{\gamma}}\|}{\|\dot{\boldsymbol{\gamma}}\|^3} = \frac{1}{(2t^2+1)^{3/2}}$$

Thus $\kappa = -\kappa_n$. Since $\kappa_g = 0$, we conclude that $\kappa^2 = \kappa_n^2 + \kappa_g^2$.

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 4.224. Hence, we can study the local shape of $\mathcal{S}$ in function of κ_1 and κ_2.

Theorem 4.229: 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

$$z = \frac{1}{2}\left(x^2\kappa_1(\mathbf{p}) + y^2\kappa_2(\mathbf{p})\right) , \tag{4.30}$$

where $\kappa_1(\mathbf{p}), \kappa_2(\mathbf{p})$ are the principal curvatures of $\mathcal{S}$ at $\mathbf{p}$.

Proof

By Theorem 4.208 the principal vectors $\{\mathbf{t}_1, \mathbf{t}_2\}$ form 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

$$\mathbf{t}_1 = (1, 0, 0) , \quad \mathbf{t}_2 = (0, 1, 0) , \quad \mathbf{N} = (0, 0, 1) . \tag{4.31}$$

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

$$(x, y, 0) = s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v , \tag{4.32}$$

where $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are evaluated at $(0, 0)$. The Taylor approximation of $\boldsymbol{\sigma}$ at $(0, 0)$ is

$$\begin{aligned}\boldsymbol{\sigma}(s, t) &= \boldsymbol{\sigma}(0, 0) + s\boldsymbol{\sigma}_u + t\boldsymbol{\sigma}_v \\ &\quad + \frac{1}{2}\left(s^2\boldsymbol{\sigma}_{uu} + 2st\boldsymbol{\sigma}_{uv} + t^2\boldsymbol{\sigma}_{vv}\right) + R , \\ &= (x, y, 0) + \frac{1}{2}\left(s^2\boldsymbol{\sigma}_{uu} + 2st\boldsymbol{\sigma}_{uv} + t^2\boldsymbol{\sigma}_{vv}\right) + R ,\end{aligned}$$

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 4.196 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 (4.31) and (4.32) 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 4.228 we can distinguish between 4 approximating shapes.

Definition 4.230: Local shape types

Let $\mathcal{S}$ be a regular surface, with $\kappa_1(\mathbf{p})$ and $\kappa_2(\mathbf{p})$ the 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 (4.30) 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 (4.30) 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 (4.30) is the equation of a **parabolic cylinder**.

- **Planar** if
$$\kappa_1(\mathbf{p}) = \kappa_2(\mathbf{p}) = 0$$
Then (4.30) is the equation of a **plane**.

Sometimes it is not easy to compute the principal curvatures κ_1, κ_2 explicitly. However, the Gaussian curvature K is simpler to compute, as it is just the determinant of the matrix of the Weingarten map. As $K = \kappa_1\kappa_2$, we can still infer some information about the local shape type from the knowledge of K.

Proposition 4.231: Gaussian curvature and local shape

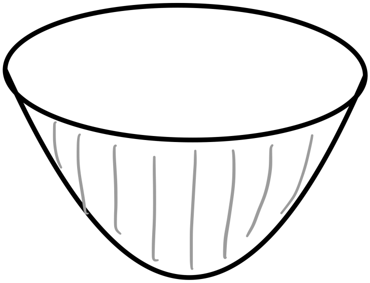
Let $\mathcal{S}$ be a regular surface, with $K(\mathbf{p})$ the Gaussian curvature at $\mathbf{p}$. The point $\mathbf{p}$ is

- **Elliptic** if $K(\mathbf{p}) > 0$,
- **Hyperbolic** if $K(\mathbf{p}) < 0$,
- **Parabolic** or **Planar** if $K(\mathbf{p}) = 0$.

Proof

The Gaussian curvature satisfies $K = \kappa_1\kappa_2$ where κ_1 and κ_2 are the principal curvatures of $\mathcal{S}$.

- If $K > 0$, then we must have either $\kappa_1, \kappa_2 > 0$ or $\kappa_1, \kappa_2 < 0$. Therefore, the point is elliptic.
- If $K < 0$, then we must have either $\kappa_1 < 0 < \kappa_2$ or $\kappa_2 < 0 < \kappa_1$. Therefore, the point is hyperbolic.
- If $K = 0$, then we might have

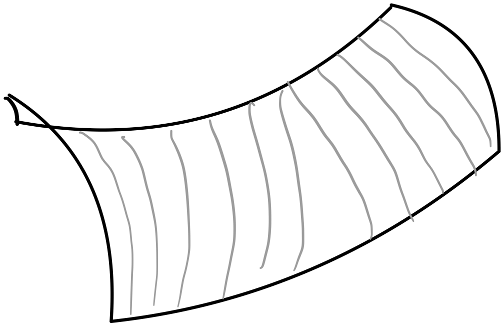


ELLIPTIC

$k_1, k_2 > 0$

OR

$k_1, k_2 < 0$

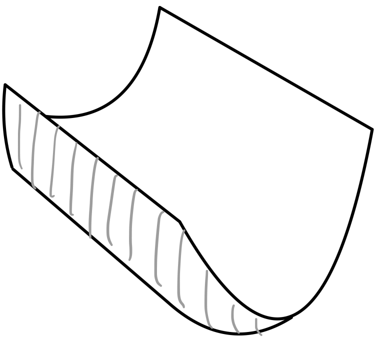


HYPERBOLIC

$k_1 < 0 < k_2$

OR

$k_2 < 0 < k_1$

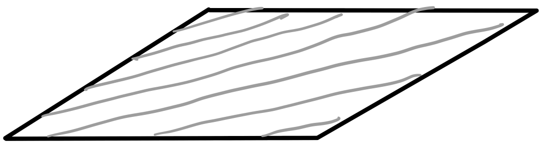


PARABOLIC

$k_1 = 0 \,,\; k_2 \neq 0$

OR

$k_2 = 0 \,,\; k_1 \neq 0$



PLANAR

$k_1 = k_2 = 0$

- $\kappa_1 \neq 0$ and $\kappa_2 = 0$, or $\kappa_1 = 0$ and $\kappa_2 \neq 0$. In both cases the point is parabolic.
- $\kappa_1 = \kappa_2 = 0$, in which case the point is planar.

Example 4.232: Analysis of local shape

Question. Consider the surface chart

$$\boldsymbol{\sigma}(u,v) = \left(u - v, u + v, u^2 + v^2\right) .$$

1. Compute the first fundamental form of $\boldsymbol{\sigma}$.
2. Compute the second fundamental form of $\boldsymbol{\sigma}$.
3. Compute the matrix of the Weingarten map.
4. Show that $\mathbf{p} = \boldsymbol{\sigma}(1,0)$ is an elliptic point.
5. Can there be points which are not elliptic?

Solution.

1. The FFF of $\boldsymbol{\sigma}$ is

$$\boldsymbol{\sigma}_u = (1,1,2u) \qquad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 4uv$$
$$\boldsymbol{\sigma}_v = (-1,1,2v) \qquad G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 2(1+2v^2)$$
$$E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 2(1+2u^2) \qquad \mathcal{F}_1 = 2\begin{pmatrix} 1+2u^2 & 2uv \\ 2uv & 1+2v^2 \end{pmatrix}$$

2. The standard unit normal is

$$\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v = 2(v-u, -u-v, 1)$$
$$\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| = 2\left(1+2u^2+2v^2\right)^{\frac{1}{2}}$$
$$\mathbf{N} = \frac{(v-u,-u-v,1)}{\left(1+2u^2+2v^2\right)^{\frac{1}{2}}}$$

The SFF of $\boldsymbol{\sigma}$ is

$$\boldsymbol{\sigma}_{uu} = (0,0,2) \qquad L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} = 2\left(1+2u^2+2v^2\right)^{-\frac{1}{2}}$$
$$\boldsymbol{\sigma}_{uv} = (0,0,0) \qquad M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} = 0$$
$$\boldsymbol{\sigma}_{vv} = (0,0,2) \qquad N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} = 2\left(1+2u^2+2v^2\right)^{-\frac{1}{2}}$$
$$\mathcal{F}_2 = \left(1+2u^2+2v^2\right)^{-\frac{1}{2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .$$

3. The inverse of $\mathscr{F}_1$ is

$$\begin{aligned}\mathscr{F}_1^{-1} &= \frac{1}{\det(\mathscr{F}_1)}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \\ &= \frac{1}{2(1+2u^2+2v^2)}\begin{pmatrix} 1+2v^2 & -2uv \\ -2uv & 1+2u^2 \end{pmatrix}.\end{aligned}$$

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. For $u = 1$ and $v = 0$ 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

$$\kappa_1(\mathbf{p}) = 3^{-\frac{3}{2}} > 0\,, \quad \kappa_2(\mathbf{p}) = 3^{-\frac{1}{2}} > 0\,.$$

Therefore $\mathbf{p}$ is an elliptic point.

5. No. This is because the Gaussian curvature is

$$K = \det(\mathscr{W}) = \frac{1}{(1+2u^2+2v^2)^2} > 0\,.$$

By Proposition 4.231 we conclude that every point is elliptic.

