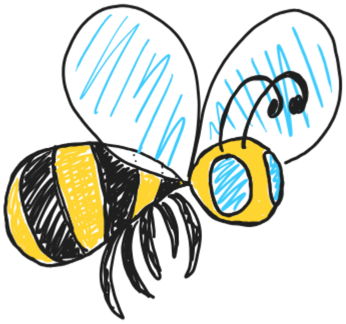
DALHOUSIE UNIVERSITY
DEPARTMENT OF MATHEMATICS & STATISTICS
FACULTY OF SCIENCES

Multivariable Calculus I

NOTES OF THE COURSE

Marcello Lanfranchi * 21st May 2026



Dear reader,

These are the notes I write for you. I don't really like the textbook presentation of the topics, so here we go: this is how I would like to present the course. You will probably immediately realize that in these notes there's a lot going on: vector spaces, a lot of stuff blablabla!

One thing I always wanted as a student was to have an organic story that related different parts of mathematics and things altogether. This is why I decided here to present the topics, adding some reference to more advanced concepts. To me this approach is very important: often, students of calculus have the perception that mathematics is just about solving numerical problems. This is not the case and I want to stress this fact! Math is about understanding abstract objects, describing them using definitions and theorems, applying the theories to real-world problems, or creating more mathematics. As a Ph.D. student in mathematics, one of the key things for me is to understand what's going on in a large context. For example, introducing a system of coordinates cannot be done, in my opinion, without talking about vector spaces, and talking about the Hessian criterion cannot be done without saying what a matrix is.

The effect of not doing this is that the student will have access to the final formulas, without really understanding where they come from. This gives the wrong perception that mathematics is something obscure and mysterious, that only geniuses can understand and develop. But this is not true! We just need to be honest and present the material in the context required to understand where things come from!

The downside of this approach is that in the end, the student has to study so many things!!! To solve this issue, I decided to clearly separate things that are important for the exam and the ones that are not required to pass the course. So, I indicated with a *, everything that is not mandatory for the exam, things that you should read at least one time, but that will not be tested in the course.

One final note: my first language is not English, but Italian. I did my best to write correct and understandable sentences, but I'm sure that some of them aren't. If you find anything that is not clear or if you find mistakes, you are very welcome to tell me about them, so that I can update the notes and make them better for future students.

TABLE OF CONTENTS

MODULE 1

Vectors & the geometry of space

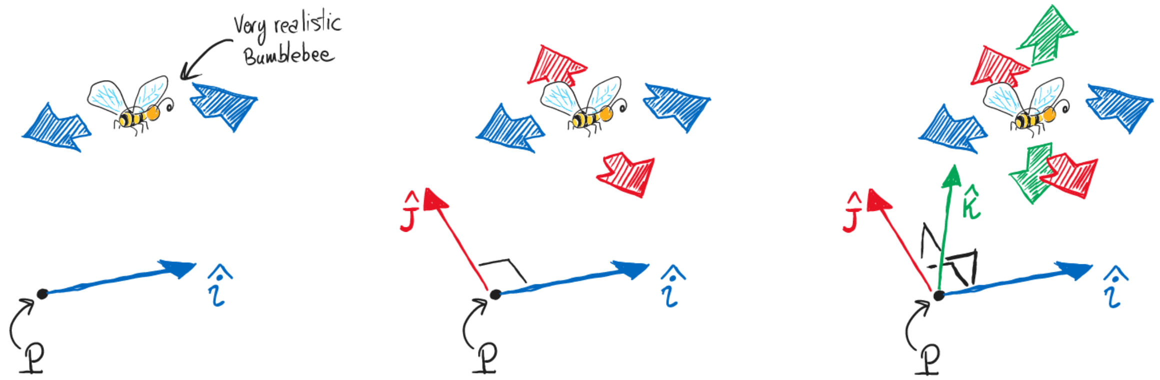
1.1 Introduction

Imagine we want to study the motion of a bumblebee in the air. For this goal, we need to find a way to keep track of the position of the bumblebee at every instant of time t. A way to do so is to employ the concept of a system of coordinates, also known as a system of reference. To build a system of reference, the first thing to do is to pick a point P in the space. There is no better choice here, so let's pick a point P, that we are going to call the origin of the system of coordinates.

With this choice, we can draw an arrow from P to the position of the bumblebee. Note that, since the bumblebee is not just a point, but an actual three-dimensional body, we also need to identify the bumblebee with a point. This choice can be made by using the notion of barycentre, which represents the centre of mass of the bumblebee. Any other point of the bumblebee works too, as long as we remember to keep that point as a reference for the position of the bumblebee. So, the problem now is to describe an arrow that starts from the point P and ends at the point we chose to represent the bumblebee.

An important observation, is that the bumblebee can move in three independent ways: up-down, left-right, and back-forth. These are called the degrees of freedom of the system. So, from this consideration, we expect to be able to uniquely describe the position of the bumblebee by using only 3 numbers: one to indicate how up or down the

Figure 1.1: A bumblebee in space and its degrees of freedom



bumblebee is, one for how much is on the left or the right, and one for how much is back or forth. Since in the following, we are going to employ the word *arrow* many times, it is good practice to make this concept precise with a definition.

Definition 1.1. *An **arrow** from a point P is a segment lying on a line called the **direction** of the arrow, equipped with an **orientation**, which is indicated by a little triangle on the tip. We call the **zero arrow**, denoted by* $\overrightarrow{0}$*, the arrow of length* 0*. We also call the direction of a non-zero arrow, the **line generated** by the arrow. We denote arrows by* $\overrightarrow{v}$*.*

*We call **modulus**, or **magnitude** or **length**, of* $\overrightarrow{v}$ *the length of* $\overrightarrow{v}$ *and we denote this by* $|\overrightarrow{v}|$*. Given two non-zero arrows* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *(not necessarily starting from the same point), we say that* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *are **parallel** if they generate parallel lines. We use the symbol* $\overrightarrow{u} \parallel \overrightarrow{v}$ *to indicate that* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *are parallel and* $\overrightarrow{u} \nparallel \overrightarrow{v}$ *if they are not.*

When $\overrightarrow{u} \nparallel \overrightarrow{v}$*, the **plane generated** by* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *is the unique plane where both* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *lie and the **angle between*** $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *is the smallest angle they form. When the angle between* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *is* $90°$*, we say that* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *are **orthogonal** and we use the symbol* $\overrightarrow{u} \perp \overrightarrow{v}$*. Finally, we denote by* $\mathbb{V}_3$ *the set of arrows in the three-dimensional space.*

The goal is to describe each arrow $\overrightarrow{v} \in \mathbb{V}_3$ which starts from P, with the least amount of data. Since the bumblebee has three degrees of freedom, we expect to identify each arrow $\overrightarrow{v} \in \mathbb{V}_3$ with a triple of numbers. To do so, we would like to introduce an operation between arrows which allows one to generate new arrows from simpler ones.

This idea of finding operations for mathematical objects is transversal in mathematics and allows one to define complex objects from simpler ones. Let's think for example of natural numbers. Every natural number is a product of prime numbers. For example, the number 46574 can be decomposed into the product of $2, 11, 29$, and 73. So, for natural numbers, the "simple" objects are prime numbers and the operation is multiplication.

As a general approach, it is always useful when a new concept is introduced to see what kinds of operations are well-defined for this class of objects. With this spirit, let's introduce an interesting operation.

Consider two arrows $\overrightarrow{u}$ and $\overrightarrow{v}$, both starting from the origin P. Imagine first that $\overrightarrow{u}$ and $\overrightarrow{v}$ are parallel. If we want to define a *sum of arrows*, a natural choice, in this case, is to imagine moving rigidly $\overrightarrow{v}$ in front of $\overrightarrow{u}$. Moving an arrow rigidly means translating the arrow to a parallel line without changing its modulus or orientation. By moving $\overrightarrow{v}$ at the end of $\overrightarrow{u}$ we obtain a new arrow that starts again in P and ends where $\overrightarrow{v}$ ends after it is moved. Because in this example, $\overrightarrow{u} \parallel \overrightarrow{v}$, the modulus of the arrow we obtained is precisely the sum of the moduli of $\overrightarrow{u}$ and $\overrightarrow{v}$.

Let's now consider the general case. Suppose that $\overrightarrow{u}$ and $\overrightarrow{v}$ are two arrows which start from the same point P. Let's define their sum as the arrow $\overrightarrow{u} + \overrightarrow{v}$ obtained by moving rigidly $\overrightarrow{v}$ in front of $\overrightarrow{u}$.

Definition 1.2. *Given two arrows* $\overrightarrow{u}$ *and* $\overrightarrow{v}$ *that both start from P, the **sum of arrows*** $\overrightarrow{u} + \overrightarrow{v}$ *is the arrow that starts from P and ends at the end point of* $\overrightarrow{v}$*, after* $\overrightarrow{v}$ *is rigidly moved in front of* $\overrightarrow{u}$*.*

A useful practice when we define a new operation is to study its properties. In particular, we look for properties of the sum of arrows which are similar to the properties of the usual algebraic sum of numbers: unitality, associativity, the existence of inverses, and commutativity. In mathematics, a set of objects with a binary operation satisfying these properties is called a commutative group. Let's introduce this concept.

Definition 1.3. *A **group** is a set G equipped with a binary operation* $\mu\colon G \times G \to G$ *such that the following properties are satisfied:*

1. *The binary operation* μ *is **unital**. This means there is a distict element* $1_G \in G$*, called the **unit**, such that for every* $g \in G$*:*

$$\mu(g, 1_G) = g = \mu(1_G, g)$$

2. *Every element* $g \in G$ *admits an inverse. This means for every* $g \in G$ *there is another element* $g^{-1} \in G$ *such that:*

$$\mu(g, g^{-1}) = 1_G = \mu(g^{-1}, g)$$

3. *The binary operation* μ *is associative, that is for every* $g, h, i \in G$*:*

$$\mu(\mu(g, h), i) = \mu(g, \mu(h, i))$$

*A group G is **commutative**, or Abelian, if for every* $g, h \in G$*,* $\mu(g, h) = \mu(h, g)$*.*

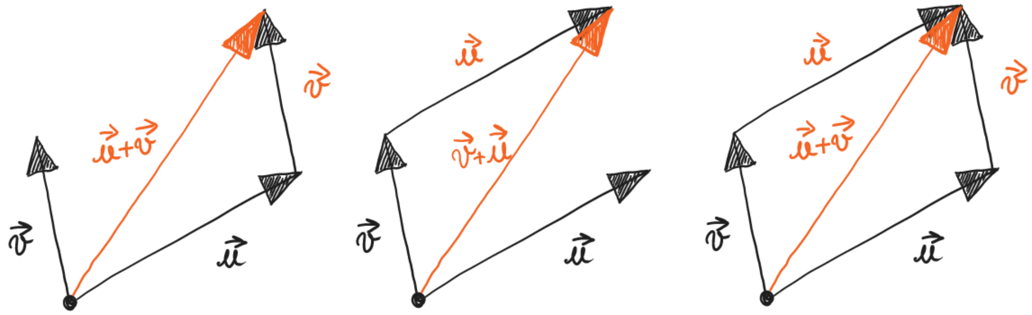


Figure 1.2: The sum of arrows

Example 1.1. An example of a group is the set of integers $\mathbb{Z}$. The binary operation μ of this group is the sum, $+$, so we write $g + h$ for $\mu(g, h)$, for every $g, h \in \mathbb{Z}$. The number zero is the unit of $\mathbb{Z}$, since $0 + g = g = g + 0$ for every g. Notice also that for each $g \in \mathbb{Z}$, $-g$ is the inverse of g, since $g + (-g) = g - g = 0 = (-g) + g$. Finally, $+$ is associative since $(g + h) + i = g + (h + i)$. $\mathbb{Z}$ is also commutative, since $g + h = h + g$. ■

Example 1.2. Let's consider the set of integers $\mathbb{Z}$ and let μ be the multiplication of two numbers. So, $\mu(g, h) := g \cdot h$. Note that 1 is the unit! This is because $1 \cdot g = g = g \cdot 1$. However, $(\mathbb{Z}, \cdot)$ is not a group. The problem is that for some numbers of $\mathbb{Z}$ there is no inverse of the multiplication. For example, there is no integer number h for which $h \cdot 2 = 1$. To define a group in which the multiplication is the binary operation, we need to consider the set of rational numbers $\mathbb{Q}$ and remove the zero. $\mathbb{Q}^*$ denotes this set, i.e. $\mathbb{Q}^* = \mathbb{Q} \setminus \{0\}$. Then, $\mathbb{Q}^*$ with the multiplication is a commutative group. ■

Example 1.3. Let's consider the set $\mathbb{Q}^*$ which contains all non-zero rational numbers. In Example 1.2 we noticed that $\mathbb{Q}^*$ with the multiplication is a commutative group. What happens if we take μ to be the division between two rational numbers? So, $\mu(a, b) := a/b$, for $a, b \in \mathbb{Q}^*$. Is this operation associative? ■

We want to see if the sum of arrows satisfies the properties of a commutative group. Let's start with unitality. This means that there should be a special arrow which, when summed by any other arrow $\vec{u}$, gives back the same arrow $\vec{u}$. Can you guess which arrow is it?

Remember that $\vec{0}$ is the zero arrow, i.e. the arrow with zero modulus. Moving rigidly $\vec{0}$ does not change its length, so $\vec{u} + \vec{0} = \vec{u} = \vec{0} + \vec{u}$. So, the sum of arrows is indeed unital. Let's show the existence of inverses. Take an arrow $\vec{u}$ and let's define the inverse arrow to be the unique arrow with the same length and direction of $\vec{u}$, but opposite orientation, denoted by $-\vec{u}$. Now, imagine to rigidly move $-\vec{u}$ in front of $\vec{u}$. We obtain an arrow that starts from P and ends in P. This arrow has zero length, thus $\vec{u} + (-\vec{u}) = \vec{0}$. Similarly, by rigidly moving $\vec{u}$ in front of $-\vec{u}$ we obtain the same result, so we have $(-\vec{u}) + \vec{u} = \vec{0}$. So, every arrow has an inverse.

The next step is to show the commutativity. Consider Picture 1.2. From the picture, we can see that, by moving first $\vec{v}$ in front of $\vec{u}$ or first moving $\vec{u}$ in front of $\vec{v}$ we obtain the same result. Thus, we have that $\vec{u} + \vec{v} = \vec{v} + \vec{u}$. Finally, let's show the associativity. Consider Picture 1.3. From this picture, we see that moving first $\vec{v}$ in front of $\vec{u}$ and then moving $\vec{u} + \vec{v}$ in front of $\vec{w}$, is the same as first moving $\vec{w}$ in front of $\vec{v}$ and then moving $\vec{v} + \vec{w}$ in front of $\vec{u}$, i.e. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$. Let's briefly collect these results in a proposition.

Proposition 1.1. *The set $\mathbb{V}_3$ of arrows in the three-dimensional space equipped with the sum of arrows forms a commutative group, i.e. the binary operation $+$ is unital, associative, commutative and admits inverses.*

1.2 Vector spaces

So far, we introduced the operation sum of arrows; the next step is to ask if we can multiply arrows together. We will see that in the special case of arrows in the three-dimensional space, there is actually a binary operation which plays the role of a particular multiplication, called vector product, however, this is a very special case that works only in three

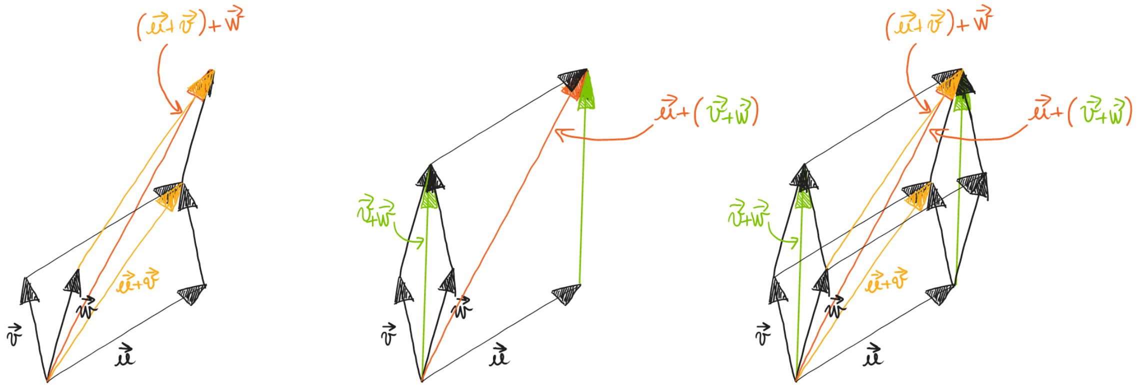


Figure 1.3: The associativity of the sum of arrows

---

dimensions (to be precise it works in dimensions $1, 3$ and 7 only), so for now, we will not introduce this operation. Instead, we focus our attention on a simpler and more general operation that looks like a multiplication, but instead of being an operation between arrows, it is an operation between a real number and an arrow: the scalar action.

Definition 1.4. *Given an arrow $\vec{u}$ and a real number $a \in \mathbb{R}$, the arrow $a \cdot \vec{u}$ is the arrow with the same direction of $\vec{u}$, modulus $|a \cdot u| = |a||\vec{u}|$ and orientation equal to the orientation of $\vec{u}$ if $a \geq 0$ and opposite orientation if $a < 0$. In the following, when is clear from the context, we drop the $\cdot$ notation and instead, we will simply write $a\vec{u}$ for $a \cdot \vec{u}$.*

As we did for the sum of arrows, we now introduce the abstract definition of scalar action.

Definition 1.5. *Given a set $\mathbb{V}$, a* ***scalar action*** *over $\mathbb{V}$ is a function:*

$$\cdot : \mathbb{R} \times \mathbb{V} \to \mathbb{V}$$

that satisfies the following properties:

1. *For every $\vec{u} \in \mathbb{V}$:*

   $$1 \cdot \vec{u} = \vec{u}$$

2. *For every $a, b \in \mathbb{R}$ and every $\vec{u} \in \mathbb{V}$:*

   $$(ab) \cdot \vec{u} = a \cdot (b \cdot \vec{u})$$

   When $(\mathbb{V}, +)$ is a commutative group, we say that the scalar action $\cdot$ is ***compatible*** *with the commutative group structure of $\mathbb{V}$ if for every $a, b \in \mathbb{R}$ and $\vec{u}, \vec{v} \in \mathbb{V}$:*

   $$\begin{aligned} a \cdot (\vec{u} + \vec{v}) &= a \cdot \vec{u} + a \cdot \vec{v} \\ (a + b) \cdot \vec{u} &= a \cdot \vec{u} + b \cdot \vec{u} \end{aligned}$$

The function $\cdot : \mathbb{R} \times \mathbb{V}_3 \to \mathbb{V}_3$ defined in Definition 1.4, is indeed a scalar action and is also compatible with the commutative structure of $\mathbb{V}_3$. This construction has a fancy name, and its very important in mathematics. We are going to work with this class of structures quite a lot, so let's introduce this concept formally.

Definition 1.6. *A (real)* ***vector space*** *is a set $\mathbb{V}$ equipped with a binary operation $+\colon \mathbb{V}\times\mathbb{V}\to\mathbb{V}$ and with a function $\cdot\colon \mathbb{R}\times\mathbb{V}\to\mathbb{V}$ such that $(\mathbb{V},+)$ is a commutative group, $\cdot$ is a scalar action and $\cdot$ is compatible with $+$. Concretely, saying that $(\mathbb{V},+)$ is a commutative group means that there exists a distinct element $\vec{0}\in\mathbb{V}$ such that for every $\vec{u},\vec{v},\vec{w}\in\mathbb{V}$:*

$$\begin{aligned}\vec{u}+\vec{0}&=\vec{u}\\ \vec{u}+\vec{v}&=\vec{v}+\vec{u}\\ (\vec{u}+\vec{v})+\vec{w}&=\vec{u}+(\vec{v}+\vec{w})\end{aligned}$$

and that every $\vec{u}\in\mathbb{V}$ has a element $-\vec{u}$, called the inverse of $\vec{u}$, such that:

$$\vec{u}+(-\vec{u})=\vec{0}$$

Moreover, saying that $\cdot$ is a scalar action compatible with the sum means that for every $a,b\in\mathbb{R}$ and every $\vec{u},\vec{v}\in\mathbb{V}$:

$$\begin{aligned}1\cdot\vec{u}&=\vec{u}\\ (ab)\cdot\vec{u}&=a\cdot(b\cdot\vec{u})\\ a\cdot(\vec{u}+\vec{v})&=a\cdot\vec{u}+a\cdot\vec{v}\\ (a+b)\cdot\vec{u}&=a\cdot\vec{u}+b\cdot\vec{u}\end{aligned}$$

In the following, we will simplify the notation by writing $a\vec{u}$ instead of $a\cdot\vec{u}$. Finally, elements of a vector space are called ***vectors****.*

So, $(\mathbb{V}_3,+,\cdot)$ is a vector space. In the following, we will use arrows and vectors interchangeably, but it should be clear that not every vector space contains arrows. Here, we want to give an example of a vector space where the relationship with arrows makes less sense.

Cool Example* 1.1. Consider the set $\mathbb{R}[x]$ of polynomials in one variable. Let's recall that polynomials can be summed together up, so there is a binary operation $+$ defined for polynomials. For example, if $p(x):=3x^2+2x-1$ and $q(x):=x^6-16x^2$,

$$p(x)+q(x)=(3x^2+2x-1)+(x^6-16x^2)=3x^2+2x-1+x^6-16x^2=x^6-13x^2+2x-1$$

In particular, $(\mathbb{R}[x],+)$ is a commutative group. Moreover, a polynomial can also be multiplied by a real number. So, let's define $a\cdot p(x)$ to be the product between the number a with the polynomial $p(x)$. For example, if $p(x):=3x^2+2x-1$ and $a:=4$, we have:

$$a\cdot p(x)=4(3x^2+2x-1)=12x^2+8x-4$$

$\cdot$ is a scalar action for $\mathbb{R}[x]$ and $\mathbb{R}[x]$ with $+$ and $\cdot$ is a vector space. ■

Another interesting example is the space $\mathbb{R}^n$, for any integer n.

Cool Example* 1.2. Take an integer n and let's define $\mathbb{R}^n$ to be the Cartesian product of $\mathbb{R}$ with itself n times. So, an element of $\mathbb{R}^n$ is a n-tuple $(a_1,\ldots,a_n)\in\mathbb{R}^n$, where $a_1,\ldots,a_n\in\mathbb{R}$. Let's define a sum as follows:

$$(a_1,\ldots,a_n)+(b_1,\ldots,b_n):=(a_1+b_1,\ldots,a_n+b_n)$$

Moreover, given any real number $a\in\mathbb{R}$, let's define the scalar action as follows:

$$a\cdot(a_1,\ldots,a_n):=(aa_1,\ldots,aa_n)$$

Thus, $\mathbb{R}^n$ is a vector space. The zero vector of $\mathbb{R}^n$ is the n-tuple $(0,\ldots,0)$. When $n=2$ we call $\mathbb{R}^2$ the Cartesian plane and when $n=3$, $\mathbb{R}^3$ is called the Euclidean space. ■

Let's prove some elementary (but important) facts about vectors.

Proposition 1.2. *Let $\mathbb{V}$ be a vector space and $\vec{u}\in\mathbb{V}$ and $a\in\mathbb{R}$, then the following equations hold:*

$$\begin{aligned}0\vec{u}&=\vec{0}\\ (-1)\vec{u}&=-\vec{u}\\ a\vec{0}&=\vec{0}\end{aligned}$$

Moreover, if $a\vec{u}=\vec{0}$ then $a=0$ or $\vec{u}=\vec{0}$ or both.

Proof. Let's start by showing that $0\vec{u} = \vec{0}$.We know that $\vec{0}$ is the unique element so that $\vec{0} + \vec{u} = \vec{u}$. So, if we prove that $0\vec{u} + \vec{u} = \vec{u}$, it follows that $0\vec{u} = \vec{0}$:

$$\begin{aligned} & 0\vec{u} + \vec{u} = \\ = \quad & 0\vec{u} + 1\vec{u} = \\ = \quad & (0+1)\vec{u} = \\ = \quad & 1\vec{u} = \\ = \quad & \vec{u} \end{aligned}$$

where we used that $1\vec{u} = \vec{u}$ and that $a\vec{u} + b\vec{u} = (a+b)\vec{u}$, for every $a, b \in \mathbb{R}$ and $\vec{u} \in \mathbb{V}$. Let's show the second equation, i.e. $(-1)\vec{u} = -\vec{u}$. To prove that, recall that, by definition, $-\vec{u}$ is the unique vector such that $\vec{u} + (-\vec{u}) = \vec{0}$, so if we prove that $\vec{u} + (-1)\vec{u} = \vec{0}$ we conclude the expected result:

$$\begin{aligned} & \vec{u} + (-1)\vec{u} = \\ = \quad & 1\vec{u} + (-1)\vec{u} = \\ = \quad & (1 + (-1))\vec{u} = \\ = \quad & 0\vec{u} = \\ = \quad & \vec{0} \end{aligned}$$

where we used that $1\vec{u} = \vec{u}$, that $a\vec{u} + b\vec{u} = (a+b)\vec{u}$ and that $0\vec{u} = \vec{0}$. Let's now prove that $a\vec{0} = \vec{0}$, for $a \in \mathbb{R}$. To do that, remember we proved that for any $\vec{u} \in \mathbb{V}$, $0\vec{u} = \vec{0}$, and this is true also when $\vec{u} = \vec{0}$, so we have $\vec{0} = 0\vec{0}$. Therefore:

$$\begin{aligned} & a\vec{0} = \\ = \quad & a(0\vec{0}) = \\ = \quad & (a0)\vec{0} = \\ = \quad & 0\vec{0} = \\ = \quad & \vec{0} \end{aligned}$$

where we used that $a(b\vec{u}) = (ab)\vec{u}$ for any $a, b \in \mathbb{R}$ and $\vec{u} \in \mathbb{V}$. To give proof, imagine that $a \neq 0$ and let's show that, in this case, $\vec{u}$ has to be the zero vector. Because $a \neq 0$, we have that $1/a$ is a well-defined real number. Therefore:

$$\frac{1}{a}(a\vec{u}) = \frac{1}{a}\vec{0} = \vec{0}$$

where we used that for any $b \in \mathbb{R}$, $b\vec{0} = \vec{0}$ and the hypothesis that $a\vec{u} = \vec{0}$. However, for every $a, b \in \mathbb{R}$, $a(b\vec{u}) = (ab)\vec{u}$, therefore:

$$\begin{aligned} & \frac{1}{a}(a\vec{u}) = \\ = \quad & \left(\frac{1}{a}a\right)\vec{u} = \\ = \quad & 1\vec{u} = \\ = \quad & \vec{u} \end{aligned}$$

where we used that $1\vec{u} = \vec{u}$. Therefore, $\vec{u} = \vec{0}$. This concludes the proof. □

This is all cool, however so far we just introduced operations for arrows and lots of concepts. We still need to understand how to describe the position of the bumblebee in the space with the minimum number of parameters possible. What we need is a basis!

Definition 1.7. *A **basis** of a vector space $\mathbb{V}$ is family $B := \{\vec{v}_i\}$ of vectors of $\mathbb{V}$, where i is an index, with the following property: if $\vec{u}$ is any vector of $\mathbb{V}$ there is a non-negative integer n^* and n^* real numbers $u_1, \ldots, u_{n^*}$ such that:*

$$\vec{u} = u_1\vec{v}_1 + u_2\vec{v}_2 + \cdots + u_{n^*}\vec{v}_{n^*}$$

and $u_1, \dots, u_{n^*}$ are the only real numbers with this property. Concretely, this means that if $u'_1, \dots, u'_{n'}$ are such that:

$$\vec{u} = u'_1 \vec{v}_1 + u'_2 \vec{v}_2 + \cdots + u'_{n'} \vec{v}_{n'}$$

then $n' = n^*$ and $u'_1 = u_1, u'_2 = u_2, \dots, u'_{n^*} = u_{n^*}$.

The definition of a basis is quite abstract and, even if we gave this definition here, we don't need to remember all the details of it. The main point is that a basis allows one to representing every vector $\vec{u}$ as a tuple of numbers $(u_1, \dots, u_{n^*})$ so that:

$$\vec{u} = u'_1 \vec{v}_1 + u'_2 \vec{v}_2 + \cdots + u'_{n'} \vec{v}_{n'}$$

In general, a basis of a vector space can contain infinite vectors. For example, the vector space $\mathbb{R}[x]$ of polynomials in one variable has an infinite basis, which is the set $\{x^n, n \in \mathbb{N}_0\}$. Indeed, every polynomial can be written as:

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x^1 + a_0 x^0$$

When the number of vectors in a basis is finite, every other basis has the same number of vectors. This number is called dimension of the vector space.

Cool Definition* 1.1. *The **dimension** of a vector space $\mathbb{V}$ is the number of vectors in a basis of $\mathbb{V}$. In case a basis is infinite, we simply say that the vector space is infinite-dimensional.*

It's time to build a basis for the vector space $\mathbb{V}_3$ of arrows in the three-dimensional space. To do that, let's come back to the example of the bumblebee and consider again the point P that we called the origin. Let's choose an arrow with modulus > 0, so the arrow is non-zero, which starts from P. To simplify the future computations we decide to take an arrow with a modulus equal to 1. Let's give a name to this arrow: let's call it $\hat{\mathbf{i}}$.

The next step is to choose another arrow, that we call $\hat{\mathbf{j}}$. This time we have to be careful to choose $\hat{\mathbf{j}}$ so that it is not the zero arrow, so $\left|\hat{\mathbf{j}}\right| > 0$, and not parallel to $\hat{\mathbf{i}}$. Again, we decide to choose $\left|\hat{\mathbf{j}}\right| = 1$ and we also decide, again for the sake of simplicity, to choose $\hat{\mathbf{j}}$ so that $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are orthogonal. Note that we still have infinite choices of $\hat{\mathbf{j}}$: setting the orthogonality of $\hat{\mathbf{j}}$ and $\hat{\mathbf{i}}$ we are simply restricting our choice of $\hat{\mathbf{j}}$ on the plane that is orthogonal to $\hat{\mathbf{i}}$.

The final step is to choose a third arrow $\hat{\mathbf{k}}$. The important is that $\hat{\mathbf{k}}$ does not lie on the plane generated by $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$. Again we have infinite choices but now we decide to choose $\hat{\mathbf{k}}$ so that $\left|\hat{\mathbf{k}}\right| = 1$ and that $\hat{\mathbf{i}}$ and $\hat{\mathbf{k}}$ are orthogonal and also $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ are orthogonal.

With these restrictions, we discover that we only have two possible choices for $\hat{\mathbf{k}}$. In particular, $\hat{\mathbf{k}}$ lies on the line which is orthogonal to the plane generated by $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$, and it has a modulus equal to 1. Therefore $\hat{\mathbf{k}}$ can only point upward or downward w.r.t. the plane $\hat{\mathbf{i}}\hat{\mathbf{j}}$. To choose the orientation (the orientation) of $\hat{\mathbf{k}}$ we introduce a useful convention, known as the right-hand rule.

Definition 1.8. *Consider three non-zero arrows $\vec{u}, \vec{v}, \vec{w}$ and suppose also that $\vec{u}$ and $\vec{v}$ are not parallel and that $\vec{w}$ is orthogonal to both $\vec{u}$ and $\vec{v}$. The ordered triple $(\vec{u}, \vec{v}, \vec{w})$ satisfies the **right-hand rule** if the following procedure works:*

1. *Start by opening your right hand and by aligning the palm in the same direction and orientation of the arrow $\vec{u}$;*
2. *Now, curl the four fingers of your right hand and close your hand leaving the thumb pointing straight, so that at some point, your four fingers point in the same direction and orientation of the arrow $\vec{v}$;*
3. *While you are curling the other fingers of your right hand, your right thumb should point in the direction and orientation of the arrow $\vec{w}$.*

In Picture 1.4, you can see explained the right-hand rule. Note that, in the same picture, we also added an equivalent way to understand the r.h.r.: point your right index finger along the first vector $\vec{u}$ of the triple, your right middle finger along the second vector $\vec{v}$, and check if the thumb is pointing in the direction of the third vector $\vec{w}$ or not.

For the right-hand rule to make sense, the order of the vectors is very important. Indeed, it is always true that if $\vec{u}$, $\vec{v}$ and $\vec{w}$ are three non-zero arrows and $\vec{u} \nparallel \vec{v}$, $\vec{u} \perp \vec{w}$ and $\vec{v} \perp \vec{w}$, then either the triple $(\vec{u}, \vec{v}, \vec{w})$ or the triple $(\vec{v}, \vec{u}, \vec{w})$ satisfies the right-hand rule (never both). Similarly, either $(\vec{u}, \vec{v}, \vec{w})$ or $(\vec{u}, \vec{v}, -\vec{w})$ satisfies the right-hand rule (never both).

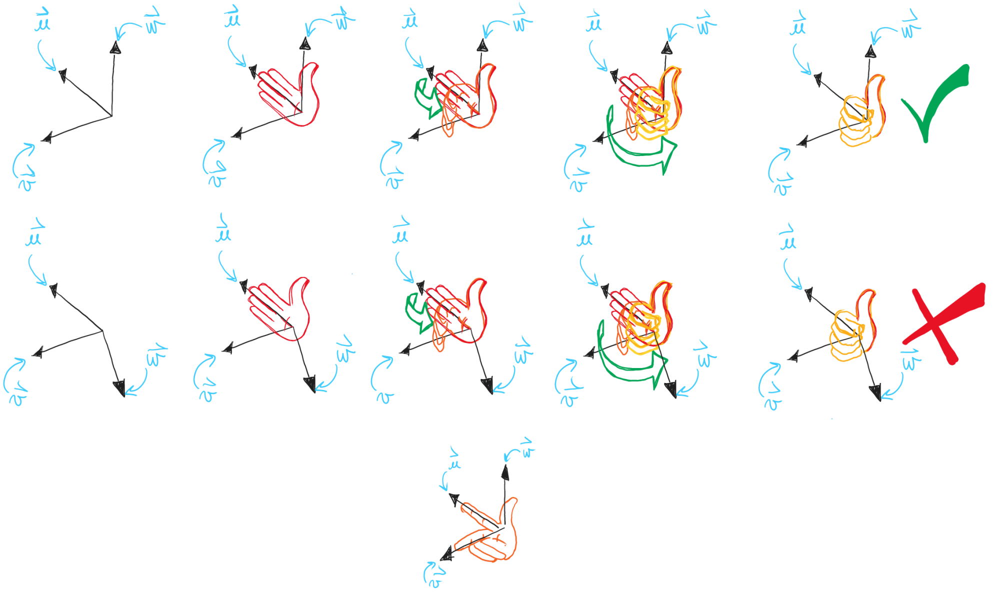


Figure 1.4: The right-hand rule

We decide to choose $\hat{\mathbf{k}}$ so that the triple $(\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}})$ satisfies the r.h.r.. So, to recap we take an arbitrary point P and an arrow $\hat{\mathbf{i}}$ of modulus 1, then we take another arrow $\hat{\mathbf{j}}$, still of modulus 1 and orthogonal to $\hat{\mathbf{i}}$. Finally, we take a third arrow $\hat{\mathbf{k}}$ again of modulus 1, orthogonal to both $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ and oriented so that $(\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}})$ satisfies the r.h.r. We call this construction a system of coordinates for the three-dimensional space.

The goal is to show that a system of coordinates provides a basis for the vector space $\mathbb{V}_3$. In Picture 1.5 we choose a generic arrow $\vec{u} \in \mathbb{V}_3$. Then we project $\vec{u}$ over the plane $\hat{\mathbf{i}}\hat{\mathbf{j}}$ and so we define two new arrows: $\vec{v}$ is the projection of $\vec{u}$ over the plane $\hat{\mathbf{i}}\hat{\mathbf{j}}$ and $\vec{z}$ is the arrow that starts where $\vec{v}$ ends and ends where $\vec{u}$ ends. We call $\vec{z}$ the vertical component of $\vec{u}$. Now, we project $\vec{v}$ along $\hat{\mathbf{i}}$ and along $\hat{\mathbf{j}}$, obtaining other two arrows that we call $\vec{x}$ and $\vec{y}$, respectively. Note that $\vec{x} + \vec{y} = \vec{v}$ and that $\vec{v} + \vec{z} = \vec{u}$, so we obtain that:

$$\vec{u} = \vec{x} + \vec{y} + \vec{z}$$

Notice that $\vec{x}$, $\vec{y}$ and $\vec{z}$ are parallel to $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$, respectively. This means that there are three real numbers x, y and z so that $\vec{x} = x\hat{\mathbf{i}}, \vec{y} = y\hat{\mathbf{j}}$ and $\vec{z} = z\hat{\mathbf{k}}$. To describe these numbers let's consider the case of $\vec{x}$. We define x to be the real number so that $|x| = |\vec{x}|$ and it has a positive sign if $\vec{x}$ and $\hat{\mathbf{i}}$ have the same orientation and negative sign if $\vec{x}$ and $\hat{\mathbf{i}}$ has opposite orientations. The same procedure is used to define y and z, so we finally obtain that:

$$\vec{u} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$$

This is a good start: we showed that every arrow in $\mathbb{V}_3$ can be written as a combination of three numbers x, y, z and the three vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$. However, this is not sufficient to prove that $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ is a basis for $\mathbb{V}_3$: we need to show that x, y, z are the unique numbers that do the job. So, consider three real numbers x', y' and z' so that:

$$\vec{u} = x'\hat{\mathbf{i}} + y'\hat{\mathbf{j}} + z'\hat{\mathbf{k}}$$

by projecting $\vec{u}$ along the $\hat{\mathbf{i}}\hat{\mathbf{j}}$ plane first, and then along $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$, we obtain that $x' = x$, $y' = y$ and that $z' = z$, so $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ is indeed a basis.

It is important to realize that $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ is a very special basis. A generic basis of $\mathbb{V}_3$ can be simply built by choosing three non-zero vectors of any length so that the first two are not parallel and the third one does not lie in the plane generated by the first ones. We didn't need to ask for orthogonality or that the modulus of arrows to be equal to 1, nor

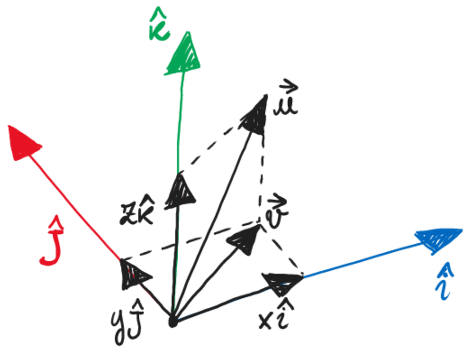


Figure 1.5: The components of an arrow

that the three arrows needed to satisfy the r.h.r.! The basis $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ is sometimes called an orthonormal basis, which is a basis with all the vectors orthogonal and with modulus equal to 1. Orthonormal bases make computations way easier than other bases, that's why we prefer them. We conclude this section by introducing an important definition.

Definition 1.9. *Given a vector* $\vec{u} \in \mathbb{V}$ *in a vector space* $\mathbb{V}$ *with a basis* $B = \{\vec{v}_i\}$*, the* ***components*** *of* $\vec{u}$ *are the unique numbers* $x_1, \ldots, x_{n^*} \in \mathbb{R}$ *such that:*

$$\vec{u} = x_1 \vec{v}_1 + \cdots + x_{n^*} \vec{v}_{n^*}$$

We adopt the following notation $\vec{u} = \langle x_1, \ldots, x_{n^*} \rangle$ *to indicate that the components of* $\vec{u}$ *are* $x_1, \ldots, x_{n^*}$*. Sometimes, we also use the other notation which is* $\vec{u} = (x_1, \ldots, x_{n^*})$*. We also point out that sometimes components are also called Cartesian coordinates.*

When $\mathbb{V} = \mathbb{V}_3$, the components of an arrow $\vec{u} \in \mathbb{V}_3$ are the numbers x, y, z so that $\vec{u} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$ and we indicate that as $\vec{u} = \langle x, y, z \rangle$ or $\vec{u} = (x, y, z)$. We note that sometimes, people refer to components as the vectors $x\hat{\mathbf{i}}$, $y\hat{\mathbf{j}}$ and $z\hat{\mathbf{k}}$ and not simply the numbers x, y and z. However, these two uses are both accepted. We finally note that $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ in components become $\langle 1, 0, 0 \rangle$, $\langle 0, 1, 0 \rangle$ and $\langle 0, 0, 1 \rangle$, respectively.

1.3 The scalar product

In the previous section, we discovered that the set of arrows we can draw in the three-dimensional space is a vector space. We also described a procedure to define an orthonormal basis in that space and we found out that the space has dimension 3, since we only need three arrows, $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$, to describe every other arrow $\vec{u}$ using three numbers x, y, z, $\vec{u} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$.

To build the orthonormal basis $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ we made extensive use of two notions: the modulus of an arrow, which corresponds to its length, and the angle between two arrows. As pointed out at the end of the section, these two notions are not required to construct a generic basis: we just need to find vectors that are "independent" and that "generate" the whole space (these two notions, independence and generators, are key concepts of linear algebra and if you are curious about I suggest you take a look at the Wikipedia page, for example).

However, with a well-defined notion of the angle between arrows and the length of arrows, we can define an orthonormal basis which is a basis with nice properties for computations. In general, in an arbitrary vector space, we cannot construct orthonormal bases because we don't have a way to measure lengths and angles.

Usually, in mathematics, it is important to understand what minimal structures we need in order to make a specific construction. It's like when you have a DIY project in your head and you need to buy the right tools to make it. Buying tools is expensive and you wanna minimize the cost of your project, so you want to find just the right number of tools you need to do the job.

We want to understand the required tools to build an orthonormal basis. The scalar product is the magic tool that allows us to measure both lengths and angles. In this section, we introduce this mathematical tool and discuss its properties.

Definition 1.10. *The scalar product between two arrows $\vec{u} = \langle x, y, z\rangle$ and $\vec{v} = \langle x', y', z'\rangle$ is the real number denoted by $\vec{u} \cdot \vec{v}$ defined as follows:*

$$\vec{u} \cdot \vec{v} := xx' + yy' + zz'$$

The operation just introduced defines a function:

$$\begin{aligned} \cdot : \ & \mathbb{V}_3 \times \mathbb{V}_3 \to \mathbb{R} \\ & (\vec{u}, \vec{v}) \mapsto \vec{u} \cdot \vec{v} \end{aligned}$$

Note that we use the same notation for the scalar action and for this new operation $\cdot$, however, these two are not the same thing! This function satisfies some interesting properties.

Proposition 1.3. *The function $\cdot : \mathbb{V}_3 \times \mathbb{V}_3 \to \mathbb{R}$, has the following properties:*

Symmetry *For every $\vec{u}, \vec{v} \in \mathbb{V}_3$:*

$$\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$$

Right linearity *For every $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}_3$ and $a, b \in \mathbb{R}$:*

$$\vec{u} \cdot (a\vec{v} + b\vec{w}) = a(\vec{u} \cdot \vec{v}) + b(\vec{u} \cdot \vec{w})$$

Left linearity *For every $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}_3$ and $a, b \in \mathbb{R}$:*

$$(a\vec{u} + b\vec{v}) \cdot \vec{w} = a(\vec{u} \cdot \vec{w}) + b(\vec{v} \cdot \vec{w})$$

Positive-definiteness *For every $\vec{u} \in \mathbb{V}_3$, $\vec{u} \cdot \vec{u} \geq 0$ and $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$*

Proof. Let's start by showing the symmetry. Let $\vec{u} = \langle x, y, z\rangle$ and $\vec{v} = \langle x', y', z'\rangle$, then:

$$\begin{aligned} & \vec{u} \cdot \vec{v} = \\ = \ & xx' + yy' + zz' = \\ = \ & x'x + y'y + z'z = \\ = \ & \vec{v} \cdot \vec{u} \end{aligned}$$

where we just used that the multiplication of numbers is commutative. Let's prove the right linearity. Let's first prove that, for every $a \in \mathbb{R}$, $\vec{u} \cdot (a\vec{v}) = a(\vec{u} \cdot \vec{v})$. Note that $a\vec{v} = \langle ax', ay', az'\rangle$:

$$\begin{aligned} & \vec{u} \cdot (a\vec{v}) = \\ = \ & x(ax') + y(ay') + z(az') = \\ = \ & a(xx') + a(yy') + a(zz') = \\ = \ & a(xx' + yy' + zz') = \\ = \ & a(\vec{u} \cdot \vec{v}) \end{aligned}$$

Let's now show that $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$, where $\vec{w}$ is a third arrow $\vec{w} = \langle x'', y'', z''\rangle$. Note that $\vec{v} + \vec{w} = \langle x' + x'', y' + y'', z' + z''\rangle$:

$$\begin{aligned} & \vec{u} \cdot (\vec{v} + \vec{w}) = \\ = \ & \vec{u} \cdot (\vec{v} + \vec{w}) = \\ = \ & x(x' + x'') + y(y' + y'') + z(z' + z'') = \\ = \ & xx' + xx'' + yy' + yy'' + zz' + zz'' = \\ = \ & xx' + yy' + zz' + xx'' + yy'' + zz'' = \\ = \ & \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w} \end{aligned}$$

To show left linearity we just need the symmetry and the right linearity, indeed, if $a, b \in \mathbb{R}$ and $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}_3$ we have:

$$\begin{aligned} & (a\vec{u} + b\vec{v}) \cdot \vec{w} = \\ = \quad & \vec{w} \cdot (a\vec{u} + b\vec{v}) = \\ = \quad & a(\vec{w} \cdot \vec{u}) + b(\vec{w} \cdot \vec{v}) = \\ = \quad & a(\vec{u} \cdot \vec{w}) + b(\vec{v} \cdot \vec{w}) \end{aligned}$$

Finally, consider $\vec{u} = \langle x, y, z \rangle$, then:

$$\vec{u} \cdot \vec{u} = xx + yy + zz = x^2 + y^2 + z^2$$

and a square of a number is always non-negative, thus $\vec{u} \cdot \vec{u} \geq 0$. Moreover, if $\vec{u} \cdot \vec{u} = 0$, means that $x^2 + y^2 + z^2 = 0$ and since $x^2, y^2, z^2 \geq 0$ they need all three to be zero, so we have that $x = y = z = 0$, i.e. $\vec{u} = \vec{0}$. So, $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$. □

Thanks to the right and left linearity, for every $a \in \mathbb{R}$ and $\vec{u}, \vec{v} \in \mathbb{V}_3$ we have:

$$(a\vec{u}) \cdot \vec{v} = a(\vec{u} \cdot \vec{v}) = \vec{u} \cdot (a\vec{v})$$

In the following, we simply write $a\vec{u} \cdot \vec{v}$ for any of the three equivalent forms above.

A function $\cdot : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ which satisfies the properties we showed in Proposition 1.3 has a fancy name: is called a scalar product, or sometimes, inner product or also dot product, for the use of the notation $\cdot$. $\mathbb{V}_3$ is not the only vector space that has a scalar product so let's introduce this concept.

Definition 1.11. *An **inner product space** is a vector space $\mathbb{V}$ equipped with a function $\cdot : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ called the **inner product** or **scalar product** or **dot product**, satisfying the following properties:*

Symmetry *For every $\vec{u}, \vec{v} \in \mathbb{V}$:*

$$\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$$

Right linearity *For every $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}$ and $a, b \in \mathbb{R}$:*

$$\vec{u} \cdot (a\vec{v} + b\vec{w}) = a(\vec{u} \cdot \vec{v}) + b(\vec{u} \cdot \vec{w})$$

Positive-definiteness *For every $\vec{u} \in \mathbb{V}$, $\vec{u} \cdot \vec{u} \geq 0$ and $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$*

Left linearity is a direct consequence of right linearity and symmetry, so we didn't include it in the definition of the scalar product. We introduced the scalar product to have a tool to measure the length of vectors and angles between two vectors, but it is still not clear why this should be the case. The next proposition gives us an answer.

Proposition 1.4. *Consider two arrows $\vec{u}$ and $\vec{v}$ in $\mathbb{V}_3$. If θ is the angle between $\vec{u}$ and $\vec{v}$ then:*

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos\theta$$

Moreover:

$$\sqrt{\vec{u} \cdot \vec{u}} = |\vec{u}|$$

Proof. Consider $\vec{u} = \langle x, y, z \rangle$ then, by Pitagora's theorem we have that:

$$|\vec{u}| = \sqrt{x^2 + y^2 + z^2}$$

However, $x^2 + y^2 + z^2 = \vec{u} \cdot \vec{u}$, so the second part is proved. To show the first claim, consider the triangle formed by $\vec{u}$, $\vec{v}$ and by the arrow that goes from the endpoint of $\vec{v}$ to the endpoint of $\vec{u}$. This arrow is actually $\vec{v} - \vec{u} = \vec{v} + (-\vec{u})$. The cosine law establishes that if θ is the angle between two sides s_1 and s_2 of a triangle, then:

$$s_3^2 = s_1^2 + s_2^2 - 2s_1s_2\cos\theta$$

where s_3 is the third side. You can think of this as a generalized Pitagora's theorem because when $\theta = \pi/2$ then $\cos\theta = \cos(\pi/2) = 0$ and then the cosine law restricts to Pitagora's theorem. By using this result we have that:

$$|\vec{v} - \vec{u}|^2 = |\vec{u}|^2 + |\vec{v}|^2 - 2|\vec{u}||\vec{v}|\cos\theta$$

However, we just showed that for every arrow $\vec{u}$, $|\vec{u}|^2 = \vec{u} \cdot \vec{u}$, therefore:

$$\begin{aligned} & |\vec{v} - \vec{u}|^2 = \\ = \quad & (\vec{v} - \vec{u}) \cdot (\vec{v} - \vec{u}) = \\ = \quad & \vec{v} \cdot \vec{v} - \vec{v} \cdot \vec{u} - \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{u} = \\ = \quad & \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2(\vec{u} \cdot \vec{v}) = \\ = \quad & |\vec{u}|^2 + |\vec{v}|^2 - 2(\vec{u} \cdot \vec{v}) \end{aligned}$$

where we used that $\cdot$ is right and left linear and that is symmetric. So, by comparing $|\vec{u}|^2 + |\vec{v}|^2 - 2(\vec{u} \cdot \vec{v})$ with $|\vec{u}|^2 + |\vec{v}|^2 - 2|\vec{u}||\vec{v}|\cos\theta$ and cancelling the terms $|\vec{u}|^2$ and $|\vec{v}|^2$ and simplifying the -2, we conclude that:

$$\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}|\cos\theta$$

as expected. □

Cool Stuff* 1.1. Proposition 1.4 proves an important thing about the scalar product of arrows. It shows that the scalar product does not depend on a particular choice of the basis. This is something quite subtle, but very important: to define $\cdot$ we used the components of the arrows, but the components are defined by choosing a basis of the vector space. We have chosen an orthonormal basis, but we could have used any other basis.

In mathematics, every time we make a choice to define a new concept in mathematics, we need to prove that that choice wasn't special, that we could have obtained the same thing with any other choice, in this case of a basis. Otherwise, we need to keep track of the basis we are working on and remember it all the time.

We proved that $\vec{u} \cdot \vec{v}$ is equal to $|\vec{u}||\vec{v}|\cos\theta$. The modulus of the arrows and the angle between them is an intrinsic property of the arrows which does not depend on any basis, so the scalar product of $\mathbb{V}_3$ is independent of the choice of basis. ■

The second important consequence of Proposition 1.4 is that with $\cdot$ we can easily measure the length of vectors. In particular, we showed that $\vec{u} \cdot \vec{u} \geq 0$ and that $\sqrt{\vec{u} \cdot \vec{u}} = |\vec{u}|$. This result has an important generalization.

Cool Stuff* 1.2. Measuring the length of vectors is usually done by a **norm**. A normed space is a vector space equipped with a function, called the norm, $\|-\| \colon \mathbb{V} \to \mathbb{R}$, which sends every vector $\vec{v}$ to a real number $\|\vec{v}\|$, interpreted as the length of $\vec{v}$. For $\|\vec{v}\|$ to correctly represent the notion of a length, the function $\|-\|$ must satisfy the following properties:

1. For every vector $\vec{v}$, $\|\vec{v}\| \geq 0$;
2. $\|\vec{v}\| = 0$ if and only if $\vec{v} = \vec{0}$;
3. For every vector $\vec{v}$ and every scalar a, $\|a\vec{v}\| = |a|\|\vec{v}\|$;
4. For every $\vec{u}, \vec{v} \in \mathbb{V}$ and $a \in \mathbb{R}$, $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$.

An example of a norm is the modulus of arrows, that's why sometimes, $|\vec{u}|$ is denoted by $\|\vec{u}\|$. Every time a vector space $\mathbb{V}$ has a scalar product $\cdot$, we can always define a norm by setting $\|\vec{v}\| := \sqrt{\vec{v} \cdot \vec{v}}$. This is the case of the modulus of arrows because $|\vec{u}| = \sqrt{\vec{u} \cdot \vec{u}}$. In this case, we say that **the scalar product induces a norm**. However, not every norm is induced by a scalar product.

For example, the so-called p-norms are not induced by a scalar product when $p \neq 2$. Let's define these family of norms as follows. Consider any real number $p \geq 1$ and take the vector space $\mathbb{V}_3$, with the usual basis $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$. For a vector $\vec{v} = \langle x, y, z \rangle$:

$$\|\vec{v}\|_p := (x^p + y^p + z^p)^{\frac{1}{p}}$$

When $p = 2$, we have exactly that $\|\vec{v}\|_p = |\vec{v}|$, but in general, for other p, $\|-\|_p$ is not induced by a scalar product. This fact is actually very useful because some vector spaces have no well-defined scalar products. This is a consequence of the fact that these vector spaces are incredibly large, i.e. they contain an incredible amount of vectors. So, in the absence of a scalar product one can try to define a norm and it turns out that defining norms is easier than defining scalar products. As a final note, there is an interesting part of mathematics, called functional analysis whose goal is to study these spaces. ■

Thanks to the scalar product we can now measure the length of vectors. But we also want to measure the angle between two vectors! Lucky for us, Proposition 1.4 addresses this problem, too. If $\vec{u}$ and $\vec{v}$ are two non-zero arrows, we can divide by $|\vec{u}|$ and $|\vec{v}|$ and obtain:

$$\cos\theta = \frac{\vec{u}\cdot\vec{v}}{|\vec{u}||\vec{v}|}$$

This does not give exactly the angle θ, however, if $\theta \geq 0$, we have $0 \leq \theta \leq \pi$ and $\cos$ is invertible in this interval, so we can take $\arccos((\vec{u}\cdot\vec{v})/(|\vec{u}||\vec{v}|))$ and obtain θ. Similarly, if $\theta < 0$, then we have that $-\pi \leq \theta < 0$ and again, $\cos$ is invertible in this interval, so again we can take $\arccos((\vec{u}\cdot\vec{v})/(|\vec{u}||\vec{v}|))$ and find θ. So, the scalar product gives both angles and lengths!

There is an interesting geometric interpretation of the scalar product. Let's first introduce a useful definition.

Definition 1.12. *Let $\vec{u} \in \mathbb{V}$ be a non-zero vector. The **versor** of $\vec{u}$ is the vector with the same direction and orientation of $\vec{u}$, but modulus equal to 1. More precisely, this is the vector $\hat{u}$ so defined:*

$$\hat{u} := \frac{\vec{u}}{|\vec{u}|}$$

*Let $\vec{u}, \vec{v} \in \mathbb{V}_3$ and suppose $\vec{u} \neq \vec{0}$. Then, the **scalar projection** of $\vec{v}$ onto $\vec{u}$ is the number:*

$$\mathsf{comp}_{\vec{u}}\vec{v} := \frac{\vec{u}\cdot\vec{v}}{|\vec{u}|} = \hat{u}\cdot\vec{v}$$

*The **vector projection** of $\vec{v}$ onto $\vec{u}$ is the vector:*

$$\mathsf{proj}_{\vec{u}}\vec{v} := (\mathsf{comp}_{\vec{u}}\vec{v})\hat{u} = \frac{\vec{u}\cdot\vec{v}}{|\vec{u}|^2}\vec{u}$$

We interpret the versor $\hat{u}$ as the direction and orientation of $\vec{u}$, the scalar projection of $\vec{v}$ onto $\vec{u}$ as the scalar component of $\vec{v}$ along $\vec{u}$, and the vector projection of $\vec{v}$ onto $\vec{u}$ as the arrow obtained by projecting $\vec{v}$ along the direction of $\vec{u}$. With these definitions, we can express $\vec{u}\cdot\vec{v}$ as the the scalar projection of $\vec{v}$ onto $\vec{u}$ times the modulus of $\vec{u}$, whenever $\vec{u} \neq \vec{0}$:

$$\vec{u}\cdot\vec{v} = |\vec{u}|\mathsf{comp}_{\vec{u}}\vec{v}$$

or, equivalently, the scalar projection of $\vec{u}$ onto $\vec{v}$ times the modulus of $\vec{v}$, whenever $\vec{v} \neq \vec{0}$:

$$\vec{u}\cdot\vec{v} = |\vec{v}|\mathsf{comp}_{\vec{v}}\vec{u}$$

Note that when $\vec{u}\perp\vec{v}$, the projection of $\vec{v}$ along $\vec{u}$ is just zero, so we have the following interesting result.

Proposition 1.5. *Suppose that $\vec{u}, \vec{v} \in \mathbb{V}_3$ and that $\vec{u}, \vec{v} \neq \vec{0}$. Then $\vec{u}\cdot\vec{v} = 0$ if and only if $\vec{u}\perp\vec{v}$. Finally, for every $\vec{u} \in \mathbb{V}_3$ we have that:*

$$\vec{u} = \langle \vec{u}\cdot\hat{\imath}, \vec{u}\cdot\hat{\jmath}, \vec{u}\cdot\hat{k}\rangle$$

Proof. Suppose that $\vec{u}\perp\vec{v}$, then $\theta = \pi/2$, thus $\cos\theta = 0$ and then $\vec{u}\cdot\vec{v} = 0$. Now, let's prove the converse. Suppose that $\vec{u}\cdot\vec{v} = 0$. Since $\vec{u}, \vec{v} \neq \vec{0}$, $|\vec{u}|, |\vec{v}| \neq 0$, so it has to be that $\cos\theta = 0$. But because $-\pi\pi$, $\cos\theta$ means that $\theta = \pi/2$ or $\theta = -\pi/2$, which means in both cases that $\vec{u}\perp\vec{v}$. To show that the components of $\vec{u}$ are determined by the scalar product with $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ let $\vec{u} = \langle x, y, z\rangle$, then:

$$\begin{aligned} & \vec{u}\cdot\hat{\mathbf{i}} = \\ = \quad & (x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}})\cdot\hat{\mathbf{i}} = \\ = \quad & x\hat{\mathbf{i}}\cdot\hat{\mathbf{i}} + y\hat{\mathbf{j}}\cdot\hat{\mathbf{i}} + z\hat{\mathbf{k}}\cdot\hat{\mathbf{i}} \end{aligned}$$

but $\hat{\mathbf{i}}\cdot\hat{\mathbf{i}} = \left|\hat{\mathbf{i}}\right| = 1$ and because $\hat{\mathbf{i}}\perp\hat{\mathbf{j}}$ and $\hat{\mathbf{i}}\perp\hat{\mathbf{k}}$, $\hat{\mathbf{j}}\perp\hat{\mathbf{i}} = \hat{\mathbf{k}}\perp\hat{\mathbf{i}} = 0$, so $\vec{u}\cdot\hat{\mathbf{i}} = x$. Similarly, $\vec{u}\cdot\hat{\mathbf{j}} = y$ and $\vec{u}\cdot\hat{\mathbf{k}} = \vec{z}$. ☐

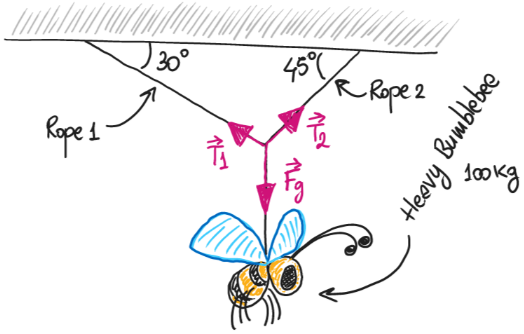


Figure 1.6: A giant bumblebee kept suspended with ropes

Cool Stuff* 1.3. Usually, a vector space does not have an intrinsic notion of the angle between vectors. So, to talk about orthogonality, we require a vector space $\mathbb{V}$ to have a scalar product $\cdot$. With such a structure, we can say that two vectors $\vec{u}$ and $\vec{v}$ are orthogonal if $\vec{u} \cdot \vec{v} = 0$. For $\mathbb{V}_3$ this corresponds to the usual notion of orthogonality. However, in general, this is a definition. This is something that mathematicians love to do: a result in a special case becomes a definition in the general case! ■

Before we conclude the section, we give an application.

Example 1.4. Imagine that a giant bumblebee of mass $100kg$ (which is about $220.5lb$) is kept suspended by a rope as in Picture As illustrated, the rope is divided into two ropes, one forming an angle of $30°$ and the other one, an angle of $45°$ with the ceiling. What is the tension the two pieces of ropes have to tolerate to sustain the weight of the bumblebee?

OK, so there are a few things you need to know about ropes and tensions. First, what is tension? Tension is the force that the rope has to tolerate to resist traction. If the tension is too strong the rope could break. The second thing we need is the second Netwon law that establishes that a force is proportional to the acceleration of the body. More precisely, $\vec{F} = m\,\vec{a}$, where $\vec{F}$ is the force, m is the mass of the body and $\vec{a}$ is the acceleration. It turns out that the gravitational acceleration doesn't change too quickly (it does, but it changes very slowly with the height) so we usually assume it is constant and equal to $g := 9.8m/s^2$.

Now, this is the magnitude of the acceleration, and the direction is the same as $\hat{\mathbf{k}}$, but it has the opposite orientation, because $\hat{\mathbf{k}}$ points upward by convention, while gravity points downward (at least on the surface of a planet like ours). So, the acceleration that the body perceives is $\vec{a} = -g\hat{\mathbf{k}}$, where the negative sign indicates that it's pointing downward.

So far so good. Now, the next thing we need to know is that in equilibrium, so when everything is stable and the bumblebee stays attached to the ropes and the ropes do not break, the forces cancel out. This is also a consequence of the second Netwon law because when everything is in equilibrium the acceleration is null and so the total force $\vec{F}$ has to be zero. There are three forces: gravity, which is $\vec{F}_g := -mg\hat{\mathbf{k}}$, the tension on the first rope $\vec{T}_2$ and the tension on the second piece of rope $\vec{T}_1$ and on equilibrium we have the equation:

$$\vec{F}_g + \vec{T}_2 + \vec{T}_1 = \vec{0}$$

Now, $\vec{T}_2$ has the direction of the first piece of rope and is pointing upward, while $\vec{T}_1$ has the same direction of the second piece of rope and is pointing upward, too. One approach to solve this problem is to find all the components of all the vectors. We already found out that $\vec{F}_g = -mg\hat{\mathbf{k}}$, so in components $\vec{F}_g = \langle 0, 0, -mg \rangle$. Let's find the components of $\vec{T}_2$ and $\vec{T}_1$. Thanks to Proposition 1.5, we have that the components of an arrow can be found by multiplying by $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$, so:

$$x_1 := \vec{T}_2 \cdot \hat{\mathbf{i}} = |\vec{T}_2| \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}|\vec{T}_2|$$

$$z_1 := \vec{T}_2 \cdot \hat{\mathbf{k}} = |\vec{T}_2| \cos\frac{\pi}{4} = \frac{1}{\sqrt{2}}|\vec{T}_2|$$

$$x_2 := \vec{T}_1 \cdot \hat{\mathbf{i}} = |\vec{T}_1| \cos \frac{5\pi}{6} = -\frac{\sqrt{3}}{2} |\vec{T}_1|$$
$$z_2 := \vec{T}_1 \cdot \hat{\mathbf{k}} = |\vec{T}_1| \cos \frac{\pi}{3} = \frac{1}{2} |\vec{T}_1|$$

where $\vec{T}_2 = \langle x_1, 0, z_1 \rangle$ and $\vec{T}_1 = \langle x_2, 0, z_2 \rangle$. Therefore, we now have the equation:

$$\langle 0,0,0 \rangle = \vec{0} = \vec{F}_g + \vec{T}_2 + \vec{T}_1 = \langle 0, 0, -mg \rangle + \langle x_1, 0, z_1 \rangle + \langle x_2, 0, z_2 \rangle = \langle x_1 + x_2, 0, -mg + z_1 + z_2 \rangle$$

where we used that $\langle x, y, z \rangle + \langle x', y', z' \rangle = \langle x + x', y + y', z + z' \rangle$. So, we have a system of two equations to solve:

$$\begin{cases} x_1 + x_2 = 0 \\ z_1 + z_2 = mg \end{cases}$$

i.e.:

$$\begin{cases} \frac{1}{\sqrt{2}} |\vec{T}_2| - \frac{\sqrt{3}}{2} |\vec{T}_1| = 0 \\ \frac{1}{\sqrt{2}} |\vec{T}_2| + \frac{1}{2} |\vec{T}_1| = mg \end{cases}$$

Here, the two variables are $|\vec{T}_2|$ and $|\vec{T}_1|$. Let's simplify a bit the notation and let's called them T_2 and T_1, respectively. So, we have:

$$\begin{cases} \frac{1}{\sqrt{2}} T_2 = \frac{\sqrt{3}}{2} T_1 \\ \frac{1}{\sqrt{2}} T_2 + \frac{1}{2} T_1 = mg \end{cases}$$

From the first equation, we find that $T_2 = \sqrt{\frac{3}{2}} T_1$, so by plugging in this into the second one we find that:

$$\frac{\sqrt{3}}{2} T_1 + \frac{1}{2} T_1 = mg$$

thus:

$$T_1 = \frac{2}{\sqrt{3} + 1} mg \sim 717.4N$$
$$T_2 = \frac{6}{\sqrt{3} + 1} mg \sim 878.6N$$

N stays for "Newton" which is the unit of measure of forces: $1N = 1kg \cdot m/s^2$. So that's it! We found that the two pieces of ropes have to tolerate $T_2 \sim 717.4N$ and $T_1 \sim 878.6N$ of tension, respectively. ■

1.4 Arrows and components: a general approach*

This section is a reflection on what we have done so far and it's not immediately necessary to understand these notes or to pass the exam and assignments. I suggest you take a look because it may clarify what's going on!

So far, we defined vector spaces and scalar products and we focused our attention on the case of $\mathbb{V}_3$. The reality is that working with arrows, even if it's very cool, it's not always the easiest. For example, proving that $\cdot$ is right linear is quite easy when we define the scalar product by using the components of the arrows, instead of using $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$. Proving that by using this as a definition for $\cdot$ is way more complicated! Other times, working only with the components does not give the whole picture of what's going on! For example, there is not an immediate intuition of the scalar projection and of the vector projection when we regard arrows as tuples of numbers, instead of as actual arrows. If we wanna be a bit philosophical, we can say that there are two main points of view on vectors: one is geometric and allows us to interpret them as arrows, and the other is algebraic and talks about the components of these vectors and their algebraic properties.

There is a way to make precise this intuition. In Example 1.2 we showed that for every integer $n > 0$, $\mathbb{R}^n$ is a vector space. In particular, this means that also $\mathbb{R}^3$ is a vector space. The vectors of this space are ordered triples of real numbers (x, y, z), the sum is defined as follows:

$$(x, y, z) + (x', y', z') := (x + x', y + y', z + z')$$

for $(x, y, z), (x', y', z') \in \mathbb{R}^3$; and the scalar action as:

$$a(x, y, z) := (ax, ay, az)$$

for $a \in \mathbb{R}$. Now, the idea is that $\mathbb{R}^3$ captures the algebraic interpretation of vectors in three-dimensional space, while the space of arrows $\mathbb{V}_3$ captures the geometric one. We just need something that translates things from $\mathbb{V}_3$ to things in $\mathbb{R}^3$ and vice versa. This translator is called an isomorphism and it's a function $\mathbb{V}_3 \to \mathbb{R}^3$ which is bijective and that compares the structures of these two spaces. Let's give a concrete definition of this isomorphism. To define it we need a basis of $\mathbb{V}_3$ and we decide to choose the usual one $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$, so now an arrow $\vec{u} \in \mathbb{V}_3$ can be described using its components $\vec{u} = \langle x, y, z \rangle$. But (x, y, z) is also an element of $\mathbb{R}^3$ so we have:

$$\gamma\colon \mathbb{V}_3 \to \mathbb{R}^3$$
$$\gamma(\langle x, y, z \rangle) := (u_1, u_2, u_3)$$

So, given a basis, we can define a function γ which maps every arrow of $\mathbb{V}_3$ in $\mathbb{R}^3$. However, also the opposite is true, given a vector $(x, y, z) \in \mathbb{R}^3$ we can find an arrow $x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}} \in \mathbb{V}_3$, so γ is a bijection. Note that this construction works more generally for every finite-dimensional vector space $\mathbb{V}$. If $B := \{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis for $\mathbb{V}$, we can define a function:

$$\gamma\colon \mathbb{V} \to \mathbb{R}^n$$
$$\gamma(x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) := (x_1, \dots, x_n)$$

So, now we have a bijection between $\mathbb{V}_3$ and $\mathbb{R}^3$. It turns out that this bijection is also linear! This means that, for every $\vec{u} = \langle x, y, z \rangle$, $\vec{v} = \langle x', y', z' \rangle$ and $a, b \in \mathbb{R}$, $\gamma(a\vec{u} + b\vec{v}) = a\gamma(\vec{u}) + b\gamma(\vec{v})$. This is the same as saying that:

$$a\vec{u} = \langle ax, ay, az \rangle$$
$$\vec{u} + \vec{v} = \langle x + x', y + y', z + z' \rangle$$

A linear bijection is often called an isomorphism of vector spaces. More generally, an isomorphism between two mathematical objects is a transformation between these two that is invertible and that preserves the specific structures of these objects. For vector spaces the important structures are the sum and the scalar action, so linearity is the right notion we need to preserve. In a way, we are saying that $\mathbb{V}_3$ and $\mathbb{R}^3$ represent the same kinds of things, but they give two different interpretations. Now, it's not the end, because γ also preserves the scalar product. First, notice that on $\mathbb{V}_3$ the scalar product is defined as $\vec{u} \cdot \vec{v} := |\vec{u}||\vec{v}|\cos\theta$, while on $\mathbb{R}^3$, the scalar product is defined as $(x, y, z) \cdot (x', y', z') = xx' + yy' + zz'$. Proposition 1.4 proves exactly that $\vec{u} \cdot \vec{v} = \gamma(\vec{u}) \cdot \gamma(\vec{v})$, which reads as:

$$|\vec{u}||\vec{v}|\cos\theta = xx' + yy' + zz'$$

for $\vec{u} = \langle x, y, z \rangle$ and $\vec{v} = \langle x', y', z' \rangle$. Now, because $|\vec{u}| = \sqrt{\vec{u} \cdot \vec{u}}$ and that γ preserves the scalar product (we sometimes say that γ is an isomorphism of inner product spaces), it turns out that γ preserves also the modulus! First, the modulus (or norm) in $\mathbb{R}^3$ is so defined:

$$|(x, y, z)| := \sqrt{x^2 + y^2 + z^2}$$

It is sometimes called the Euclidean scalar product. In the proof of Proposition 1.4, we showed that, if $\vec{u} = \langle x, y, z \rangle$, then $|\vec{u}|^2 = x^2 + y^2 + z^2$. We can translate this result by saying that $|\vec{u}| = |\gamma(\vec{u})|$ (i.e. γ is an isomorphism of normed spaces). We can collect these results in a little table. We conclude this section with an interesting observation. We said that the isomorphism $\gamma : \mathbb{V}_3 \to \mathbb{R}^3$ is defined once we choose a basis. This is absolutely correct and indeed one could actually show that every such isomorphism corresponds precisely to a basis on $\mathbb{V}_3$. This is actually stronger. Since γ also preserves the scalar product, one can show that an isomorphism like γ gives exactly an orthonormal basis for $\mathbb{V}_3$! The idea of this proof comes from invertibility. If γ is a linear bijective function from $\mathbb{V}_3$ to $\mathbb{R}^3$ that preserves the scalar product, then its inverse γ^{-1} is a linear bijective function from $\mathbb{R}^3$ to $\mathbb{V}_3$ that preserves the scalar product. So, by taking $\vec{v}_1 := \gamma^{-1}(1, 0, 0)$, $\vec{v}_2 := \gamma^{-1}(0, 1, 0)$ and $\vec{v}_3 := \gamma^{-1}(0, 0, 1)$, one proves that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is an orthonormal basis of $\mathbb{V}_3$.

Operation	$\mathbb{V}_3$	$\mathbb{R}^3$
Zero vector	$\vec{0}$	$(0, 0, 0)$
Sum	$\vec{u} + \vec{v}$	$(x + x', y + y', z + z')$
Scalar action	$a\vec{u}$	(ax, ay, az)
Scalar product	$\lvert\vec{u}\rvert\lvert\vec{v}\rvert\cos\theta$	$xx' + yy' + zz'$
Norm	$\lvert\vec{u}\rvert$	$\sqrt{x^2 + y^2 + z^2}$

This observation generates a new question: how many different orthonormal bases there are for $\mathbb{V}_3$? Well, this is a good question, but the answer is not really interesting, simply because there are too many possible orthonormal bases: they are infinite indeed. A better question is this one here: if we have two orthonormal bases B_1 and B_2, what kinds of geometrical transformations there are between them?

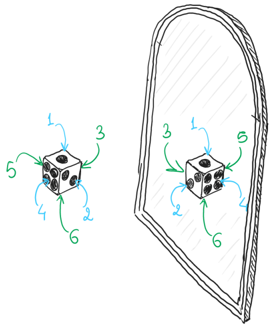


Figure 1.7: Two dice that display chiral properties

It turns out that some special geometrical transformations help to answer this question: translation, rotation and inversion. Imagine we want to move our bumblebee in space. If we move it with a translation, this means that we move it rigidly in one direction, without changing its orientation and without squeezing or deforming it. If we instead rotate the bumblebee, well, we rotate it, so we change its orientation essentially. An inversion is a kind of transformation that is less common in nature. It corresponds to a reflection, like using a flat mirror. It turns out that by using all three of these transformations and combining them we can transform every orthonormal basis into any other. Imagine to have an orthonormal basis $B := \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ and $B' := \{\vec{v}'_1, \vec{v}'_2, \vec{v}'_3\}$ of $\mathbb{V}_3$, the first one with origin in the point P and the second in P'.

The first thing to do, is to translate rigidly B so that the point P overlaps with P'. Then, with a rotation, we can overlap $\vec{v}_1$ with $\vec{v}'_1$ and similarly, with a second rotation we can overlap $\vec{v}_2$ with $\vec{v}'_2$. It turns out that at this point there are only two possibilities: either $\vec{v}_3$ is already overlapping completely $\vec{v}'_3$, or $\vec{v}_3$ has the same direction but opposite orientation of $\vec{v}'_3$. In the first case, we just needed 1 translation and 2 rotations to transform B into B', in the second we need to reflect $\vec{v}_3$ into its opposite arrow, i.e. we need also a reflection.

This suggests dividing orthonormal bases of $\mathbb{V}_3$ into two main categories: in one class we put every orthonormal basis B which is obtained by transforming $\{\hat{\imath}, \hat{\jmath}, \hat{\mathbf{k}}\}$ by only using translations and rotations. In the second class, we put every orthonormal basis B' which is obtained by transforming $\{\hat{\imath}, \hat{\jmath}, \hat{\mathbf{k}}\}$ by using translations and rotations an odd number of inversions. Note that by using an even number of inversions we don't do anything.

This classification can actually be simplified a little bit: the fact that an inversion is required for some of these bases, it's because they do not satisfy the right-hand rule. So, we could simply say that the first class of orthonormal bases satisfy the r.h.r. and the orthonormal bases of the second one do not.

The fact that there are two main classes of orthonormal bases in $\mathbb{V}_3$ has a very interesting effect in nature, called chirality. Look at Picture 1.7. These two dice are fundamentally different because one is the reflection of the other. In nature, things can translate and rotate quite easily but rarely can be inverted. Another way to say that is that there is no way to completely overlap these two dice because they are one the mirror image of the other one!

Chirality plays an important role in chemistry. Some molecules, especially organic molecules, have chiral properties. This means that for the same molecule, there are two different species, called enantiomers, each the perfect reflection of the other and the two enantiomers stay separate: one cannot become the other and vice versa. The two enantiomers can have fundamentally very different effects and properties. For example, some chiral molecules have one enantiomer which is toxic for humans and the other which is actually beneficial. If you are curious, take a look at the Wiki page on chirality.

1.5 The vector product

In the previous sections, we introduced the scalar product, which is an operation $\mathbb{V} \times \mathbb{V} \to \mathbb{R}$. A good question is if there is any binary operation for vector spaces, i.e. operations of the type $\mathbb{V} \times \mathbb{V} \to \mathbb{V}$, which send two vectors to a new vector in a compatible way with the sum and the scalar action. Of course, there are some trivial choices, like the constant map that takes any pair of vectors $(\vec{u}, \vec{v})$ and returns the zero vector $\vec{0}$, but this is not really interesting.

There is an interesting operation, but it is only a special property of the three-dimensional vector space (and any other isomorphic to it). To be honest, the situation is a bit more general: these kinds of operations can be defined only in dimensions 1, 3 and 7. Since we are interested in three-dimensional space, we focus

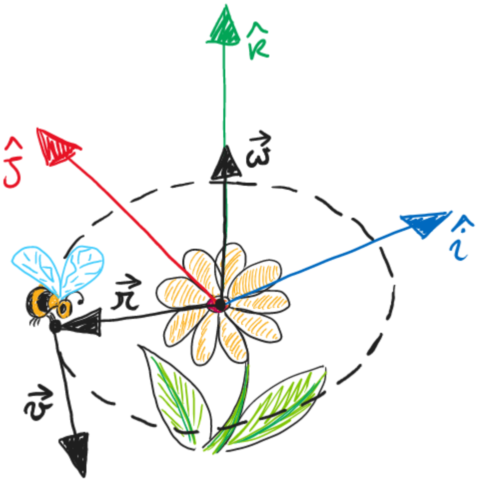


Figure 1.8: A bumblebee who's flying in circle around a flower

our attention only on this case. So, let's start!

We first introduce this operation in a geometric fashion and then we review this from an algebraic point of view. In the following, we do to distinguish anymore between vectors in $\mathbb{R}^3$ and arrows in $\mathbb{V}_3$ and we will simply talk about vectors (see the previous section for a complete discussion about the difference between these two spaces).

Imagine that our bumblebee is moving in a circular motion around a big yellow flower at a constant speed. To be clear, the speed is the modulus of the velocity, which is the vectorial quantity that measures speed, direction and orientation of the motion. So, with the velocity we keep track of the direction, orientation and speed of the bumblebee. This is particularly useful when the bumblebee is moving on a straight line because the only change in the velocity will be the modulus and possibly the orientation.

However, if the bumblebee is moving in a circle, the velocity vector could be a bit annoying, because the direction changes all the time. Let's see how we can define a related quantity, which describes with the minimum amount of information possible the circular motion of the bumblebee around the flower. First, let's take a system of coordinates centred in the centre of the circle and with the vertical vector $\hat{\mathbf{k}}$, pointing upward and in the same direction as the axis of the circle. See Picture 1.8 for reference.