Definition 4.233: Umbilical point

Let $\mathscr{S}$ be a regular surface, with $\kappa_1(\mathbf{p})$ and $\kappa_2(\mathbf{p})$ the principal curvatures at $\mathbf{p}$. We say that $\mathbf{p}$ is an **umbilical point** if

$$\kappa_1(\mathbf{p}) = \kappa_2(\mathbf{p})\,.$$

Remark 4.234

Umbilical points might be **planar** or **elliptic**.

Example 4.235: Plane and Sphere

1. From Example 4.205, the principal curvatures of the plane are
$$\kappa_1 = \kappa_2 = 0\,.$$
Therefore, all points are umbilical, of planar type.
2. From Example 4.214, the principal curvatures of the sphere are
$$\kappa_1 = \kappa_2 = 1\,.$$
Therefore, all points are umbilical, of elliptic type.

Suppose that $\mathbf{p}$ is an umbilic, that is,
$$\kappa_1(\mathbf{p}) = \kappa_2(\mathbf{p})\,.$$
Let κ_n be the normal curvature of a unit-speed curve $\boldsymbol{\gamma}$ passing through $\mathbf{p}$. By Theorem 4.223 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 4.236: Structure theorem at umbilics

Let $\mathcal{S}$ be a regular surface such that every point $\mathbf{p} \in \mathcal{S}$ is umbilic. Then $\mathcal{S}$ is an open subset of plane or a sphere.

Proof

By assumption, we have
$$\kappa_1(\mathbf{p}) = \kappa_2(\mathbf{p}) = \kappa(\mathbf{p})\,, \quad \forall\, \mathbf{p} \in \mathcal{S}\,. \tag{4.33}$$
Step 1. κ is constant.
By Theorem 4.208 the principal vectors $\{\mathbf{t}_1, \mathbf{t}_2\}$ are an orthonormal basis of $T_{\mathbf{p}}\mathcal{S}$. Hence, for each $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$ there exist $\lambda, \mu \in \mathbb{R}$ such that
$$\mathbf{v} = \lambda \mathbf{t}_1 + \mu \mathbf{t}_2\,.$$
Using the linearity of $\mathscr{W}_{\mathbf{p},\mathcal{S}}$ and (4.33) we obtain
$$\begin{aligned} \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) &= \lambda \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) + \mu \mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) \\ &= \lambda\kappa \mathbf{t}_1 + \mu\kappa \mathbf{t}_2 \\ &= \kappa \mathbf{v}\,, \end{aligned}$$
showing that
$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) = \kappa \mathbf{v}\,, \quad \forall\, \mathbf{v} \in T_{\mathbf{p}}\mathcal{S}\,. \tag{4.34}$$

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a chart of $\mathcal{S}$. Up to restricting $\boldsymbol{\sigma}$, we can assume that U is connected. By Lemma 4.198 we have

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) = -\mathbf{N}_u\,, \quad \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v) = -\mathbf{N}_v\,.$$

On the other hand, by (4.34) we infer

$$\mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_u) = \kappa\boldsymbol{\sigma}_u\,, \quad \mathscr{W}_{\mathbf{p},\mathcal{S}}(\boldsymbol{\sigma}_v) = \kappa\boldsymbol{\sigma}_v\,,$$

from which

$$\mathbf{N}_u = -\kappa\boldsymbol{\sigma}_u\,, \quad \mathbf{N}_v = -\kappa\boldsymbol{\sigma}_v\,. \tag{4.35}$$

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

$$\kappa_v\boldsymbol{\sigma}_u = \kappa_u\boldsymbol{\sigma}_v\,. \tag{4.36}$$

Recall that $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ are linearly independent, being $\mathcal{S}$ regular. Hence the linear combination at (4.36) 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 (4.35) 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 (4.35) 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 $1/\kappa$.

Proposition 4.237: Criterion for umbilics

Let $\mathcal{S}$ be a regular surface. The point $\mathbf{p}$ is umbilical if and only if

$$H^2(\mathbf{p}) = K(\mathbf{p}) .$$

In particular, $\mathbf{p}$ cannot be umbilical if

$$K(\mathbf{p}) < 0 .$$

Proof

Part 1. By Proposition 4.212, the principal curvatures are

$$\kappa_1 = H + \sqrt{H^2 - K} , \quad \kappa_2 = H - \sqrt{H^2 - K} .$$

By definition, $\mathbf{p}$ is umbilic if and only if $\kappa_1 = \kappa_2$ at $\mathbf{p}$, which is equivalent to $H^2 - K = 0$.
Part 2. If $K(\mathbf{p}) < 0$, then we cannot have that $K(\mathbf{p}) = H^2(\mathbf{p})$. Therefore, by Part 1, $\mathbf{p}$ cannot be umbilical.

Proposition 4.238: Chart criterion for umbilics

Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ be a regular chart and $\mathcal{S} = \boldsymbol{\sigma}(U)$. A point $\mathbf{p}$ is umbilic if and only if there exists a scalar κ such that

$$\mathcal{F}_2 = \kappa \mathcal{F}_1 .$$

Proof

Part 1. By Theorem 4.208, there exists a basis $\{\mathbf{t}_1, \mathbf{t}_2\}$ of $T_{\mathbf{p}}\mathcal{S}$ such that the Weingarten map satisfies

$$\mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) = \kappa_1 \mathbf{t}_1 , \quad \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) = \kappa_2 \mathbf{t}_2 ,$$

where κ_1 and κ_2 are the principal curvatures of $\mathcal{S}$ at $\mathbf{p}$. If $\mathbf{p}$ is umbilical, then

$$\kappa_1 = \kappa_2 = \kappa\,.$$

Let $\mathbf{v} \in T_{\mathbf{p}}\mathcal{S}$. Then $\mathbf{v} = \lambda \mathbf{t}_1 + \mu \mathbf{t}_2$ for some $\lambda, \mu \in \mathbb{R}$. By linearity of $\mathcal{W}_{\mathbf{p},\mathcal{S}}$, we get

$$\begin{aligned} \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{v}) &= \mathcal{W}_{\mathbf{p},\mathcal{S}}(\lambda \mathbf{t}_1 + \mu \mathbf{t}_2) \\ &= \lambda \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_1) + \mu \mathcal{W}_{\mathbf{p},\mathcal{S}}(\mathbf{t}_2) \\ &= \lambda \kappa \mathbf{t}_1 + \mu \kappa \mathbf{t}_2 \\ &= \kappa \mathbf{v}\,, \end{aligned}$$

showing that $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ is a multiple of the identity map. Therefore, the matrix representation of $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ with respect to any basis of $T_{\mathbf{p}}\mathcal{S}$ is a multiple of the identity matrix. In particular,

$$\mathcal{W} = \kappa I\,,$$

where $\mathcal{W}$ is the matrix of $\mathcal{W}_{\mathbf{p},\mathcal{S}}$ with respect to the basis $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v\}$ of $T_{\mathbf{p}}\mathcal{S}$. Recalling that $\mathcal{W} = \mathcal{F}_1^{-1}\mathcal{F}_2$, we obtain

$$\mathcal{W} = \kappa I \quad \Longrightarrow \quad \mathcal{F}_1^{-1}\mathcal{F}_2 = \kappa I \quad \Longrightarrow \quad \mathcal{F}_2 = \kappa \mathcal{F}_1\,.$$

Example 4.239: Plane and Sphere

1. If the plane is charted as in Example 4.205, the FFF and SFF are

$$\mathcal{F}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathcal{F}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

   Therefore $\mathcal{F}_2 = \kappa \mathcal{F}_1$ with $\kappa = 0$, and all points are umbilical.

2. If the sphere is charted as in Example 4.214, the FFF and SFF are

$$\mathcal{F}_1 = \mathcal{F}_2 = \begin{pmatrix} \cos^2(v) & 0 \\ 0 & 1 \end{pmatrix}.$$

   Since $\mathcal{F}_2 = \mathcal{F}_1$, all points on the sphere are umbilical.

Remark 4.240: How to find umbilics

Condition $\mathcal{F}_2 = \kappa \mathcal{F}_1$ is equivalent to

$$(E, F, G) \times (L, M, N) = \mathbf{0}\,.$$

In practice, umbilics can be found by solving the above equations. Common factors may be discarded, if convenient.

Proof

By the properties of vector product, we have that

$$(E, F, G) \times (L, M, N) = \mathbf{0}$$

if and only if the vectors (E, F, G) and (L, M, N) are parallel. Therefore, there exists a constant κ such that

$$(L, M, N) = \kappa(E, F, G) \qquad \Longleftrightarrow \qquad \mathcal{F}_2 = \kappa \mathcal{F}_1 \,.$$

Example 4.241: Local shape of the Monkey Saddle

Question. Consider the *Monkey Saddle* surface $\mathcal{S}$ described by

$$z = x^3 - 3xy^2 \,.$$

1. Compute the Gaussian curvature of $\mathcal{S}$.
2. Does $\mathcal{S}$ contain any hyperbolic point?
3. Prove that the origin is the only umbilical point.