We can start by defining the angular speed $\omega := v/r$, which measures how rapidly the bumblebee moves around. More precisely, ω is the angle, usually in radians, that the bumblebee covers in its fly in a unit of time, 1 second. Ok, ω keeps track of the angular speed, but this is not enough to describe the motion: we are still missing the information about the position in the space of the circle. However, we realize that any circle lies on a plane and that every plane can be completely identified with a point where the plane passes through and a line through that point, which is orthogonal to the plane.

So, by fixing a point and a line we can identify the plane where the circle lies. Finally, we also want to keep track of the orientation of the circular motion: clockwise or anticlockwise. Note that the notion of clockwise orientation depends if we look at the bumblebee from above or from below, so we have to be a bit careful there.

To keep track of the orientation we can use the right-hand rule. We say that the circular motion of the bumblebee is oriented upward if the thumb of our right hand points upward when we curl it in the same direction as the motion of the bumblebee and downward in the opposite case.

So, essentially we need a number ω, a point P, which is the centre of the circle, a line passing through P

which is orthogonal to the plane where the circle lies and an orientation, upward or downward to establish if the bumblebee is moving on one sense or the other. But these are precisely the data of a vector: $|\omega|$ is the modulus, P is the starting point, the line is the direction and the orientation is the orientation! Let $\vec{\omega}$ be this vector and let's call it the angular velocity of the bumblebee. But what's the relationship between the usual velocity $\vec{v}$ and $\vec{\omega}$? To answer this question, we first need to introduce the vector product

Definition 1.13. *Let $\vec{u}, \vec{v} \in \mathbb{V}_3$. The **vector product** between $\vec{u}$ and $\vec{v}$ is the vector $\vec{u} \times \vec{v}$ with modulus equal to:*

$$|\vec{u} \times \vec{v}| := |\vec{u}||\vec{v}||\sin\theta|$$

where θ is the oriented angle from $\vec{u}$ to $\vec{v}$. When $|\vec{u} \times \vec{v}| \neq 0$ (i.e. when $\vec{u}, \vec{v} \neq \vec{0}$ and $\vec{u} \nparallel \vec{v}$), the direction of $\vec{u} \times \vec{v}$ is orthogonal to both $\vec{u}$ and $\vec{v}$ and finally, the orientation of $\vec{u} \times \vec{v}$, whenever $|\vec{u} \times \vec{v}| \neq 0$, is so that the triple $(\vec{u}, \vec{v}, \vec{u} \times \vec{v})$ satisfies the right-hand rule. The vector product is also called the cross product.

We can now express $\vec{\omega}$ as follows:

$$\vec{\omega} := \frac{\vec{r} \times \vec{v}}{|\vec{r}|^2}$$

where $\vec{v}$ is the velocity of the bumblebee and $\vec{r}$ is the vector that indicates the position of the bumblebee. $\vec{r}$ and $\vec{v}$, as vectors, lie both on the plane of the circular orbit. By definition, $\vec{\omega}$ is orthogonal to that plane, so it is also orthogonal to $\vec{r}$ and $\vec{v}$. The modulus of $\vec{\omega}$ is, by definition:

$$\left|\vec{\omega}\right| = \left|\frac{\vec{r} \times \vec{v}}{|\vec{r}|^2}\right| = \frac{|\vec{r} \times \vec{v}|}{|\vec{r}|^2} = \frac{|\vec{r}||\vec{v}|\sin\theta}{|\vec{r}|^2} = \frac{v}{r} = \omega$$

where $v := |\vec{v}|$ is the speed and $r := |\vec{r}|$ is the distance of the bumblebee from the centre and where we used that in a circle $\vec{r} \perp \vec{v}$, then $\theta = \pi/2$ and $\sin\theta = 1$. Let's now compute the vector product of two vectors from the components.

Proposition 1.6. *Let $\vec{u} = \langle x, y, z\rangle$ and $\vec{v} = \langle x', y', z'\rangle$ be two vectors. Then:*

$$\vec{u} \times \vec{v} = \langle yz' - y'z, x'z - xz', xy' - x'y\rangle$$

Proof. to prove this result, we wanna show that the vector $\vec{w} := \langle yz' - y'z, xz' - x'z, xy' - x'y\rangle$ is orthogonal to both $\vec{u}$ and $\vec{v}$ and that it's modulus is $\left|\vec{w}\right| = |\vec{u}||\vec{v}||\sin\theta|$. We don't show that $(\vec{u}, \vec{v}, \vec{w})$ satisfies the r.h.r. because it's a tedious geometric problem but keep in mind that this is also true. To show the orthogonality we can use Proposition 1.5 and show that $\vec{w} \cdot \vec{u} = \vec{w} \cdot \vec{v} = 0$. Let's start with $\vec{w} \cdot \vec{u}$:

$$\begin{aligned} & \vec{w} \cdot \vec{u} = \\ = \quad & \langle yz' - y'z, x'z - xz', xy' - x'y\rangle \cdot \langle x, y, z\rangle = \\ = \quad & (yz' - y'z)x + (x'z - xz')y + (xy' - x'y)z = \\ = \quad & yz'x - y'zx + x'zy - xz'y + xy'z - x'yz = \\ = \quad & 0 \end{aligned}$$

Similarly, we can also show that $\vec{w} \cdot \vec{v} = 0$:

$$\begin{aligned} & \vec{w} \cdot \vec{v} = \\ = \quad & \langle yz' - y'z, x'z - xz', xy' - x'y\rangle \cdot \langle x', y', z'\rangle = \\ = \quad & (yz' - y'z)x' + (x'z - xz')y' + (xy' - x'y)z' = \\ = \quad & yz'x' - y'zx' + x'zy' - xz'y' + xy'z' - x'yz' = \end{aligned}$$

$$= \quad 0$$

So, $\vec{w}\perp\vec{u}$ and $\vec{w}\perp\vec{v}$. Let's compute the modulus of $\vec{w}$. To do that, let's use Proposition 1.5 so $\left|\vec{w}\right|^2 = \vec{w}\cdot\vec{w}$:

$$\begin{aligned}
&\vec{w}\cdot\vec{w} = \\
= \quad &\langle yz' - y'z, x'z - xz', xy' - x'y\rangle \cdot \langle yz' - y'z, x'z - xz', xy' - x'y\rangle = \\
= \quad &(yz' - y'z)^2 + (x'z - xz')^2 + (xy' - x'y)^2 = \\
= \quad &y^2z'^2 + y'^2z^2 - 2yy'zz' + x^2z'^2 + x'^2z^2 - 2xx'zz' + x^2y'^2 + x'^2y^2 - 2xx'yy'
\end{aligned}$$

Let's now take a closer look at $|\vec{u}|^2|\vec{v}|^2 - (\vec{u}\cdot\vec{v})^2$. The trick is to realize that, by using the trigonometry identity $1 - \cos^2\theta = \sin^2\theta$ we can rewrite this as follows:

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

However, by using the components we also obtain:

$$\begin{aligned}
&|\vec{u}|^2|\vec{v}|^2 - (\vec{u}\cdot\vec{v})^2 = \\
= \quad &(x^2 + y^2 + z^2)(x'^2 + y'^2 + z'^2) - (\langle x, y, z\rangle \cdot \langle x', y', z'\rangle)^2 = \\
= \quad &x^2x'^2 + x^2y'^2 + x^2z'^2 + y^2x'^2 + y^2y'^2 + y^2z'^2 + z^2x'^2 + z^2y'^2 + z^2z'^2 - (xx' + yy' + zz')^2 = \\
= \quad &x^2x'^2 + x^2y'^2 + x^2z'^2 + y^2x'^2 + y^2y'^2 + y^2z'^2 + z^2x'^2 + z^2y'^2 + z^2z'^2 + \\
&-x^2x'^2 - y^2y'^2 - z^2z'^2 - 2xx'yy' - 2xx'zz' - 2yy'zz' = \\
= \quad &x^2y'^2 + x^2z'^2 + y^2x'^2 + y^2z'^2 + z^2x'^2 + z^2y'^2 - 2xx'yy' - 2xx'zz' - 2yy'zz'
\end{aligned}$$

So, we proved that $\left|\vec{w}\right|^2 = \vec{w}\cdot\vec{w} = |\vec{u}|^2|\vec{v}|^2 - (\vec{u}\cdot\vec{v})^2 = |\vec{u}|^2|\vec{v}|^2\sin^2\theta$, thus $\left|\vec{w}\right| = |\vec{u}||\vec{v}||\sin\theta|$. □

Now that we have a better understanding of the vector product, we can show the main properties of this operation.

Proposition 1.7. *Consider three vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}_3$ and $a, b \in \mathbb{R}$, then:*

Antisymmetry *For any pairs of vector $\vec{u}, \vec{v} \in \mathbb{V}_3$:*

$$\vec{u}\times\vec{v} = -\vec{v}\times\vec{u}$$

Right linearity *For every $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}_3$ and $a, b \in \mathbb{R}$:*

$$\vec{u}\times(a\vec{v} + b\vec{w}) = a(\vec{u}\times\vec{v}) + b(\vec{u}\times\vec{w})$$

Left linearity *For every $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}_3$ and $a, b \in \mathbb{R}$:*

$$(a\vec{u} + b\vec{v})\times\vec{w} = a(\vec{u}\times\vec{w}) + b(\vec{v}\times\vec{w})$$

Jacobi identity *For every $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}_3$:*

$$\vec{u}\times(\vec{v}\times\vec{w}) + \vec{w}\times(\vec{u}\times\vec{v}) + \vec{v}\times(\vec{w}\times\vec{u}) = \vec{0}$$

Moreover, the vector product is compatible with the scalar product:

$$\begin{aligned}
&\vec{u}\cdot(\vec{v}\times\vec{w})=(\vec{u}\times\vec{v})\cdot\vec{w}\\
&\vec{u}\times(\vec{v}\times\vec{w})=(\vec{u}\cdot\vec{w})\vec{v}-(\vec{u}\cdot\vec{v})\vec{w}
\end{aligned}$$

Proof. Let's show the antisymmetry. So the idea is that, by definition, $(\vec{u},\vec{v},\vec{u}\times\vec{v})$ satisfies the r.h.r.. As pointed out in the discussion that followed Definition 1.8, a triple of vectors $(\vec{u},\vec{v},\vec{w})$ so that $\vec{w}\perp\vec{u}$ and $\vec{w}\perp\vec{v}$ satisfies the r.h.r. or not, there is no other choice. We sometimes say that a triple can be right-handed or left-handed and nothing else. We also said that, if $(\vec{u},\vec{v},\vec{w})$ is right-handed, then $(\vec{u},\vec{v},-\vec{w})$ and $(\vec{v},\vec{u},\vec{w})$ are not and vice versa, if one of these latter is right-handed, then the former is left-handed. So, we have that $(\vec{u},\vec{v},\vec{u}\times\vec{v})$ is right-handed, so $(\vec{v},\vec{u},\vec{u}\times\vec{v})$ is not. But then, $(\vec{v},\vec{u},\vec{u}\times\vec{v})$ is right-handed! But also $(\vec{v},\vec{u},\vec{v}\times\vec{u})$ is right-handed and $|-\vec{u}\times\vec{v}|=|\vec{u}||\vec{v}||\sin\theta|=|\vec{v}\times\vec{u}|$, so $\vec{v}\times\vec{u}=-\vec{u}\times\vec{v}$, as expected.

To prove right linearity we are going to use Proposition 1.6, so let $\vec{u}=\langle x,y,z\rangle$, $\vec{v}=\langle x',y',z'\rangle$ and $\vec{w}=\langle x'',y'',z''\rangle$. Then we have:

$$\begin{aligned}
&\vec{u}\times(a\vec{v})=\\
=\ &\langle x,y,z\rangle\times\langle ax',ay',az'\rangle=\\
=\ &\langle yaz'-y'az,x'az-xaz',xay'-x'ay\rangle=\\
=\ &\langle a(yz'-y'z),a(x'z-xz'),a(xy'-x'y)\rangle=\\
=\ &a\langle yz'-y'z,x'z-xz',xy'-x'y\rangle=\\
=\ &a(\vec{u}\times\vec{v})
\end{aligned}$$

Moreover:

$$\begin{aligned}
&\vec{u}\times(\vec{v}+\vec{w})=\\
=\ &\langle x,y,z\rangle\times\langle x'+x'',y'+y'',z'+z''\rangle=\\
=\ &\langle y(z'+z'')-(y'+y'')z,(x'+x'')z-x(z'+z''),x(y'+y'')-(x'+x'')y\rangle=\\
=\ &\langle yz'+yz''-y'z-y''z,x'z+x''z-xz'-xz'',xy'+xy''-x'y-x''y\rangle=\\
=\ &\langle (yz'-y'z)+(yz''-y''z),(x'z-xz')+(x''z-xz''),(xy'-x'y)+(xy''-x''y)\rangle=\\
=\ &\langle yz'-y'z,x'z-xz',xy'-x'y\rangle+\langle yz''-y''z,x''z-xz'',xy''-x''y\rangle=\\
=\ &\vec{u}\times\vec{v}+\vec{u}\times\vec{w}
\end{aligned}$$

To show left linearity we can just use right linearity and the antisymmetry as follows:

$$\begin{aligned}
&(a\vec{u}+b\vec{v})\times\vec{w}=\\
=\ &-\vec{w}\times(a\vec{u}+b\vec{v})=\\
=\ &-(a(\vec{w}\times\vec{u})+b(\vec{w}\times\vec{v}))=\\
=\ &-a(\vec{w}\times\vec{u})-b(\vec{w}\times\vec{v})=\\
=\ &a(\vec{u}\times\vec{w})+b(\vec{v}\times\vec{w})
\end{aligned}$$

We are not going to show the Jacobi identity, simply because the proof is very tedious, but not hard. The Jacobi identity is not a property that we are going to use in this course, however, it is conceptually very important because this makes $(\mathbb{V}_3,\times)$ into a Lie algebra. We'll dedicate two words to this after this proof. Let's keep going. We are now gonna prove the first one and we leave it to the reader to show the second one:

$$\begin{aligned}
&\vec{u}\cdot(\vec{v}\times\vec{w})=\\
=\ &\langle x,y,z\rangle\cdot\langle y'z''-y''z',x''z'-x'z'',x'y''-x''y'\rangle=\\
=\ &xy'z''-xy''z'+yx''z'-yx'z''+zx'y''-zx''y'=
\end{aligned}$$

$$
\begin{aligned}
&= (yz' - y'z)x'' + (x'z - xz')y'' + (xy' - x'y)z'' = \\
&= \langle yz' - y'z, x'z - xz', xy' - x'y\rangle \cdot \langle x'', y'', z''\rangle = \\
&= (\vec{u} \times \vec{v}) \cdot \vec{w}
\end{aligned}
$$

as expected. □

Cool Stuff* 1.4. Proposition 1.7 establishes that the vector product satisfies antisymmetry, left linearity, right linearity, and the so-called Jacobi identity. This is (probably) the first time you encounter a cool mathematical animal: a Lie algebra Tadaaaa!!!

A Lie algebra is a set $\mathfrak{g}$ equipped with a binary operation, often denoted by $[,]$, called Lie bracket, which is antisymmetric, bilinear, i.e. left and right linear, and that satisfies the Jacobi identity. So, the vector product is a Lie bracket and $(\mathbb{V}_3, \times)$ is a Lie algebra. Lie algebras are important in group theory and geometry.

We already talked about groups, but we didn't talk about a special class of groups, called Lie groups. We cannot go into details because this topic is quite advanced, but the idea is that a Lie group is a geometrical space, called a smooth manifold, equipped with a group structure. It turns out that every Lie group has an associated Lie algebra.

The interpretation of these two objects sounds like this: a group contains the symmetries of a system, which are transformations that leave the system invariant. For example, rotating a cube of $90°$ on one face is a symmetry of a cube. The Lie algebra associated with a Lie group contains the infinitesimal symmetries of the system. In physics, these two concepts play a fundamental role in describing the symmetries of many things! They are for example used to describe the standard model, which is the theory that we currently use to describe the subatomic particles. ■

So far, we have seen the algebraic properties of the vector product. Now, we want to give a geometric interpretation of this operation. We start by noticing that if $\vec{u}$ and $\vec{v}$ are parallel vectors, then $\vec{u} \times \vec{v} = \vec{0}$.

Proposition 1.8. *Suppose that $\vec{u}$ and $\vec{v}$ are non-zero vectors in $\mathbb{V}_3$. Then, $\vec{u} \times \vec{v} = \vec{0}$ if and only if $\vec{u}$ and $\vec{v}$ are parallel.*

Proof. Suppose first that $\vec{u} \parallel \vec{v}$, then the angle θ between them is 0 or π. In both cases, $\sin\theta = 0$, thus $|\vec{u} \times \vec{v}| = 0$, so $\vec{u} \times \vec{v} = \vec{0}$. Let's now suppose that $\vec{u} \times \vec{v} = \vec{0}$, then $|\vec{u}||\vec{v}||\sin\theta| = 0$, but because we assumed that $\vec{u}, \vec{v} \neq \vec{0}$, we conclude that $\sin\theta = 0$. Since $-\pi \leq \theta < \pi$, we conclude that $\theta = 0$ or $\theta = \pi$, which means that $\vec{u} \parallel \vec{v}$. □

We can see Proposition 1.8 as the analogous of Proposition 1.5, for the vector product. So, $\vec{u} \cdot \vec{v} = 0$ means that $\vec{u} \perp \vec{v}$ and $\vec{u} \times \vec{v} = \vec{0}$ that $\vec{u} \parallel \vec{v}$. Before we continue, let's quickly evaluate the vector products of $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$.

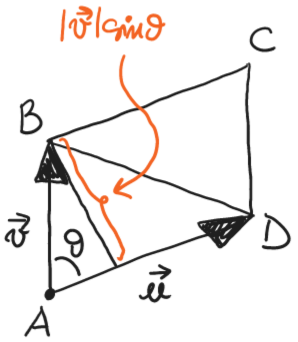


Figure 1.9: The area of the parallelogram given by two vectors and its relation with the vector product

---

Proposition 1.9. *The following formulas work:*

$$\begin{aligned}
&\hat{\imath} \times \hat{\jmath} = \hat{k}\\
&\hat{k} \times \hat{\imath} = \hat{\jmath}\\
&\hat{\jmath} \times \hat{k} = \hat{\imath}\\
&\hat{\jmath} \times \hat{\imath} = -\hat{k}\\
&\hat{\imath} \times \hat{k} = -\hat{\jmath}\\
&\hat{k} \times \hat{\jmath} = -\hat{\imath}\\
&\hat{\imath} \times \hat{\imath} = \hat{\jmath} \times \hat{\jmath} = \hat{k} \times \hat{k} = \vec{0}
\end{aligned}$$

Proof. To prove it, note that the angle from $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ and from $\hat{\mathbf{i}}$ and $\hat{\mathbf{k}}$ is $\pi/2$, and $\sin(\pi/2) = 1$. Moreover, recall that $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ are unit vectors. So, $\left|\hat{\mathbf{i}} \times \hat{\mathbf{j}}\right| = 1$. Moreover, $\hat{\mathbf{k}}$ is orthogonal to both $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ and $(\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}})$, by definition, satisfies the r.h.r., thus $\hat{\mathbf{k}} = \hat{\mathbf{i}} \times \hat{\mathbf{j}}$. Similarly, one can show that $\hat{\mathbf{k}} \times \hat{\mathbf{i}} = \hat{\mathbf{j}}$ and that $\hat{\mathbf{j}} \times \hat{\mathbf{k}} = \hat{\mathbf{i}}$. To show the next three equations, remember that $\times$ is antisymmetric, thus $\hat{\mathbf{j}} \times \hat{\mathbf{i}} = -\hat{\mathbf{i}} \times \hat{\mathbf{j}} = -\hat{\mathbf{k}}$ and similarly, for $\hat{\mathbf{i}} \times \hat{\mathbf{k}} = -\hat{\mathbf{j}}$ and $\hat{\mathbf{k}} \times \hat{\mathbf{j}} = -\hat{\mathbf{i}}$. Finally, $\hat{\mathbf{i}} \parallel \hat{\mathbf{i}}, \hat{\mathbf{j}} \parallel \hat{\mathbf{j}}$ and $\hat{\mathbf{k}} \parallel \hat{\mathbf{k}}$, so by Proposition 1.8, they are equal to zero. □

Proposition 1.10. *Consider two vectors $\vec{u} = \langle x, y, z\rangle$ and $\vec{v} = \langle x', y', z'\rangle$ in $\mathbb{V}_3$ and let A, B, C and D the points $A = \vec{0}$, i.e. the origin, $B = \vec{u}$, i.e. the endpoint of $\vec{u}$, $C = \vec{u} + \vec{v}$, i.e. the endpoint of $\vec{u} + \vec{v}$, and $D = \vec{v}$, i.e. the endpoint of $\vec{v}$. Then the area of the parallelogram $ABCD$ is equal to the modulus of the vector product between $\vec{u}$ and $\vec{v}$, i.e. $|\vec{u} \times \vec{v}|$.*

Proof. Take a look at Picture 1.9. As you can see, $|\vec{v}||\sin\theta|$ corresponds precisely to the height of the triangle ABD, so $|\vec{u}||\vec{v}||\sin\theta|/2$ is equal to the area of the triangle ABD. However, the area of $ABCD$ is exactly equal to the double of the area of ABD, so the area of $ABCD$ is $|\vec{u}||\vec{v}||\sin\theta| = |\vec{u} \times \vec{v}|$. □

Proposition 1.10 shows the geometrical meaning of the vector product.

Example 1.5. Consider the triangle with vertexes $A = \langle 1, 0, 0\rangle, B = \langle 0, 1, 0\rangle$ and $D = \langle 0, 0, 1\rangle$. One way to compute the area of this triangle is to use the vector product. First, we need to write two of the three sides of the triangle as vectors. But note that the vector from A to B is precisely given by $\vec{u} := B - A = \langle -1, 1, 0\rangle$. To be precise, $\langle -1, 1, 0\rangle$ is the vector that starts from the origin so that, translated to A, will end in B. Because translation doesn't affect the area we can use this trick and do the same for the side AD. So, we also have the vector $\vec{v} := D - A = \langle -1, 0, 1\rangle$. Let's now evaluate the vector product between $\vec{u}$ and $\vec{v}$, so we have:

$$\vec{u} \times \vec{v} = \langle yz' - y'z, x'z - xz', xy' - x'y\rangle = \langle 1, 1, 1\rangle$$

and its modulus is $\sqrt{3}$. This is the area of the parallelogram $ABCD$, where $C = \vec{u} + \vec{v}$. To find the area of ABD we just have to divide by 2, so we obtain $\sqrt{3}/2$. We could have also argued geometrically and concluded that the triangle ABD is an equilateral triangle with sides long $\sqrt{2}$. So, the height of ABD is equal to $\sqrt{\sqrt{2}^2 - \sqrt{2}^2/2} = \sqrt{3/2}$. Therefore, the area of ABD is $(\sqrt{2} \cdot \sqrt{3/2})/2 = \sqrt{3}/2$, in agreement with our result. ■

Cool Stuff* 1.5. There is a third way to see the vector product. Let's define the so-called Levi-Civita symbol $\varepsilon_{i,j,k}$. First, i, j and k are three indexes that run from 1 to 3. A permutation is a bijective function over a finite set. This means that a permutation is a transformation that shuffles the elements of a set, without creating duplicates or cancelling any element.

Now, every permutation can be obtained by swapping two elements at a time. Let's do an example. Suppose that we shuffle $(1, 2, 3)$ and obtain $(3, 1, 2)$. To do that, we can first swap 1 with 3, so we now have $(3, 2, 1)$ and then we can swap 2 with 1, so we have $(3, 1, 2)$. Now, interestingly, to permute a set you can use different combinations of swapping, but the number of swaps required is always even or odd.

This means that there are only two classes of permutations: the even permutations are the permutations obtained with an even number of swaps and the odd permutations are the ones obtained with an odd number of swaps. So, now we can define the Levi-Civita symbol as follows:

$$\varepsilon_{i,j,k} := \begin{cases} 1 & \text{if } (i, j, k) \text{ is an even permutation of } (1, 2, 3) \\ -1 & \text{if } (i, j, k) \text{ is an odd permutation of } (1, 2, 3) \\ 0 & \text{if } (i, j, k) \text{ is not a permutation of } (1, 2, 3) \end{cases}$$

Now, the even permutations of $(1, 2, 3)$ are $(1, 2, 3), (2, 3, 1)$ and $(3, 1, 2)$, the odd permutations are $(3, 2, 1), (1, 3, 2)$ and $(2, 1, 3)$ and saying that (i, j, k) is not a permutation of $(1, 2, 3)$ is the same as saying that at least two of 1, j and k are equal, so here are all the possible combinations:

(i, j, k)	Parity	$\varepsilon_{i,j,k}$
$(1, 2, 3)$	Even	1
$(2, 3, 1)$	Even	1
$(3, 1, 2)$	Even	1
$(3, 2, 1)$	Odd	-1
$(1, 3, 2)$	Odd	-1
$(2, 1, 3)$	Odd	-1
$(1, 1, 1)$	Not Permutation	0
$(1, 1, 2)$	Not Permutation	0
$(1, 2, 1)$	Not Permutation	0
$(2, 1, 1)$	Not Permutation	0
$(1, 2, 2)$	Not Permutation	0
$(2, 1, 2)$	Not Permutation	0
$(2, 2, 1)$	Not Permutation	0
$(2, 2, 2)$	Not Permutation	0

(i, j, k)	Parity	$\varepsilon_{i,j,k}$
$(1, 1, 3)$	Not Permutation	0
$(1, 3, 1)$	Not Permutation	0
$(3, 1, 1)$	Not Permutation	0
$(1, 3, 3)$	Not Permutation	0
$(3, 1, 3)$	Not Permutation	0
$(3, 3, 1)$	Not Permutation	0
$(2, 2, 3)$	Not Permutation	0
$(2, 3, 2)$	Not Permutation	0
$(3, 2, 2)$	Not Permutation	0
$(2, 3, 3)$	Not Permutation	0
$(3, 2, 3)$	Not Permutation	0
$(3, 3, 2)$	Not Permutation	0
$(3, 3, 3)$	Not Permutation	0

Now that we know that is the Levi-Civita symbol we can see that, if $\vec{u} = \langle u_1, u_2, u_3\rangle$ and $\vec{v} = \langle v_1, v_2, v_3\rangle$, then:

$$\vec{u} \times \vec{v} = \langle \sum_{j,k=1}^{3} \varepsilon_{1,j,k} u_i v_j, \sum_{j,k=1}^{3} \varepsilon_{2,j,k} u_i v_j, \sum_{j,k=1}^{3} \varepsilon_{3,j,k} u_i v_j \rangle$$

■

The vector product is also related to matrix theory. Appendix 5.1 is dedicated to exploring some introductory topics of this theory. Suppose that $\vec{u} = \langle x, y, z\rangle$ and $\vec{v} = \langle x', y', z'\rangle$. Then, consider the matrix:

$$\begin{pmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ x & y & z \\ x' & y' & z' \end{pmatrix}$$

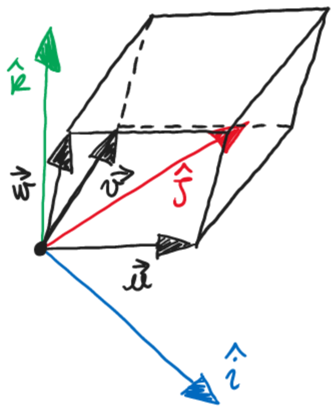


Figure 1.10: A parallelepiped associated with three vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$

Then if we compute the determinant of this matrix we obtain:

$$\begin{aligned}
&\begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ x & y & z \\ x' & y' & z' \end{vmatrix} = \\
= \;& \begin{vmatrix} y & z \\ y' & z' \end{vmatrix} \hat{\mathbf{i}} - \begin{vmatrix} x & z \\ x' & z' \end{vmatrix} \hat{\mathbf{j}} + \begin{vmatrix} x & y \\ x' & y' \end{vmatrix} = \\
= \;& (yz' - y'z)\hat{\mathbf{i}} - (xz' - x'z)\hat{\mathbf{j}} + (xy' - x'y)\hat{\mathbf{k}} = \\
= \;& (yz' - y'z)\hat{\mathbf{i}} + (x'z - xz')\hat{\mathbf{j}} + (xy' - x'y)\hat{\mathbf{k}} = \\
= \;& \vec{u} \times \vec{v}
\end{aligned}$$

so we have a useful formula.

Proposition 1.11. *Given $\vec{u} = \langle x, y, z \rangle$ and $\vec{v} = \langle x', y', z' \rangle$, the vector product between $\vec{u}$ and $\vec{v}$ is given by the following formula:*

$$\vec{u} \times v = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ x & y & z \\ x' & y' & z' \end{vmatrix}$$

As a consequence of this, the scalar triple product is related to the determinant of 3×3 matrix. Let's introduce this operation, first.

Definition 1.14. *The scalar triple product of three vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{V}_3$ is the real number $\vec{u} \cdot (\vec{v} \times \vec{w})$.*

Now, using Proposition 1.11, we obtain that:

$$\begin{aligned}
&\vec{u} \cdot (\vec{v} \times \vec{w}) = \\
= \;& \langle x, y, z \rangle \cdot \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ x' & y' & z' \\ x'' & y'' & z'' \end{vmatrix} = \\
= \;& \langle x, y, z \rangle \cdot \left\langle \begin{vmatrix} y' & z' \\ y'' & z'' \end{vmatrix}, - \begin{vmatrix} x' & z' \\ x'' & z'' \end{vmatrix}, \begin{vmatrix} x' & y' \\ x'' & y'' \end{vmatrix} \right\rangle =
\end{aligned}$$

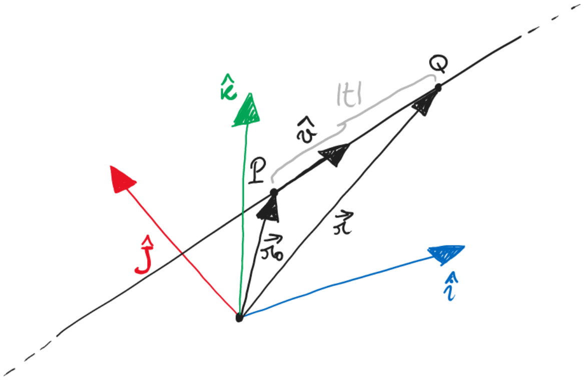


Figure 1.11: Equation of a line in three-dimensional space

$$
\begin{aligned}
&= x\begin{vmatrix} 'y & z' \\ y'' & z'' \end{vmatrix} - y\begin{vmatrix} x' & z' \\ x'' & z'' \end{vmatrix} + z\begin{vmatrix} x' & y' \\ x'' & y'' \end{vmatrix} = \\
&= \begin{vmatrix} x & y & z \\ x' & y' & z' \\ x'' & y'' & z'' \end{vmatrix}
\end{aligned}
$$

Proposition 1.12. *Let $\vec{u} = \langle x, y, z\rangle, \vec{v} = \langle x', y', z'\rangle$ and $\vec{w} = \langle x'', y'', z''\rangle$ be three vectors in $\mathbb{V}_3$. Then:*

$$
\vec{u} \cdot (\vec{v} \times \vec{w}) = \begin{vmatrix} x & y & z \\ x' & y' & z' \\ x'' & y'' & z'' \end{vmatrix}
$$

Moreover, $\left|\vec{u} \cdot (\vec{v} \times \vec{w})\right|$ is the volume of the parallelepiped described by $\vec{u}$, $\vec{v}$ and $\vec{w}$, see Picture 1.10.

1.6 Geometry in 3d space

1.6.1 Lines

It is time to introduce some geometry! The first thing we wanna do is to see how to describe straight lines in space. We already know that a line is completely described by a point P where it passes through and a versor, which indicates its direction.

Let $\hat{v}$ be a versor and P the endpoint of a vector $\vec{r}_0$ starting from the origin. Consider any point Q along the line. Suppose that the distance between Q and P is $|t|$ and we decide that t is positive when the vector $\overrightarrow{PQ}$ has the same orientation of the versor $\hat{v}$ and negative if the orientation is opposite. So, from Picture 1.11 we can see that the vector $\vec{r}_0 + t\hat{v}$ ends precisely in Q. This means that every point on the line can be expressed in the form $\vec{r}_0 + t\hat{v}$, for some value of $t \in \mathbb{R}$.

So, there we go! We have the **vector equation** of the line:

$$
\vec{r} = \vec{r}_0 + t\hat{v}
$$

Now, suppose that the components of $\vec{r}_0$ and $\hat{v}$ are the following: $\vec{r}_0 = \langle x_0, y_0, z_0\rangle$, $\hat{v} = \langle a, b, c\rangle$, then the equation for the generic point $\vec{r} = \langle x, y, z\rangle$ on the line becomes:

$$
\langle x, y, z\rangle = \langle x_0, y_0, z_0\rangle + t\langle a, b, c\rangle = \langle x_0 + at, y_0 + bt, z_0 + ct\rangle
$$

This is equivalent to a system of three equations:

$$\begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = z_0 + ct \end{cases}$$

Here, we have three variables, x, y, z, and 1 parameter, t. x_0, y_0, z_0, a, b and c are all given data, so they will be real numbers. What we can do is to remove the parameter, as follows. First, note that because $\hat{v}$ is a versor, i.e. $|v| = 1$, at least one of the three components a, b or c has to be non-zero otherwise, $\hat{v}$ would be the zero vector.

Suppose that $a \neq 0$. Then, by solving for t the first equation we find that $t = (x - x_0)/a$. So, now we can replace t with this expression in the other two equations and obtain:

$$\begin{cases} y = y_0 + b\frac{x-x_0}{a} \\ z = z_0 + c\frac{x-x_0}{a} \end{cases}$$

Multiplying both by a we obtain:

$$\begin{cases} ay = ay_0 + b(x - x_0) \\ az = az_0 + c(x - x_0) \end{cases}$$

and finally reordering we have:

$$\begin{cases} ay - bx = ay_0 - bx_0 \\ az - cx = az_0 - cx_0 \end{cases}$$

If $b \neq 0$ and we can choose $t = (y - y_0)/b$ and obtain:

$$\begin{cases} ay - bx = ay_0 - bx_0 \\ bz - cy = bz_0 - cy_0 \end{cases}$$

and if $c \neq 0$, and choosing $t = (z - z_0)/c$:

$$\begin{cases} ay - bx = ay_0 - bx_0 \\ bz - cy = bz_0 - cy_0 \end{cases}$$

Finally, if all $a, b, c \neq 0$, then we can write the **symmetric equations**:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

since each of these three terms is equal to t.

Example 1.6. Consider the point $P = (1, -2, 3)$ and the point $Q = (5, 2, -3)$. These two points are distinct since their coordinates don't match. Therefore, there is only a single line which passes through both. We want to find the equation of this line. Let's start with the vector equation. We already know that the line passes through P, we need to find the versor $\hat{v}$. Notice that, we decided to choose $\hat{v}$ to be a unit vector, i.e. a versor, but we could have taken any non-zero vector. Let $\vec{r}_0$ and $\vec{r}_1$ denote the vectors which point to P and Q, respectively. So, $\vec{r}_0 = \langle 1, -2, 3 \rangle$ and $\vec{r}_1 = \langle 5, 2, -3 \rangle$. Now, consider the vector $\vec{v} := \vec{r}_1 - \vec{r}_0$. By drawing this vector you realize that $\vec{v}$ is parallel to the line we are looking for!

So, we have:

$$\vec{v} = \vec{r}_1 - \vec{r}_0 = \langle 5, 2, -3 \rangle - \langle 1, -2, 3 \rangle = \langle 4, 4, -6 \rangle$$

Therefore, the vector equation of the line reads as follows:

$$\vec{r} = \langle 1, -2, 3 \rangle + t\langle 4, 4, -6 \rangle = \langle 1 + 4t, -2 + 4t, 3 - 6t \rangle$$

Let's now find the parametric equation. First, let's denote the coordinates of $\vec{r}$ by $x(t)$, $y(t)$, and $z(t)$. So, we can rewrite:

$$\begin{cases} x(t) = 1 + 4t \\ y(t) = -2 + 4t \\ z(t) = 3 - 6t \end{cases}$$

Let's now eliminate the parameter t. Notice that we can solve the first equation by t as follows:

$$t = \frac{x-1}{4}$$

Thus:

$$y = -2 + 4\frac{x-1}{4} = 2 + x - 1 = x + 1 \qquad z = 3 - 6t = 3 - 6\frac{x-1}{4} = 3 - \frac{3}{2}(x-1)$$

So, we can rewrite:

$$\begin{cases} y = x + 1 \\ z = 3 = \frac{3}{2}(x-1) \end{cases}$$

So, we expressed the equation of the line in three ways: using the vector form, the parametric form, and the system form. ■

Note that all of these are equivalent ways to describe a line in space. We'll have a better understanding of why we need a system of two equations to describe a line once we introduce planes. But first, let's talk about segments.

1.6.2 Segments

A segment is a connected part of a line of finite length, between two extreme points. Suppose that $\vec{r}_0 = \langle x_0, y_0, z_0 \rangle$ and $\vec{r}_1 = \langle x_1, y_1, z_1 \rangle$ are the two extreme points of the segment we are interested in. We also suppose that $\vec{r}_0$ and $\vec{r}_1$ are not equal, otherwise, we will just have a single point. The equation of the line that passes through both $\vec{r}_0$ and $\vec{r}_1$ is given by:

$$\vec{r} = \vec{r}_0 + t\vec{v}$$

where $\vec{v}$ is a vector with the same direction as this line. Now, note that before we used a versor, but the reality is that we could have used any non-zero vector. So, we just need to find a good choice for $\vec{v}$. However, $\vec{r}_1 - \vec{r}_0$ is exactly the vector we need. Note also that, because $\vec{r}_1 \neq \vec{r}_0$, we also have that $\vec{r}_1 - \vec{r}_0 \neq \vec{0}$. So, the equation now becomes:

$$\vec{r} = \vec{r}_0 + t(\vec{r}_1 - \vec{r}_0) = (t-1)\vec{r}_0 + t\vec{r}_1$$

However, this is the equation of the line, not of the segment. So to limit $\vec{r}$ between $\vec{r}_0$ and $\vec{r}_1$, we just need to ask t to be between 0 and 1, so we have:

$$\vec{r} = (t-1)\vec{r}_0 + t\vec{r}_1 \qquad 0 \leq t \leq 1$$

1.6.3 Planes

A plane is completely specified by a normal vector $\vec{n} = \langle a, b, c \rangle$, which is a non-zero vector that is orthogonal to the plane, and a point $\vec{r}_0 = \langle x_0, y_0, z_0 \rangle$ where the plane passes through. This means that if $\vec{v}$ is any vector which lies on the plane, then $\vec{n} \perp \vec{v}$, i.e. $\vec{n} \cdot \vec{v} = 0$. But given any point $\vec{r} = \langle x, y, z \rangle$ on the plane, the vector $\vec{r} - \vec{r}_0$ lies on the plane so we obtain:

$$\vec{n} \cdot (\vec{r} - \vec{r}_0) = 0$$

Now, using the components, we obtain:

$$\begin{aligned} & 0 = \langle a, b, c \rangle \cdot (\langle x, y, z \rangle - \langle x_0, y_0, z_0 \rangle) = \\ = \quad & = \langle a, b, c \rangle \cdot \langle x - x_0, y - y_0, z - z_0 \rangle = \\ = \quad & a(x - x_0) + b(y - y_0) + c(z - z_0) \end{aligned}$$

Reorganizing the terms we have the equation:

$$ax + by + cz = d$$

where $d := ax_0 + by_0 + cz_0$. So, the equation of a plane in space is precisely a first-order polynomial equation in three variables. Note that this generalizes the fact that in two dimensions the equation of a line is a polynomial equation in two variables.

Notice that to define a line, we needed a system of two first-order polynomial equations. Geometrically, each of these equation represents a plane where the line lies. Taking the system of two equations means taking the intersection of the corresponding geometrical figures, so there we go: a line is an intersection of two non-parallel planes. This also clarifies why we had different systems to represent the same line: there are infinite pairs of non-parallel planes that intersect precisely along the same line.

Example 1.7. Euclid established, as a fundamental postulate of his plane geometry, that from two distinct points passes a unique line. In three dimensions, we can say that from three distinct, not aligned (i.e. not on the same line) points, passes a unique plane. To see that, suppose that $A = \langle x_0, y_0, z_0 \rangle, B = \langle x_1, y_1, z_1 \rangle$ and $C = \langle x_2, y_2, z_2 \rangle$ are three distinct, not aligned points.

Therefore, the vectors $\vec{u} := B - A = \langle x_1 - x_0, y_1 - y_0, z_1 - z_0 \rangle$ and $\vec{v} := C - A = \langle x_2 - x_0, y_2, y_0, z_2 - z_0 \rangle$ are non-zero non-parallel vectors that lie on the plane which passes through A, B and C. Note also that $\vec{u}$ and $\vec{v}$ starts both from the same point A, so to find a normal vector to the plane, we can find a vector that is orthogonal to both $\vec{u}$ and $\vec{v}$. A good candidate is $\vec{u} \times \vec{v}$, so:

$$\begin{aligned} \vec{n} \quad & := \vec{u} \times \vec{v} = \\ = \quad & \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2, y_0 & z_2 - z_0 \end{vmatrix} = \\ = \quad & \langle (y_1 - y_0)(z_2 - z_0) - (y_2 - y_0)(z_1 - z_0), \\ & (x_2 - x_0)(z_1 - z_0) - (x_1 - x_0)(z_2 - z_0), \\ & (x_1 - x_0)(y_2 - y_0) - (x_2 - x_0)(y_1 - y_0) \rangle \end{aligned}$$

You don't need to remember this horrible formula, instead, it's important you understand the procedure and you are able to apply it in concrete cases. Once we find the components of $\vec{n} = \langle a, b, c \rangle$, then we know that the equation of the plane will be of the form $ax + by + cz = d$, for some number $d \in \mathbb{R}$. To find d, it suffices to plug in the coordinates of one of the three points and conclude that $d = ax_0 + bx_0 + cz_0$. ■

Example 1.8. In this example, we apply the abstract procedure of Example 1.7 on a concrete case. We want to determine the equation of a plane passing through the points $A = \langle 1, 0, 0\rangle$, $B = \langle 0, -1, 0\rangle$ and $C = \langle 1/2, 0, 1/2\rangle$. The idea is to define first the vectors $\vec{u} = B - A = \langle -1, -1, 0\rangle$ and $\vec{v} = C - A = \langle -1/2, 0, 1/2\rangle$. These are the vectors that start from A and end in B and C, respectively.

The next step is to define a normal vector to the plane. But because both $\vec{u}$ and $\vec{v}$ lie on the plane, such normal vector, that we call $\vec{n}$, needs to be orthogonal to both $\vec{u}$ and $\vec{v}$, therefore a good definition is to take the vector product between $\vec{u}$ and $\vec{v}$:

$$\vec{n} = \vec{u} \times \vec{v} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ -1 & -1 & 0 \\ -1/2 & 0 & 1/2 \end{vmatrix} = -\frac{1}{2}\hat{\mathbf{i}} + \frac{1}{2}\hat{\mathbf{j}} - \frac{1}{2}\hat{\mathbf{k}}$$

So, $\vec{n} = \langle -1/2, 1/2, -1/2\rangle$. But we know that the components of $\vec{n}$ give the a, b, c coefficients in the equation of the plane. So we have:

$$-1/2x + 1/2y - 1/2z = d$$

Finally, $d = -1/2x_0 + 1/2y_0 - 1/2z_0$, where $A = \langle x_0, y_0, z_0\rangle$ is the point where $\vec{n}$ starts from. So, $d = -1/2$ and therefore:

$$-1/2x + 1/2y - 1/2z = -1/2$$

which we can also rewrite as:

$$x - y + z = 1$$

where we just multiplied on both sides by -2.

Alternatively, without using the vectors, we could have simply plugged in the coordinates of A, B and C into the equation $ax + by + cz = d$ and solved the system:

$$\begin{cases} a = d \\ -b = d \\ 1/2a + 1/2c = d \end{cases}$$

Therefore, we can rewrite the equation as:

$$dx - dy + dz = d$$

d cannot be 0, otherwise $ax + by + cz$ wouldn't be a first-order polynomial, therefore, we can divide by d and obtain exactly $x - y + z = 1$, as expected. ∎

Example 1.9. Consider the unit sphere $\mathbb{S}^2$ in the three-dimensional space. This is the set of points distant 1 from the origin. The superscript 2 indicates that $\mathbb{S}^2$ is a two-dimensional surface. So, the equation that defines $\mathbb{S}^2$ is:

$$\mathbb{S}^2 := \{\langle x, y, z\rangle \in \mathbb{R}^3, x^2 + y^2 + z^2 = 1\}$$

Note that, given a point $P = \langle x_0, y_0, z_0\rangle$, the square distance between the generic point $Q = \langle x, y, z\rangle$ from P is precisely $\vec{u} \cdot \vec{u}$, where $\vec{u} = Q - P$, so the square distance between Q and P is $(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2$. Now, when P is the origin, $x_0 = y_0 = z_0 = 0$, so the set of points distant 1 from the origin is precisely given by the condition $x^2 + y^2 + z^2 = 1^2 = 1$.

Ok, now that we know what is the unit sphere, we wonder: how can we define the tangent plane of the unit sphere on a generic point Q of $\mathbb{S}^2$? In the third chapter, we will have a better understanding of how to find the equation of the tangent plane for a generic two-dimensional surface, but for now, we don't have the tools to describe this concept in full generality. So, we need to use some geometric intuition of what is going on here. It is not too hard to believe that if $Q = \langle x_1, y_1, z_1\rangle$ is a point on $\mathbb{S}^2$, the tangent plane of $\mathbb{S}^2$ in Q will have a normal vector parallel to the vector $\langle x_1, y_1, z_1\rangle$. This is a special property of the sphere, so this is a rather ad hoc construction. Therefore, the equation of the tangent plane of $\mathbb{S}^2$ in Q will be:

$$x_1 x + y_1 y + z_1 z = d$$

To find d, we just need to plug-into the equation the components of the point Q, so we have:

$$d = x_1 x_1 + y_1 y_1 + z_1 z_1 = x_1^2 + y_1^2 + z_1^2 = 1$$

where we used that Q is a point of the sphere and then $x_1^2 + y_1^2 + z_1^2 = 1$. Thus we find the equation of the tangent plane of the unit sphere in a generic point $Q = \langle x_1, y_1, z_1\rangle$ of $\mathbb{S}^2$ to be:

$$x_1 x + y_1 y + z_1 z = 1$$

Let's do a concrete example: let $Q = \langle 1, 0, 0\rangle = \hat{\mathbf{i}}$, then we are expecting the tangent plane to be the plane $\hat{\mathbf{j}}\hat{\mathbf{k}}$ translated in front of $\hat{\mathbf{i}}$. According to the formula we have found we obtain:

$$x = 1$$

which is precisely the equation of that plane. ■

1.6.4 Distance between a plane and a point

Suppose we want to measure the distance between a plane and a point $Q = \langle x_0, y_0, z_0\rangle$. The plane is given by the equation $ax + by + cz = d$ and the vector $\vec{n} = \langle a, b, c\rangle$, is orthogonal to the plane. Now, consider the generic point $P = \langle x_1, y_1, z_1\rangle$ on the plane and let $\vec{v} = Q - P = \langle x_0 - x_1, y_0 - y_1, z_0 - z_1\rangle$ be the vector that starts in P and ends in Q. Now, the distance between the plane and the point Q is precisely, equal to the modulus of the vector projection of the vector $\vec{v}$ along the normal vector $\vec{n}$, so we have:

$$\begin{aligned} D = |\mathrm{proj}_{\vec{n}}\vec{v}| = \frac{|\vec{v}\cdot\vec{n}|}{|\vec{n}|} &= \\ = \frac{|\langle x_0 - x_1, y_0 - y_1, z_0 - z_1\rangle \cdot \langle a, b, c\rangle|}{\sqrt{a^2+b^2+c^2}} &= \\ = \frac{|ax_0 + by_0 + cz_0 - ax_1 - by_1 - cz_1|}{\sqrt{a^2+b^2+c^2}} & \end{aligned}$$

However, $P = \langle x_1, y_1, z_1\rangle$ is a point of the plane, thus $ax_1 + by_1 + cz_1 = d$, thus we obtain the final equation:

$$D = \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2+b^2+c^2}}$$

This gives a formula for the distance between a generic point Q and a plane.

1.6.5 Cylinders & quadric surfaces

In this section we briefly explore two classes of two-dimensional surfaces. The first one is the class of the so-called cylinders. A cylinder is a surface that can be obtained in the following way. Imagine considering

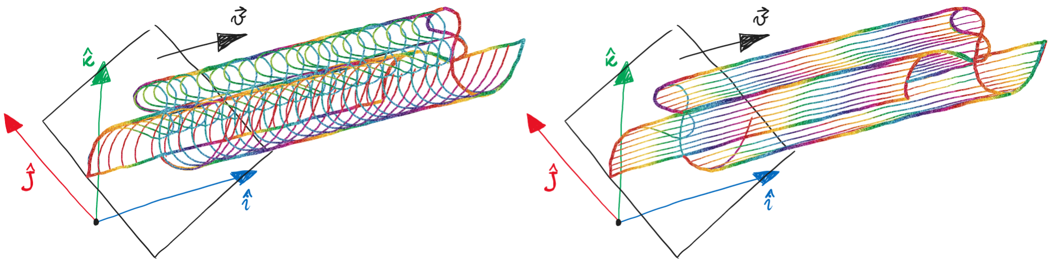


Figure 1.12: A cylinder surface

---

a plane and a curve that lies over it. In Picture 1.12 we have an example of this situation. Now, consider any vector $\vec{v}$ that does not lie on the plane and imagine translating rigidly the plane with the curve in the direction of the vector $\vec{v}$ and imagine that in doing that we are leaving a track of the passage of the curve. The whole track is the cylinder.

An equivalent way to describe cylinders is to say that such surface is obtained by moving by translation a line that is pointing in some direction along the curve. In Picture 1.12 we also added this second point of view. The easiest cylinders are obtained when the vector $\vec{v}$ is orthogonal to the plane where the line lies. In this case, we can always rotate and translate our system of coordinates $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$ to obtain a new one $\{\hat{\mathbf{i}}', \hat{\mathbf{j}}', \hat{\mathbf{k}}'\}$ so that the vector $\vec{v}$ is parallel to one of the unit vectors and the plane lies on the opposite plane. For example, we can make $\{\hat{\mathbf{i}}', \hat{\mathbf{j}}', \hat{\mathbf{k}}'\}$ so that $\vec{v}$ is parallel to $\hat{\mathbf{k}}'$. In this new system of coordinates the cylinder will be described by an equation in only the two variables x and y: the z disappears because the cylinder does not depend on the z coordinate anymore. Let's do an example.

Example 1.10. Suppose we want to obtain an equation of a cylinder in the usual sense of the term, i.e. a circular pipe. The first thing is to choose a system of coordinates where the vertical axis, i.e. the one generated by $\hat{\mathbf{k}}$, coincides with the axis of the cylinder. Then, if we slice the surface along a horizontal plane, i.e. a plane parallel to the plane $\hat{\mathbf{i}}\hat{\mathbf{j}}$, we obtain a perfect circle. Suppose that the radius is r, then the equation of the circle in two dimensions is just $x^2 + y^2 = r^2$. In three coordinates this is precisely the equation of our cylinder. ■

Example 1.11. Consider now the equation $z = x^2$. Can you picture this surface in three dimensions? In two dimensions is a parabola, so in three dimensions will be a cylinder obtained by translating the parabola along the y-axis. ■

Let's move to quadric surfaces. By definition, a quadric surface is a second-order polynomial equation in three variables. The whole expression is pretty messy and complicated. However, as we did for cylinders, if we cleverly choose the system of coordinates, rotating and translating the one we start from, we can always simplify the whole equations and obtain precisely two possible equations:

$$Ax^2 + By^2 + Cz^2 + D = 0$$

or:

$$Ax^2 + By^2 + Cz = 0$$

The difference is that in the first equation, there is no linear term, i.e. a first-order term, while in the second one z has order 1. Two surfaces that belong to this category of quadric surfaces are the unit sphere $\mathbb{S}^2$ and the ellipsoid, which has the equation:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

where a, b and c are the three axes of the ellipsoid. In the third category, we find the "pringle-shaped" surface pictured in Picture 3.12a, which has the equation:

$$z = y^2 - x^2$$

We conclude this section and module with a reflection on the meaning of vectors and geometry.

Cool Stuff* 1.6. We used vectors in two fundamentally distinct ways: sometimes we used them to indicate the **position** of a point; other times, we used vectors to indicate **velocities** of things that are moving. From a mathematical point of view, the first use of vectors gives a system of coordinates for the space; while with the second, we indicate the direction, the orientation and the intensity of a variation, which could be a velocity, a force, an acceleration and so on. These two ways of using vectors, even if they seem both very legit (and they are, don't get me wrong!) are fundamentally different.

It turns out that, in more general contexts, only the second use of vectors is actually well-defined, while the first one, i.e. the "position" vectors, loose meaning. The key to understanding this fact is that $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$ and more generally in $\mathbb{R}^n$, are very special spaces. Let's consider, for example, a universe with a different geometry. Imagine being on a cylinder. In this scenario, to indicate a position, we can decide to pick a point, that will be the origin of our system. The second step is to realize that a cylinder is two-dimensional, so we only need two coordinates, but now, if we pick two non-parallel vectors, from the chosen point, we realize that what we are doing is impossible, eventually, for at least one of the two vectors, the line that this vector generates will be detached from the surface.

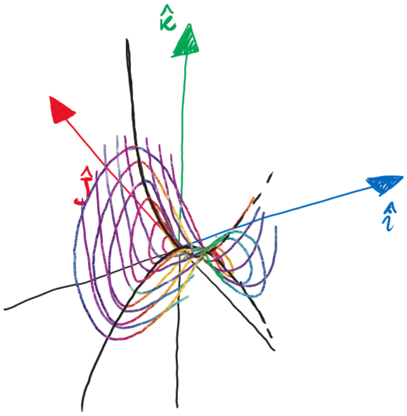
If we are very little w.r.t. the cylindrical surface where we live, we could say that textbflocally, our universe looks like a plane and we can still use locally a system of coordinates with two vectors, but as soon as we move a bit further, our local system of coordinate will not work anymore and we will need a new local system of coordinates. This phenomenon is not something special about the cylinder. For example, the same is true for the sphere and many other geometrical spaces. The impossibility, for some spaces, to define a global system of coordinates makes not well defined the use of vectors to indicate the position of something.

However, as we said before, locally, we can still make use of our system of coordinates defined with vectors. This idea is formalized with the notion of the tangent plane and, more generally, of tangent space. The tangent space is a local approximation of the geometrical space as something that is easy to work with: some space that looks like $\mathbb{R}^n$, for some integer n. So, for example, the tangent plane is a flat approximation of a surface. Now, luckily, concepts like velocity, forces, accelerations, and every concept that involves some notion of dynamics, or from a mathematical point of view of infinitesimal variations, are local concepts, so we can still describe them in our local system of coordinates, even in curved spaces. This intuition is the fundamental idea of differential geometry and the link between tangent spaces and derivatives is the core of this fascinating part of mathematics.

Now, let's come back to $\mathbb{R}^3$ and take any point $P \in \mathbb{R}^3$. Then the tangent space of $\mathbb{R}^3$ in P is $\mathbb{R}^3$ itself!!! This feature is something very peculiar to $\mathbb{R}^n$ and it's the fundamental reason why in $\mathbb{R}^n$ we can use vectors to indicate positions.

One final comment. There are geometrical spaces where even local systems of coordinates are impossible: this is the case for a cube. The problem with the cube is that on the edges, it's impossible to define a tangent plane. This is because the edges are sharp and not smooth! ■

Figure 1.13: "Pringle-shaped" surface



1.7 Review

Here is the list of important concepts of this module:

1. System of coordinates
    (a) Arrows
    (b) Vector spaces
    (c) Construction of the basis $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$
    (d) Components of a vector in the basis $\{\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}\}$
2. The scalar product
    (a) Definition in components (see Definition 1.11
    (b) Geometrical interpretation: $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}| \cos \theta$
    (c) Properties: Symmetry, left & right linearity, positive-definiteness
    (d) Orthogonality: $\vec{u} \cdot \vec{v} = 0$ iff $\vec{u} \perp \vec{v}$
    (e) Relationship with modulus: $\vec{u} \cdot \vec{u} = |\vec{u}|^2$
    (f) Scalar and vector projections: $\mathsf{comp}_{\vec{u}} \vec{v}, \mathsf{proj}_{\vec{u}} \vec{v}$
3. The vector product
    (a) Geometrical definition (see Definition 1.13
    (b) Components of $\vec{u} \times \vec{v}$
    (c) Properties: Antisymmetry, left & right linearity
    (d) Parallel vectors have null vector product: $\vec{u} \times \vec{v} = \vec{0}$ iff $\vec{u} \parallel \vec{v}$
    (e) Geometrical interpretation: $|\vec{u} \times \vec{v}|$ is the area of the parallelogram generated by $\vec{u}$ and $\vec{v}$
    (f) Vector product between the vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$
    (g) Relationship with the determinant of a 3×3 matrix
    (h) Scalar-triple product gives the volume of the parallelepiped
4. Geometry in 3d space
    (a) Vector equation of the line
    (b) Algebraic equation of the line as a system of two linear equations
    (c) The line that passes through two distinct points
    (d) The equation of a plane
    (e) Normal vector of a plane
    (f) The plane that passes through three distinct non-aligned points
    (g) Distance between a point and a plane
    (h) Cylinders

MODULE 2

Continuity & total derivatives

2.1 Some abstract non-sense

In this chapter, we start looking at multivariable functions. We're gonna mostly focus on two classes:

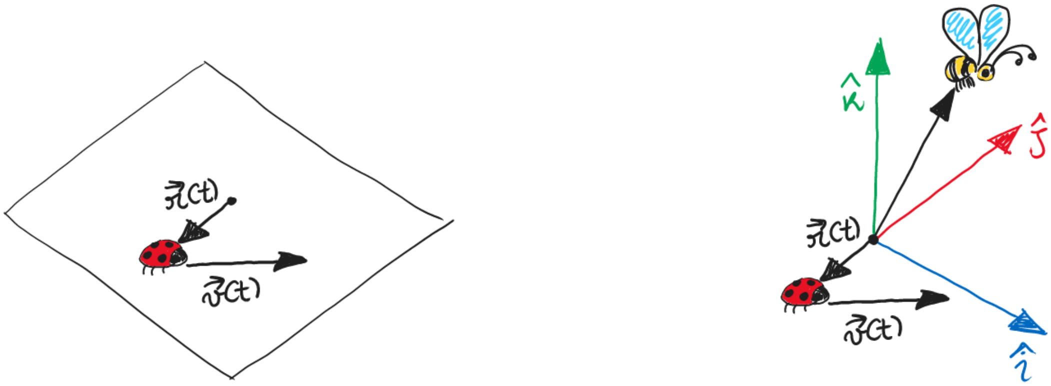
1. functions of the type $\vec{f} : \mathbb{R} \to \mathbb{R}^3$
2. functions of the type $f : \mathbb{R}^2 \to \mathbb{R}$

This is because these two cases have an easier geometrical description. However, it is very important to realize that the concepts we are introducing in this course do not restrict to only these two special cases, but they apply to every kind of multivariable function. To stress this aspect, we're gonna introduce each new concept in the general case and only after it, we will talk about the two special cases of vector functions and scalar functions.

Let's start by saying why someone would need multivariable and what are multivariable functions. Imagine considering a ladybug, as in Picture 2.1a and suppose we want to record its trajectory. This means recording at each time t, its position, which we denote by $\vec{r}(t)$, and its velocity $\vec{v}(t)$. Now, we can simplify the problem and assume that the ladybug will stay on a flat surface the whole time, so we assume that it doesn't jump, nor fly. So, to record the position, we need two coordinates, the coordinates of a plane. So, $\vec{r}(t) = \langle x(t), y(t) \rangle$, so we need 2 variables to describe the position. If we also allowed the ladybug to fly, or to jump, we needed a third variable, so we would have $\vec{r}(t) = \langle x(t), y(t), z(t) \rangle$.

The second step, is to define the variables we need to record the velocity of the ladybug. Recall that the velocity, is a vector quantity because it takes into account the speed, the direction and the orientation of the motion. If the ladybug moves only on a plane, $\vec{v}(t)$ needs other 2 coordinates, so we have $\vec{v}(t) = \langle u(t), v(t) \rangle$. Note that if we let the ladybug fly, we need a third component for the velocity too, then we would have $\vec{v}(t) = \langle u(t), v(t), w(t) \rangle$.

We realize that the trajectory of the ladybug is a function $\vec{f}$ that associates to every value t of time, 4 numbers, which are $x(t), y(t), u(t), v(t)$. In case we leave the ladybug fly, $\vec{f}$ associates to $t \in \mathbb{R}$, 6 numbers,



(a) We need four independent variables to describe the trajectory of a ladybug on a plane

(b) We need ten variables to describe the system ladybug-bumblebee

Figure 2.1: Bugs and multivariables

---

i.e. $x(t), y(t), z(t), u(t), v(t), w(t)$. But now, the elements of $\mathbb{R}^4$ are quadruples of numbers, so we can define the trajectory of a ladybug on a plane as a function $\vec{f} : \mathbb{R} \to \mathbb{R}^4$.

Unfortunately, it's hard to have a geometrical intuition of objects in $\mathbb{R}^4$, simply because our mind developed to move in a three-dimensional space. But there is nothing mysterious about $\mathbb{R}^4$, from a mathematical point of view: this is a vector space of dimension 4, whose elements are quadruples of real numbers.

Now, imagine also that we wanna keep track of the trajectory of the ladybug on the plane, but also of a bumblebee that flies around. In Picture 2.1b I drew this situation. Here, we need 4 components for the ladybug, 2 for the position and 2 for the velocity, and other 6 components for the bumblebee, 3 for the position, 3 for the velocity.

So, now, the describe the dynamics of the system ladybug-bumblebee we need a function $\vec{f} : \mathbb{R} \to \mathbb{R}^{10}$. It can seem too abstract and complicated to work with more than three dimensions, but it turns out that, even if the geometrical intuition fails, the algebraic properties of these spaces give us powerful tools to work with. The other scenario, where multivariable plays an important role is when the domain of a function has more than 1 dimension. This is the case of the temperature in a room. The temperature is a function of the three-dimensional space that associates with each point in the space, so to each triple $\langle x, y, z \rangle$, a corresponding number which indicates the temperature at that point. So, the temperature can be regarded as a function $T : \mathbb{R}^3 \to \mathbb{R}$. This is an example of a scalar function, called in this way because the codomain contains "scalars", i.e. numbers.

A vector function is a function with vector values. So, for example, the trajectory of the ladybug, $\vec{f} : \mathbb{R} \to \mathbb{R}^4$, is an example of a vector function. Another example is the function $\vec{E}$, which represents the electric field. The electric field in a point is a vector, so $\vec{E}$ is a function that associates to every point in the three-dimensional space a vector, so we have $\vec{E} : \mathbb{R}^3 \to \mathbb{R}^3$.
Let's give a definition.

Definition 2.1. *A multivariable function is a function* $\vec{f} : \Omega \to \mathbb{R}^m$ *with domain a subset* Ω *of the space* $\mathbb{R}^n$ *(we write* $\Omega \subseteq \mathbb{R}^n$*) and with codomain the space* $\mathbb{R}^m$*, where* n, m *are two positive integers. Concretely,* $\vec{f}$ *associates to every n-uple* $(x_1, x_2, \ldots, x_n) \in \Omega$ *an m-tuple* $\vec{f}(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), f_2(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$.

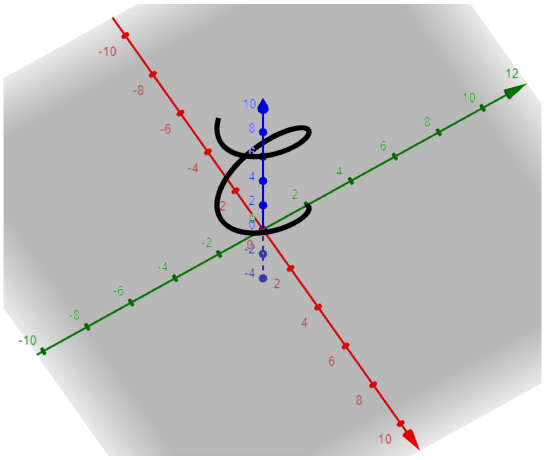


Figure 2.2: Representation of the vector function $\langle \sin t, 2\cos t, t \rangle$

When $n = 1$ and $m > 1$ we also call these functions, vector functions and we denote them by $\vec{f}$. When $n > 1$ and $m = 1$, we call these functions, scalar functions.

Example 2.1. Let's give an example of a multivariable function: the scalar product. The scalar product is an example of a multivariable function, because it takes two vectors and returns a number, so we can think of it as a function $f : \mathbb{R}^6 \to \mathbb{R}$, so defined:

$$f(x, y, z, x', y', z') := \langle x, y, z \rangle \cdot \langle x', y', z' \rangle = xx' + yy' + zz'$$

Also the vector product is a multivariable function:

$$f : \mathbb{R}^6 \to \mathbb{R}^3$$
$$f(x, y, z, x', y', z') := \langle x, y, z \rangle \times \langle x', y', z' \rangle = \langle yz' - y'z, x'z - xz', xy' - x'y \rangle$$

Let's now do a more classical example. Consider the function so defined:

$$\vec{f}(t) := \sin t\hat{\mathbf{i}} + 2\cos t\hat{\mathbf{j}} + t\hat{\mathbf{k}}$$

Let's try to understand what is this function. First, note that we can think about this function as a path in the three-dimensional space. With this interpretation, for any time t, $\vec{f}(t)$ represents the position of something in the three-dimensional space. To understand the shape of this path, we will look at the projections of $\vec{f}$ along the three planes $\hat{\mathbf{i}}\hat{\mathbf{j}}$, $\hat{\mathbf{i}}\hat{\mathbf{k}}$ and $\hat{\mathbf{j}}\hat{\mathbf{k}}$.

Projecting along one of the planes, means forgetting about one of the three components of $\vec{f}(t)$. So, if we wanna project along the $\hat{\mathbf{i}}\hat{\mathbf{j}}$ plane we obtain $\langle x(t), y(t) \rangle = \langle \sin t, 2\cos t \rangle$. This is the equation of an ellipse. To see this, we want to remove the parameter t and write this equation in an algebraic form, i.e. with only the variables x and y.

To do that, remember the important relation $\sin^2 t + \cos^2 t = 1$, thus we have:

$$x(t) = \sin t$$
$$\frac{y(t)}{2} = \cos t$$

therefore:

$$x^2 + \frac{y^2}{4} = 1$$

which is the equation of an ellipse with x-axis equal 1 and y-axis equal 2. Now, let's see what is the projection along the $\hat{\mathbf{i}}\hat{\mathbf{k}}$ plane. This is $\langle x(t), z(t)\rangle = \langle \sin t, t\rangle$. Because $t = z$, we have the equation $x = \sin z$. Similarly, we can $y = \cos z$ for the projection along the $\hat{\mathbf{j}}\hat{\mathbf{k}}$ plane. In Picture 2.2 I have shown this situation with the projection on the plane $\hat{\mathbf{i}}\hat{\mathbf{j}}$. ■

2.2 Limits & continuity

"*Natura non facit saltus*", this is a Latin quote from Leibniz which can be translated as "*Nature does not make jumps*". At that time, when Leibniz wrote this sentence, people didn't have a formal concept of limits and continuity (well, was actually Leibniz that started the whole process of introducing a formal definition of differential calculus), but they correctly understood that things in nature tend to not change too suddenly. This general notion of "things don't make sudden variations of their statuses" in modern mathematichese is called regularity. So, when we say that a function f is a regular function, we mean that f satisfies some condition of regularity, that is it doesn't change too abruptly. There are different levels of regularity and the simplest one is called continuity. Intuitively speaking, something that is continuous doesn't make jumps, so now the fancy Latin quote from Leibniz, is translated as "functions that represent natural systems, are continuous" (less catchy, but more precise).

To extend the definition of continuous functions for multivariable functions, first, let's briefly recap the definition of limits for functions in 1 variable. So, let's take our favourite function $f : \mathbb{R} \to \mathbb{R}$ and a point $x_0 \in \mathbb{R}$. So:

> *the limit of f for x approaching x_0 is a number $l \in \mathbb{R}$ with the following property: if we take any positive number $\varepsilon > 0$, we find a number $\delta > 0$ so that, there is at least an $x \in \Omega$ that belongs to the interval $(x_0 - \delta, x_0 + \delta)$ and it's not equal to x_0, i.e. $x \in \Omega \cap (x_0 - \delta, x_0 + \delta) \setminus \{x_0\}$, and for any of these x, $f(x)$ is an element of the interval $(l - \varepsilon, l + \varepsilon)$.*

This sounds very complicated. And it actually looks more complicated if you see the formal definition that uses the logic symbols and all the rest! So, let's try to unpack what's going on here.
First, when we say that we take a positive number ε or δ and we look at intervals of the form $(l - \varepsilon, l + \varepsilon)$ or $(x_0 - \delta, x_0 + \delta)$, what we actually want is to look close enough to our point l or x_0.
Intervals seem to play quite an important role in this definition, so the first step is to understand what is the correct generalization of an interval for multidimensions. Notice also that we only used open intervals, so what we actually need is a good notion of whatever is a multidimensional open interval. Now, there are two main approaches to answering this question and, it turns out, that they are equivalent: we can define cubes or balls. We're gonna use the second approach here. So, the idea is to note that an open interval of the type $(x_0 - r, x_0 + r)$ corresponds to an open ball centred in x_0 and of radius r, in 1 dimension.

So, the idea is to replace, in a multidimensional space, open intervals with open balls. But first, we need to define the notion of balls.

Definition 2.2. *Let $\mathbb{V}$ be an inner product space with scalar product $\cdot$. Consider also a point $\vec{x}_0 \in \mathbb{V}$ and a positive number $r > 0$. The* ***open ball*** *of centre $\vec{x}_0$ and radius r is the set so defined:*

$$\mathbb{B}(\vec{x}_0, r) := \{\vec{x} \in \mathbb{V} \mid d(\vec{x}, \vec{x}_0) < r\}$$

where $d(\vec{x}, \vec{x}_0)$ is the distance between $\vec{x}$ and $\vec{x}_0$ and it's defined as follows:

$$d(\vec{x}_1, \vec{x}_2) = |\vec{x}_1 - \vec{x}_2| = \sqrt{(\vec{x}_1 - \vec{x}_2) \cdot (\vec{x}_1 - \vec{x}_2)}$$

*Moreover, the **closed ball** of centre $\vec{x}_0$ and radius r is the set so defined:*

$$\overline{\mathbb{B}}(\vec{x}_0, r) := \{\vec{x} \in \mathbb{V} \mid d(\vec{x}, \vec{x}_0) \leq r\}$$

*Finally, the **sphere** of centre $\vec{x}_0$ and radius r is the set so defined:*

$$\mathbb{S}(\vec{x}_0, r) := \{\vec{x} \in \mathbb{V} \mid d(\vec{x}, \vec{x}_0) = r\}$$

Let's see what is an open ball, a closed ball and a sphere in the case of $\mathbb{V}_3$. Let $\vec{x}_0 = \langle x_0, y_0, z_0 \rangle$ and $r > 0$:

$$\begin{aligned}
\mathbb{B}(\vec{x}_0, r) &= \{\vec{x} = \langle x, y, z \rangle \in \mathbb{V}_3 \mid \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} < r\} = \\
&= \{\vec{x} = \langle x, y, z \rangle \in \mathbb{V}_3 \mid (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 < r^2\} \\
\overline{\mathbb{B}}(\vec{x}_0, r) &= \{\vec{x} = \langle x, y, z \rangle \in \mathbb{V}_3 \mid (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \leq r^2\} \\
\mathbb{S}(\vec{x}_0, r) &= \{\vec{x} = \langle x, y, z \rangle \in \mathbb{V}_3 \mid (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2\}
\end{aligned}$$

Let's understand what these definitions are saying. Let's start by saying what is the function d. The function d, called the distance, is a function that takes two vectors $\vec{u}$ and $\vec{v}$ and returns a real number $d(\vec{u}, \vec{v})$ which is the distance between $\vec{u}$ and $\vec{v}$.

Cool Stuff* 2.1. As you may have suspected at this point that, in the same way, inner product spaces generalize vector spaces with a scalar product and normed spaces, vector spaces with a norm, there is also a generalization for spaces with a notion of distance between points. These are called metric spaces because the distance sometimes is also called the metric, even if there is also another use for the word metric in mathematics.

Let's give here a brief definition of what is a metric space. A metric space is a set A equipped with a function d, called the distance, or sometimes the metric, so defined:

$$d \colon A \times A \to \mathbb{R}$$

that takes a pair of points a, b of A and returns a real number. To be a distance d needs to satisfy a few property that we now list:

1. Positive-definiteness: for every $a, b \in A$, $d(a, b) \geq 0$ and, moreover, $d(a, b) = 0$ if and only if $a = b$
2. Symmetry: for every $a, b \in A$, $d(a, b) = d(b, a)$
3. Triangle inequality: for every $a, b, c \in A$, $d(a, c) \leq d(a, b) + d(b, c)$

The first property says that the distance is always non-negative. This is simply because we don't want to measure negative distances. It also says that the distance between two points is equal to zero if and only if the two points coincide. The second assumption, symmetry, says that measuring the distance from a to b is the same as measuring the distance from b to a. Finally, the triangle inequality says that if you walk from a to c, it takes less or equal time than walking first from a to b and then from b to c.

It turns out that, if a vector space $\mathbb{V}$ is equipped with a norm $\|-\|$ (see Cool Example 1.2), then we can define a distance as follows:

$$d(\vec{u}, \vec{v}) := \|\vec{u} - \vec{v}\|$$

The norm measures the lengths of vectors, therefore, by taking the norm of the difference between two vectors $\vec{u}$ and $\vec{v}$, we are measuring the length of the vector $\vec{u} - \vec{v}$, which, geometrically, is the same as measuring the distance between the endpoints of $\vec{u}$ and $\vec{v}$. Note that, as we already said, if your vector space has a scalar product $\cdot$, then you also have a norm, defined by $\|\vec{u}\| := \sqrt{\vec{u} \cdot \vec{u}}$. Thus, having a scalar product allows to define also a distance, as follows:

$$d(\vec{u}, \vec{v}) := \sqrt{(\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v})}$$

Note that to define balls in our context, we adopted this precise distance. However, in the same way, there are norms that are not induced by scalar products, there are also distances that are not defined by norms. An example of such distance is the so-called discrete distance, sometimes denoted by δ, which is 0 if a and b are the same point, and 1 if not:

$$\delta(a, b) := \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \end{cases}$$

This is a very extreme case! In such metric space every point is far 1 from any other point. ■

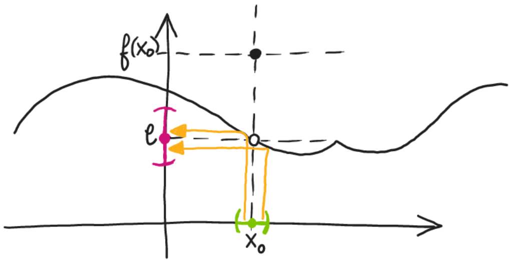


Figure 2.3: Definition of limits in one dimension

So, we have introduced a distance d between vectors. With that, we defined the open ball to be the set of points within a distance r from the centre $\vec{x}_0$. Note that the inequality $d(\vec{x}_0, \vec{x}) < r$ is strict because for the open ball we want to exclude the boundary (this is why we call it open). A suggestive way to think about an open ball is thinking of an apple, which represents a ball, at which we peeled off the skin. The skin represents the boundary.

Similarly, a closed ball is an open ball plus the boundary. So, in our metaphor, will be the whole apple with the skin on. Finally, the sphere is the boundary, i.e. the set of points that are far from the centre exactly r. So, this would correspond to the skin of the apple.
Note that $\mathbb{S}^2$, introduced in Example 1.9, the unit sphere, is a sphere of radius 1, centred in $\vec{0}$, according to our definition, so we have that $\mathbb{S}^2 = \mathbb{S}(\vec{0}, 1)$, in $\mathbb{V}_3$. Now, coming back to open intervals, it is not too hard to see that an open interval $(x_0 - r, x_0 + r)$ is precisely the set of points $x \in \mathbb{R}$ within a distance of r from x_0, so $(x_0 - r, x_0 + r)$ is an open ball in 1 dimension. Can you guess what are a closed ball and a sphere in 1 dimension?

So, coming back to our definition of limits in 1 dimension, we can now rephrase it as follows:
We say that:

> *the limit of a function $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}$, for x approaching $x_0 \in \mathbb{R}$ is the number l, whenever, taking any open ball $\mathbb{B}(l, \varepsilon)$ centred in l and of radius $\varepsilon > 0$, we can find another open ball $\mathbb{B}(x_0, \delta)$ centred in x_0 and or radius δ, so that, if we take any $x \in \mathbb{B}(x_0, \delta) \cap \Omega$, excluding x_0 itself (so $x \neq x_0$), we have that $f(x) \in \mathbb{B}(l, \varepsilon)$.*

In Picture 2.3 we have a function $f : \mathbb{R} \to \mathbb{R}$, we are taking a point $x_0 \in \mathbb{R}$ and we are showing that the limit of f for x approaching x_0 is the value l. As you can see, we chose a little open ball around l and we found another ball around x_0, so that every x in that ball, excluding x_0, is mapped by f into the little ball around l; and this works for any ball around l.
Now, we wanna finally move to multivariable functions. Before doing so, we need some technical concepts.

Definition 2.3. *Let $\Omega \subseteq \mathbb{R}^n$ be a subset of $\mathbb{R}^n$. We say that Ω is **open** if, given any point $\vec{x}_0 \in \Omega$ we can find a positive number $\delta > 0$ so that the open ball $\mathbb{B}(\vec{x}_0, \delta)$ is entirely included inside Ω, i.e. $\mathbb{B}(\vec{x}_0, \delta) \subseteq \Omega$.*
*We say that Ω is **closed** if, the complementary of Ω, i.e. $\mathbb{R}^n \setminus \Omega$ is open.*
*We also call the **interior** of Ω the largest open set $\mathring{\Omega}$ included in Ω, the **closure** of Ω is the smallest closed set $\overline{\Omega}$ that includes Ω, and we call the **boundary** of Ω, the difference between the closure and the interior. More precisely:*

$$\mathring{\Omega} := \bigcup_{A \text{ open } \& \, A \subseteq \Omega} A$$

$$\overline{\Omega} := \bigcap_{C \text{ closed } \& \, \Omega \subseteq C} C$$

$$\partial\Omega := \overline{\Omega} \setminus \mathring{\Omega}$$

Example 2.2. We can think of the boundary of a set as the skin of this set, the interior as the flesh of the set and the closure as the body with all skin. So, imagine taking an apple. The boundary of the apple is the skin of the apple, its interior is the pulp and the closure is the whole apple. So, if the apple has a little scar, so part of the skin is removed, and the closure will be the apple with the skin covering also the scar.

We already have examples of open sets: every open ball is an open set, every closed ball is a closed set. Given a closed ball $\overline{\mathbb{B}}(\vec{x}_0, r)$, its interior is the open ball $\mathbb{B}(\vec{x}_0, r)$ and the boundary is the sphere $\mathbb{S}(\vec{x}_0, r)$. ■

We're ready to define the notion of limits.