Solution. The Monkey Saddle is charted by

$$\boldsymbol{\sigma}(u, v) = (u, v, u^3 - 3uv^2) \,.$$

The FFF of $\boldsymbol{\sigma}$ is

$$\boldsymbol{\sigma}_u = (1, 0, 3(u^2 - v^2)) \qquad F = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = -18uv(u^2 - v^2)$$

$$\boldsymbol{\sigma}_v = (0, 1, -6uv) \qquad G = \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1 + 36u^2v^2$$

$$E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = 1 + 9(u^2 - v^2)^2$$

The SFF of $\boldsymbol{\sigma}$ is

$$
\begin{aligned}
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= (-3(u^2 - v^2), 6uv, 1) \\
\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| &= 1 + 36u^2v^2 + 9(u^2 - v^2)^2 \\
&= 1 + 9u^4 + 9v^4 + 18u^2v^2 \\
&= 1 + 9(u^2 + v^2)^2 \\
\mathbf{N} &= \frac{(-3(u^2 - v^2), 6uv, 1)}{\sqrt{1 + 9(u^2 + v^2)^2}} \\
\boldsymbol{\sigma}_{uu} &= (0, 0, 6u) \\
\boldsymbol{\sigma}_{uv} &= (0, 0, -6v) \\
\boldsymbol{\sigma}_{vv} &= (0, 0, -6u) \\
L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} &= \frac{6u}{\sqrt{1 + 9(u^2 + v^2)^2}} \\
M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} &= \frac{-6v}{\sqrt{1 + 9(u^2 + v^2)^2}} \\
N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} &= \frac{6u}{\sqrt{1 + 9(u^2 + v^2)^2}}
\end{aligned}
$$

1. We have that

$$
\begin{aligned}
EG - F^2 &= (1 + 9(u^2 - v^2)^2)(1 + 36u^2v^2) - (-18uv(u^2 - v^2))^2 \\
&= 1 + 36u^2v^2 + 9(u^2 - v^2)^2 \\
&= 1 + 9u^4 + 9v^4 + 18u^2v^2 \\
&= 1 + 9(u^2 + v^2)^2 \\
LN - M^2 &= -\frac{36(u^2 + v^2)}{1 + 9(u^2 + v^2)^2}
\end{aligned}
$$

   Therefore the Gaussian curvature is

$$
K = \frac{LN - M^2}{EG - F^2} = -\frac{36(u^2 + v^2)}{[1 + 9(u^2 + v^2)^2]^2} .
$$

2. Note that

$$
K < 0 , \quad \forall \, (u, v) \neq (0, 0) .
$$

   By Proposition 4.231, we conclude that all the points outside of the origin are hyperbolic.

3. Since $K < 0$ everywhere except at the origin, Proposition 4.236 implies that points outside the origin cannot be umbilic. At $(0, 0)$, we have

$$
\mathscr{F}_1 = du^2 + dv^2 , \quad \mathscr{F}_2 = 0 .
$$

Therefore $\mathscr{F}_2$ is a multiple of $\mathscr{F}_1$, and by Proposition 4.238 we conclude that $(0,0)$ is an umbilical point. Note: the matrix of the Weingarten map is $\mathscr{W} = \mathscr{F}_1^{-1}\mathscr{F}_2 = 0$. Therefore the principal curvatures are $\kappa_1 = \kappa_2 = 0$, showing that $(0,0)$ is a planar point.

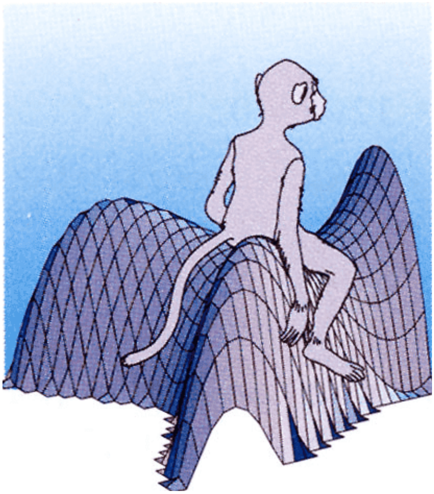


Figure 4.53.: The Monkey Saddle surface $z = x^3 - 3xy^2$.

4.15. Conclusion: FTS and Theorema Egregium

We conclude by discussing two important and powerful Theorems:

1. Fundamental Theorem of Surfaces (FTS)
2. Theorema Egregium

We proceed in analogy with curves: for each point of the surface we assign a basis of $\mathbb{R}^3$, analogous to the Frenet frame. Let $\mathcal{S}$ be an orientable surface and $\boldsymbol{\sigma}$ a chart at $\mathbf{p}$. The triple

$$\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v, \mathbf{N}\}$$

gives a basis of $\mathbb{R}^3$ (not orthonormal, but it does not matter). We can now express the derivatives of $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ with respect to such basis.

Proposition 4.242: Christoffel symbols

Let $\boldsymbol{\sigma}$ be a regular chart, with first and second fundamental forms given by

$$Edu^2 + Fdudv + Gdv^2\,, \quad Ldu^2 + Mdudv + Ndv^2\,,$$

Then

$$
\begin{aligned}
\boldsymbol{\sigma}_{uu} &= \Gamma^1_{11}\boldsymbol{\sigma}_u + \Gamma^2_{11}\boldsymbol{\sigma}_v + L\mathbf{N} \\
\boldsymbol{\sigma}_{uv} &= \Gamma^1_{12}\boldsymbol{\sigma}_u + \Gamma^2_{12}\boldsymbol{\sigma}_v + M\mathbf{N} \\
\boldsymbol{\sigma}_{vv} &= \Gamma^1_{22}\boldsymbol{\sigma}_u + \Gamma^2_{22}\boldsymbol{\sigma}_v + N\mathbf{N}
\end{aligned}
$$

where

$$
\begin{aligned}
\Gamma^1_{11} &= \frac{GE_u - 2FF_u + FE_v}{2\left(EG - F^2\right)} \\
\Gamma^2_{11} &= \frac{2EF_u - EE_v - FE_u}{2\left(EG - F^2\right)} \\
\Gamma^1_{12} &= \frac{GE_v - FG_u}{2\left(EG - F^2\right)} \\
\Gamma^2_{12} &= \frac{EG_u - FE_v}{2\left(EG - F^2\right)} \\
\Gamma^1_{22} &= \frac{2GF_v - GG_u - FG_v}{2\left(EG - F^2\right)} \\
\Gamma^2_{22} &= \frac{EG_v - 2FF_v + FG_u}{2\left(EG - F^2\right)}
\end{aligned}
$$

The six coefficients Γ^k_{ij} are called the **Christoffel symbols** of $\boldsymbol{\sigma}$.

Proof

Since $\{\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v, \mathbf{N}\}$ is a basis of $\mathbb{R}^3$, there exists scalars $\alpha_i, \beta_i, \gamma_i$ such that

$$
\begin{aligned}
\boldsymbol{\sigma}_{uu} &= \alpha_1\boldsymbol{\sigma}_u + \alpha_2\boldsymbol{\sigma}_v + \alpha_3\mathbf{N} \\
\boldsymbol{\sigma}_{uv} &= \beta_1\boldsymbol{\sigma}_u + \beta_2\boldsymbol{\sigma}_v + \beta_3\mathbf{N} \\
\boldsymbol{\sigma}_{vv} &= \gamma_1\boldsymbol{\sigma}_u + \gamma_2\boldsymbol{\sigma}_v + \gamma_3\mathbf{N}
\end{aligned}
\tag{4.37}
$$

Recall that the coefficient of the second fundamental form are

$$
L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N}\,, \quad M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N}\,, \quad N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N}\,.
$$

Therefore, taking the dot product of each equation in (4.37) with $\mathbf{N}$ gives

$$
\alpha_3 = L\,, \quad \beta_3 = M\,, \quad \gamma_3 = N\,,
$$

where we used that $\mathbf{N}$ is orthogonal to both $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$. Taking the dot product of each equation in (4.37) with $\boldsymbol{\sigma}_u$ and $\boldsymbol{\sigma}_v$ gives 6 scalar equations which determine the Christoffel symbols Γ^k_{ij}. For example, dotting the first equation in (4.37) with $\boldsymbol{\sigma}_u$ gives

$$
\boldsymbol{\sigma}_{uu} \cdot \boldsymbol{\sigma}_u = \alpha_1\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u + \alpha_2\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = \alpha_1 E + \alpha_2 F\,.
$$

On the other hand,

$$E_u = \frac{d}{du}(\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u) = 2\boldsymbol{\sigma}_{uu} \cdot \boldsymbol{\sigma}_u \,,$$

from which we get

$$\alpha_1 E + \alpha_2 F = \frac{1}{2} E_u \,. \tag{4.38}$$

Similarly, dotting the first equation in (4.37) with $\boldsymbol{\sigma}_v$ gives

$$\boldsymbol{\sigma}_{uu} \cdot \boldsymbol{\sigma}_v = \alpha_1 \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v + \alpha_2 \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = \alpha_1 F + \alpha_2 G \,.$$

On the other hand,

$$\boldsymbol{\sigma}_{uu} \cdot \boldsymbol{\sigma}_v = (\boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v)_v - \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_{uv} = F_v - \frac{1}{2} E_v \,,$$

from which we obtain

$$\alpha_1 F + \alpha_2 G = F_v - \frac{1}{2} E_v \,. \tag{4.39}$$

Equations (4.38) and (4.39) foem a 2x2 linear system in α_1 and α_2, which reads

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} E_u \\ F_v - \frac{1}{2} E_v \end{pmatrix} .$$

Inverting the matrix of the first fundamental form, we obtain

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \frac{1}{EG - F^2} \begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \begin{pmatrix} \frac{1}{2} E_u \\ F_v - \frac{1}{2} E_v \end{pmatrix} ,$$

which gives the first 2 Christoffel symbols

$$\begin{aligned} \alpha_1 &= \frac{GE_u - 2FF_v + FE_v}{2(EG - F^2)} = \Gamma^1_{11} \\ \alpha_2 &= \frac{-FE_u + 2EF_v - EE_v}{2(EG - F^2)} = \Gamma^2_{11} \end{aligned}$$

The remaining 4 Christoffel symbols are obtained in a similar manner.