Definition 2.4. *Let $\vec{f} : \Omega \to \mathbb{R}^m$ be a multivariable function, where $\Omega \subseteq \mathbb{R}^n$ is a subset of $\mathbb{R}^n$. Consider a point $\vec{x}_0 \in \mathbb{R}^n$ and suppose that $\vec{x}_0$ is a point of the closure $\overline{\Omega}$ of Ω. We say that the function $\vec{f}$ has **limit** $\vec{l} \in \mathbb{R}^m$ for $\vec{x}$ approaching $\vec{x}_0$, if, choosing any open ball $\mathbb{B}(\vec{l}, \varepsilon) \subseteq \mathbb{R}^m$, centred in $\vec{l}$ and of some arbitrary positive radius $\varepsilon > 0$, we find an open ball $\mathbb{B}(\vec{x}_0, \delta) \subseteq \mathbb{R}^n$, centred in $\vec{x}_0$ and of some small positive number $\delta > 0$, so that, given the set $B := (\mathbb{B}(\vec{x}_0, \delta) \cap \Omega) \setminus \{x_0\}$, i.e. the intersection between the ball $\mathbb{B}(\vec{x}_0, \delta)$ and Ω substracted $\vec{x}_0$, for every $\vec{x} \in B$, $\vec{f}(\vec{x}) \in \mathbb{B}(\vec{l}, \varepsilon)$ (where $B = (\mathbb{B}(\vec{x}_0, \delta) \cap \Omega) \setminus \{x_0\}$).*
We denote that $\vec{l}$ is the limit of $\vec{f}$ for $\vec{x}$ approaching $\vec{x}_0$ as follows:

$$\lim_{\vec{x} \to \vec{x}_0} \vec{f}(\vec{x}) = \vec{l}$$

We also sometimes use the notation: $\vec{f}(\vec{x}) \to \vec{l}$, for $\vec{x} \to \vec{x}_0$ and we read $\to$ as "approaches".

A few important things here: you have probably noticed that, in the definition of limits, we only take points $\vec{x}$ that belong to the intersection between the ball $\mathbb{B}(\vec{x}_0, \delta)$ and Ω and that we also exclude $\vec{x}_0$. Here, there are two things to say. First, we take the intersection between the ball and Ω, because Ω is the domain of the function, where the function is well-defined. Take for example the function $f(x) := \sqrt{x}$. This function is well-defined for $x \geq 0$, but not for negative numbers, so here its domain is $[0, +\infty)$. Because we are going to take $\vec{f}(\vec{x})$, we have to be sure that this thing is well-defined.
It can happen that $\vec{x}_0$ is on the boundary of Ω. In that case, you have to be sure that you are picking an $\vec{x}$ in the ball $\mathbb{B}(\vec{x}_0, \delta)$ that is also part of the domain of the function. That's why we take the intersection.

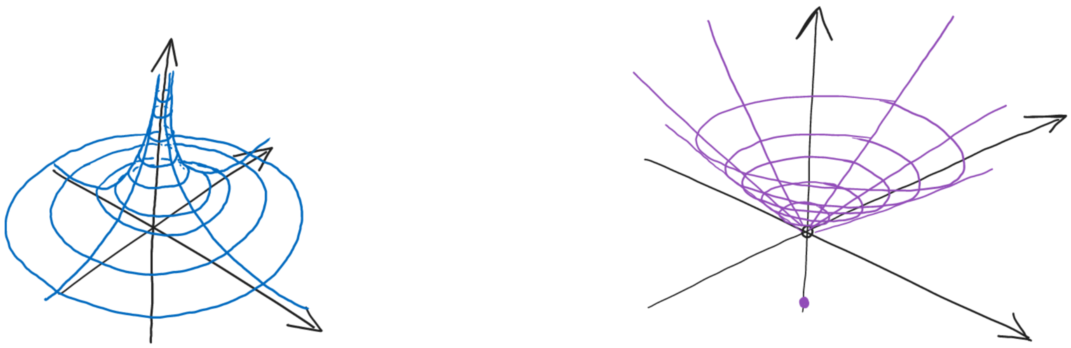
The second thing is that we are excluding $\vec{x}_0$. This is because, when $\vec{x}_0 \in \Omega$, then we also have a value $\vec{f}(\vec{x}_0)$, but this value doesn't necessarily correspond to the limit. The whole idea of limits is that the function tends to go closer and closer to a specific point whenever we approach a value $\vec{x}_0$. It can happen that at that point the function makes a sudden jump, or that the function is not defined there, so we want to exclude the point $\vec{x}_0$.

The next thing to note is that we only allowed $\vec{x}_0$ be points of the closure of Ω. This looks like a complicated assumption, but it's a very important technical hypothesis. The idea is that if $\vec{x}_0$ is not in the closure of Ω, this means that $\vec{x}_0$ is not in the interior of Ω, nor on the boundary of Ω. This means that $\vec{x}_0$ is too far from the domain of $\vec{f}$, so we can always find a little ball around $\vec{x}_0$ which does not intersect with Ω at all. Essentially, this means that we don't have any $\vec{x} \in \Omega$ that could go close to $\vec{x}_0$, because $\vec{x}_0$ is just too far.

Let's see now some concrete examples of limits.

Example 2.3. Let's define the following scalar function. First, let's define $\Omega := \mathbb{R}^2 \setminus \vec{0}$, so Ω is the whole plane without the origin $\vec{0}$. Let $f : \Omega \to \mathbb{R}$ so defined:

$$f(\vec{x}) := \frac{1}{|\vec{x}|}$$



(a) Representation of the function $f(\vec{x}) = 1/|\vec{x}|$

(b) Representation of the function $f(\vec{x}) = 2|\vec{x}|$, for $\vec{x} \neq \vec{0}$ and $f(\vec{0}) = -2$

Figure 2.4: Some interesting scalar functions

We want to see if f has a limit in $\vec{x}_0 = \vec{0}$. First, note that the boundary of Ω is precisely the singleton set $\{\vec{0}\}$, so $\vec{x}_0$ is part of the closure of Ω. In Picture 2.4a I drew a representation of the graph of the function f and, as you can see, for $\vec{x}$ approaching $\vec{0}$, from any possible direction, we always have that $f(\vec{x})$ goes to infinity. So, here we go: f has no limit in $\vec{0}$.
Let's now do a positive example. Suppose now that $f : \Omega \to \mathbb{R}$ is now defined as follows:

$$f(\vec{x}) := \begin{cases} 2|\vec{x}| & \text{if } \vec{x} \in \Omega \\ -2 & \text{if } \vec{x} = \vec{0} \end{cases}$$

where again $\Omega = \mathbb{R}^2 \setminus \{\vec{0}\}$. This case is represented in Picture 2.4b. As you can see if we take a little ball $\mathbb{B}(0, \varepsilon) = (-\varepsilon, +\varepsilon) \subseteq \mathbb{R}$ around $\vec{0}$, choosing $\delta := \frac{\varepsilon}{2}$, we have that if $\vec{x} \in \mathbb{B}(\vec{0}, \delta)$ and $\vec{x} \neq \vec{0}$, then $|\vec{x}| < \varepsilon/2$, so moving 2 on the other side we find that $2|\vec{x}| < \varepsilon$. But for $\vec{x} \neq \vec{0}$, $f(\vec{x}) = 2|\vec{x}|$, so we find that $d(f(\vec{x}), \vec{0}) = \left|f(\vec{x}) - \vec{0}\right| = f(\vec{x}) \leq \varepsilon$, so $f(\vec{x}) \in \mathbb{B}(0, \varepsilon)$. Because this is true for every $\varepsilon > 0$, f has limit 0 for $\vec{x}$ approaching $\vec{0}$.
As a third example, consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$ so defined:

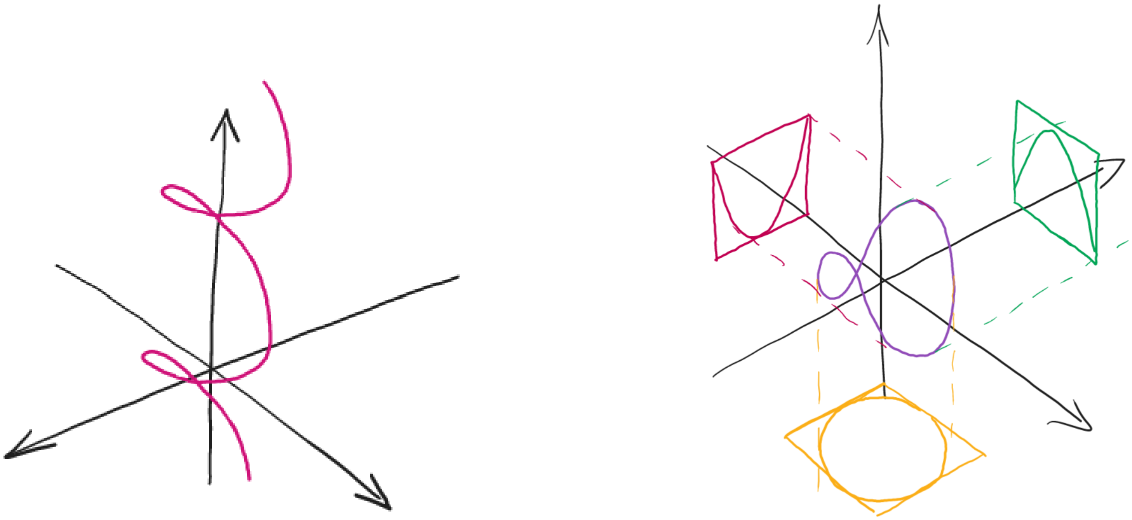
$$f(x, y) := \begin{cases} x^2 + y^2 & \text{if } x, y \neq 0 \\ -1 & \text{if } x = y = 0 \end{cases}$$

What is the limit of $\vec{x}$ approaching $\vec{0}$ of $f(\vec{x})$? ■

Let's now see how can we evaluate limits for vector functions, i.e. functions of the type $\mathbb{R} \to \mathbb{R}^3$. In this case, we can think of the points in the domain as a parameter time t and for different values of t we have, for example, the position $\vec{f}(t)$ in $\mathbb{R}^3$ of something that is flying around.
So, using the components, we find out that such a function is the form:

$$\begin{aligned} \vec{f} : \mathbb{R} &\to \mathbb{R}^3 \\ t &\mapsto \vec{f}(t) := \langle x(t), y(t), z(t) \rangle \end{aligned}$$

or, using the vectorial notation, $\vec{f}(t) = x(t)\hat{\mathbf{i}} + y(t)\hat{\mathbf{j}} + z(t)\hat{\mathbf{k}}$. It turns out that, to evaluate the limit of such a function we just need to evaluate the limits of $x(t)$, $y(t)$ and $z(t)$, so we have the following result.



(a) Representation of the vector function $\vec{f}(t) = \langle \cos t, 2\sin t, t \rangle$

(b) Reprerentation of the vector function $\vec{f}(t) = \langle \cos t, \sin t, \cos(2t) \rangle$

Figure 2.5: Visualization of two vector functions

Proposition 2.1. *If $\vec{f} : \Omega \to \mathbb{R}^3$, with $\Omega \subseteq \mathbb{R}$ is a vector function and $t_0 \in \overline{\Omega}$, is a point of the closure of Ω, then the limit of $\vec{f}(t) = \langle x(t), y(t), z(t) \rangle$ for t approaching t_0 is as follows:*

$$\lim_{t \to t_0} \vec{f}(t) = \langle \lim_{t \to t_0} x(t), \lim_{t \to t_0} y(t), \lim_{t \to t_0} z(t) \rangle$$

More generally, if $\vec{f} : \Omega \to \mathbb{R}^m$ and $\Omega \subseteq \mathbb{R}$, is such that $\vec{f}(t) = \langle x_1(t), \ldots, x_m(t) \rangle$ we have that:

$$\lim_{t \to t_0} \vec{f}(t) = \langle \lim_{t \to t_0} x_1(t), \ldots, \lim_{t \to t_0} x_m(t) \rangle$$

Example 2.4. Consider the vector function $\vec{f}(t) := \langle \cos t, 2\sin t, t \rangle$. In Picture 2.5a you see a representation of the path $\vec{f}$. We want to compute the limit of $\vec{f}$ for $t \to \pi$:

$$\lim_{t \to \pi} \vec{f}(t) = \langle \lim_{t \to \pi} \cos t, \lim_{t \to \pi} 2\sin t, \lim_{t \to \pi} t \rangle = \langle -1, 0, \pi \rangle$$

Consider now the vector function $\vec{f}(t) := \langle \cos t, \sin t, \cos(2t) \rangle$. In Picture 2.5b you see a representation of the path $\vec{f}$. We want to compute the limit of $\vec{f}$ for $t \to 0$:

$$\lim_{t \to 0} \vec{f}(t) = \langle \lim_{t \to 0} \cos t, \lim_{t \to 0} \sin t, \lim_{t \to 0} \cos(2t) \rangle = \langle 1, 0, 1 \rangle$$

Note that the projections of $\vec{f}$ along the plane $\hat{\mathbf{i}}\hat{\mathbf{j}}$ (which means that we forget about the vertical component) gives $\langle x(t), y(t) \rangle = \langle \cos t, \sin t \rangle$. But because $\cos^2 t + \sin^2 t = 1$, we have $x^2 + y^2 = 1$, which is the equation of a circle. Projecting along $\hat{\mathbf{i}}\hat{\mathbf{k}}$ instead, we obtain $\langle x(t), z(t) \rangle = \langle \cos t, \cos(2t) \rangle$. But $\cos(2t) = \cos^2 t - \sin^2 t = 2\cos^2 t - 1$, so we obtain the algebraic equation $z = 2x^2 - 1$, which is the equation of a parabola with negative concavity. ■

Finally, we can discuss the concept of continuity for multivariable functions.

Definition 2.5. *A function $\vec{f} : \Omega \to \mathbb{R}^m$, with $\Omega \subseteq \mathbb{R}^n$, is continuous in $\vec{x}_0$, if three things are true:*

1. *$\vec{x}_0 \in \Omega$, i.e. $\vec{f}$ is well-defined in $\vec{x}_0$*
2. *The limit of $\vec{f}(\vec{x})$ for $\vec{x}$ approaching $\vec{x}_0$ exists*
3. *The limit of $\vec{f}(\vec{x})$ for $\vec{x} \to \vec{x}_0$ is equal to $\vec{f}(\vec{x}_0)$, i.e. $\lim_{\vec{x}\to\vec{x}_0} \vec{f}(\vec{x}) = \vec{f}(\vec{x}_0)$.*

Moreover, we say that the function $\vec{f}$ is continuous, meaning that is continuous in every point of its domain.

Every polynomial function is continuous, every trigonometric function is continuous, and so are exponential and logarithmic functions. Moreover, linear combinations, multiplications and compositions of continuous functions are all continuous.

2.3 Differentiability & total derivatives

Finally, we are able to introduce some differential calculus!
First of all, let's understand what we wanna capture with the notion of differentiability. We introduced the previous section with the Leibniz quote "*nature does not make jumps*" and we formalized this intuitive idea with the notion of continuity. But in reality, things follow more regularity than simple continuity. Imagine for example you are running a race and at some point, you have to turn really abruptly in another direction. This change of direction, at first glance, looks like you are making a sharp angle in your path. For sure, you will not suddenly make a jump from one point to another (this is called teleportation and so far we are not achieving great accomplishments in this way of travelling). So, for sure your path will be continuous.

However, you can suspect that you can make a non-smooth path. The reality is that a sharp change of direction is not possible for bodies with a mass. The reason why this is true is because of inertia: when you try to suddenly make a sharp turn, your body, for inertia, will continue to move forward, the floor where you are running will contrast your force bending in an imperceptible way and at the end of the day, your path will still be smooth.

This is not the case for light beams! When reflected on a mirror, they actually make sharp corners (at least, if we don't take a look at the complicated interaction between the beam and the mirror on a microscopic level). So, the smoothness of paths is a reasonable assumption for trajectories of bodies with a mass. This idea of smooth paths is called differentiability.

However, this new regularity condition, applies to a broader class of functions. Imagine having a two-dimensional surface, described as the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$. Now, imagine to be a little ant, moving on this surface. The continuity of f means that as an ant, you will not have to jump to move around. Smoothness is about the roughness of the surface. Over a smooth surface, you will not need to jump and travelling is quite easy because the surface is not rough, it has no sharp corners coming out.

In this section, we're gonna introduce this notion and we'll do it for every multivariable function. However, we will mostly focus, for now, on the case of vector functions.

This idea of smoothness can be translated as saying that the function can be locally approximated as a flat space. Flat spaces are vector spaces, so what we are actually asking is to be able to locally approximate a function with a linear function. Let's recall this definition here.

Definition 2.6. *Given two vector spaces $\mathbb{V}$ and $\mathbb{V}'$, a function $\vec{L} : \mathbb{V} \to \mathbb{V}'$ is* ***linear*** *if the following property, called linearity, holds. For every real numbers $a, b \in \mathbb{R}$ and every vectors $\vec{u}, \vec{v} \in \mathbb{V}$:*

$$\vec{L}(a\vec{u} + b\vec{v}) = a\vec{L}(\vec{u}) + b\vec{L}(\vec{v})$$

We notice that we dedicated an entire section of the appendix to discuss in detail the notion of linearity. Take a look at Appendix 5.1. Let's define the main concept of this section.

Definition 2.7. *Let $\vec{f} : \Omega \to \mathbb{R}^m$, with $\Omega \subseteq \mathbb{R}^n$, be a multivariable function and let $\vec{x}_0 \in \Omega$. We say that $\vec{f}$ is* ***differentiable*** *in $\vec{x}_0$ if there is a linear function $\vec{L}_{\vec{x}_0} : \mathbb{R}^n \to \mathbb{R}^m$ and an open ball $\mathbb{B}(\vec{x}_0, \delta)$ centred in $\vec{x}_0$ and of some small radius $\delta > 0$, so that, for every $\vec{x} \in \mathbb{B}(\vec{x}_0, \delta) \cap \Omega$:*

$$\vec{f}(\vec{x}) \simeq \vec{f}(\vec{x}_0) + \vec{L}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$$

We also say that $\vec{f}$ is differentiable if is differentiable in any point of its domain. Finally, we call the linear function $\vec{L}_{\vec{x}_0}$ the ***differential*** *of $\vec{f}$ in $\vec{x}_0$.*
The differential is often denoted by $\mathrm{d}\vec{f}_{\vec{x}_0}$.

In the expression:

$$\vec{f}(\vec{x}) \simeq \vec{f}(\vec{x}_0) + \vec{L}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$$

we are saying that the function $\vec{f}$ can be approximated to the function $\vec{f}(\vec{x}_0) + \vec{L}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$ whenever $\vec{x}$ is close enough to $\vec{x}_0$. Note that $\vec{f}(\vec{x}_0)$ is a fixed vector of $\mathbb{R}^m$, while with the expression $\vec{L}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$ we mean that the linear function $\vec{L}_{\vec{x}_0}$ takes the vector $\vec{x} - \vec{x}_0$ of $\mathbb{R}^n$ and maps it to the vector $\vec{L}_{\vec{x}_0}(\vec{x} - \vec{x}_0) \in \mathbb{R}^m$.

Cool Stuff* 2.2. Mathematicians don't like writing $\simeq$ at all. What does it even mean? Luckily, we have a precise definition of when we say the $\vec{f}(\vec{x})$ can be locally approximated as $\vec{f}(\vec{x}_0) + \vec{L}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$. We say that a function $\vec{f}$ can be locally approximated with another function $\vec{g}$ around the point $\vec{x}_0$, if the following limit exists and is zero:

$$\lim_{\vec{x} \to \vec{x}_0} \frac{\vec{f}(x) - \vec{g}(x)}{|\vec{x} - \vec{x}_0|} = \vec{0}$$

Mathematicians used to write this using a fancy notation called the little-o notation:

$$\vec{f}(\vec{x}) = \vec{g}(\vec{x}) + o(|\vec{x} - \vec{x}_0|)$$

The idea is to think of the symbol $o(|\vec{x} - \vec{x}_0|)$ as a very small quantity, as long as $\vec{x}$ is close enough to $\vec{x}_0$. ■

Before we continue, let's see what happens when we take 1 dimensional functions.

Example 2.5. Imagine now to consider a function $f : (a, b) \to \mathbb{R}$ and let's see what means for this function to be differentiable in $x_0 \in (a, b)$. By definition, there is a linear function $L_{x_0} : \mathbb{R} \to \mathbb{R}$ and so that:

$$f(x) \simeq f(x_0) + L_{x_0}(x - x_0)$$

whenever x is close enough to x_0. But in one dimension, linear functions are only multiplicative functions, i.e. there is a number $c_{x_0} \in \mathbb{R}$ such that, for any number $x \in \mathbb{R}$, $L_{x_0}(x) = c_{x_0} \cdot x$, so L_{x_0} simply multiplies c_{x_0} to x. So, with this in mind, we can rewrite the expression as:

$$f(x) \simeq f(x_0) + c_{x_0} \cdot (x - x_0)$$

But now, when x is close enough to x_0, we can rewrite this expression as:

$$f(x) - f(x_0) \simeq c_{x_0} \cdot (x - x_0)$$

and diving by $x - x_0$ we find that:

$$c_{x_0} \simeq \frac{f(x) - f(x_0)}{x - x_0}$$

But this symbol $\simeq$ means precisely the following:

$$c_{x_0} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Does this remind you of anything? This is the definition of the derivative of f in x_0!!! So, we find out that the constant c_{x_0} is precisely $f'(x_0)$! So, now rewriting the whole expression we find that:

$$f(x) \simeq f(x_0) + f'(x_0)(x - x_0)$$

The function $f(x_0) + f'(x_0)(x - x_0)$ in the variable x is known as the linear approximation of f in x_0 and it gives the equation of the tangent line of f in x_0, which is $y = f(x_0) + f'(x_0)(x - x_0)$. So, the definition we gave of differentiability is equivalent to the definition you have already studied for functions in 1 dimension. ■

Now, giving a definition is one thing, learning how to compute something is another story. We're gonna split this problem into two main classes: first, we study differentiability for vector functions and, in the next chapter, we apply this concept to multivariable scalar functions. The second class requires an entire discussion because we will need partial derivatives. However, the first class looks a bit more familiar, because it allows the definition of total derivatives. So, let's introduce this concept.

Definition 2.8. *Let $\vec{f} : \Omega \to \mathbb{R}^3$ be a vector function with $\Omega \subseteq \mathbb{R}$. We say that $\vec{f}$ has total derivative in $t_0 \in \Omega$ if the following limit is well-defined:*

$$\lim_{t \to t_0} \frac{\vec{f}(t) - \vec{f}(t_0)}{t - t_0} = \lim_{h \to 0} \frac{\vec{f}(t_0 + h) - \vec{f}(t_0)}{h}$$

In this case, we call this limit the ***total derivative*** *of $\vec{f}$ in t_0 and we denote it by $\vec{f}'(t_0)$ or, using Leibniz notation, by $df(t_0)/dt$.*

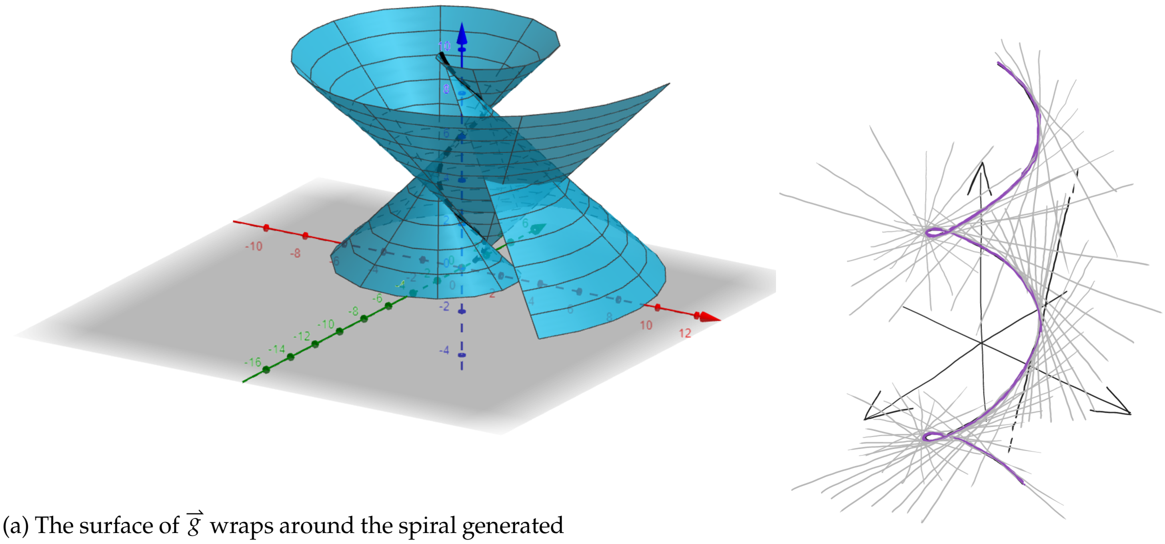
It turns out that, for vector functions, having total derivatives is the same as being differentiable. Moreover, in this case, the differential $\mathsf{d}\vec{f}_{t_0}$ of $\vec{f}$ in t_0 is the multiplicative function $\mathsf{d}\vec{f}_{t_0}(t) := t \cdot \vec{f}'(t_0)$, where $\cdot$ here denotes the scalar action.

Example 2.6. Consider the following vector function:

$$\vec{f} : \mathbb{R} \to \mathbb{R}^3$$
$$\vec{f}(t) := \sin t\,\hat{\mathbf{i}} + \cos t\,\hat{\mathbf{j}} + t\,\hat{\mathbf{k}}$$

Let's compute the total derivative of $\vec{f}$. By definition, we have that:

$$\begin{aligned} &\frac{d\vec{f}}{dt}(t) = \\ = \quad &\lim_{h \to 0} \frac{\vec{f}(t+h) - \vec{f}(t)}{h} = \\ = \quad &\lim_{h \to 0} \frac{\sin(t+h)\hat{\mathbf{i}} + \cos(t+h)\hat{\mathbf{j}} + (t+h)\hat{\mathbf{k}} - \sin t\,\hat{\mathbf{i}} - \cos t\,\hat{\mathbf{j}} - t\hat{\mathbf{k}}}{h} = \end{aligned}$$



(a) The surface of $\vec{g}$ wraps around the spiral generated by the collection of all the tangent lines

(b) A representation of the tangent lines of the spiral

Figure 2.6: Tangent lines of the spiral

$$
\begin{aligned}
&= \lim_{h\to 0} \frac{(\sin(t+h)-\sin t)\hat{\mathbf{i}} + (\cos(t+h)-\cos t)\hat{\mathbf{j}} + (t+h-t)\hat{\mathbf{k}}}{h} = \\
&= \lim_{h\to 0} \frac{(\sin(t+h)-\sin t)}{h}\hat{\mathbf{i}} + \lim_{h\to 0} \frac{(\cos(t+h)-\cos t)}{h}\hat{\mathbf{j}} + \lim_{h\to 0} \frac{t+h-t}{h}\hat{\mathbf{k}} = \\
&= \frac{d}{dt}(\sin t)\hat{\mathbf{i}} + \frac{d}{dt}(\cos t)\hat{\mathbf{j}} + \frac{d}{dt}(t)\hat{\mathbf{k}} = \\
&= \cos t\hat{\mathbf{i}} - \sin t\hat{\mathbf{j}} + \hat{\mathbf{k}}
\end{aligned}
$$

In Picture 2.7 I draw the graph of the function $\vec{f}'$. To interpret the meaning of this function, we see that at time $t \in \mathbb{R}$, $\vec{f}'(t)$ is the vector $< \cos t, -\sin t, 1\rangle$. If we rigidly move this vector to the point $\vec{f}(t)$, we see that this is tangent to the curve. We can interpret this vector as the velocity at which a point is moving along the curve.

Let's now find the function of the tangent line at $t_0 \in \mathbb{R}$ for the function $\vec{f}$. We know that the linear approximation of $\vec{f}$ in t_0 is as follows:

$$
\begin{aligned}
&\vec{L}(t) = \\
&= \vec{f}(t_0) + \mathrm{d}\vec{f}_{t_0}(t-t_0) = \\
&= \vec{f}(t_0) + (t-t_0)\cdot\frac{d\vec{f}}{dt}(t_0) = \\
&= \sin t_0\hat{\mathbf{i}} + \cos t_0\hat{\mathbf{j}} + t_0\hat{\mathbf{k}} + (t-t_0)(\cos t_0\hat{\mathbf{i}} - \sin t_0\hat{\mathbf{j}} + \hat{\mathbf{k}}) = \\
&= \sin t_0\hat{\mathbf{i}} + \cos t_0\hat{\mathbf{j}} + t_0\hat{\mathbf{k}} + (t-t_0)\cos t_0\hat{\mathbf{i}} - (t-t_0)\sin t_0\hat{\mathbf{j}} + (t-t_0)\hat{\mathbf{k}} = \\
&= (\sin t_0 + (t-t_0)\cos t_0)\hat{\mathbf{i}} + (\cos t_0 - (t-t_0)\sin t_0)\hat{\mathbf{j}} + t\hat{\mathbf{k}}
\end{aligned}
$$

This is the linear approximation of the function $\vec{f}$ at the point t_0. Here, there are two variables, t_0 and t. t_0 is the time at which we look at the tangent line, so $(t_0, \vec{f}(t_0))$ is precisely the point in the graph of the function, at which we are interested to calculate the tangent line. The variable t instead, gives the points

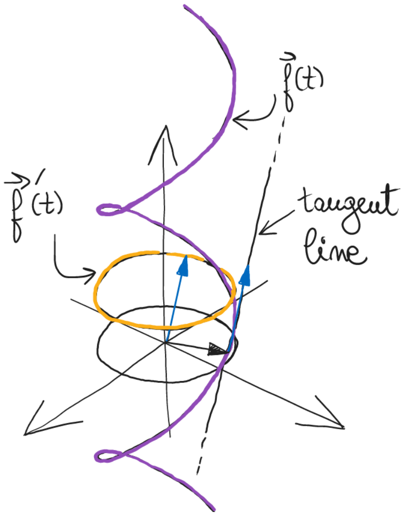


Figure 2.7: Representation of the total derivative of the spiral function

along the tangent line. However, if we consider t and t_0 as two independent variables x and y than we obtain an interesting function $g : \mathbb{R}^2 \to \mathbb{R}$ so defined:

$$\vec{g} : \mathbb{R}^2 \to \mathbb{R}^3$$
$$\vec{g}(x, y) := (\sin x + (y - x)\cos x)\hat{\mathbf{i}} + (\cos x - (y - x)\sin x)\hat{\mathbf{j}} + y\hat{\mathbf{k}}$$

In Picture 2.6a I plot the surface given by $\vec{g}$, obtained by varying both the variables x and y in the interval $[0, 10]$, together with the curve given by $\vec{f}$. As you can see, the surface of $\vec{g}$ wraps around the spiral curve. This is because this surface is the collection of all the tangent lines of the spiral for every point of the curve. In Picture 2.6b you can see the collection of the tangent lines of the spiral. ■

In Example 2.6, we computed the total derivative of the function $\vec{f}(t) = \sin t\hat{\mathbf{i}} + \cos t\hat{\mathbf{j}} + t\hat{\mathbf{k}}$. To do that, we used a little trick: we reordered the terms of the rate of change quotient to split it into the derivatives of the components. This is actually not a trick that works only in this case, but an actual result.

Proposition 2.2. *Given a vector function $\vec{f} : \Omega \to \mathbb{R}^3$, with $\Omega \subseteq \mathbb{R}$, if $\vec{f}$ is differentiable in t_0, then:*

$$\frac{d\vec{f}}{dt}(t_0) = \left\langle \frac{dx(t_0)}{dt}, \frac{dy(t_0)}{dt}, \frac{dz(t_0)}{dt} \right\rangle$$

where $\vec{f}(t_0) = \langle x(t_0), y(t_0), z(t_0)\rangle$

Let's now see the properties of total derivatives.

Proposition 2.3. *Let $\vec{f}, \vec{g} : \Omega \to \mathbb{R}^3$ be two differentiable vector functions, with $\Omega \subseteq \mathbb{R}$, $h : \Omega \to \mathbb{R}$ a differentiable real-valued function, and let $a, b \in \mathbb{R}$ be two real numbers. Then:*

1. *Linearity:*

$$\frac{d}{dt}(a\vec{f}(t) + b\vec{g}(t)) = a\frac{d\vec{f}(t)}{dt} + b\frac{d\vec{g}(t)}{dt}$$

2. *Scalar action rule:*

$$\frac{d}{dt}(h(t)\vec{f}(t)) = \frac{dh(t)}{dt}\vec{f}(t) + h(t)\frac{d\vec{f}(t)}{dt}$$

3. *Scalar product rule:*

$$\frac{d}{dt}(\vec{f}(t)\cdot\vec{g}(t)) = \frac{d\vec{f}(t)}{dt}\cdot\vec{g}(t) + \vec{f}(t)\cdot\frac{d\vec{g}(t)}{dt}$$

4. *Vector product rule:*

$$\frac{d}{dt}(\vec{f}(t)\times\vec{g}(t)) = \frac{d\vec{f}(t)}{dt}\times\vec{g}(t) + \vec{f}(t)\times\frac{d\vec{g}(t)}{dt}$$

5. *Chain rule:*

$$\frac{d}{dt}\vec{f}(h(t)) = \frac{dh(t)}{dt}\frac{d\vec{f}}{dt}(h(t)) = h'(t)\vec{f}'(h(t))$$

Example 2.7. Let's evaluate the total derivative of the following function:

$$\langle e^{\sin t}, \ln t, t^4+2t\rangle \cdot \langle 2t^2-1, \sqrt{t}, e^t\rangle$$

as follows:

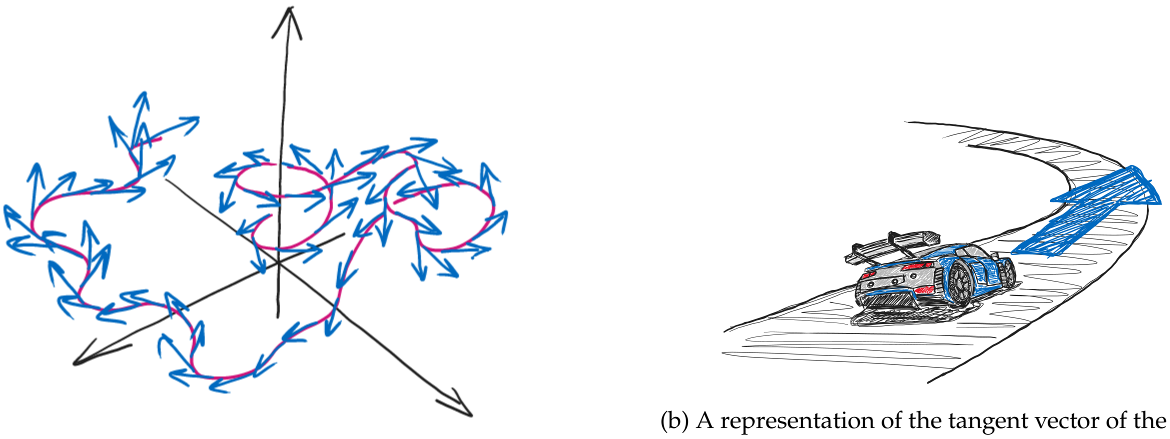
$$\begin{aligned}
&\frac{d}{dt}\left(\langle e^{\sin t}, \ln t, t^4+2t\rangle \cdot \langle 2t^2-1, \sqrt{t}, e^t\rangle\right) = \\
= \;& \frac{d}{dt}\left(\langle e^{\sin t}, \ln t, t^4+2t\rangle\right) \cdot \langle 2t^2-1, \sqrt{t}, e^t\rangle + \langle e^{\sin t}, \ln t, t^4+2t\rangle \cdot \frac{d}{dt}\left(\langle 2t^2-1, \sqrt{t}, e^t\rangle\right) = \\
= \;& \left\langle \frac{d}{dt}e^{\sin t}, \frac{d}{dt}\ln t, \frac{d}{dt}(t^4+2t)\right\rangle \cdot \langle 2t^2-1, \sqrt{t}, e^t\rangle + \langle e^{\sin t}, \ln t, t^4+2t\rangle \cdot \left\langle \frac{d}{dt}(2t^2-1), \frac{d}{dt}\sqrt{t}, \frac{d}{dt}e^t\right\rangle = \\
= \;& \left\langle \cos t e^{\sin t}, \frac{1}{t}, 4t^3+2\right\rangle \cdot \langle 2t^2-1, \sqrt{t}, e^t\rangle + \langle e^{\sin t}, \ln t, t^4+2t\rangle \cdot \left\langle 4t, \frac{1}{2\sqrt{t}}, e^t\right\rangle = \\
= \;& \cos t e^{\sin t}(2t^2-1) + \frac{1}{t}\sqrt{t} + 4t^3 + 2e^t + 4te^{\sin t} + \frac{\ln t}{2\sqrt{t}} + (t^4+2t)e^t
\end{aligned}$$

where we used the scalar product rule and Proposition 2.2. ■

2.4 Curves and the geometrical meaning of total derivatives

Imagine playing a video game where your goal is to drive a little race car along a road and do your best time. To make it difficult, the road has some curves and when it goes up the car goes slower, because it takes more energy to go on a hill, and it accelerates when it goes down. So, here we go, we have this fun situation and we want to understand if we can find some mathematical tools that help us to drive the car.

We can model the road as the path of a function $\vec{f} : [0,1] \to \mathbb{R}^3$, where we decide to take as the domain of this function the interval $[0,1]$, so that at time $t = 0$, the car is at the start of the road and at time 1 to reaches the finish line.



(a) The tangent vectors of a smooth path

(b) A representation of the tangent vector of the road the race car has to follow

Figure 2.8: The notion of tangent vectors

The first information we need is the direction and orientation of the road. This information is given by the total derivative of $\vec{f}$. However, we don't really need the modulus of $\vec{f}'$, so we decide to remove by diving by $\left|\vec{f}'(t)\right|$. Note that it makes sense to do that, because if the derivative is $\vec{0}$, $\vec{f}$ is constant, i.e. $\vec{f}$ is only a point. We are assuming that the road is not pointwise, so $\vec{f}' \neq \vec{0}$.

Definition 2.9. *Given a vector function $\vec{f} : \Omega \to \mathbb{R}^3$, with $\Omega \subseteq \mathbb{R}$, if $\vec{f}$ is differentiable in $t_0 \in \Omega$ and $\vec{f}'(t_0) \neq \vec{0}$, then the* ***tangent vector*** *of $\vec{f}$ at t_0 is the versor so defined:*

$$\vec{\mathsf{T}}(t_0) := \frac{\vec{f}'(t_0)}{\left|\vec{f}'(t_0)\right|}$$

In Picture 2.8a you can see a representation of tangent vectors for a vector function. In our situation, (see Picture 2.8b), we can interpret the tangent vector as a road sign added at each point, so that, just by looking that $\vec{\mathsf{T}}(t)$, we exactly know where to go to follow the road.

However, if you are like me in driving, you wanna have someone that tells you when to turn in time and knowing only the direction and orientation of the road at each time is just too hard because you could have a sudden curve and fly away. We need something that tells us when to make a turn.

The problem we need to solve is to find some indicator of a variation in the direction of the road. But the direction is given by $\vec{\mathsf{T}}(t)$ and to measure variations we usually use derivatives. Note that also $\vec{\mathsf{T}} \colon [0, 1] \to \mathbb{R}^3$ is a vector function, whenever $\vec{f}$ is differentiable. Note however, that, even if $\vec{\mathsf{T}}(t)$ is a unit vector, there is no reason why $\vec{\mathsf{T}}'$ should be a unit vector as well.

Definition 2.10. *Suppose that $\vec{f} : \Omega \to \mathbb{R}^3$ is a vector function, with $\Omega \subseteq \mathbb{R}$, which is differentiable, $\vec{f}'$ is differentiable in t_0 as well, and $\vec{f}'(t_0), \vec{f}''(t_0) \neq 0$. Then, the* ***normal vector*** *of $\vec{f}$ at t_0 is the versor so defined:*

$$\vec{\mathsf{N}}(t_0) := \frac{\vec{\mathsf{T}}'(t_0)}{|\vec{\mathsf{T}}'(t_0)|}$$

where $\vec{\mathsf{T}}(t_0)$ is the tangent vector of $\vec{f}$ at t_0.

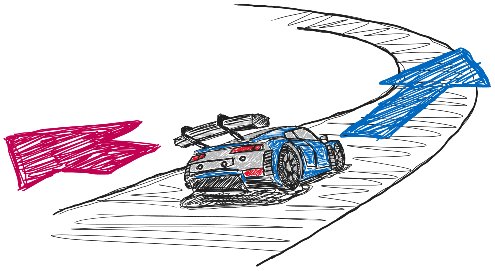


Figure 2.9: The tangent vector (in blue) and the normal vector (in red) of the road the race car has to follow

Contrary to the tangent vector, the normal vector of a function is not always defined: even when $\vec{f}$ is twice-differentiable, i.e. $\vec{f}$ is differentiable and so is its total derivative $\vec{f}'$, in order to define $\vec{\mathsf{N}}(t_0)$, $\vec{\mathsf{T}}'(t_0)$ needs to be not null. But $\vec{\mathsf{T}}'(t)$ is null if and only if $\vec{\mathsf{T}}(t)$ is constant around t, i.e. when there is no turn. This makes sense because if there is no turn, we don't need someone that tells us to go straight.

Now, imagine being the pilot of the race car. In front of you, on the windshield, you see a big arrow that tells you to go in a specific direction. That is $\vec{\mathsf{T}}(t)$. On the passenger side, there's a person that is telling you when to turn and they also tell you if you should turn left or right and they don't say anything when you need to go straight. This is $\mathbb{N}(t)$. To do that, they are pointing their index finger on the left or on the right and to be sure that there is no mistake of misinterpretation, the direction they point their finger is always orthogonal to the direction of $\vec{\mathsf{T}}(t)$.

Proposition 2.4. *Suppose that the tangent vector $\vec{\mathsf{T}}(t_0)$ and the normal vector $\vec{\mathsf{N}}(t_0)$ of a vector function $\vec{f}$ at time t_0 are both well-defined. Then, they are orthogonal.*

Proof. Proposition 1.5 tells that two non-zero vectors $\vec{u}$ and $\vec{v}$ are orthogonal if and only if $\vec{u} \cdot \vec{v} = 0$, so let's show that $\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{N}}(t_0) = 0$. To do that, let's start by taking the derivative of $\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{T}}(t_0)$:

$$\begin{aligned} & \frac{d}{dt}(\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{T}}(t_0)) = \\ = \quad & \vec{\mathsf{T}}'(t_0) \cdot \vec{\mathsf{T}}(t_0) + \vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{T}}'(t_0) = \\ = \quad & 2\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{T}}'(t_0) \end{aligned}$$

However, $\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{T}}(t_0) = \left|\vec{\mathsf{T}}(t_0)\right|^2 = 1$, thus $d/dt(\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{T}}(t_0)) = 0$, thus $\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{T}}'(t_0) = 0$. But, this is also true, even if we multiply by $1/\left|\vec{\mathsf{T}}'(t_0)\right|$, which is just a number, thus:

$$\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{N}}(t_0) = \vec{\mathsf{T}}(t_0) \cdot \frac{\vec{\mathsf{T}}'(t_0)}{\left|\vec{\mathsf{T}}'(t_0)\right|} = \frac{\vec{\mathsf{T}}(t_0) \cdot \vec{\mathsf{T}}'(t_0)}{\left|\vec{\mathsf{T}}'(t_0)\right|} = 0$$

This proves the result. □

In Picture 2.9 we drew the race car on the path, with $\vec{\mathsf{T}}(t)$ and $\vec{\mathsf{N}}(t)$. We need a third piece of the puzzle: so far we know where to go thanks to $\vec{\mathsf{T}}(t)$, we need when to stir and in which direction, having defined $\vec{\mathsf{N}}(t)$, we still don't know how quick we need to stir. A curve can be very wide or sharp. This information is contained in the curvature.

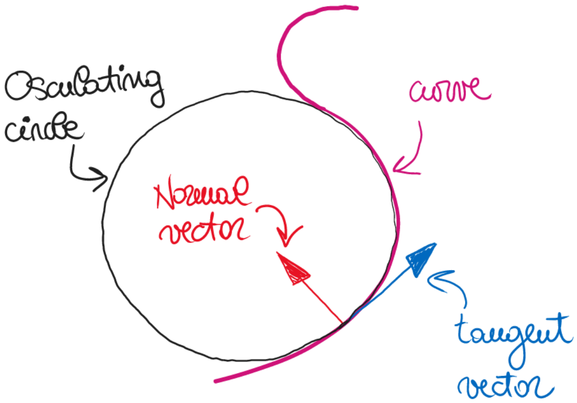


Figure 2.10: The osculating circle is the best circle that approximates a curve

---

Definition 2.11. *Suppose that $\vec{f} : \Omega \to \mathbb{R}^3$ is a vector function, with $\Omega \subseteq \mathbb{R}$, which is twice-differentiable in t_0 and that $\vec{f}'(t_0) \neq 0$. Then, the* ***curvature*** *of $\vec{f}$ at t_0 is the non-negative number $\kappa(t_0)$ so defined:*

$$\kappa(t_0) := \frac{\left|\vec{\mathsf{T}}'(t_0)\right|}{\left|\vec{f}'(t_0)\right|}$$

where $\vec{\mathsf{T}}(t_0)$ is the tangent vector of $\vec{f}$ at t_0.

To understand this definition, let ds denote the infinitesimal length of the curve. Then, thanks to the chain rule we can write:

$$\frac{d\vec{\mathsf{T}}}{dt} = \frac{ds}{dt}\frac{d\vec{\mathsf{T}}}{ds}$$

But ds/dt is just $\left|\vec{f}'(t)\right|$, so we can rewrite:

$$\kappa(t) = \left|\frac{d\vec{\mathsf{T}}(t)}{ds}\right|$$

Intuitively, this is a measure of the variation of $\vec{\mathsf{T}}$ w.r.t. the length of the curve. Geometrically, the curvature gives the radius of the best circle that approximates the curve at time t. This circle has a fancy name.

Definition 2.12. *Whenever the tangent vector $\vec{\mathsf{T}}(t_0)$ and the normal vector $\vec{\mathsf{N}}(t_0)$ are both well-defined, the plane generated by $\vec{\mathsf{T}}(t_0)$ and $\vec{\mathsf{N}}(t_0)$ is called the* ***osculating plane*** *of $\vec{f}$ at time t_0. Moreover, the circle of radius $1/\kappa(t_0)$, on the osculating plane and its centre is $\vec{f}(t_0) + \frac{1}{\kappa(t_0)}\vec{\mathsf{N}}(t_0)$ is called* ***osculating circle*** *of the curve $\vec{f}$ at time t_0.*

On Picture 2.10 there is a representation of the osculating circle. Finally, we point out that the curvature can also be obtained as follows:

$$\kappa(t) = \frac{\left|\vec{f}'(t) \times \vec{f}''(t)\right|}{\left|\vec{f}'(t)\right|^3}$$

One final concept: the binormal vector.

Definition 2.13. *Whenever the tangent vector* $\vec{\mathsf{T}}(t_0)$ *and the normal vector* $\vec{\mathsf{N}}(t_0)$ *are both well-defined, the* ***binormal vector*** *is the vector so defined:*

$$\vec{\mathsf{B}}(t_0) := \vec{\mathsf{T}}(t_0) \times \vec{\mathsf{N}}(t_0)$$

Example 2.8. Let's take into account again the spiral $\vec{f}(t) := \sin t\hat{\mathbf{i}} + \cos t\hat{\mathbf{j}} + t\hat{\mathbf{k}}$ and let's compute $\vec{\mathsf{T}}(t), \vec{\mathsf{N}}(t), \vec{\mathsf{B}}(t)$ and $\kappa(t)$.

$$\vec{f}'(t) = \cos t\hat{\mathbf{i}} - \sin t\hat{\mathbf{j}} + \hat{\mathbf{k}}$$

$$\left|\vec{f}'(t)\right| = \sqrt{\cos^2 t + \sin^2 t + 1^2} = \sqrt{2}$$

$$\vec{\mathsf{T}}(t) = \frac{\vec{f}'(t)}{\left|\vec{f}'(t)\right|} = \frac{1}{\sqrt{2}}\cos t\hat{\mathbf{i}} - \frac{1}{\sqrt{2}}\sin t\hat{\mathbf{j}} + \frac{1}{\sqrt{2}}\hat{\mathbf{k}}$$

$$\vec{\mathsf{T}}'(t) = -\frac{1}{\sqrt{2}}\sin t\hat{\mathbf{i}} - \frac{1}{\sqrt{2}}\cos t\hat{\mathbf{j}}$$

$$\left|\vec{\mathsf{T}}'(t)\right| = \frac{1}{\sqrt{2}}\sqrt{\sin^2 t + \cos^2 t} = \frac{1}{\sqrt{2}}$$

$$\vec{\mathsf{N}}(t) = \frac{\vec{\mathsf{T}}'(t)}{\left|\vec{\mathsf{T}}'(t)\right|} = -\frac{1}{2}\sin t\hat{\mathbf{i}} - \frac{1}{2}\cos t\hat{\mathbf{j}}$$

$$\kappa(t) = \frac{\left|\vec{\mathsf{T}}'(t)\right|}{\left|\vec{f}'(t)\right|} = \frac{\frac{1}{\sqrt{2}}}{\sqrt{2}} = \frac{1}{2}$$

Finally, the osculating circle of the spiral, is the circle of radius $1/\kappa(t) = 2$ and centre $\vec{f}(t) + (1/\kappa(t))\vec{\mathsf{N}}(t) = \sin t\hat{\mathbf{i}} + \cos t\hat{\mathbf{j}} + t\hat{\mathbf{k}} - \sin t\hat{\mathbf{i}} - \cos t\hat{\mathbf{j}} = t\hat{\mathbf{k}}$. ■

2.5 Review

Here is the list of important concepts of this module:

1. Limits & continuity
    - (a) Open, closed balls and spheres
    - (b) Open and closed sets
    - (c) Interior, closure and boundary of a set
    - (d) Definition of limits for multivariable functions
    - (e) The limit of a vector function
    - (f) Continuity for multivariable functions
2. Differentiability
    - (a) Linear functions
    - (b) Differentiability
    - (c) Differential
    - (d) Total derivative of a vector function

    (e) Properties: linearity, scalar action, scalar product, vector product and chain rules

3. Curves
    (a) Tangent vector
    (b) Normal vector
    (c) Orthogonality between normal and tangent vectors
    (d) Curvature
    (e) Osculating circle
    (f) Binormal vector

MODULE 3

Partial derivatives

3.1 Differentiability, directional & partial derivatives

In the previous chapter, we defined the notion of differentiability for a multivariable function. We said that a function $\vec{f}: \ : \Omega \to \mathbb{R}^m$, with $\Omega \subseteq \mathbb{R}^n$, is differentiable at a point $\vec{x}_0 \in \Omega$ if there is a linear application $\mathsf{d}\vec{f}_{\vec{x}_0}: \mathbb{R}^n \longrightarrow \mathbb{R}^m$, called the differential of $\vec{f}$ at $\vec{x}_0$, so that the function $\vec{f}$ can be approximated as follows:

$$\vec{f}(\vec{x}) \simeq \vec{f}(\vec{x}_0) + \mathsf{d}\vec{f}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$$

for $\vec{x}$ close enough to $\vec{x}_0$. Now, recall that a linear function $\vec{L} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a function that behaves well with the scalar action and the sum, i.e. given $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$:

$$\vec{L}(a\vec{x} + b\vec{y}) = a\vec{L}(\vec{x}) + b\vec{L}(\vec{y})$$

and, by definition, the differential of $\vec{f}$ at $\vec{x}_0$ is linear, therefore we have:

$$\mathsf{d}\vec{f}_{\vec{x}_0}(a\vec{x} + b\vec{y}) = a\mathsf{d}\vec{f}_{\vec{x}_0}(\vec{x}) + b\mathsf{d}\vec{f}_{\vec{x}_0}(\vec{y})$$

Now, when $n > 1$, i.e. the domain of the function is multidimensional, the notion of total derivative doesn't work anymore. The issue is that, in higher dimensions, the direction from which we approach a point matters. So, instead of defining a total derivative, we instead select a direction, represented by a versor $\hat{u}$, and we slice the graph of the function with a plane that passes through $\vec{x}_0$ and has direction $\hat{u}$ and then we evaluate the derivative along that plane. In Picture 3.1 you can see a representation of this idea. Let's introduce this concept once and for all.

Definition 3.1. *Let $\vec{f} : \Omega \to \mathbb{R}^m$ be a multivariable function with $\Omega \subseteq \mathbb{R}^n$ which is differentiable at $\vec{x}_0 \in \Omega$. Consider now a versor $\hat{u} \in \mathbb{R}^n$. The* ***directional derivative*** *of $\vec{f}$ at $\vec{x}_0$ along the direction of $\hat{u}$ is defined as follows:*

$$\frac{\partial \vec{f}}{\partial u}(\vec{x}_0) := \lim_{h \to 0} \frac{\vec{f}(\vec{x}_0 + h\hat{u}) - \vec{f}(\vec{x}_0)}{h}$$

We denote the directional derivative of $\vec{f}$ along $\hat{u}$ by $\partial\vec{f}/\partial u$, or $\partial_u \vec{f}$. We note that it is also common to denote directional derivatives by $\vec{f}_u$, but we decide to stick with the first two. instead.

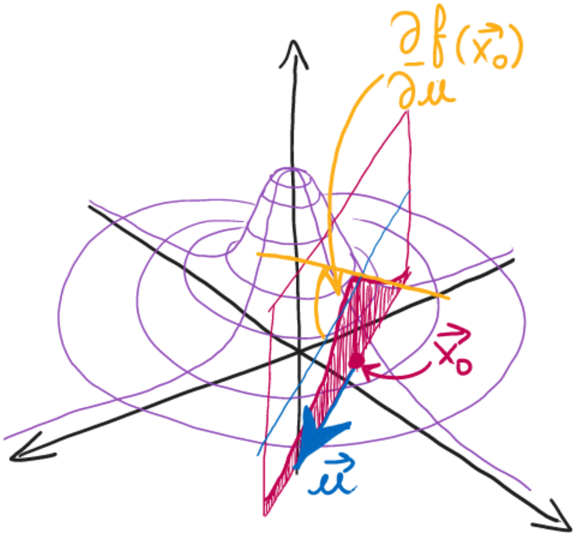


Figure 3.1: A representation of the directional derivative $\partial_u f(\vec{x}_0)$ in a point $\vec{x}_0$ along the direction of the versor $\hat{u}$

Example 3.1. Consider the so-called bivariate Gaussian function, which is the function so defined:

$$\begin{aligned} & g : \mathbb{R}^2 \to \mathbb{R} \\ & g(x,y) := e^{-\frac{x^2+y^2}{2}} \end{aligned}$$

The usual one-dimensional Gaussian function, known also as the bell curve, is a bell-shaped strictly positive function that has many many applications in mathematics, especially statistics. The bivariate Gaussian function is the two-dimensional version of this function and it can be seen as the surface obtained by rotating the Gaussian around the centre.

We would like to calculate the directional derivative of g at the point $(x_0, y_0) = (1/2, 1/2)$ along the direction $\hat{u} = (u, v) = (1/\sqrt{5}, 2/\sqrt{5})$:

$$\begin{aligned} & \partial_u g(x_0,y_0) = \\ = \ & \lim_{h\to 0} \frac{g(x_0+hu, y_0+hv) - g(x_0,y_0)}{h} = \\ = \ & \lim_{h\to 0} \frac{g\left(\frac{1}{2}+\frac{h}{\sqrt{5}}, \frac{1}{2}+h\frac{2}{\sqrt{5}}\right) - g\left(\frac{1}{2},\frac{1}{2}\right)}{h} = \\ = \ & \lim_{h\to 0} \frac{e^{-\frac{\left(\frac{1}{2}+\frac{h}{\sqrt{5}}\right)^2+\left(\frac{1}{2}+\frac{2h}{\sqrt{5}}\right)^2}{2}} - e^{-\frac{1}{4}}}{h} \end{aligned}$$

Now, let's compute the first exponent separately:

$$\begin{aligned} & -\frac{\left(\frac{1}{2}+\frac{h}{\sqrt{5}}\right)^2+\left(\frac{1}{2}+\frac{2h}{\sqrt{5}}\right)^2}{2} = \\ = \ & -\frac{1}{2}\left(\frac{1}{4}+\frac{h^2}{5}+\frac{h}{\sqrt{5}}+\frac{1}{4}+\frac{4h^2}{5}+\frac{2h}{\sqrt{5}}\right) = \end{aligned}$$

$$\begin{aligned} &= -\frac{1}{2}\left(h^2+\frac{3}{\sqrt{5}}h+\frac{1}{2}\right)= \\ &= -\frac{1}{2}h^2-\frac{3}{2\sqrt{5}}h-\frac{1}{4} \end{aligned}$$

So, we obtain the following limit:

$$\partial_u g(x_0,y_0)=\lim_{h\to 0}\frac{e^{-\frac{1}{2}h^2-\frac{3}{2\sqrt{5}}h-\frac{1}{4}}-e^{-\frac{1}{4}}}{h}$$

To solve this limit we can use l'Hopital:

$$\begin{aligned} &\lim_{h\to 0}\frac{e^{-\frac{1}{2}h^2-\frac{3}{2\sqrt{5}}h-\frac{1}{4}}-e^{-\frac{1}{4}}}{h}= \\ \overset{\mathsf{H}}{=}\ &\lim_{h\to 0}\frac{\left(-h-\frac{3}{2\sqrt{5}}\right)e^{-\frac{1}{2}h^2-\frac{3}{2\sqrt{5}}h-\frac{1}{4}}}{1}= \\ =\ &-\frac{3}{2\sqrt{5}}e^{-\frac{1}{4}} \end{aligned}$$

So, $\partial_u g(1/2,1/2)=-3e^{-1/4}/(2\sqrt{5})$.

■

But what is the relationship between directional derivatives and differential? In the previous chapter, we saw that, for vector functions, i.e. when the domain is one-dimensional, the differential is the multiplicative function $\mathsf{d}\vec{f}_{t_0}(t)=t\cdot d\vec{f}/dt(t_0)$, so the differential provides the total derivative. What about the directional derivatives?

Proposition 3.1. *Suppose that $\vec{f}$ is differentiable in $\vec{x}_0$ and that $\hat{u}$ is a versor of $\mathbb{R}^n$. Then the directional derivative of $\vec{f}$ at $\vec{x}_0$ along $\hat{u}$ exists and is equal to $\mathsf{d}\vec{f}_{\vec{x}_0}(\hat{u})$, i.e.:*

$$\partial_u\vec{f}(\vec{x}_0)=\mathsf{d}\vec{f}_{\vec{x}_0}(\hat{u})$$

Proof. Remember that, because $\vec{f}$ is differentiable in $\vec{x}_0$ we can write:

$$\vec{f}(\vec{x})\simeq\vec{f}(\vec{x}_0)+\mathsf{d}\vec{f}_{\vec{x}_0}(\vec{x}-\vec{x}_0)$$

for any $\vec{x}$ close enough to $\vec{x}_0$. Let's then imagine taking $\vec{x}=\vec{x}_0+h\hat{u}$, where $h>0$ is a very small strictly positive number. Note that $\vec{x}$ so defined is very close to $\vec{x}_0$ and it has the direction of $\vec{u}$. Intuitively speaking, $\vec{x}$ is one h far from $\vec{x}_0$ is the direction $\hat{u}$. Then, we have:

$$\vec{f}(\vec{x}_0+h\hat{u})\simeq\vec{f}(\vec{x}_0)+\mathsf{d}\vec{f}_{\vec{x}_0}(\vec{x}_0+h\hat{u}-\vec{x}_0)=\vec{f}(\vec{x}_0)+\mathsf{d}\vec{f}_{\vec{x}_0}(h\hat{u})$$

But now, remember that $\mathsf{d}\vec{f}_{\vec{x}_0}$ is linear, therefore we have:

$$\vec{f}(\vec{x}_0+h\hat{u})\simeq\vec{f}(\vec{x}_0)+h\mathsf{d}\vec{f}_{\vec{x}_0}(\hat{u})$$

Now, rearranging, we obtain:

$$\mathsf{d}\vec{f}_{\vec{x}_0}(\hat{u})\simeq\frac{\vec{f}(\vec{x}_0+h\hat{u})-\vec{f}(\vec{x}_0)}{h}$$

But now, recall that $\simeq$ means that the two expressions are equal as long as h is very small, so we have:

$$\mathrm{d}\vec{f}_{\vec{x}_0}(\hat{u}) = \lim_{h \to 0} \frac{\vec{f}(\vec{x}_0 + h\hat{u}) - \vec{f}(\vec{x}_0)}{h}$$

But now we can see that the right term is precisely the directional derivative of $\vec{f}$ at $\vec{x}_0$ along $\hat{u}$. So we conclude that $\partial_u \vec{f}(\vec{x}_0) = \mathrm{d}\vec{f}_{\vec{x}_0}(\hat{u})$. □

Now we have a nice relationship between the differential and directional derivatives. However, the differential is a linear function, so we don't really need to know all the directional derivatives to know what is the differential.
To understand that, take the standard basis of $\mathbb{R}^n$, which is so defined:

$$\begin{aligned}
\vec{v}_1 &:= (1, 0, 0, \ldots, 0) \\
\vec{v}_2 &:= (0, 1, 0, \ldots, 0) \\
&\vdots \\
\vec{v}_n &:= (0, 0, \ldots, 0, 1)
\end{aligned}$$

So, $\vec{v}_k$ is the n-tuple with zero everywhere but in the kth position, where it has a 1. Remember also that a basis is a collection of vectors in a vector space so that every other vector can be written as a linear combination of these vectors in a unique way. So, if $\vec{u}$ is any vector in $\mathbb{R}^n$, there is a unique n-tuple of real numbers $u_1, u_2, \ldots, u_n \in \mathbb{R}$, so that:

$$\vec{u} = u_1\vec{v}_1 + u_2\vec{v}_2 + \cdots + u_n\vec{v}_n$$

So, when we have a linear function $\vec{L} : \mathbb{R}^n \to \mathbb{R}^m$, we can write:

$$\vec{L}(\vec{u}) = \vec{L}(u_1\vec{v}_1 + u_2\vec{v}_2 + \cdots + u_n\vec{v}_n) = u_1\vec{L}(\vec{v}_1) + u_2\vec{L}(\vec{v}_2) + \ldots u_n\vec{L}(\vec{v}_n)$$

So, once we have determined the vectors $\vec{L}(\vec{v}_1), \vec{L}(\vec{v}_2), \ldots, \vec{L}(\vec{v}_n) \in \mathbb{R}^m$, we know what is $\vec{L}(\vec{u})$ for any possible vector $\vec{u} \in \mathbb{R}^n$. Let's see what this discussion implies for the differential. Consider a versor $\hat{u} = \langle u_1, u_2, \ldots, u_n \rangle \in \mathbb{R}^n$. Therefore:

$$\mathrm{d}\vec{f}_{\vec{x}_0}(\hat{u}) = \mathrm{d}\vec{f}_{\vec{x}_0}(u_1\vec{v}_1 + u_2\vec{v}_2 + \cdots + u_n\vec{v}_n) = u_1\mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_1) + u_2\mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_2) + \cdots + u_3\mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_n)$$

The vectors $\mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_1), \mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_2), \ldots, \mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_n)$ have a special name.

Definition 3.2. *The* ***partial derivatives*** *of a multivariable function $\vec{f} : \Omega \to \mathbb{R}^m$ at $\vec{x}_0$ are the vectors in $\mathbb{R}^m$ so defined:*

$$\begin{aligned}
\frac{\partial \vec{f}}{\partial x_1}(\vec{x}_0) &:= \mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_1) \\
\frac{\partial \vec{f}}{\partial x_2}(\vec{x}_0) &:= \mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_2) \\
&\vdots \\
\frac{\partial \vec{f}}{\partial x_n}(\vec{x}_0) &:= \mathrm{d}\vec{f}_{\vec{x}_0}(\vec{v}_n)
\end{aligned}$$

We are going to use the following notations $\partial\vec{f}/\partial x_k$ or $\partial_{x_k}\vec{f}$ to denote the kth partial derivative of $\vec{f}$.

Note that, thanks to Proposition 3.1, we know that $\mathsf{d}\vec{f}_{\vec{x}_0}(\vec{v}_k)$ corresponds to the directional derivative of the function $\vec{f}$ at $\vec{x}_0$ along the direction $\vec{v}_k$. So, we have the formula:

$$\frac{\partial \vec{f}}{\partial x_k}(\vec{x}_0) = \lim_{h\to 0} \frac{\vec{f}(\vec{x}_0 + h\vec{v}_k) - \vec{f}(\vec{x}_0)}{h}$$

Example 3.2. Let's evaluate the partial derivatives of the bivariate Gaussian function $g(x, y) = \exp -(x^2 + y^2)/2$:

$$\begin{aligned}
& \partial_x g(x, y) = \\
= \quad & \lim_{h\to 0} \frac{g\left((x, y) + h(1, 0)\right) - g(x, y)}{h} = \\
= \quad & \lim_{h\to 0} \frac{g(x + h, y) - g(x, y)}{h} = \\
= \quad & \lim_{h\to 0} \frac{e^{-\frac{(x+h)^2+y^2}{2}} - e^{-\frac{x^2+y^2}{2}}}{h} = \\
= \quad & \lim_{h\to 0} \frac{e^{-\frac{(x+h)^2}{2}} e^{-\frac{y^2}{2}} - e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}}}{h} = \\
= \quad & e^{-\frac{y^2}{2}} \lim_{h\to 0} \frac{e^{-\frac{(x+h)^2}{2}} - e^{-\frac{x^2}{2}}}{h}
\end{aligned}$$

Do you recognize what is this limit?
Well, that's the definition of the derivative of the function $\exp(-x^2/2)$, so we have:

$$\begin{aligned}
& \partial_x g(x, y) = \\
= \quad & e^{-\frac{y^2}{2}} \frac{d}{dx} e^{-\frac{x^2}{2}} = \\
= \quad & -x e^{-\frac{y^2}{2}} e^{-\frac{x^2}{2}} = \\
= \quad & -x e^{-\frac{x^2+y^2}{2}}
\end{aligned}$$

From a similar computation, we also find that:

$$\partial_y g(x, y) = -y e^{-\frac{x^2+y^2}{2}}$$

Let's now consider a versor $\hat{u} = \langle u, v\rangle$. We find out that we can calculate the directional derivatives of g at any point $(x, y) \in \mathbb{R}^2$ along $\hat{u}$ as follows:

$$\begin{aligned}
& \partial_u g(x, y) = \\
= \quad & \mathsf{d}g_{(x,y)}(\hat{u}) = \\
= \quad & \mathsf{d}g_{(x,y)}(u\hat{\mathbf{i}} + v\hat{\mathbf{j}}) = \\
= \quad & u\,\mathsf{d}g_{(x,y)}(\hat{\mathbf{i}}) + v\,\mathsf{d}g_{(x,y)}(\hat{\mathbf{j}}) = \\
= \quad & u\partial_x g(x, y) + v\partial_y(x, y)
\end{aligned}$$

Let's now consider the versor $\hat{u} = (1/\sqrt{5}, 2/\sqrt{5})$ and the point $(x_0, y_0) = (1/2, 1/2)$ of Example 3.1, so we have:

$$\partial_u g(x_0, y_0) =$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{5}}\partial_x g\left(\frac{1}{2},\frac{1}{2}\right) + \frac{2}{\sqrt{5}}\partial_y g\left(\frac{1}{2},\frac{1}{2}\right) = \\
&= \frac{1}{\sqrt{5}}(-xe^{-\frac{x^2+y^2}{2}})\bigg|_{x,y=1/2} + \frac{2}{\sqrt{5}}(-ye^{-\frac{x^2+y^2}{2}})\bigg|_{x,y=1/2} = \\
&= -\frac{1}{2\sqrt{5}}e^{-\frac{1}{4}} - \frac{2}{2\sqrt{5}}e^{-\frac{1}{4}} = \\
&= -\frac{3}{2\sqrt{5}}e^{-\frac{1}{4}}
\end{aligned}
$$

This corresponds to the result we found in Example 3.1, as expected. ■

Before we continue, we want to give a third way to think about partial derivatives. Consider, for example, a scalar function $f : \mathbb{R}^2 \to \mathbb{R}$. So, f is a function of the two variables x and y. Now, imagine we want to evaluate the partial derivatives of f in a point $(x_0, y_0) \in \mathbb{R}^2$. Let's start with $\partial_x f(x_0, y_0)$. By definition, the idea is that we are going to increment slightly the coordinate x_0 of a small quantity h, look at the difference $f(x_0 + h, y_0) - f(x_0, y_0)$ and divide this by h.

So, what we are doing here is to vary slightly the variable x and keep fixed the variable y. We can think at this process as defining a new function:

$$
\begin{aligned}
&{f_1}^{(y_0)} : \mathbb{R} \to \mathbb{R} \\
&{f_1}^{(y_0)}(x) := f(x, y_0)
\end{aligned}
$$

Note that now this is a one-variable function. So, if we take the derivative of ${f_1}^{(y_0)}$ in x_0 we obtain exactly $\partial_x f(x_0, y_0)$, i.e.:

$$\partial_x f(x_0, y_0) = \frac{d}{dx}{f_1}^{(y_0)}(x_0)$$

Similarly, we can introduce a function:

$$
\begin{aligned}
&{f_2}^{(x_0)} : \mathbb{R} \to \mathbb{R} \\
&{f_2}^{(x_0)}(y) := f(x_0, y)
\end{aligned}
$$

in the variable y and notice that:

$$\partial_y f(x_0, y_0) = \frac{d}{dy}{f_2}^{(x_0)}(y_0)$$

So, essentially, taking partial derivatives is equivalent to keeping fixed all the other variables and deriving only for a specific variable.

Example 3.3. Consider the following two functions:

$$
\begin{aligned}
&f(x, y) = xe^{-xy} + 2y \\
&\vec{g}(x, y, z) = \langle yz, 2xz, 3xy \rangle
\end{aligned}
$$

Let's evaluate the partial derivatives of them.
Let's start with f:

$$\partial_x f(x, y) = \partial_x(xe^{-xy} + 2y) = \partial_x xe^{-xy} = e^{-xy} - ye^{-xy}$$

$$\partial_y f(x,y) = \partial_y(xe^{-xy} + 2y) = -x^2 e^{-xy} + 2$$

Let's now evaluate the partial derivatives of $\vec{g}$:

$$\begin{aligned}
\partial_x \vec{g}(x,y,z) &= \langle \partial_x(yz), \partial_x(2xz), \partial_x(3xy)\rangle = \langle 0, 2z, 3y\rangle \\
\partial_x \vec{g}(x,y,z) &= \langle \partial_y(yz), \partial_y(2xz), \partial_y(3xy)\rangle = \langle z, 0, 3x\rangle \\
\partial_x \vec{g}(x,y,z) &= \langle \partial_z(yz), \partial_z(2xz), \partial_z(3xy)\rangle = \langle y, 2x, 0\rangle
\end{aligned}$$

Note that in the case of a function $\mathbb{R}^3 \to \mathbb{R}^3$, partial derivatives are vectors in the codomain of the function. ■

Example 3.4. The body-mass index is a number that relates the weight with the height of a person as follows; if m is the mass and h is the height, the body-mass index $b(m,h)$ is given by the formula:

$$b(m,h) := \frac{m}{h^2}$$

We want to understand, for a person tall 1.68 metres and weighting 64 kilograms, what happens to their body-mass index if they gain a small amount of weight: does $b(m,h)$ increase or decrease? And how fast does it vary?
Mathematically, this means that we need to calculate the partial derivative of the function b w.r.t. m:

$$\partial_m b(m,h) = \partial_m \frac{m}{h^2} = \frac{1}{h^2}$$

Note that $\partial_m b(m,h) > 0$. This means that the variation of b w.r.t. the mass is always positive. So, if the person is gaining a small amount of weight, their body-mass index will increase. Quantitatively, we have that $\partial_m b(64, 1.68) = 1/(1.68)^2 \simeq 0.351/m^2$. This means that, if the person gains for example 1000 grams of mass, their body-mass index varies of $\partial_m b(64,1.68) \cdot 1000g = 0.351m^{-2} \cdot 1kg = 0.351kg/m^2$. Since $b(64,1.68) \simeq 22.68kg/m^2$ we have that, the final body-mass index is approximately equal to $b(64,1.68) + \partial_m b(64,1.68) \cdot 1000g = 22.68 + 0.351 \simeq 23.03kg/m^2$. Note that this formula is the linear approximation of $b(m,h)$ so this works only for small variation of the mass.

Similarly, if a person grows of $5cm = 0.05m$, we can evaluate the variation of their body-mass index as follows:

$$\partial_h b(m,h) = \partial_h \frac{m}{h^2} = -2\frac{m}{h^3}$$

So we have:

$$\partial_h b(64,1.68) = -2\frac{64}{(1.68)^3} \simeq -26.99kg/m$$

So, the variation this time is negative. This means that if the person grows 5 centimetres, without changing their weight, the final body-mass index is about (again using the linear approximation) $b(64,1.68) + \partial_h b(64,1.68) \cdot 0.05m = 22.68 - 26.99 \cdot 0.05 \simeq 21.33kg/m^2$. ■

3.2 Differentiability & regularity

In the previous chapter, we studied differentiability for vector functions and we found out that being differentiable is the same as having total derivatives. When the domain of the function is not one-dimensional, e.g. $f : \mathbb{R}^2 \to \mathbb{R}$, then the situation is more complicated.

We saw in the previous section that in this case, we can define partial derivatives and one could think that having all partial derivatives is equivalent to be differentiable. However, this is not the case as the next example shows.

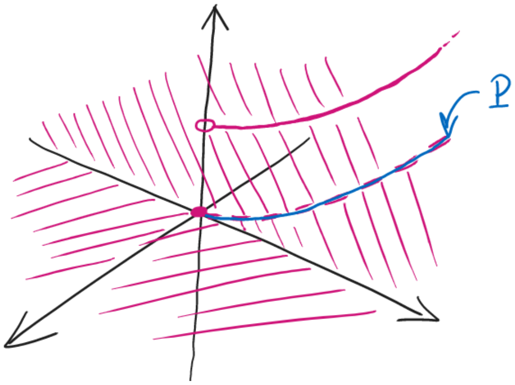


Figure 3.2: Representation of the graph of the function $f(x, y) = 1$, when $(x, y) \in P$ and $f(x, y) = 0$ when $(x, y) \notin P$.

Cool Example* 3.1. Consider the following subset $P \subseteq \mathbb{R}^2$ of the plane:

$$P := \{(x, y) \in \mathbb{R}^2, x > 0, y = x^2\}$$

and let's define the function f as follows:

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

$$f(x, y) := \begin{cases} 1 & \text{if } (x, y) \in P \\ 0 & \text{else} \end{cases}$$

In Picture 3.2 there is a representation of the graph of the function f. We want to see if the partial derivatives of f in $(0, 0)$ are well-defined. So we have:

$$\partial_x f(0, 0) = \lim_{h \to 0} \frac{f(h, 0) - f(0, 0)}{h}$$

However, for any $h \in \mathbb{R}$, $(h, 0) \neq P$, this is because if $(h, 0) \in P$, this means that $h > 0$ and $0 = h^2$, but the second equation means that $h = 0$, so we have a contradiction. Therefore, by definition of f, $f(h, 0) = 0$ and also $f(0, 0) = 0$, so we have that the limit is just zero, i.e. $\partial_x f(0, 0) = 0$. Let's now evaluate the other partial derivative:

$$\partial_y f(0, 0) = \lim_{h \to 0} \frac{f(0, h) - f(0, 0)}{h}$$

But again, $(0, h) \neq P$ because the x-component needs to be strictly positive to be a point of P, so again $\partial_y f(0, 0) = 0$. Note that this also shows that every directional derivative $\partial_u f(0, 0) = 0$. So, we obtained that all partial derivatives of f are well-defined and null. However, the function f is not even continuous in $(0, 0)$!!! To see that, notice that there is a jump in the graph in $(0, 0)$. Therefore, we conclude that f cannot be differentiable in $(0, 0)$ since differentiability implies continuity.
This shows that there are functions that are not differentiable, but they have all partial derivatives. ■

The previous example shows that, in order for a multivariable function to be differentiable is not sufficient to ask for the existence of all partial derivatives. However, if the partial derivatives are well-defined and also continuous then the function is differentiable. First, let's introduce an important definition.

Cool Definition* 3.1. *Let $\vec{f} : \Omega \to \mathbb{R}^m$ be a multivariable function, with $\Omega \subseteq \mathbb{R}^n$ and let $\vec{x}_0 \in \Omega$. We say that the function is of **class** C^0 in $\vec{x}_0$ if $\vec{f}$ is continuous at $\vec{x}_0$, of **class** C^1 at $\vec{x}_0$ if $\vec{f}$ has all the partial derivatives and each of them is a continuous function. Moreover, we say that a function is of **class** C^k in $\vec{x}_0$ if $\vec{f}$ has all partial derivatives, each of them has all partial derivatives, and all of these have partial derivatives, this k-times and all the k-th partial derivatives are all continuous functions.*

An equivalent way to say that a function is of class C^k is to say that the function is differentiable and all its partial derivatives are of class C^{k-1}. With this definition, we are able to give the main result of this section.

Proposition 3.2. *Given a multivariable function $\vec{f} : \Omega \to \mathbb{R}^m$, with $\Omega \subseteq \mathbb{R}^n$, if the function $\vec{f}$ is of class C^1 at $\vec{x}_0 \in \Omega$, then $\vec{f}$ is differentiable at $\vec{x}_0$.*

3.3 The tangent plane (or space)

What is the geometrical meaning of the differential and of the partial derivatives of a multivariable function? If we consider a function $f : \mathbb{R} \to \mathbb{R}$, we already know that the (total) derivative of f measures the slope of the tangent line of f in a particular point. We also know that the equation of the tangent line of f in a point x_0 is given by the equation:

$$y = f(x_0) + f'(x_0)(x - x_0)$$

The function $\lambda(x) := f(x_0) + f'(x_0)(x - x_0)$ is called the linear approximation of f in x_0, or linearization of f in x_0.

Similarly, for functions $f : \mathbb{R}^2 \to \mathbb{R}$, partial derivatives represent the slope of the tangent plane of f in a point $\vec{x}_0$, along the directions of the axes.

Definition 3.3. *Consider a multivariable function $\vec{f} : \Omega \to \mathbb{R}^m$, with $\Omega \subseteq \mathbb{R}^n$, differentiable in $\vec{x}_0 \in \Omega$. The linearization of $\vec{f}$ in $\vec{x}_0$, also called the* ***linear-approximation*** *of $\vec{f}$ in $\vec{x}_0$, is the function:*

$$\begin{aligned} &\vec{\lambda} : \mathbb{R}^n \to \mathbb{R}^m \\ &\vec{\lambda}(\vec{x}) := \vec{f}(\vec{x}_0) + \mathsf{d}\vec{f}_{\vec{x}_0}(\vec{x} + \vec{x}_0) \end{aligned}$$

If $\vec{\lambda}$ is the linearization of the function $\vec{f}$ in $\vec{x}_0$, then the equation of the tangent space of $\vec{f}$ in $\vec{x}_0$ is given by:

$$\vec{y} = \vec{\lambda}(\vec{x})$$

In the case of a function $f : \mathbb{R}^2 \to \mathbb{R}$ the equation becomes:

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

This is the equation of the tangent plane of f in (x_0, y_0).