Note that the Christoffel symbols depend only on the first fundamental form of $\boldsymbol{\sigma}$.

The question is whether there are relations between the first and second fundamental forms. As it turns out, all the existing relations are encoded in two sets of equations:

- Codazzi-Mainardi Equations
- Gauss Equations

Proposition 4.243: Codazzi-Mainardi and Gauss Equations

Let $\boldsymbol{\sigma}$ be a regular chart, with first and second fundamental forms

$$Edu^2 + Fdudv + Gdv^2 \,, \quad Ldu^2 + Mdudv + Ndv^2 \,,$$

and Christoffel symbols as in Proposition 4.242. They are satisfied:

1. The **Codazzi-Mainardi Equations**

$$\begin{aligned} L_v - M_u &= L\Gamma^1_{12} + M\left(\Gamma^2_{12} - \Gamma^1_{11}\right) - N\,\Gamma^2_{11} \\ M_v - N_u &= L\Gamma^1_{22} + M\left(\Gamma^2_{22} - \Gamma^1_{12}\right) - N\,\Gamma^2_{12} \end{aligned}$$

2. The **Gauss Equations**

$$\begin{aligned} EK &= \left(\Gamma^2_{11}\right)_v - \left(\Gamma^2_{12}\right)_u + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - \left(\Gamma^2_{12}\right)^2 \\ FK &= \left(\Gamma^1_{12}\right)_u - \left(\Gamma^1_{11}\right)_v + \Gamma^2_{12}\Gamma^1_{12} - \Gamma^2_{11}\Gamma^1_{22} \\ FK &= \left(\Gamma^2_{12}\right)_v - \left(\Gamma^2_{22}\right)_u + \Gamma^1_{12}\Gamma^2_{12} - \Gamma^1_{22}\Gamma^2_{11} \\ GK &= \left(\Gamma^1_{22}\right)_u - \left(\Gamma^1_{12}\right)_v + \Gamma^1_{22}\Gamma^1_{11} + \Gamma^2_{22}\Gamma^1_{12} - \left(\Gamma^1_{12}\right)^2 - \Gamma^2_{12}\Gamma^1_{22} \end{aligned}$$

   where K denotes the Gaussian curvature of $\boldsymbol{\sigma}$.

The proof involves a lot of calculations, and we decide to omit it. For a reference, see Propositions 10.1.1 and 10.1.2 in [7].

The Codazzi-Mainardi and Gauss equations are necessary and sufficient to completely determine a surface, up to rigid motions. This is the statement of the *Fundamental Theorem of Surfaces*, which can be seen as the surfaces analogue of the Fundamental Theorem of Space Curves: Curvature and torsion completely characterize a regular curve, up to rigid motions. The equivalent Theorem for surfaces states that First and Second Fundamental Forms, with coefficients satisfying the Codazzi-Mainardi and Gauss equations, completely determine a surface, up to rigid motions.

Theorem 4.244: Fundamental Theorem of Surfaces (FTS)

1. Let $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ and $\tilde{\boldsymbol{\sigma}} : U \to \mathbb{R}^3$ be regular surface charts with the same first and second fundamental form. Then, there exists a rigid motion $M : \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$\tilde{\boldsymbol{\sigma}} = M(\boldsymbol{\sigma}) \,.$$

2. Let $V \subseteq \mathbb{R}^3$ be open, and

$$E, F, G, L, M, N : V \to \mathbb{R}$$

   be smooth functions on V, such that

$$E > 0, \quad G > 0, \quad EG - F^2 > 0 \,,$$

and satisfying the Codazzi-Mainardi and Gauss equations in Proposition 4.243, with

$$K = \frac{LN - M^2}{EG - F^2} \,.$$

Then, if $(u_0, v_0) \in V$, there exists

- an open set $U \subseteq V$ containing (u_0, v_0)
- a regular surface chart $\boldsymbol{\sigma} : U \to \mathbb{R}^3$ with first and second fundamental forms given by

$$Edu^2 + Fdudv + Gdv^2 \,, \quad Ldu^2 + Mdudv + Ndv^2 \,.$$

The proof of the FTS is very complicated, as it involves PDEs, and is found at page 239 of [3]. The FTS is the reason why the differential geometry of surfaces is still an active field of research today:

- If one wishes to construct a surface with prescribed first and second fundamental form, then one needs to solve the Codazzi-Mainardi and Gauss Equations
- These are very complicated PDEs
- Examples of active research directions are
  - Minimal Surfaces
  - Costant Mean Curvature Surfaces
  - Geometric Flows: a surface evolves following the direction of steepest descent of some energy
  - An example of Geometric flow is the Mean Curvature Flow, in which $\mathcal{S}$ minimizes

$$F(\mathcal{S}) = \int_{\mathcal{S}} H \, dA \,,$$

where H is the mean curvature of $\mathcal{S}$.

The other major result we want to talk about is the Theorema Egregium (which means remarkable) by Gauss.

Theorem 4.245: Theorema Egregium

The Gaussian curvature is invariant under local isometries.

Proof

Let $\boldsymbol{\sigma}$ be a regular surface chart. The first Gauss equation in 4.243 gives that

$$KE = (\Gamma_{11}^2)_v - (\Gamma_{12}^2)_u + \Gamma_{11}^1\Gamma_{12}^2 + \Gamma_{11}^2\Gamma_{22}^2 - \Gamma_{12}^1\Gamma_{11}^2 - (\Gamma_{12}^2)^2 \,,$$

where K is the Gauss curvature , E one of the coefficients of the first fundamental form, and Γ_{ij}^k the Christoffel symbols of $\boldsymbol{\sigma}$. As $\boldsymbol{\sigma}$ is regular, we have that $E = \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u > 0$. Therefore, we can divide by E and obtain

$$K = \frac{(\Gamma^2_{11})_v - (\Gamma^2_{12})_u + \Gamma^1_{11}\Gamma^2_{12} + \Gamma^2_{11}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^2_{11} - \left(\Gamma^2_{12}\right)^2}{E} . \tag{4.40}$$

Note that the Christoffel symbols depend only on the coefficients of the first fundamental form. Therefore, the RHS of (4.40) depends only on the coefficients of the first fundamental form. Since the first fundamental form is invariant under local isometries, we conclude that the Gaussian curvature K is invariant under local isometries.

The Theorem is remarkable because the Gaussian curvature is defined in terms of both first and second fundamental forms

$$K = \det(\mathscr{W}) = \frac{LN - M^2}{EG - F^2} .$$

Being a curvature, K should depend on how the surface bends in space. Instead, the Theorema Egregium shows that K can be computed using only the first fundamental form, which is a quantity intrinsic to the surface.

As an immediate application, we obtain that there is no perfect World Map:

Proposition 4.246

There is no isometry between and open set of the unit sphere $\mathbb{S}^2$ and the plane.

Proof

Suppose there was an isometry

$$\boldsymbol{\sigma} : U \to \mathbb{S}^2$$

for some open set $U \subseteq \mathbb{R}^2$. The Theorema Egregium implies that $\mathbb{S}^2$ and U have the same Gaussian curvature. However, the Gaussian curvature of the unit sphere is $K = 1$, while the one of the plane is $K = 0$.

In particular, we have have proven that a map

$$\boldsymbol{\sigma} : U \subseteq \mathbb{R}^2 \to \mathbb{S}^2$$

cannot be equiareal and conformal at the same time, otherwise it would be an isometry. Therefore, world maps will always distort areas or angles, and there is no world map which preserves both.

The converse of the Theorema Egregium is false: there are surfaces which are not isometric but have the same Gaussian curvature, see the next Example.

Example 4.247

Question. Define the open set $U = (0, 2\pi) \times (0, \infty)$ and let $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ the surfaces defined by $\mathcal{S} = \boldsymbol{\sigma}(U)$, $\widetilde{\mathcal{S}} = \tilde{\boldsymbol{\sigma}}(U)$ with

$$\begin{aligned}
\boldsymbol{\sigma}(u,v) &= (\cos(u)v, \sin(u)v, \log(v))\,, \\
\tilde{\boldsymbol{\sigma}}(u,v) &= (\cos(u)v, \sin(u)v, u)\,.
\end{aligned}$$

Note that $\mathcal{S}$ is the surface of revolution obtained by rotating the curve $(v, 0, \log(v))$, while $\widetilde{\mathcal{S}}$ is a portion of Helicoid.

1. Prove that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ are not locally isometric.
2. Prove that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ have the same Gaussian curvature.

Solution.

1. Compute the first fundamental form of $\boldsymbol{\sigma}$

$$\begin{aligned}
\boldsymbol{\sigma}_u &= (-\sin(u)v, \cos(u)v, 0) \\
\boldsymbol{\sigma}_v &= (\cos(u), \sin(u), 1/v) \\
E &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_u = v^2 \\
F &= \boldsymbol{\sigma}_u \cdot \boldsymbol{\sigma}_v = 0 \\
G &= \boldsymbol{\sigma}_v \cdot \boldsymbol{\sigma}_v = 1 + 1/v^2 \\
\mathcal{F}_1 &= \begin{pmatrix} v^2 & 0 \\ 0 & 1 + 1/v^2 \end{pmatrix}.
\end{aligned}$$

The first fundamental form of $\tilde{\boldsymbol{\sigma}}$ is

$$\begin{aligned}
\tilde{\boldsymbol{\sigma}}_u &= (-\sin(u)v, \cos(u)v, 1) \\
\tilde{\boldsymbol{\sigma}}_v &= (\cos(u), \sin(u), 0) \\
\widetilde{E} &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_u = 1 + v^2 \\
\widetilde{F} &= \tilde{\boldsymbol{\sigma}}_u \cdot \tilde{\boldsymbol{\sigma}}_v = 0 \\
\widetilde{G} &= \tilde{\boldsymbol{\sigma}}_v \cdot \tilde{\boldsymbol{\sigma}}_v = 1 \\
\widetilde{\mathcal{F}}_1 &= \begin{pmatrix} 1 + v^2 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned}$$

Suppose by contradiction there was a local isometry $f : \mathcal{S} \to \widetilde{\mathcal{S}}$. Therefore, the charts $\boldsymbol{\sigma}$ and $f \circ \boldsymbol{\sigma}$ would have the same first fundamental form. Since $f \circ \boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are both charts for $\widetilde{\mathcal{S}}$, there exists a diffeomorphism Φ such that

$$f \circ \boldsymbol{\sigma} = \Phi \circ \tilde{\boldsymbol{\sigma}}\,.$$

In particular, the first fundamental form of $f \circ \boldsymbol{\sigma}$ is the same as the one of $\Phi \circ \tilde{\boldsymbol{\sigma}}$. In conclusion, $\boldsymbol{\sigma}$ and $\Phi \circ \tilde{\boldsymbol{\sigma}}$ have the same first fundamental form. By Proposition 4.129, the first fundamental form of $\Phi \circ \tilde{\boldsymbol{\sigma}}$ is $(J\Phi)^T \widetilde{\mathcal{F}}_1 J\Phi$. Therefore, we have

$$\mathcal{F}_1 = (J\Phi)^T \widetilde{\mathcal{F}}_1 J\Phi \,. \tag{4.41}$$

Taking the determinant of both sides, we get

$$\det(\mathcal{F}_1) = \det(J\Phi)^2 \det(\widetilde{\mathcal{F}}_1) \,.$$

We compute that

$$\det(\mathcal{F}_1) = \det(\widetilde{\mathcal{F}}_1) = 1 + v^2 \,,$$

and thus

$$\det J\Phi = \pm 1 \,. \tag{4.42}$$

On the other hand,

$$\begin{aligned} (J\Phi)^T \widetilde{\mathcal{F}}_1 J\Phi &= \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} 1+v^2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \begin{pmatrix} a^2(1+v^2)+c^2 & * \\ * & * \end{pmatrix} . \end{aligned}$$

By (4.41), we get

$$\begin{pmatrix} a^2(1+v^2)+c^2 & * \\ * & * \end{pmatrix} = \begin{pmatrix} v^2 & 0 \\ 0 & 1+1/v^2 \end{pmatrix} .$$

Equating the first entries, we obtain

$$a^2(1+v^2)+c^2 = v^2 \,, \quad \forall\, v > 0 \,.$$

Taking the limit for $v \to 0^+$ gives

$$\lim_{v \to 0^+} \left[a^2(u,v) + c^2(u,v) \right] = 0 \,,$$

which implies

$$\lim_{v \to 0^+} a(u,v) = 0 \,, \qquad \lim_{v \to 0^+} c(u,v) = 0 \,.$$

Therefore, we have that

$$\lim_{v \to 0^+} J\Phi(u,v) = \lim_{v \to 0^+} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} .$$

By continuity of the determinant, we infer

$$\lim_{v \to 0^+} \det J\Phi(u,v) = \det \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} = 0 \,.$$

This contradicts (4.42). Hence, $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ cannot be locally isometric.

2. Note that we could compute the Gaussian curvature from the first fundamental form, by first computing the Christoffel symbols, and then using the first Gauss Equation. However, we proceed as usual, and compute K as the determinant of the Weingarten map. To this end, compute the second fundamental forms of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$

$$
\begin{aligned}
\boldsymbol{\sigma}_{uu} &= (-\cos(u)v, -\sin(u)v, 0) \\
\boldsymbol{\sigma}_{uv} &= (-\sin(u), \cos(u), 0) \\
\boldsymbol{\sigma}_{vv} &= (0, 0, -1/v^2) \\
\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v &= (\cos(u), \sin(u), -v) \\
\|\boldsymbol{\sigma}_u \times \boldsymbol{\sigma}_v\| &= (1+v^2)^{1/2} \\
\mathbf{N} &= (1+v^2)^{-1/2}(\cos(u), \sin(u), -v) \\
L = \boldsymbol{\sigma}_{uu} \cdot \mathbf{N} &= -v(1+v^2)^{-1/2} \\
M = \boldsymbol{\sigma}_{uv} \cdot \mathbf{N} &= 0 \\
N = \boldsymbol{\sigma}_{vv} \cdot \mathbf{N} &= (1+v^2)^{-1/2}/v \\
\tilde{\boldsymbol{\sigma}}_{uu} &= (-\cos(u)v, -\sin(u)v, 0) \\
\tilde{\boldsymbol{\sigma}}_{uv} &= (-\sin(u), \cos(u), 0) \\
\tilde{\boldsymbol{\sigma}}_{vv} &= (0, 0, 0) \\
\tilde{\boldsymbol{\sigma}}_u \times \tilde{\boldsymbol{\sigma}}_v &= (-\sin(u), \cos(u), -v) \\
\|\tilde{\boldsymbol{\sigma}}_u \times \tilde{\boldsymbol{\sigma}}_v\| &= (1+v^2)^{1/2} \\
\widetilde{\mathbf{N}} &= (1+v^2)^{-1/2}(-\sin(u), \cos(u), -v) \\
\tilde{L} = \tilde{\boldsymbol{\sigma}}_{uu} \cdot \widetilde{\mathbf{N}} &= 0 \\
\widetilde{M} = \tilde{\boldsymbol{\sigma}}_{uv} \cdot \widetilde{\mathbf{N}} &= (1+v^2)^{-1/2} \\
\widetilde{N} = \tilde{\boldsymbol{\sigma}}_{vv} \cdot \widetilde{\mathbf{N}} &= 0
\end{aligned}
$$

Therefore, the Gaussian curvature of $\boldsymbol{\sigma}$ is

$$
K = \frac{LN - M^2}{EG - F^2} = -\frac{1}{(1+v^2)^2} ,
$$

while the one of $\tilde{\boldsymbol{\sigma}}$ is

$$
\widetilde{K} = \frac{\tilde{L}\widetilde{N} - \widetilde{M}^2}{\widetilde{E}\widetilde{G} - \widetilde{F}^2} = -\frac{1}{(1+v^2)^2} ,
$$

showing that $\mathcal{S}$ and $\widetilde{\mathcal{S}}$ have the same Gaussian curvature.

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] Chern, Shiing-shen. "An elementary proof of the existence of isothermal parameters on a surface". In: *Proceedings of the American Mathematical Society* 6 (1955), pp. 771–782. DOI: 10.1090/S0002-9939-1955-0074856-1.

[3] M. P. do Carmo. *Differential Geometry of Curves and Surfaces*. Second Edition. Dover Books on Mathematics, 2017.

[4] R. Johansson. *Numerical Python. Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib*. Second Edition. Apress, 2019.

[5] Kong, Qingkai, Siauw, Timmy, and Bayen, Alexandre. *Python Programming and Numerical Methods*. Academic Press, 2020.

[6] M. Manetti. *Topology*. Second Edition. Springer, 2023.

[7] A. Pressley. *Elementary Differential Geometry*. Second Edition. Springer, 2010.

[8] W. Walter. *Ordinary differential equations*. Springer-Verlag, 1998.

[9] V. A. Zorich. *Mathematical Analysis I*. Second Edition. Springer, 2015.

[10] V. A. Zorich. *Mathematical Analysis II*. Second Edition. Springer, 2016.