Example 3.5. Consider the paraboloid:

$$z = 2x^2 + y^2$$

Let's find the tangent plane of the paraboloid at $(1, 1)$. To do that, let's first define the function:

$$f(x, y) := 2x^2 + y^2$$

and let's evaluate the partial derivatives of f:

$$\begin{aligned} &\frac{\partial f}{\partial x}(x, y) = 4x \\ &\frac{\partial f}{\partial y}(x, y) = 2y \end{aligned}$$

In particular, in $(1, 1)$, we have that $\partial_x f(1, 1) = 4$ and $\partial_y f(1, 1) = 2$. Therefore, the equation of the tangent plane of the paraboloid at $(1, 1)$ is given by:

$$z = f(1, 1) + \frac{\partial f}{\partial x}(1, 1)(x - 1) + \frac{\partial f}{\partial y}(1, 1)(y - 1) = 3 + 4(x - 1) + 2(y - 1) = 4x + 2y - 3$$

which can be rewritten as follows:

$$4x + 2y - z - 3 = 0$$

Note also that this equation tells also what is the normal vector of the paraboloid in $(1, 1)$, i.e. $\vec{n} = \langle 4, 2, -1 \rangle$. ■

Example 3.6. Let's take into account again the bivariate Gaussian function $g(x, y) = \exp(-(x^2 + y^2)/2)$ and let's find the equation of the tangent plane of g in the generic point (x_0, y_0):

$$\frac{\partial g}{\partial x}(x_0, y_0) = -x_0 e^{-\frac{x_0^2+y_0^2}{2}}$$
$$\frac{\partial g}{\partial y}(x_0, y_0) = -y_0 e^{-\frac{x_0^2+y_0^2}{2}}$$

Thus, the equation of the tangent plane of g at (x_0, y_0) is:

$$z = e^{-\frac{x_0^2+y_0^2}{2}} - x_0 e^{-\frac{x_0^2+y_0^2}{2}}(x - x_0) - y_0 e^{-\frac{x_0^2+y_0^2}{2}}(y - y_0)$$

Multiplying on both sides by $\exp((x_0^2 + y_0^2)/2)$, we obtain:

$$x_0 x + y_0 y + e^{\frac{x_0^2+y_0^2}{2}} z = x_0^2 + y_0^2$$

Thus, $\vec{n} = \langle x_0, y_0, \exp((x_0^2 + y_0^2)/2) \rangle$ is the normal vector of the graph of g at (x_0, y_0). ■

3.4 The chain rule

In this section, we are going to study the chain rule for multivariable functions. The chain rule describes the relationship between two operations of functions: composition and derivation. The composition is a binary operation that takes two functions $f : B \to C$ and $g : A \to B$ and returns a new function $f \circ g : A \to C$ defined as follows:

$$(f \circ g)(a) := f[g(a)]$$

for every $a \in A$.

Proposition 3.3. *Let's consider two multivariable functions $\vec{g} : \mathbb{R}^n \to \mathbb{R}^k$ and $\vec{f} : \mathbb{R}^k \to \mathbb{R}^l$. Suppose that $\vec{g}$ is differentiable in $\vec{x}_0 \in \mathbb{R}^n$ and that $\vec{f}$ is differentiable in $\vec{g}(\vec{x}_0) \in \mathbb{R}^k$. Thus, the function $\vec{f} \circ \vec{g} : \mathbb{R}^n \to \mathbb{R}^l$ is differentiable in $\vec{x}_0$ and the differential of such function is given by the following formula, known as the chain rule:*

$$\mathsf{d}(\vec{f} \circ \vec{g})_{\vec{x}_0} = \mathsf{d}\vec{f}_{\vec{y}_0} \circ \mathsf{d}\vec{g}_{\vec{x}_0}$$

where $\vec{y}_0 = \vec{g}(\vec{x}_0)$.

Let's now see in practice what this formula means. Let's start by considering a function $f : \mathbb{R}^2 \to \mathbb{R}$ and a function $\vec{g} : \mathbb{R} \to \mathbb{R}^2$. In this scenario, the function $\vec{g}$ gives two functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$ defined as follows:

$$\vec{g}(t) = \langle g_1(t), g_2(t) \rangle$$

Let's find the differential of f:

$$\mathsf{d}f_{(x_0,y_0)}(x, y) = \frac{\partial f}{\partial x}(x_0, y_0)x + \frac{\partial f}{\partial y}(x_0, y_0)y$$

Let's find the differential of $\vec{g}$:

$$\mathrm{d}\vec{g}_{t_0}(t) = \frac{d\vec{g}}{dt}(t_0)t = t\left\langle \frac{dg_1}{dt}(t_0), \frac{dg_2}{dt}(t_0)\right\rangle$$

Proposition 3.3 establishes then that the differential of the function $f \circ \vec{g}$ is given by the following formula:

$$\begin{aligned}
&\mathrm{d}(f \circ \vec{g})_{t_0}(t) = \left(\mathrm{d}f_{\vec{g}(t_0)} \circ \mathrm{d}g_{t_0}\right)(t) = \mathrm{d}f_{\vec{g}(t_0)}\left(\mathrm{d}g_{t_0}(t)\right) = \\
&= \mathrm{d}f_{(g_1(t_0), g_2(t_0))}\left(tg_1'(t_0), tg_2'(t_0)\right) = \\
&= \frac{\partial f}{\partial x}(g_1(t_0), g_2(t_0))\frac{dg_1}{dt}(t_0)t + \frac{\partial f}{\partial y}(g_1(t_0), g_2(t_0))\frac{dg_2}{dt}(t_0)t
\end{aligned}$$

However:

$$\mathrm{d}(f \circ \vec{g})_{t_0}(t) = \frac{d(f \circ \vec{g})}{dt}(t_0)t$$

Therefore, we obtain the formula:

$$\frac{d}{dt}\left[f(g_1(t), g_2(t))\right] = \frac{\partial f}{\partial x}(g_1(t_0), g_2(t_0))\frac{dg_1}{dt}(t_0) + \frac{\partial f}{\partial y}(g_1(t_0), g_2(t_0))\frac{dg_2}{dt}(t_0)$$

If we call $z = f(x, y)$ and $x(t) = g_1(t)$ and $y(t) = g_2(t)$ we can rewrite this expression as follows:

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

Example 3.7. Consider the following:

$$\begin{aligned}
&z = x^2y + 2xy^4 \\
&x = \sin(2t) \\
&y = \cos t
\end{aligned}$$

Then we have:

$$z(t) = \sin^2(2t)\cos t + 3\sin(2t)\cos^4 t$$

Let's take the total derivative of z by using the chain rule:

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}(\sin 2t, \cos t)\frac{dx}{dt} + \frac{\partial z}{\partial y}(\sin 2t, \cos t)\frac{dy}{dt}$$

Let's find the partial derivatives of z:

$$\begin{aligned}
&\partial_x z = 2xy + 3y^4 \\
&\partial_y z = x^2 + 12xy^3
\end{aligned}$$

Let's find the total derivatives of x and y:

$$\begin{aligned}
&x'(t) = 2\cos(2t) \\
&y'(t) = -\sin t
\end{aligned}$$

Thus:

$$\begin{aligned}
&z'(t) = \\
= \quad & \frac{\partial z}{\partial x}(\sin 2t, \cos t)\frac{dx}{dt} + \frac{\partial z}{\partial y}(\sin 2t, \cos t)\frac{dy}{dt} = \\
= \quad & (2\sin(2t)\cos t + 3\cos^4 t)2\cos(2t) - (\sin^2(2t) + 12\sin(2t)\cos^3 t)\sin t
\end{aligned}$$

This concludes the example. ■

Let's now consider the case when we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ and another function $\vec{g} : \mathbb{R}^2 \to \mathbb{R}^2$. In this case we obtain from $\vec{g}$ two functions $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$ defined as follows:

$$\vec{g}(s,t) := \langle g_1(s,t), g_2(s,t) \rangle$$

Now, the function $f \circ \vec{g} : \mathbb{R}^2 \to \mathbb{R}$ can be partially derived along the variable s or along t. Therefore, from the chain rule, we find the following two formulas:

$$\frac{\partial(f \circ \vec{g})}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial g_1}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial g_2}{\partial s}$$

$$\frac{\partial(f \circ \vec{g})}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial g_1}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial g_2}{\partial t}$$

and, if we call $z = f(x,y)$ and $x = g_1(t), y = g_2(t)$ we have:

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

Example 3.8. The equation of an ideal gas relates the pressure (P), volume (V) and temperature (T) as follows:

$$PV = nRT$$

where n is the number of moles of the gas (which means the number of molecules of the gas) and $R \simeq 8.31 L \cdot kPa/kMol$ is a constant. We will assume $n = 1mol$. If we decide to isolate the pressure (P) in the formula we obtain the following equation:

$$P(V,T) := R\frac{T}{V}$$

Suppose that the temperature at time $t = 0$ is $300K$ and that it's increasing at a rate of $0.1K/s$. So, $T(t) \simeq 300 + 0.1t$. Moreover, the volume V at time $t = 0$ is $100L$ and it's also increasing, at a rate of $0.2L/s$. So, $V(t) \simeq 100 + 0.2t$. We deduce that:

$$\frac{dT}{dt} = 0.1$$

$$\frac{dV}{dt} = 0.2$$

We want to find the rate of change of the pressure at time $t = 0$:

$$\frac{dP}{dt}(0) = \frac{\partial P}{\partial T}\frac{dT}{dt} + \frac{\partial P}{\partial V}\frac{dV}{dt}$$

Let's find out what are the partial derivatives of the pressure:

$$\frac{\partial P}{\partial T} = \frac{R}{V}$$
$$\frac{\partial P}{\partial V} = -\frac{RT}{V^2}$$

Therefore:

$$\begin{aligned}
&\frac{dP}{dt}(0) = \\
= \quad &\frac{\partial P}{\partial T}\frac{dT}{dt} + \frac{\partial P}{\partial V}\frac{dV}{dt} = \\
= \quad &\frac{R}{V}\frac{dT}{dt} - \frac{RT}{V^2}\frac{dV}{dt} = \frac{R}{V}\left(\frac{dT}{dt} - \frac{T}{V}\frac{dV}{dt}\right) = \\
= \quad &\frac{8.31}{200L}\left(0.1K/s - \frac{300K}{200L}\cdot 0.2L/s\right) \simeq 0.042kPa/s
\end{aligned}$$

Where one Pascal corresponds to $1Pa = 1N/m^2$ and it's the international unit of measure for pressure. ■

Finally, let's consider the general case. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $\vec{g} : \mathbb{R}^k \to \mathbb{R}^n$ as follows:

$$\vec{g}(t_1, \ldots, t_k) = \langle g_1(t_1, \ldots, t_k), \ldots, g_n(t_1, \ldots, t_k)\rangle$$

Thus:

$$\frac{\partial(f \circ \vec{g})}{\partial t_i} = \frac{\partial f}{\partial x_1}\frac{\partial g_1}{\partial t_i} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial g_n}{\partial t_i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\frac{\partial g_j}{\partial t_i}$$

Example 3.9. Consider the following function:

$$u(x, y, z) := x^4 y + y^2 z^3$$

and let:

$$x(r, s, t) = rse^t$$
$$y(r, s, t) = rs^2 e^{-t}$$
$$z(r, s, t) = r^2 s \sin t$$

Find $\partial_s u$ when $r = 2, s = 1$ and $t = 0$. From the chain rule we have:

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial s}$$

We have:

$$\frac{\partial u}{\partial x} = 4x^3 y \qquad \frac{\partial x}{\partial s} = re^t$$
$$\frac{\partial u}{\partial y} = x^4 + 2yz^3 \qquad \frac{\partial y}{\partial s} = 2rse^{-t}$$
$$\frac{\partial u}{\partial z} = 3y^2 z^2 \qquad \frac{\partial z}{\partial s} = r^2 \sin t$$

Moreover, for $r = 2, s = 1, t = 0$ we have:

$$x(2,1,0) = 2 \qquad \frac{\partial x}{\partial s}(2,1,0) = 2 \qquad \frac{\partial u}{\partial x}(2,2,0) = 64$$
$$y(2,1,0) = 2 \qquad \frac{\partial y}{\partial s}(2,1,0) = 4 \qquad \frac{\partial u}{\partial y}(2,2,0) = 16$$
$$z(2,1,0) = 0 \qquad \frac{\partial z}{\partial s}(2,1,0) = 0 \qquad \frac{\partial u}{\partial z}(2,2,0) = 0$$

Therefore:

$$\frac{\partial u}{\partial s}(2,1,0) = 64 \cdot 2 + 16 \cdot 4 + 0 \cdot 0 = 192$$

This concludes the example. ■

3.5 Implicit differentiation

Implicit differentiation is used when we want to derive an expression w.r.t. one of the variables but we cannot fully isolate the variable we are interested in. For example, consider the equation of the unit sphere $\mathbb{S}^2$:

$$x^2 + y^2 + z^2 = 1$$

Suppose we want to isolate the variable z, so we end up with:

$$z^2 = 1 - x^2 - y^2$$

If we want to express z as a function of x and y, we need to take the square root, but this lead to two separate cases:

$$z = \sqrt{1 - x^2 - y^2}$$
$$z = -\sqrt{1 - x^2 - y^2}$$

This is because the sphere is not the graph of any function $z = f(x, y)$.
Instead of splitting the problem in two or more subproblems, we can use implicit differentiation. The idea is to consider the expression we have as an equation of the following form:

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

Where $F : \mathbb{R}^3 \to \mathbb{R}$ is now a scalar function in the three variables x, y and z. In the example of the sphere, we have:

$$F(x, y, z) := x^2 + y^2 + z^2 - 1$$

So, the equation $F(x, y, z) = 0$ is precisely the equation of the sphere. Note that such an equation is sometimes called a constraint, because it imposes a constraint to the three variables x, y, z, in this case, to represent a point of the unit sphere.
Now, we are interested in understanding the rate of change of the variable z w.r.t. the variable x, for example, so what we can do is to partially differentiate F along the variable x, i.e.:

$$0 = \frac{\partial F}{\partial x}(x, y, z) = \frac{\partial F}{\partial x}\frac{dx}{dx} + \frac{\partial F}{\partial y}\frac{dy}{dx} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x}$$

We are assuming that y does not depend from x and that z depends from x and y, so we have $dy/dx = 0$ and, of course, $dx/dx = 1$. So, when $\partial_z F \neq 0$, we can solve the equation w.r.t. $\partial_x z$:

$$\frac{\partial z}{\partial x} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial z}}$$

Similarly, when $\partial_z F \neq 0$ we have:

$$\frac{\partial z}{\partial y} = -\frac{\frac{\partial F}{\partial y}}{\frac{\partial F}{\partial z}}$$

In the case of the sphere, we obtain that:

$$\frac{\partial F}{\partial x} = 2x$$
$$\frac{\partial F}{\partial z} = 2z$$
$$\frac{\partial z}{\partial x} = -\frac{x}{z}$$

Note that, by approaching this problem by first solving the initial equation w.r.t. z, we would have obtained the two equations:

$$z = \sqrt{1 - x^2 - y^2}$$
$$z = -\sqrt{1 - x^2 - y^2}$$

Now, by deriving both of them w.r.t. x we obtain:

$$\frac{\partial z}{\partial x} = -\frac{x}{\sqrt{1 - x^2 - y^2}} = -\frac{x}{z}$$
$$\frac{\partial z}{\partial x} = -\frac{x}{-\sqrt{1 - x^2 - y^2}} = -\frac{x}{z}$$

So, the two approaches are equivalent.

3.6 The gradient of a scalar function

In the previous section, we discovered that the differential of a function $\vec{f}$ in a point $\vec{x}_0$ is the map $\mathsf{d}\vec{f}_{\vec{x}_0}$ that takes any vector $\vec{u}$ and returns $\mathsf{d}\vec{f}_{\vec{x}_0}(\vec{u})$ that is the directional derivative of $\vec{f}$ at $\vec{x}_0$ along the direction of $\vec{u}$. We also showed the relationship between directional derivatives and partial derivatives:

$$\partial_u \vec{f}(\vec{x}_0) = \mathsf{d}\vec{f}_{\vec{x}_0}(\vec{u}) = u_1 \partial_{x_1} \vec{f}(\vec{x}_0) + u_2 \partial_{x_2} \vec{f}(\vec{x}_0) + \cdots + u_n \partial_{x_n} \vec{f}(\vec{x}_0)$$

Now, let's assume that $f : \Omega \to \mathbb{R}$ is a scalar function with $\Omega \subseteq \mathbb{R}^n$. Then, the partial derivatives of f are also scalar functions $\partial_{x_1} f, \ldots, \partial_{x_n} f : \Omega \to \mathbb{R}$, thus we have:

$$\mathsf{d}f_{\vec{x}_0}(\vec{u}) = u_1 \partial_{x_1} f(x_0) + u_2 \partial_{x_2} f(x_0) + \cdots + \partial_{x_n} f(\vec{x}_0)$$

where $u_1, u_2, \ldots, u_n$ are the components of the vector $\vec{u}$. Does this expression remind you of anything? To better understand what we are saying, let's consider the simpler case when $f : \mathbb{R}^2 \to \mathbb{R}$. In this case, for any vector $\vec{u} = \langle u, v \rangle$ we have:

$$\mathsf{d}f_{(x_0,y_0)}(\vec{u}) = u\partial_x f(x_0, y_0) + v\partial_y f(x_0, y_0)$$

What expression is the one on the left of this equation?
Exactly! This is the scalar product between the vector $\vec{u}$ and the vector $\big\langle \partial_x f(x_0, y_0), \partial_y f(x_0, y_0) \big\rangle$.

Proposition 3.4. *Let $f : \Omega \to \mathbb{R}$, with $\Omega \subseteq \mathbb{R}^n$, be a differentiable scalar function. Let's introduce the following vector:*

$$\vec{\nabla} f(x_1, \ldots, x_n) := \partial_{x_1} f(x_1, \ldots, x_n)\vec{v}_1 + \cdots + \partial_{x_n} f(x_1, \ldots, x_n)\vec{v}_n$$

Then, if $\vec{u}$ is any vector in $\mathbb{R}^n$, we have the equality:

$$\mathsf{d}f_{(x_1,\ldots,x_n)}(\vec{u}) = \vec{u} \cdot \vec{\nabla} f(x_1, \ldots, x_n)$$

The vector $\vec{\nabla} f$ has a fancy name:

Definition 3.4. *The **gradient** of a scalar function $f : \Omega \to \mathbb{R}$, with $\Omega \subseteq \mathbb{R}^n$ in a point $\vec{x}_0$ is the vector whose components are the partial derivatives of f at $\vec{x}_0$, that is the vector so defined:*

$$\vec{\nabla} f(\vec{x}_0) = \big\langle \partial_{x_1} f(\vec{x}_0), \ldots, \partial_{x_n} f(\vec{x}_0) \big\rangle$$

In the case of a function $f : \mathbb{R}^2 \to \mathbb{R}$ the gradient of f is the vector:

$$\vec{\nabla} f(x_0, y_0) := \big\langle \partial_x f(x_0, y_0), \partial_y f(x_0, y_0) \big\rangle$$

Example 3.10. Let's consider the bivariate Gaussian function:

$$g(x, y) := e^{-\frac{x^2+y^2}{2}}$$

We want to compute the gradient of g in a point (x_0, y_0). In Picture 3.3 we represented the following situation: we sliced the graph of g along the directions of the two axes $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ and we represented the two partial derivatives as the slopes of the curves along the two planes. In the domain of the function, that is the plane $\hat{\mathbf{i}}\hat{\mathbf{j}}$, we represented the two partial derivatives as components of the gradient of f. Remember that:

$$\partial_x g(x_0, y_0) = -x_0 e^{-\frac{x_0^2+y_0^2}{2}}$$

$$\partial_y g(x_0, y_0) = -y_0 e^{-\frac{x_0^2+y_0^2}{2}}$$

So, the gradient of g at (x_0, y_0) is the vector:

$$\vec{\nabla} g(x_0, y_0) = \left\langle -x_0 e^{-\frac{x_0^2+y_0^2}{2}}, -y_0 e^{-\frac{x_0^2+y_0^2}{2}} \right\rangle$$

Let's now study the sign of the partial derivatives. Note that $g(x, y) > 0$ for every $x, y \in \mathbb{R}$ and that $\partial_x g(x, y) = -xg(x, y)$ and $\partial_y g(x, y) = -yg(x, y)$. Then we have:

$$\begin{cases} \partial_x g(x, y) > 0 & \text{if } x < 0 \\ \partial_x g(x, y) < 0 & \text{if } x > 0 \\ \partial_x g(x, y) = 0 & \text{if } x = 0 \end{cases}$$

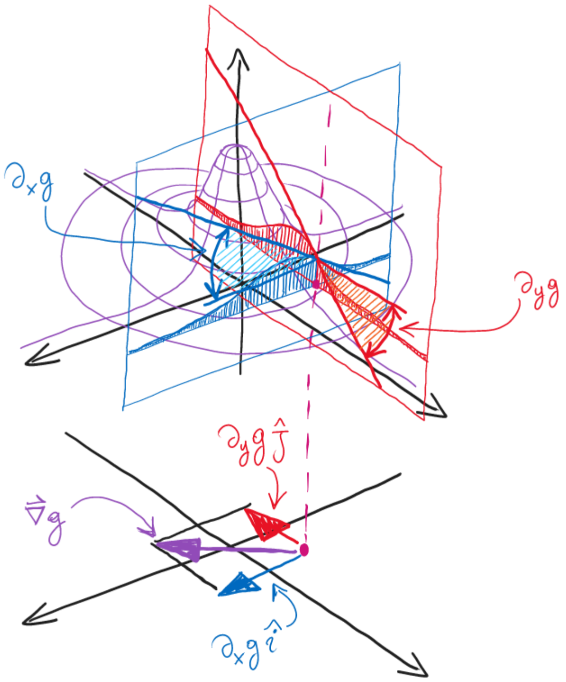


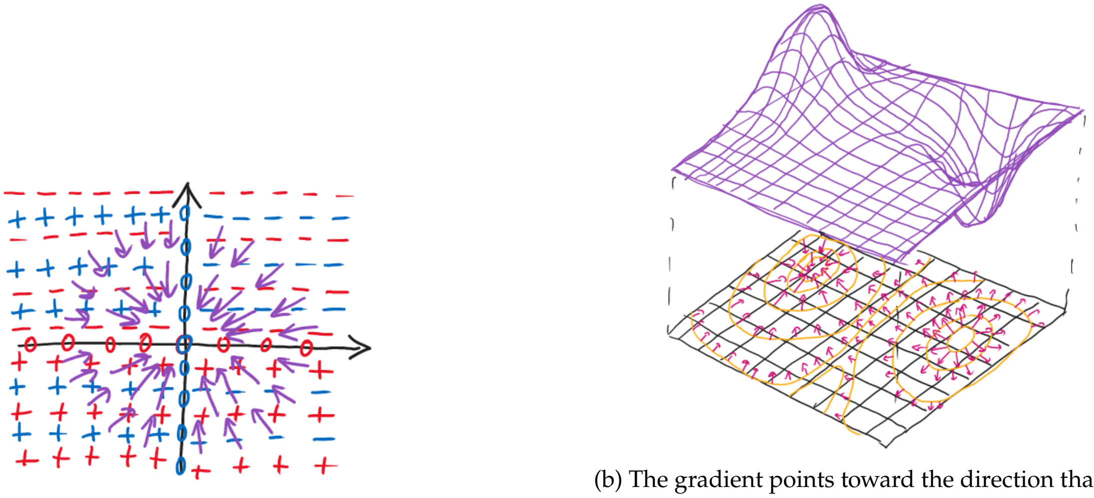
Figure 3.3: The gradient of the bivariate Gaussian function in a point with a representation of the partial derivatives

$$\begin{cases} \partial_y g(x,y) > 0 & \text{if } y < 0 \\ \partial_y g(x,y) < 0 & \text{if } y > 0 \\ \partial_y g(x,y) = 0 & \text{if } y = 0 \end{cases}$$

But:

1. saying that $x > 0$ means that (x, y) belongs to the first and the fourth quadrants, excluding the vertical axis;
2. saying that $x < 0$ means that (x, y) belongs to the second and third quadrants, excluding the vertical axis;
3. saying that $x = 0$ means that (x, y) belongs to the vertical axis;
4. saying that $y > 0$ means that (x, y) belongs to the first and the second quadrants; excluding the horizontal axis;
5. saying that $y < 0$ means that (x, y) belongs to the third and the fourth quadrants; excluding the horizontal axis;
6. saying that $y = 0$ means that (x, y) belongs to the horizontal axis.

In Picture 3.4a there is a representation of this situation. As you can see, this means that the gradient of g points always to the centre of the plane. This is not a coincidence, but rather an important property of the gradient: the gradient points always in the direction of a local maximum, or said differently, it points in the direction with maximum variation. In Picture 3.4b there is a representation of the gradient of a scalar function. As you can see the gradient is pointing towards the picks and in the opposite direction of a sink. ■



(a) In blue the sign of $\partial_x g$ and in red the sign of $\partial_y g$; the gradient is indicated by the purple vectors

(b) The gradient points toward the direction that maximizes the directional derivative

Figure 3.4: The gradient

The next proposition formalizes the property of the gradient vector to point towards the local maximum of the function.

Proposition 3.5. *Consider a differentiable scalar function $f : \Omega \to \mathbb{R}$ with $\Omega \subseteq \mathbb{R}^n$. Then, by considering a generic versor $\hat{u}$, the directional derivative of f along $\hat{u}$ is maximal when $\hat{u}$ has the same direction and orientation of the gradient of $\vec{f}$. In particular, for every point $\vec{x}_0 \in \Omega$ and any versor $\hat{u}$:*

$$\frac{\partial f}{\partial u}(\vec{x}_0) \leq \left|\vec{\nabla} f(\vec{x}_0)\right|$$

Proof. Consider a versor $\hat{u}$, then we have:

$$\frac{\partial f}{\partial u}(\vec{x}_0) = \hat{u} \cdot \vec{\nabla} f(\vec{x}_0) = |\hat{u}|\left|\vec{\nabla} f(\vec{x}_0)\right| \cos\theta \leq \left|\vec{\nabla} f(\vec{x}_0)\right|$$

where θ is the angle between $\hat{u}$ and $\vec{\nabla} f(\vec{x}_0)$ and where we used that $|\hat{u}| = 1$ and that $\cos\theta \leq 1$. Moreover, if $\hat{u}$ has same direction and orientation of $\vec{\nabla} f(\vec{x}_0)$, then $\theta = 0$, thus $\cos\theta = 1$ and in this case we have that:

$$\frac{\partial f}{\partial u}(\vec{x}_0) = \left|\vec{\nabla} f(\vec{x}_0)\right|$$

Thus, the direction and orientation of the gradient are the direction of orientation towards which the directional derivative of f is maximal. □

A good way to understand what Proposition 3.5 tells us is to imagine hiking on a mountain. Imagine stopping at a point and leaving a ball rolling on the ground. The ball will fall in the direction where the ground is steeper. This is because gravity moves objects along the fastest path there is to reach the least altitude possible. The gradient is that vector that points in the direction in which the mountain is the steepest possible and orientation to the tip of the mountain. So, the gradient will represent precisely the opposite of the gravitational force.

A useful way to visualize and represent the graph of a scalar function is the so-called contour diagram of the function. To understand this concept, first, we need to introduce the notion of level-sets. Imagine, for

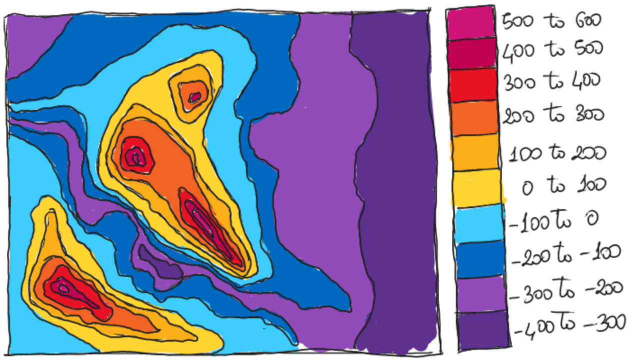


Figure 3.5: Representation of the level curves of an island and the sea that surrounds it

---

example, you want to represent on a two-dimensional map the complex geography of an island. You want to represent the mountains on the island and also the deep sea that surrounds the island, with its submarine canyons and valleys.

one we do that, is to draw the lines of points that have the same altitude. In Picture 3.5 there is an example of such a map. We can decide to do the same for scalar functions $f : \mathbb{R}^2 \to \mathbb{R}$, where we think of $f(x, y)$ as the altitude of the surface at (x, y). The lines of points that have the same value of $f(x, y) = c$ are called level-sets of the function f.

Definition 3.5. *Let $f : \Omega \to \mathbb{R}$ be a scalar function with $\Omega \subseteq \mathbb{R}^n$. Moreover, let $c \in \mathbb{R}$ be a real number. The c-**level-set** of f is the counter-image of c along f, that is the set so defined:*

$$\Gamma_c(f) := \{\vec{x} \in \Omega \mid f(\vec{x}) = c\}$$

*If $n = 2$, i.e. f is a function of type $\mathbb{R}^2 \to \mathbb{R}$, then the level sets of f are also called the level-curves of f, or isocurves and when $n = 3$, level-surfaces, or isosurfaces. Finally, the collection of all the level-sets of a function f is called the **contour diagram** of f.*

In Picture 3.6a we represented the contour diagram of the bivariate Gaussian function and the corresponding gradient vectors. Moreover, in Picture 3.6b we also represented the contour diagrams of the function $I(x, y, z) := P/(4\pi(x^2 + y^2 + z^2))$ which represents the intensity of a pointwise source of light. In this case, the level-sets of I are two-dimensional spheres centred in the position of the light source. Note that the gradient vectors are pointing at the source.

Note also that in both cases the gradient is orthogonal to the level sets. This is not a coincidence but a key property of the gradient.

Proposition 3.6. *Let $f : \Omega \to \mathbb{R}$ be a differentiable scalar function, $c \in \mathbb{R}$ and suppose that $\vec{x}_0 \in \Omega$ is a point on the $c-$level-set of f, i.e. $f(\vec{x}_0) = c$. If T is the tangent space of $\Gamma_c(f)$ at $\vec{x}_0$, then $\vec{\nabla} f(\vec{x}_0)$ is orthogonal to T.*

Proof. We are going to prove this result when $n = 3$, i.e. when f is a function of type $\mathbb{R}^3 \to \mathbb{R}$. Thus, in this case, the c-level-set is so defined:

$$\Gamma_c(f) = \{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = c\}$$

(a) Contour diagram of the bivariate Gaussian function (b) Contour diagram of the intensity of a light source

Figure 3.6: Contour diagrams

Let's suppose to take a path $\vec{r}(t) := (x(t), y(t), z(t))$ of points of $\Gamma_c(f)$ that passes through (x_0, y_0, z_0) at $t = 0$. So, by composing f with $\vec{r}(t)$ we obtain the function $f \circ \vec{r} : \mathbb{R} \to \mathbb{R}$, $(f \circ \vec{r})(t) = f(\vec{r}(t)) = f(x(t), y(t), z(t))$. However, because $\vec{r}(t) = (x(t), y(t), z(t))$ is a point of $\Gamma_c(f)$, we have that, for every $t \in \mathbb{R}$:

$$(f \circ \vec{r})(t) = f(x(t), y(t), z(t)) = c$$

Thus, because c is a constant, the total derivative of $f \circ \vec{r}$ is always equal to zero. However, by using the chain rule, we also have:

$$\begin{aligned} 0 &= \frac{d(f \circ \vec{r})}{dt}(t) = \\ &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = \\ &= \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle \cdot \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right\rangle = \\ &= \vec{\nabla} f(x(t), y(t), z(t)) \cdot \vec{r}'(t) \end{aligned}$$

Now, by taking $t = 0$, we can see that $\vec{r}'(0)$ is the vector tangent of the path $\vec{r}(t)$ in the point (x_0, y_0, z_0), and because $r(t)$ is a path in the set $\Gamma_c(f)$, the vector $\vec{r}'(t)$ is a vector of the tangent plane of the set $\Gamma_c(f)$ at (x_0, y_0, z_0). Moreover, also the opposite is true: every vector of the tangent plane of $\Gamma_c(f)$ at (x_0, y_0, z_0) is the vector $\vec{r}'(0)$ for some path $\vec{r}(t)$ of $\Gamma_c(f)$ that passes through (x_0, y_0, z_0) at time $t = 0$. Therefore, the equation we found says that every vector of the tangent plane of $\Gamma_c(f)$ at (x_0, y_0, z_0) is orthogonal to the gradient of f at (x_0, y_0, z_0). □

We conclude this section with a comment about the meaning of the gradient of a scalar function in physics.

Cool Example* 3.2. In physics, a vector field is a function that associates every point in space with a vector. An example of such a field is the gravitational field of an object. Gravity is a force that points in the centre of an object. The gravitational field of an object is a vector field that represents this force at each point in space. To better understand this concept, imagine floating in space. Suppose that in front of you, there is a giant star, so you will be gravitationally attracted toward its centre. Imagine measuring at each point this force. The function $\vec{F}_g \colon \mathbb{R}^3 \to \mathbb{R}^3$ that maps each point (x, y, z) into the corresponding vector $\vec{F}_g\ (x, y, z)$ that represents such force, is called the gravitational field of the star.

Now that we know what the gravitational field is, we can ask the following question: is there a scalar function $p : \mathbb{R}^3 \to \mathbb{R}$ that satisfies the following equation?

$$\vec{F}_g = -\vec{\nabla} p$$

And, assuming there is such a function, what is its physical meaning?
It turns out that a vector field is equal to (negative) the gradient of a scalar function only in some circumstances; whenever this is the case, we call that vector field, irrotational or conservative. It turns out that gravity is an irrotational vector field, therefore we can find a function $p : \mathbb{R}^3 \to \mathbb{R}$ so that $\vec{F}_g = -\vec{\nabla} p$.
We also call the function p a **potential** for the vector field $\vec{F}$, and for gravity, we call p the gravitational potential. We can interpret p as the potential energy that a body has in a specific position. One fundamental principle of physics is that things tend to minimize their potential energy. Now, p is smaller and smaller, closer and closer we go to the centre of the star and it's bigger further way we go. Therefore, if we want to minimize our potential energy we have to go closer and closer to the centre of the star.
This is what the gradient is telling us to do: $-\vec{\nabla} p$ is that vector that points toward the centre of the star and that tells us how our acceleration is while we are falling toward the centre of the star. This way of thinking about forces is very useful and it's used many times in physics. Instead of thinking about the vectorial forces, we instead focus on their potential and we interpret these potentials as the potential energy.

Another interesting application regards scalar fields. An example of such a field is temperature T. Temperature T is a scalar function that associates to every point in the three-dimensional space a number that represents the temperature measured at that point. One of the principles of physics is that things cool down, i.e. the temperature tends to decrease in time.
A useful conceptual tool is to consider the vector field $\vec{\nabla} T$. The gradient of T points in each point towards the direction and orientation that maximizes the variation of T. This vector field, known as the temperature gradient, describes the dynamic of the heat flow, i.e. the flow of energy that moves from warmer points toward colder ones. ■

3.7 The extreme value theorem in higher dimensions

Recall the following result you have studied from calculus in 1 dimension:

> *If $f : \Omega \to \mathbb{R}$ is a continuous function and Ω is a closed interval $[a, b]$, then f has a global maximum and a global minimum.*

We want to extend this result in multivariable calculus. The result we are looking for, will sound like "given a continuous scalar function $f : \Omega \to \mathbb{R}$, if Ω is blablabla then f has a global maximum and a global minimum", where "blablabla" means that we need to specify some conditions on Ω. First, let's introduce the notions of global maximum and global minimum in this context.

Definition 3.6. *Let $f : \Omega \to \mathbb{R}$ be a scalar function with $\Omega \subseteq \mathbb{R}^n$. We say that f has a* ***global maximum*** *at $\vec{x}_0 \in \Omega$ if $f(\vec{x}_0)$ is a global maximum for f, that is, if for every $\vec{x} \in \Omega$:*

$$f(\vec{x}) \leq f(\vec{x}_0)$$

Moreover, we say that f has a ***global minimum*** *at $\vec{x}_0 \in \Omega$ is $f(\vec{x}_0)$ is a global minimum for f, that is, if for every $\vec{x} \in \Omega$:*

$$f(\vec{x}) \geq f(\vec{x}_0)$$

We also say that $\vec{x}_0$ is a point of ***extreme value*** *for f if $f(\vec{x}_0)$ is an extreme value for f, that is if $f(\vec{x}_0)$ is a global maximum or a global minimum for f.*

Cool Stuff* 3.1. Note that to define the notions of extreme values we need that the codomain of the function f to be a totally-ordered set. A totally-ordered set is a set A equipped with a relation, usually denoted by $\geq$, which satisfies the following properties:

1. Reflexivity: for every $a \in A$, $a \geq a$;
2. Transitivity: for every $a, b, c \in A$, if $a \geq b$ and $b \geq c$ then $a \geq c$;
3. Antisymmetry: for every $a, b \in A$, if $a \geq b$ and $b \geq a$ then $a = b$;
4. Totality: for every $a, b \in A$, $a \geq b$ or $b \geq a$ or both.

Such a relation is also called total order. An example of a totally-ordered set is $\mathbb{R}$, where $\geq$ is the usual $\geq$. The relation $\geq$ is used to compare two elements of the set and it tells which one is the biggest one. In $\mathbb{R}^n$ there is no such relation because vectors cannot be compared with a total order.
In the definition of global maximum and global minimum, we had to use the relation $\geq$, which is the total order on the codomain of f. Therefore, we couldn't define these concepts for a function $f : \mathbb{R}^n \to \mathbb{R}^m$, whenever $m > 1$. ■

The next step is to understand what conditions we need for the domain Ω of f. Recall that in 1 dimension, even if the function $f : (a, b] \to \mathbb{R}$ is continuous because $(a, b]$ is not a closed interval, the theorem does not work. This is because, we can always find at least a continuous function $f : (a, b] \to \mathbb{R}$ that has no global maximum or global minimum. As an example, take $f : (0, 1] \to \mathbb{R}$ defined as follows:

$$f(x) := \frac{1}{x}$$

Clearly, this function has no global maximum. So, from this example, we understand that we need to find the right conditions on Ω to make this result work. One possible solution is to consider the obvious generalization of a closed interval in multidimensions, which is the closed ball. However, this is a very strict generalization and the theorem we are going to study can be applied and many many more cases: we need to introduce the notion of compact sets.

The full general definition of this concept is however quite abstract and requires a lot of work to be fully understood. Therefore, here are going to study this concept only in the special case of compact subsets of $\mathbb{R}^n$. First, we need to introduce the notion of bounded subsets.

Definition 3.7. *A subset Ω of $\mathbb{R}^n$ is **bounded** if there is a real number R such that the closed ball $\overline{\mathbb{B}}(\vec{0}, R)$ includes Ω, that is $\Omega \subseteq \overline{\mathbb{B}}(\vec{0}, R)$. If a subset of $\mathbb{R}^n$ is not bounded, we say it is **unbounded**.*

Example 3.11. Examples of bounded subsets of $\mathbb{R}^n$ are the following:

1. Every open ball
2. Every closed ball
3. The subset $B := [0, 1] \times (0, 1)$ of $\mathbb{R}^2$ is a bounded set

Examples of unbounded subsets of $\mathbb{R}^n$ are the following:

1. The whole $\mathbb{R}^n$
2. The first quadrant $[0, \infty) \times [0, \infty) \subseteq \mathbb{R}^2$ of the Cartesian plane
3. Any line in the space

Intuitively, bounded subsets are subsets that can be bounded by a ball. So, every unbounded subset is a subset that cannot be included in a ball of any radius. ■

In Section 2.2 we introduced the concepts of open and closed sets. Here, we briefly recall this definition. We say that a set $A \subseteq \mathbb{R}^n$ is open if, given any point $\vec{x}_0 \in A$, there is a positive number $r > 0$ so that the open ball $\mathbb{B}(\vec{x}_0, r)$ is entirely included in A, that is $\mathbb{B}(\vec{x}_0, r) \subseteq A$.
Intuitively, an open set is a set without the boundary. So, for example, the set $A := (0, 1) \times (0, 1)$ is open, because if $(x_0, y_0) \in A$, we can always find a tiny open ball centred in (x_0, y_0) that is entirely included in A. The set $B := (0, 1] \times (0, 1]$ is not open, this because if we consider for example the point $(1, 1) \in B$, we see that for any radius $r > 0$, the open ball $\mathbb{B}((1, 1), r)$ is not entirely included in B. In Picture 3.7 we represented this situation.