A. Plots with Python

A.1. Curves in Python

A.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
```

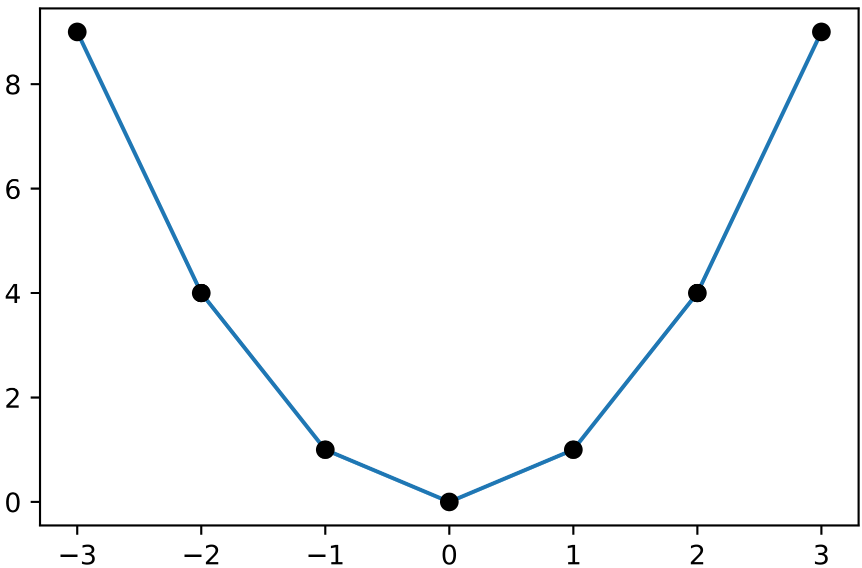
The above imports **numpy** and the module **pyplot** from **matplotlib**, and renames them to np and `plt`, respectively. These shorthands are standard in the literature, and they make code much more readable.
The function for plotting 2D graphs is called `plot(x,y)` and is contained in `plt`. As the 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

$$t = [-3, -2, -1, 0, 1, 2, 3] . \tag{A.1}$$

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 `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 A.1, as shown below

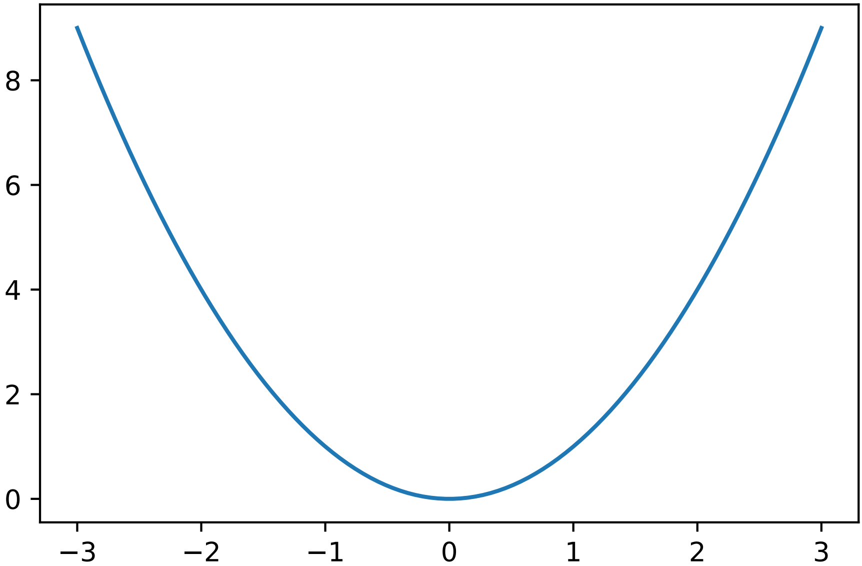
```
# Displaying output of np.linspace

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()
```

[1]The command `plt.show()` can be omitted if working in Jupyter Notebook, as it is called by default.



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

$$\boldsymbol{\gamma}(t) = (\sqrt{t}\cos(t), \sqrt{t}\sin(t)) \quad \text{for} \quad t \in [0, 50]\,. \tag{A.2}$$

```
# Plotting Fermat's spiral

import numpy as np
import matplotlib.pyplot as plt

# Divides time interval (0,50) in 500 parts
t = np.linspace(0, 50, 500)

# Computes Fermat's Spiral
x = np.sqrt(t) * np.cos(t)
y = np.sqrt(t) * np.sin(t)

# Plots the Spiral
plt.plot(x,y)
plt.show()
```

Before displaying the output of the above code, a few comments are in order. The array `t` has size 500, due to the behavior of `linspace`. You can also fact check this information by printing `np.size(t)`, which is the numpy function that returns the size of an array. We then use the numpy function `np.sqrt` to compute the square root of the array `t`. The outcome is still an array with the same size of `t`, that is,

$$t = [t_1, \dots, t_n] \quad \Longrightarrow \quad \sqrt{t} = [\sqrt{t_1}, \dots, \sqrt{t_n}]\,.$$

Similary, the call `np.cos(t)` returns the array

$$\cos(t) = [\cos(t_1), \dots, \cos(t_n)]\,.$$

The two arrays `np.sqrt(t)` and `np.cos(t)` are then multiplied, term-by-term, and saved in the array `x`. The array `y` is computed similarly. The command `plt.plot(x,y)` then yields the graph of the Fermat's spiral:

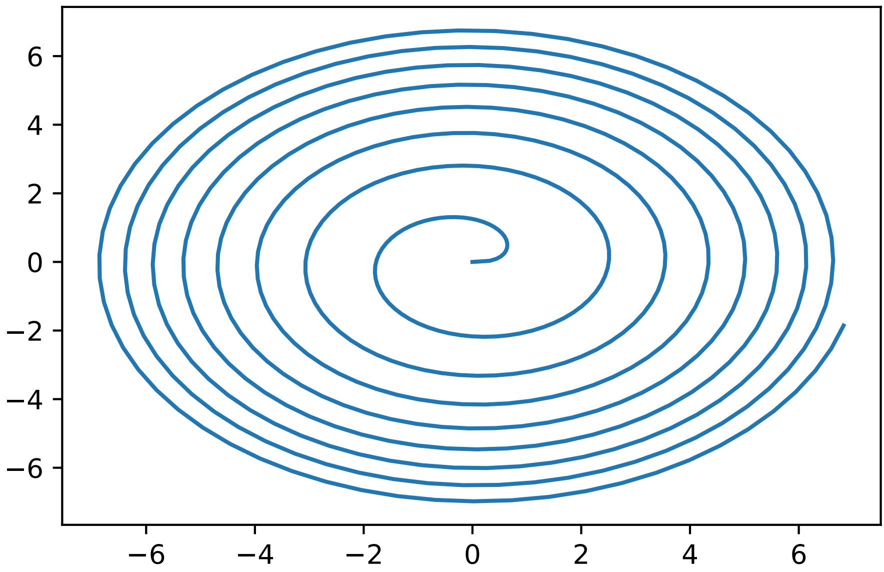


Figure A.1.: Fermat's spiral

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()
```

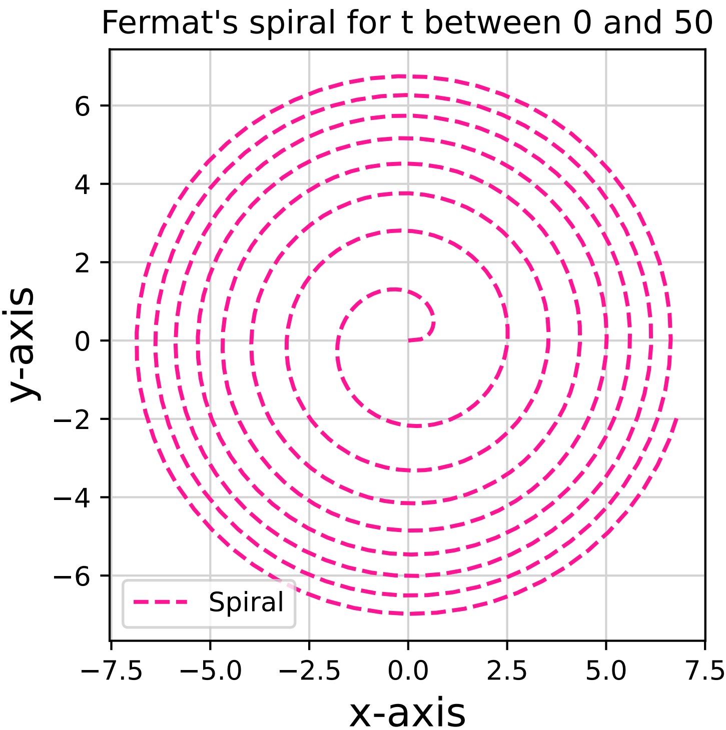


Figure A.2.: Adding a bit of style

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 aesthetic 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

```python
import numpy as np

t = np.arange(0,1, 0.2)
print("t =",t)
```

```
t = [0.  0.2 0.4 0.6 0.8]
```

A.1.2. Implicit curves 2D

A curve γ 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 γ 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

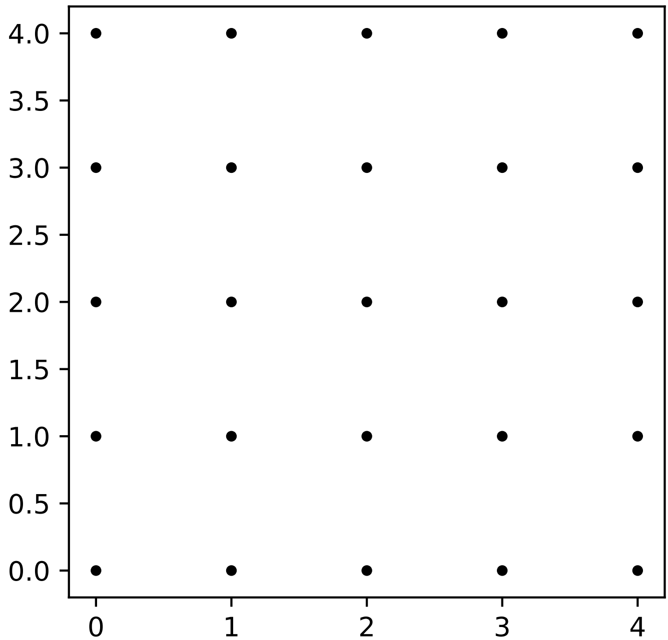


Figure A.3.: The 5 x 5 grid corresponding to the matrix A

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

$$x = [0, 1, 2, 3, 4]\,, \qquad y = [0, 1, 2, 3, 4]\,.$$

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 A.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()
```

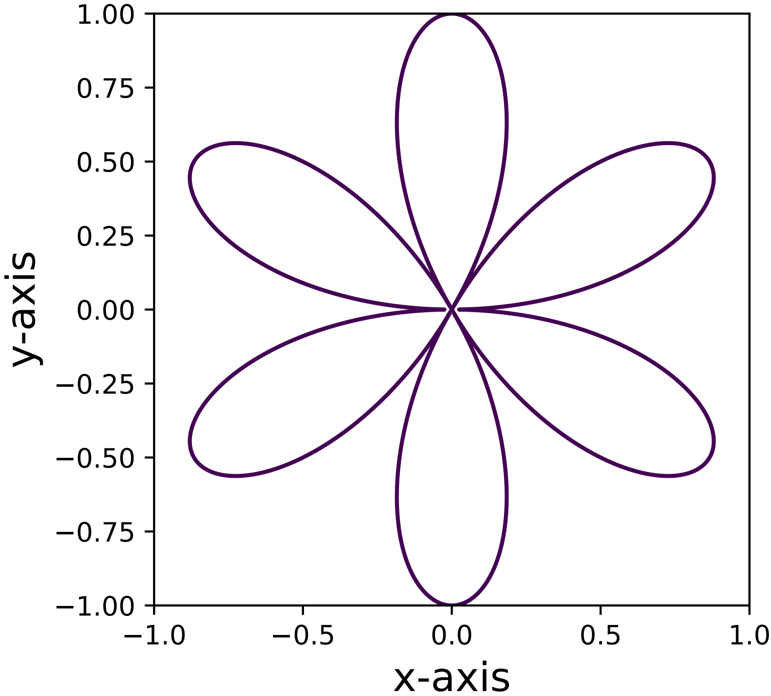


Figure A.4.: Plot of the curve defined by f=0

A.1.3. Curves in 3D

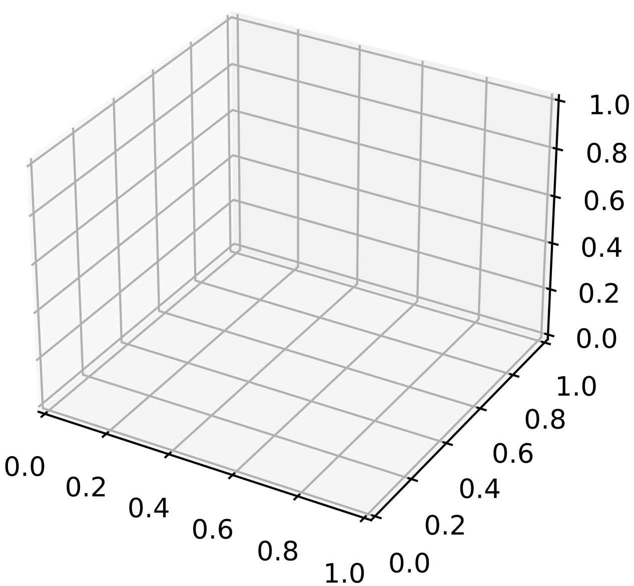
Plotting in 3D with matplotlib requires the `mplot3d` toolkit, see here for documentation. Therefore our first lines will always be

```
# 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()
```

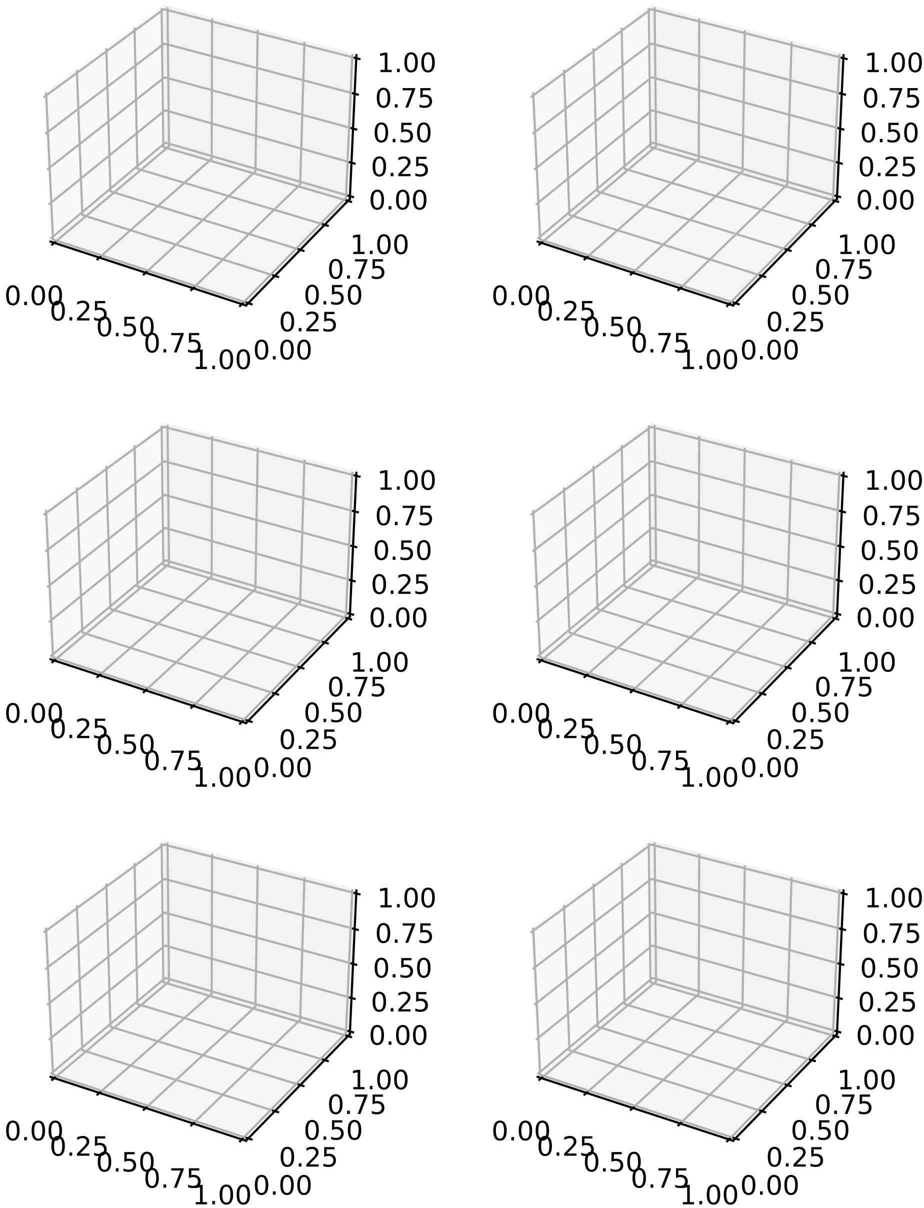


In the above code `fig` is a figure object, while `ax` is an axes object. In practice, the figure object contains the axes objects, and the actual plot information will be contained in axes. If you want multiple plots in the figure container, you should use the command

```
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

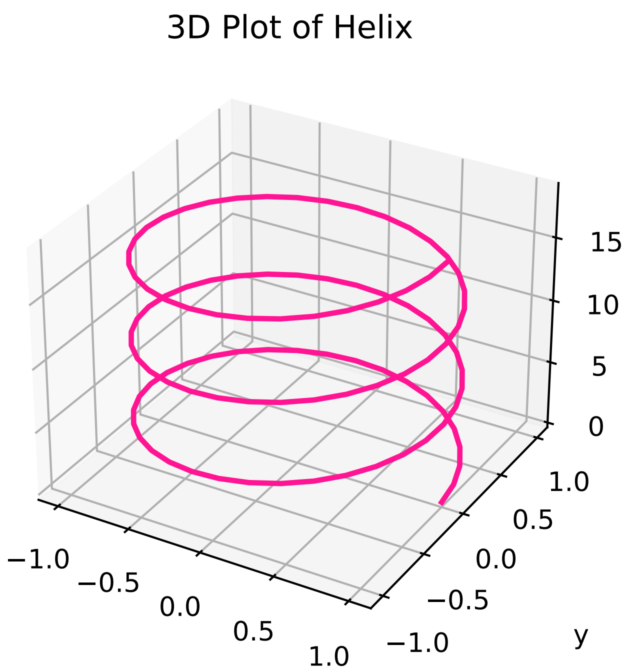
```python
# Code to plot 3D curve

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

$$x(t) = \cos(t)\,, \quad y(t) = \sin(t)\,, \quad z(t) = t \tag{A.3}$$

for $t \in [0, 6\pi]$.

```python
# 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()
```



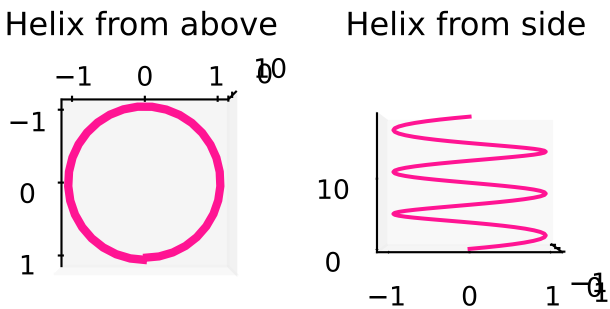
We can also change the viewing angle for a 3D plot store in ax. This is done via

```
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()
```



A.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.

A.1.4.1. 2D Plots

Say we want to plot the 2D curve $\boldsymbol{\gamma} \colon (a, b) \to \mathbb{R}^2$ 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 A.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`“.

A.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 A.3. We also add some styling. We can also use the same pre-defined templates described for `go.Scatter` in the previous section, see here for official documentation.

```
# Plotting 3D Helix

# Import libraries
import numpy as np
import plotly.graph_objects as go

```

```
# Divides time interval (0,6pi) in 100 parts
t = np.linspace(0, 6*np.pi, 100)

# Computes Helix
x = np.cos(t)
y = np.sin(t)
z = t

# Create empty figure object and saves
# it in the variable "fig"
fig = go.Figure()

# Create the line plot object
# We add options for the line width and color
data = go.Scatter3d(
    x = x, y = y, z = z,
    mode = 'lines', name = 'gamma',
    line = dict(width = 10, color = "darkblue")
    )

# Add "data" plot to the figure "fig"
fig.add_trace(data)

# Here we start with the styling options
# First we set a figure title
fig.update_layout(title_text = "Plotting 3D Helix with Plotly")

# Adjust figure size
fig.update_layout(
    autosize = False,
    width = 600,
    height = 600
    )

# Set pre-defined template
fig.update_layout(template = "seaborn")

# Options for curve line style



# Display the figure
fig.show()
```

```
Unable to display output for mime type(s): text/html
```

The above code generates an image that cannot be rendered in pdf. To see the output, please click here for the digital version of these notes. Once again, the style customizations could be listed in a single call of the function `update_layout`.

A.2. Surfaces in Python

A.2.1. Plots with Matplotlib

I will take for granted all the commands explained in Section A.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()
```

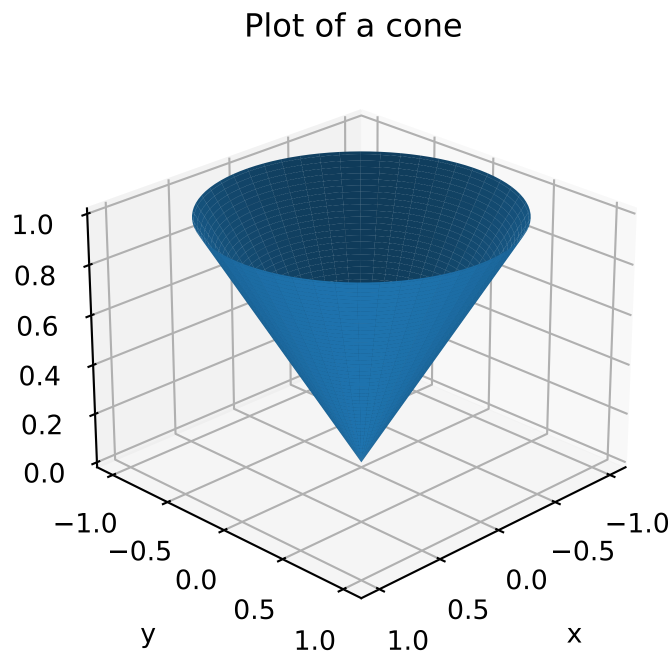
For example let us plot a cone described parametrically by:

$$x = u\cos(v)\,, \quad y = u\sin(v)\,, \quad z = u$$

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 A.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:

$$
\begin{aligned}
x &= (R + r\cos(u))\cos(v) \\
y &= (R + r\cos(u))\sin(v) \\
z &= r\sin(u)
\end{aligned}
$$

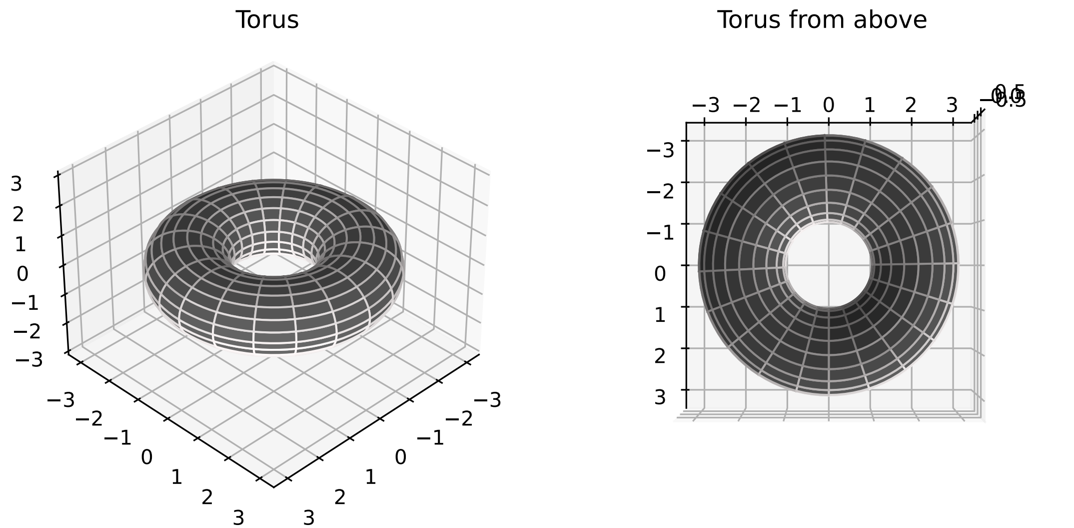
for $u, v \in (0, 2\pi)$ and with

- 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

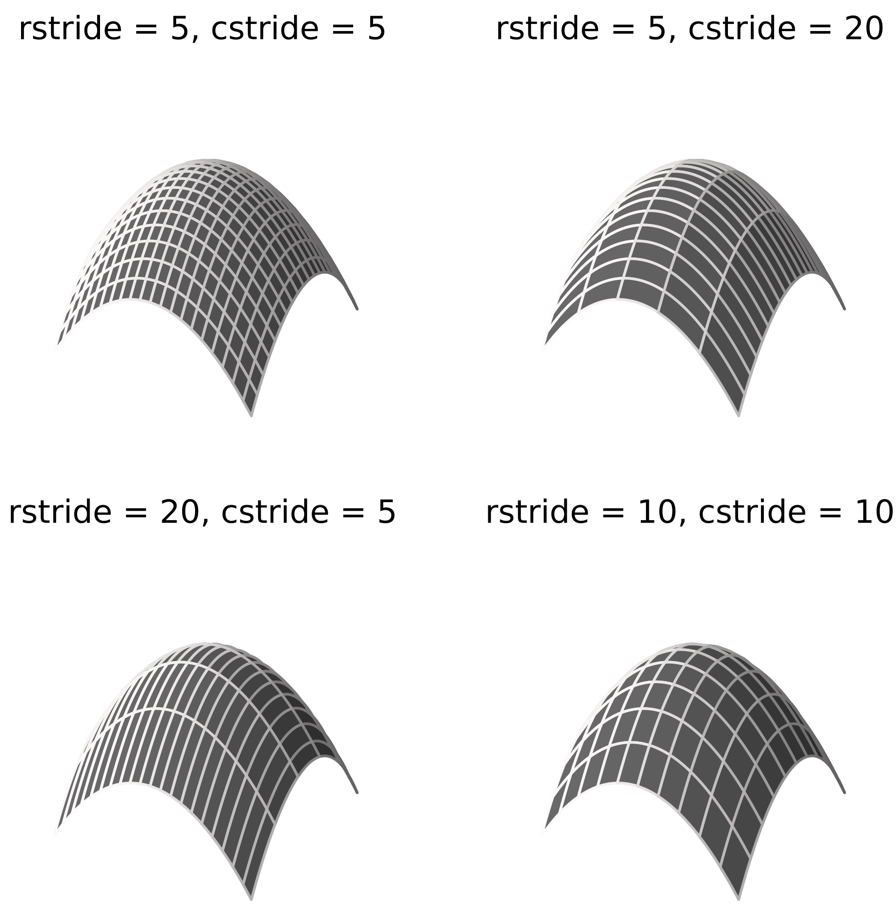
$$x = u$$
$$y = v$$
$$z = -u^2 - v^2$$

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.

A.2.2. Plots with Plotly

As done in Section A.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.