We also recall that a closed set is the complement of an open set. This means that $C \subseteq \mathbb{R}^n$ is closed if the set $\mathbb{R}^n \setminus C$ is open. Intuitively, closed sets are set with the boundary. So, for example, the set $C := [0, 1] \times \mathbb{R}$ is a closed set, because the set $\mathbb{R}^n \setminus C = ((-\infty, 0) \times \mathbb{R}) \cup ((1, \infty) \times \mathbb{R})$ is an open set.

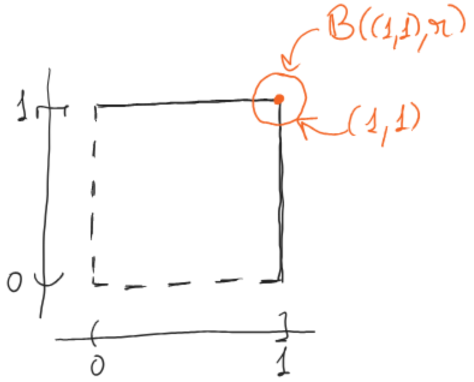


Figure 3.7: Representation of the set $B = (0,1] \times (0,1)] \subseteq \mathbb{R}^2$. We also indicated the point $(1,1)$, which is on the boundary of B since, every open ball centred in $(1,1)$ cannot be fully included into B, but it still has a non-empty intersection with B

Another example of a closed set is the cubic set $[0,1] \times [0,1] \times [0,1] \subseteq \mathbb{R}^3$. An example of a set that is not open nor closed is the set $(0,1] \times (0,1]$. We also note that in $\mathbb{R}^n$ the only sets that are both open and closed are the empty set and the whole $\mathbb{R}^n$.
Let's introduce the concept we need.

Definition 3.8. *A subset $K \subseteq \mathbb{R}^n$ of $\mathbb{R}^n$ is **compact** if K is closed and bounded.*

Example 3.12. Here is a table of subsets of $\mathbb{R}^n$:

SUBSET	OPEN	CLOSED	BOUNDED	COMPACT
Empty set $\emptyset$	✓	✓	✓	✓
$\mathbb{R}^n$	✓	✓	✗	✗
$\mathbb{B}(\vec{x}_0, r)$	✓	✗	✓	✗
$\overline{\mathbb{B}}(\vec{x}_0, r)$	✗	✓	✓	✓
$\{\vec{x}_0\}$	✗	✓	✓	✓
$(0,1) \times (0,1) \times (0,1)$	✓	✗	✓	✗
$[0,1] \times [0,1] \times [0,1]$	✗	✓	✓	✓
$(0,1] \times (0,1] \times (0,1)$	✗	✗	✓	✗
$(0,1) \times (0,1) \times \mathbb{R}$	✓	✗	✗	✗
$[0,1] \times [0,1] \times \mathbb{R}$	✗	✓	✗	✗

These are only a few examples of subsets and their classification. We note that the empty set is the unique compact subset of $\mathbb{R}^n$ that is both open and closed. ■

We can finally state the main theorem.

Theorem 3.1 (Extreme-Value theorem). *Suppose that $f : K \to \mathbb{R}$ is a scalar function. If the following conditions are true:*

1. *f is continuous*
2. *K is compact*

then there is at least a point $\vec{x}_0 \in K$ such that $f(\vec{x}_0)$ is global maximum for f and moreover there is at least a point $\vec{x}_1 \in K$ such that $f(\vec{x}_1)$ is a global minimum for f.

The best way to understand this result is by looking at counterexamples. Concretely, this means that we are going to show that when one of the two assumptions of this theorem is not true anymore, the thesis is false.

Example 3.13. In this example, we are going to show that if f is a function that is not continuous, then, even if K is compact, it is not guaranteed anymore that f has extreme values anymore. Let's introduce the following function:

$$f : [0, 1] \to \mathbb{R}$$
$$f(x) := \begin{cases} \frac{1}{2} & \text{if } x = 0 \\ x & \text{if } x \in (0, 1) \\ \frac{1}{2} & \text{if } x = 1 \end{cases}$$

So f is not continuous in 0 and in 1 because:

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} x = 0$$
$$f(0) = \frac{1}{2}$$

$$\lim_{x\to 1} f(x) = \lim_{x\to 1} x = 1$$
$$f(1) = \frac{1}{2}$$

Note also that the domain of f is the closed interval $[0, 1]$ which is bounded and closed, so it is compact. However, f has no global maximum and neither global minimum. To show that note first that:

$$0 < f(x) < 1$$

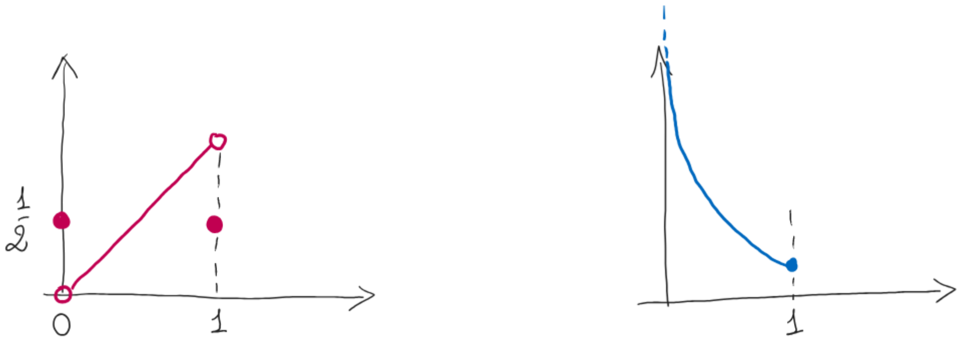
and that if $a \in (0, 1)$, then $f(a) = a$. So, if, by contradiction, f would have a global maximum, that global maximum should have been 1, but for no value $x_0 \in [0, 1]$ $f(x_0) = 1$. Similarly, if we assume that f has a global minimum, we conclude that this has to be 0, but there is no elements $x_0 \in [0, 1]$ so that $f(x_0) = 0$. In Picture 3.8a we gave a representation of this situation. ■

Example 3.14. In this example, we are assuming that $f : \Omega \to \mathbb{R}$ is a continuous function, but Ω is not compact. First, we assume that Ω is bounded but not closed. An example of this situation is the following one:

$$f : (0, 1] \to \mathbb{R}$$
$$f(x) := \frac{1}{x}$$

This is a continuous function defined over a non-closed interval. Note that, if we fix any positive real number $M > 1$, we can always find an element $x_0 := 1/M$ such that $f(x_0) = 1/(1/M) = M$. So, because every positive real number M can be obtained as the image of an element $x_0 \in (0, 1]$, there is no global maximum. In Picture 3.8b we gave a representation of this situation.
Now, let's consider the other situation, that is when Ω is closed but not bounded. For example, consider the whole $\Omega = \mathbb{R}$, thus we can choose the following function:

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



(a) An example of a discontinuous function defined over a compact set which violates the Extreme-Value theorem

(b) An example of a continuous function defined over a non-compact set which violates the Extreme-Value theorem

Figure 3.8: Counterexamples for the Extreme-Value theorem

$$x(x) := x$$

In this case, there is no global minimum nor global minimum. ■

3.8 Local maximum and local minimum

The extreme value theorem tells that any continuous scalar function defined over a compact set has a global maximum and a global minimum. What is not telling is how to find these extreme values. To solve this problem, let's first look at a similar problem: finding local extreme values. Let's first introduce this concept.

Definition 3.9. *Let $f : \Omega \to \mathbb{R}$ be a scalar function and $\vec{x}_0, \vec{x}_1 \in \Omega$. We say that $\vec{x}_0$ is a point of* ***local maximum*** *for f, if there is an open ball $\mathbb{B}(\vec{x}_0, r)$ such that, for every $\vec{x} \in \mathbb{B}(\vec{x}_0, r) \cap \Omega$:*

$$f(\vec{x}) \le f(\vec{x}_0)$$

Moreover, we say that $\vec{x}_1$ is a point of ***local minimum*** *for f, if there is an open ball $\mathbb{B}(\vec{x}_1, r)$ such that, for every $\vec{x} \in \mathbb{B}(\vec{x}_0, r) \cap \Omega$:*

$$f(\vec{x}) \ge f(\vec{x}_0)$$

Finally, if a point is a point of local maximum or local minimum for f we simply say that it is a point of ***local extreme value****.*

In Picture 3.9 we represented some points of local maximum and of local minimum for f. We notice from this picture that in all the points of extreme values that are not on the border of the domain of the function, the tangent plane of the function is horizontal. Let's formalize this idea.

Definition 3.10. *Consider a scalar function $f : \Omega \to \mathbb{R}$ with $\Omega \subseteq \mathbb{R}^n$. We say that $\vec{x}_0 \in \Omega$ is a* ***stationary point*** *for f if:*

1. *f is differentiable in $\vec{x}_0$*
2. $\vec{\nabla} f(\vec{x}_0) = \vec{0}$

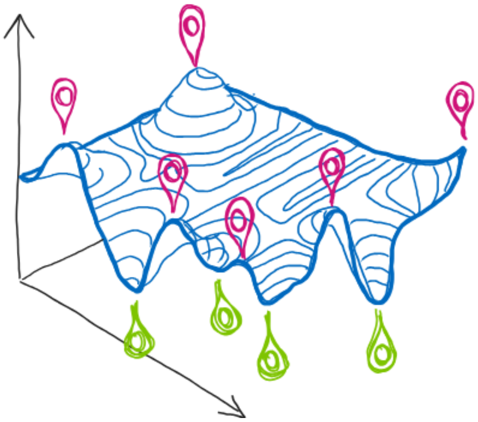


Figure 3.9: In red points of local maximum and in green points of local minimum

---

Moreover, we say that $\vec{x}_0$ is a **critical point** of f is one of the two conditions is true:

1. f is **not** differentiable in $\vec{x}_0$, or
2. $\vec{x}_0$ is a stationary point for f

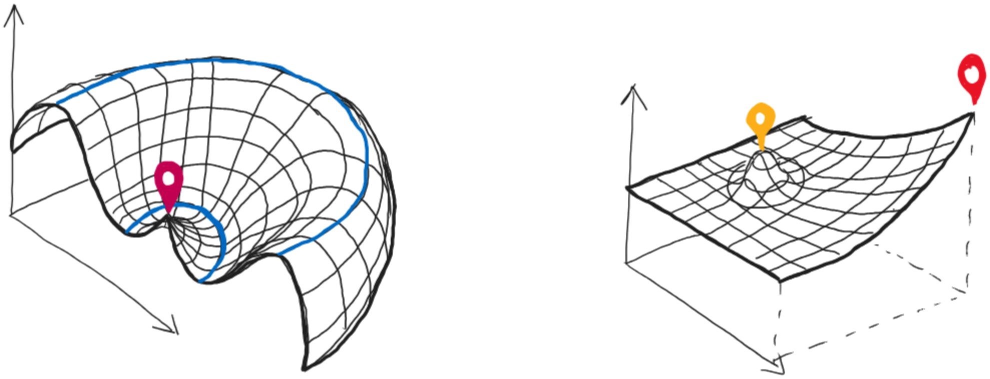
In Picture 3.10a we coloured the critical points (red and blue) and the stationary points (in blue) of a function f.
At this point, one could conclude that every point of local extreme value is a critical point. This sounds correct, however, the situation is slightly more complicated. The issue is that when $\vec{x}_0$ is on the boundary of the domain of the function, $\vec{x}_0$ could be a point of local extreme and f be differentiable in $\vec{x}_0$, but its gradient is not null.

Intuitively speaking, when $\vec{x}_0$ is internal and f is differentiable in $\vec{x}_0$, if $f(\vec{x}_0)$ is a local maximum then the function has to go up to reach $f(\vec{x}_0)$ and then it has to go down in any direction. However, when $\vec{x}_0$ is on the boundary, then f doesn't need to go down anymore, so the gradient could be non-null in $\vec{x}_0$. In Picture 3.10b we represented this situation.
To formalize this idea we first need to specify what means for a point to be internal or on the boundary of a set.

Definition 3.11. *Given a set $\Omega \subseteq \mathbb{R}^n$, a point $\vec{x}_0 \in \mathbb{R}^n$ is* ***internal*** *to Ω if there exists an open ball $\mathbb{B}(\vec{x}_0, r)$ such that $\mathbb{B}(\vec{x}_0, r)$ is entirely included in Ω, i.e. $\mathbb{B}(\vec{x}_0, r) \subseteq \Omega$. We also say that a point $\vec{x}_0 \in \mathbb{R}^n$ is within the* ***closure*** *of Ω is given any positive real number $r > 0$, the open ball $\mathbb{B}(\vec{x}_0, r)$ intersects in at least a point the set Ω, that is the set $\mathbb{B}(\vec{x}_0, r) \cap \Omega$ is not empty. Finally, we say that a point $\vec{x}_0 \in \mathbb{R}^n$ is on the* ***boundary*** *of Ω if $\vec{x}_0$ is within the closure of Ω, but it is not internal to Ω.*

Note that in Definition 2.3 we already defined the closure, the interior and the boundary of a set. It turns out that a point $\vec{x}_0$ is internal to Ω if and only if $\vec{x}_0$ is an element of the interior; that $\vec{x}_0$ is within the closure of Ω if it is an element of the closure; and that $\vec{x}_0$ is on the boundary of Ω if $\vec{x}_0$ is an element of the boundary of Ω.

Note that if $\vec{x}_0$ is internal to Ω, then $\vec{x}_0 \in \Omega$, but if $\vec{x}_0$ is in the closure of Ω, this does not guarantee that $\vec{x}_0$ is an element of Ω. For example, 1 is on the boundary (and therefore in the closure) of $[0, 1)$, but it is not an element of $[0, 1)$.
With this definition, we are now able to enounce the following result.



(a) Critical points: in red we indicated the points where the function is not differentiable, while in blue the stationary point

(b) In red a point of local maximum on the boundary, while in yellow a point of local maximum on the interior. Note that the local maximum on the boundary is not a stationary point

Figure 3.10: Critical points and local extremes values on the boundary

---

Theorem 3.2. *Let $f : \Omega \to \mathbb{R}$ be a scalar function with $\Omega \subseteq \mathbb{R}^n$. If $\vec{x}_0$ is an internal point of Ω and it is also a point of local extreme value for f then $\vec{x}_0$ is a critical point for f.*

In conclusion, to find the global maximum of a function $f : \Omega \to \mathbb{R}$ we need to split the problem in two:

1. Finding the global maximum of f on the interior of Ω
2. Finding the global maximum of f on the boundary points of Ω

To solve the second problem, i.e. finding the extreme values of f on the boundary of Ω, we will need to introduce a new technique, known as the Lagrange multipliers. We dedicate an entire section to this technique. In the next section, we will discuss the problem of finding the stationary points of f in the interior of Ω.

Example 3.15. In this example, we want to find all the critical points of the following function:

$$f(x, y) := y^2 - x^3 - x^2$$

In Picture 3.11 we plot the graph of the function f using Geogebra. Note that the function, being polynomial, is differentiable everywhere. Therefore, the only critical points are the stationary points. So we need to solve the equation $\vec{\nabla} f(x, y) = \vec{0}$:

$$\partial_x f(x, y) = -3x^2 - 2x$$
$$\partial_y f(x, y) = 2y$$

Thus, the equation becomes:

$$\vec{\nabla} f(x, y) = \langle 0, 0 \rangle$$
$$\begin{cases} -3x^2 - 2x = 0 \\ 2y = 0 \end{cases}$$

So, the solutions are $(0, 0)$ and $(-2/3, 0)$. So, the critical points of the function f are $(0, 0)$ and $(-2/3, 0)$. From the graph, we can also see that $(-2/3, 0)$ is a local minimum, but that $(0, 0)$ is a stationary point that is neither a local minimum nor a local maximum. Such a point is known as a saddle point. ■

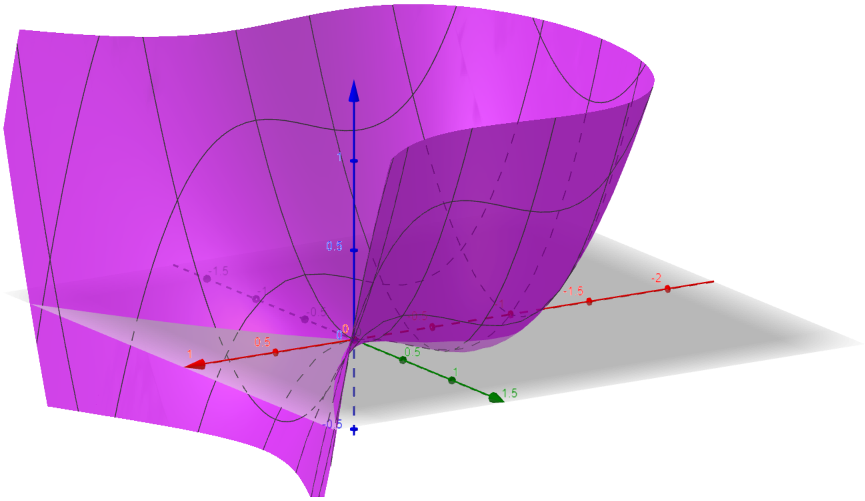


Figure 3.11: A representation of the graph of the function $f(x, y) = y^2 - x^3 - x^2$ plotted with Geogebra

---

3.9 The Hessian & its criterion

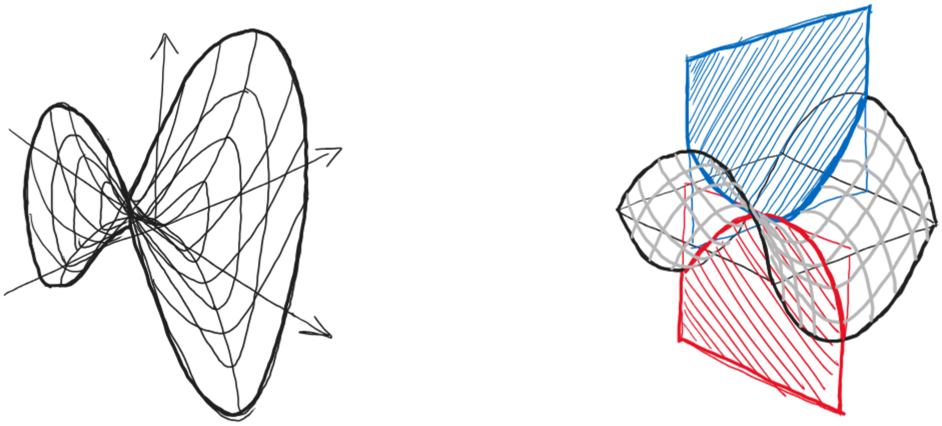
We concluded the previous section with an important result: if $\vec{x}_0$ is an internal point of the domain of f and is also a point of local extreme value for f, then $\vec{x}_0$ is a critical point of f. We could suspect that also the opposite is true, that is, if $\vec{x}_0$ is a critical point for f, then $\vec{x}_0$ is a local maximum or a local minimum for f.

It turns out that this is true only in one dimension. A counterexample of this statement in two dimensions is given by the "pringle-shaped" surface of Picture 3.12a. In the picture, you can see that in the point $\vec{x}_0$ the gradient is zero because the tangent plane is horizontal. However, the point is neither a point of local maximum nor of local minimum. Taking a closer look we realize that the issue with this surface is that in one direction $\vec{x}_0$ looks like a local maximum, but on another direction is looks like a local minimum.

In Picture 3.12b we drew the shape that the surface has when it is sectioned along the $\hat{\mathbf{i}}$-direction (in blue) and along the $\hat{\mathbf{j}}$-direction (in red). As you can see, the two sections look like two parabolas and one (the blue one) has a positive curvature, while the other one (the red one) has a negative curvature. At this point, we realize that, in order for $\vec{x}_0$ to be a point of local maximum for f, the curvature along all the directions has to be negative (or non-positive) and to be a local minimum the curvature along all the directions has to be positive (or non-negative).

For ordinary functions in 1 dimension, to measure the concavity (or the curvature) we evaluate the second derivative of that function. In multidimensions, we need to use directional second derivatives. So we have the following statement:

1. Suppose that $\partial^2_u f(\vec{x}_0) < 0$ for every versor $\hat{u}$, then $\vec{x}_0$ is a point of local maximum for f
2. Suppose that $\vec{x}_0$ is a point of local maximum for f, then for every versor $\hat{u}$, $\partial^2_u f(\vec{x}_0) \leq 0$
3. Suppose that $\partial^2_u f(\vec{x}_0) > 0$ for every versor $\hat{u}$, then $\vec{x}_0$ is a point of local minimum for f
4. Suppose that $\vec{x}_0$ is a point of local minimum for f, then for every versor $\hat{u}$, $\partial^2_u f(\vec{x}_0) \geq 0$



(a) A pringle-shaped surface: the middle point is called a saddle point and even if is a stationary point it is not a point of local extreme value

(b) The $\hat{\mathbf{i}}$-section (in blue) and the $\hat{\mathbf{j}}$-section (in red) of the pringle-shaped surface

Figure 3.12: Saddle points

This result is great! But we can do a bit better. The point is that it can be hard to evaluate every possible second directional derivative of a function $\partial_u^2 f$. However, recall that we can write:

$$\frac{\partial f}{\partial u} = u\frac{\partial f}{\partial x} + v\frac{\partial f}{\partial y}$$

where $\hat{u} = \langle u, v \rangle$. This means that to evaluate directional derivatives we just need to evaluate partial derivatives. So, if we evaluate the second directional derivative $\partial_u^2 f$ we have:

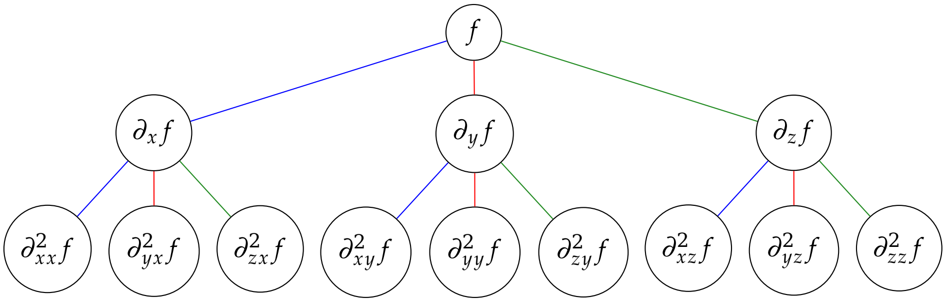
$$\begin{aligned} & \frac{\partial^2 f}{\partial u^2} = \\ = \ & \frac{\partial}{\partial u}\left(\frac{\partial f}{\partial u}\right) = \\ = \ & \frac{\partial}{\partial u}\left(u\frac{\partial f}{\partial x} + v\frac{\partial f}{\partial y}\right) = \\ = \ & u\frac{\partial}{\partial u}\frac{\partial f}{\partial x} + v\frac{\partial}{\partial u}\frac{\partial f}{\partial y} = \\ = \ & u\left(u\frac{\partial}{\partial x}\frac{\partial f}{\partial x} + v\frac{\partial}{\partial y}\frac{\partial f}{\partial x}\right) + v\left(u\frac{\partial}{\partial x}\frac{\partial f}{\partial y} + v\frac{\partial}{\partial y}\frac{\partial f}{\partial y}\right) = \\ = \ & u^2\frac{\partial^2 f}{\partial x^2} + uv\frac{\partial^2 f}{\partial y \partial x} + vu\frac{\partial^2 f}{\partial x \partial y} + v^2\frac{\partial^2 f}{\partial y^2} \end{aligned}$$

Let's introduce an important definition.

Definition 3.12. *The* ***Hessian matrix*** *of a scalar function* $f : \Omega \to \mathbb{R}$ *of class* C^2*, with* $\Omega \subseteq \mathbb{R}^n$ *is the square* *matrix $n \times n$ of functions so defined:*

$$\mathsf{H}f(\vec{x}) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

First of all, recall that a function is of class C^2 if every partial derivative of f exists and for every partial derivative $\partial_{x_i} f$, every partial derivative exists as well and the latter are all continuous functions. To understand better this concept, take a look at the following scheme:



Each edge of the graph represents the operation of taking the partial derivative of the function above. In particular, the blue edges represent ∂_x, red edges ∂_y and green edges ∂_z. So, the Hessian matrix of a function $f : \mathbb{R}^3 \to \mathbb{R}$ is given by:

$$\mathsf{H}f(\vec{x}) := \begin{pmatrix} \partial^2_{xx} f & \partial^2_{xy} f & \partial^2_{xz} f \\ \partial^2_{yx} f & \partial^2_{yy} f & \partial^2_{yz} f \\ \partial^2_{zx} f & \partial^2_{zy} f & \partial^2_{zz} f \end{pmatrix}$$

For a function $f : \mathbb{R}^2 \to \mathbb{R}$ we have instead:

$$\mathsf{H}f(\vec{x}) := \begin{pmatrix} \partial^2_{xx} f & \partial^2_{xy} \\ \partial^2_{yx} f & \partial^2_{yy} \end{pmatrix}$$

The next result simplifies a little bit the description of the Hessian matrix.

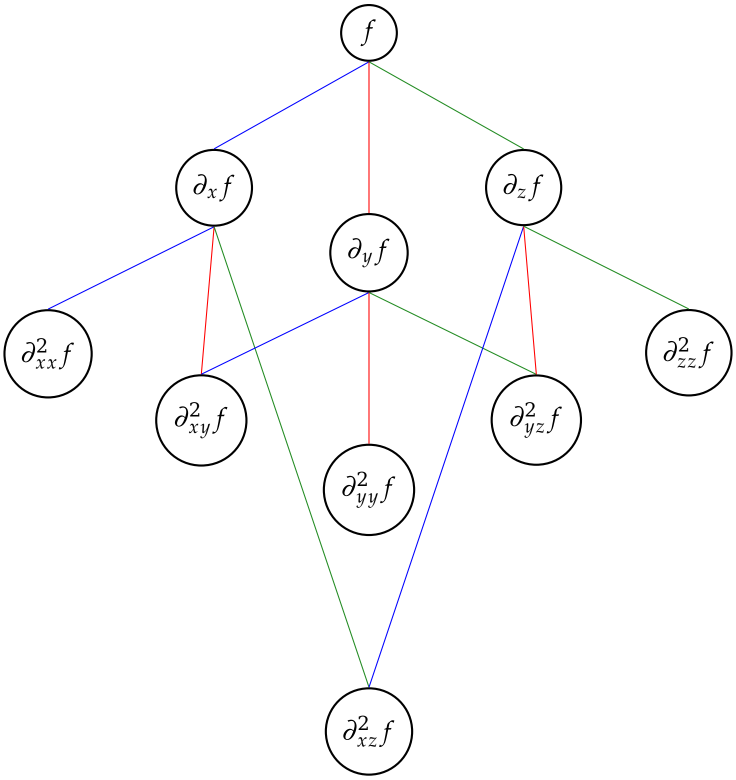
Proposition 3.7 (Schwartz' lemma). *The Hessian matrix of a function of class C^2 is symmetric. Concretely, this means that for every $i, k = 1, \ldots, n$:*

$$\frac{\partial^2 f}{\partial x_i \partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_i}$$

In 3 dimensions, this means the following:

$$\partial^2_{xy} f = \partial^2_{yx} f$$
$$\partial^2_{xz} f = \partial^2_{zx} f$$
$$\partial^2_{yz} f = \partial^2_{zy} f$$

So, actually, the scheme is as follows:



Example 3.16. Let's take into account again the bivariate Gaussian function $g(x, y) = \exp(-(x^2 + y^2)/2)$ and let's calculate the Hessian matrix of g:

$$\partial_x g(x, y) = -xe^{-\frac{x^2+y^2}{2}}$$

$$\partial_y g(x, y) = -ye^{-\frac{x^2+y^2}{2}}$$

$$\partial^2_{xx} g(x, y) = -e^{-\frac{x^2+y^2}{2}} + x^2e^{-\frac{x^2+y^2}{2}} = (x^2 - 1)e^{-\frac{x^2+y^2}{2}}$$

$$\partial^2_{xy} g(x, y) = xye^{-\frac{x^2+y^2}{2}}$$

$$\partial^2_{yx} g(x, y) = yxe^{-\frac{x^2+y^2}{2}}$$

$$\partial^2_{yy} g(x, y) = -e^{-\frac{x^2+y^2}{2}} + y^2e^{-\frac{x^2+y^2}{2}} = (y^2 - 1)e^{-\frac{x^2+y^2}{2}}$$

So we have:

$$\mathsf{H}g(x, y) = \begin{pmatrix} (x^2-1)e^{-\frac{x^2+y^2}{2}} & xye^{-\frac{x^2+y^2}{2}} \\ xye^{-\frac{x^2+y^2}{2}} & (y^2-1)e^{-\frac{x^2+y^2}{2}} \end{pmatrix} = e^{-\frac{x^2+y^2}{2}} \begin{pmatrix} x^2-1 & xy \\ xy & y^2-1 \end{pmatrix}$$

As you can see, the matrix is symmetric, as stated by Proposition 3.7. ■

Now, given an $n \times n$ square matrix A we can define an associated linear function $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ using matrix multiplication. See 5.1 for more details. For example, if A is a 2×2 matrix:

$$A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

the associated function is so defined:

$$\begin{aligned}&A : \mathbb{R}^2 \to \mathbb{R}^2\\ &A\vec{u} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} au + bv \\ cu + dv \end{pmatrix}\end{aligned}$$

There is also another way to associate a function to a square matrix A, called the quadratic-form associated to A.

Definition 3.13. *Given an $n \times n$ square matrix A, the* ***quadratic-form*** *associated to A is the function so defined:*

$$\begin{aligned}&q_A : \mathbb{R}^n \to \mathbb{R}\\ &q_A\vec{u} := \vec{u} \cdot (A\vec{u})\end{aligned}$$

Example 3.17. Let's consider a 2×2 matrix:

$$A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then its associated quadratic form is as follows:

$$\begin{aligned}&q_A(u, v) = \begin{pmatrix} u \\ v \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} \cdot \begin{pmatrix} au + bv \\ cu + dv \end{pmatrix} =\\ &u(au + bv) + v(cu + dv) = u^2a + uvb + vuc + v^2d\end{aligned}$$

Now, suppose that A is the Hessian matrix of the bivariate Gaussian function:

$$\mathsf{H}g(x, y) = \begin{pmatrix} -x^2e^{-\frac{x^2+y^2}{2}} & -xye^{-\frac{x^2+y^2}{2}} \\ -xye^{-\frac{x^2+y^2}{2}} & -y^2e^{-\frac{x^2+y^2}{2}} \end{pmatrix}$$

Then, its quadratic-form is defined as follows:

$$\begin{aligned}& q_{\mathsf{H}g}(u, v) = u^2\partial^2_{xx}g + uv\partial^2_{xy}g + vu\partial^2_{yx}g + v^2\partial^2_{yy}g =\\ = \quad & -u^2x^2e^{-\frac{x^2+y^2}{2}} - uvxye^{-\frac{x^2+y^2}{2}} - vuyxe^{-\frac{x^2+y^2}{2}} - v^2y^2e^{-\frac{x^2+y^2}{2}} =\\ = \quad & -(u^2x^2 + 2uvxy + v^2y^2)e^{-\frac{x^2+y^2}{2}}\end{aligned}$$

This is the quadratic form of the Hessian matrix of the bivariate Gaussian function. ■

Proposition 3.8. *Let $\hat{u}$ be a versor. Suppose also that $f : \Omega \to \mathbb{R}$ is a function of class C^2. Then:*

$$\frac{\partial^2 f(\vec{x})}{\partial u^2} = q_{\mathsf{H}f(\vec{x})}(\hat{u}) = \hat{u} \cdot \mathsf{H}f(\vec{x})\hat{u}$$

Definition 3.14. *An $n \times n$ square matrix A is:*

1. ***positive-semidefinite*** *if for any versor $\hat{u}$, $q_A(\hat{u}) \geq 0$*
2. ***positive-definite*** *if for any versor $\hat{u}$, $q_A(\hat{u}) > 0$*
3. ***negative-semidefinite*** *if for any versor $\hat{u}$, $q_A(\hat{u}) \leq 0$*
4. ***negative-definite*** *if for any versor $\hat{u}$, $q_A(\hat{u}) < 0$*

where q_A is the quadratic-form associated to A.

Finally, we can enounce the main result of this section.

Theorem 3.3 (Hessian criterion). *Let $f : \Omega \to \mathbb{R}$ be a scalar function of class C^2. Suppose also that $\vec{x}_0$ is an internal point to Ω.*

1. *If $\vec{x}_0$ is a point of local extreme value for f, then $\vec{x}_0$ is a stationary point for f*
2. *If $\vec{x}_0$ is a point of local maximum for f, then $\mathsf{H}f(\vec{x}_0)$ is negative-semidefinite*
3. *If $\vec{x}_0$ is a point of local minimum for f, then $\mathsf{H}f(\vec{x}_0)$ is positive-semidefinite*
4. *If $\vec{x}_0$ is a stationary point for f and $\mathsf{H}f(\vec{x}_0)$ is negative-definite then $\vec{x}_0$ is a point of local maximum for f*
5. *If $\vec{x}_0$ is a stationary point for f and $\mathsf{H}f(\vec{x}_0)$ is positive-definite then $\vec{x}_0$ is a point of local minimum for f*

Luckily, for 2×2 symmetric matrices we can determine when the matrix is positive/negative (semi)definite by looking at its determinant and at two entries. Note that this is only true for symmetric 2×2 matrices.

Proposition 3.9. *Let A be a 2×2 symmetric square matrix:*

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

Then:

1. *A is positive-semidefinite if and only if $a, c \geq 0$ and $\det A \geq 0$*
2. *A is positive-definite if and only if $a > 0$ and $\det A > 0$*
3. *A is negative-semidefinite if and only if $a, c \leq 0$ and $\det A \geq 0$*
4. *A is negative-definite if and only if $a < 0$ and $\det A > 0$*

Therefore, since the Hessian matrix is symmetric we have the following corollary.

Corollary 3.1. *Let $f : \Omega \to \mathbb{R}$ be a scalar function of class C^2 where $\Omega \subseteq \mathbb{R}^2$. Moreover, let $(x_0, y_0) \in \Omega$ be an internal and stationary point for f. Let:*

$$\mathsf{h}_f(x_0, y_0) := \det \mathsf{H}f(x_0, y_0) = \frac{\partial^2 f}{\partial x^2}(x_0, y_0)\frac{\partial^2 f}{\partial y^2}(x_0, y_0) - \left(\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)\right)^2$$

Then:

1. *If $\mathsf{h}_f(x_0, y_0) > 0$ and $\partial^2_{xx} f(x_0, y_0) > 0$ then $f(x_0, y_0)$ is a local minimum for f*
2. *If $\mathsf{h}_f(x_0, y_0) > 0$ and $\partial^2_{xx} f(x_0, y_0) < 0$ then $f(x_0, y_0)$ is a local maximum for f*

Let's come back for a second to the pringle-shaped surface. As we said, this is not a point of local maximum or minimum, because in one direction $\partial^2_u f$ is positive, while in another one, $\partial^2_u f$ is negative. This is an example of a saddle point.

Definition 3.15. *A **saddle point** of a function $f : \Omega \to \mathbb{R}$ is a stationary point for f which is not a point of local extreme value for f.*

It turns out that when $n = 2$ we can classify all the saddle points as follows.

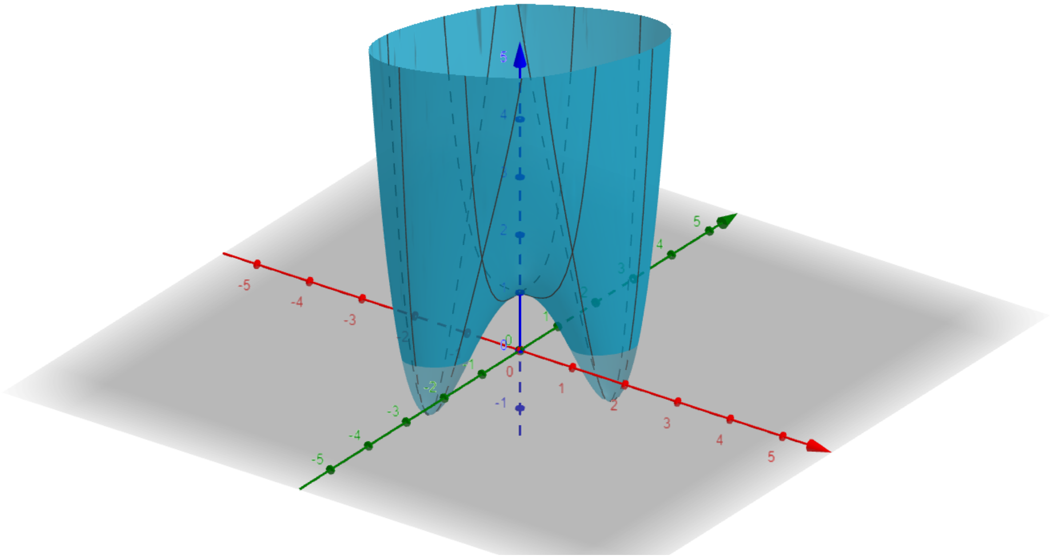


Figure 3.13: A representation of the graph of the function $f(x, y) = x^4 + y^4 - 4xy + 1$ plotted with Geogebra

---

Proposition 3.10. *Let $f : \Omega \to \mathbb{R}$ be a scalar function of class C^2 where $\Omega \subseteq \mathbb{R}^2$. Moreover, let $(x_0, y_0) \in \Omega$ be an internal and stationary point for f. Let:*

$$\mathrm{h}_f(x_0, y_0) := \det \mathrm{H}f(x_0, y_0) = \frac{\partial^2 f}{\partial x^2}(x_0, y_0)\frac{\partial^2 f}{\partial y^2}(x_0, y_0) - \left(\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)\right)^2$$

Then:

1. *If $\mathrm{h}_f(x_0, y_0) < 0$, then (x_0, y_0) is a saddle point for f*

Example 3.18. Let's find all points of local extreme value and all saddle points of the function:

$$f(x, y) := x^4 + y^4 - 4xy + 1$$

First, let's find the stationary points of f:

$$\frac{\partial f}{\partial x}(x, y) = 4x^3 - 4y$$
$$\frac{\partial f}{\partial y}(x, y) = 4y^3 - 4x$$
$$\vec{\nabla} f(x, y) = \left\langle 4x^3 - 4y, 4y^3 - 4x \right\rangle = 4\left\langle x^3 - y, y^3 - x \right\rangle$$

We need to find the points (x, y) so that $\vec{\nabla} f(x, y) = \langle 0, 0\rangle$, so we have to solve the system:

$$\begin{cases} y = x^3 \\ x = y^3 \end{cases}$$

So:

$$x = (x^3)^3 = x^9$$

that can be rewritten as:

$$x(x^8 - 1) = 0$$

Therefore, $x = 0, \pm 1$, and $y = x^3$, so the stationary points are $(0,0), (1,1), (-1,-1)$. Let's now evaluate $\mathsf{h}_f(x,y)$:

$$\frac{\partial^2 f}{\partial x^2}(x,y) = 12x^2$$
$$\frac{\partial^2 f}{\partial x \partial y}(x,y) = -4 = \frac{\partial^2 f}{\partial y \partial x}(x,y)$$
$$\frac{\partial^2 f}{\partial y^2}(x,y) = 12y^2$$
$$\mathsf{H}f(x,y) = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}$$
$$\mathsf{h}_f(x,y) = \det \mathsf{H}f(x,y) = 144x^2y^2 - 16 = 16(9x^2y^2 - 1)$$

Let's find the points where $\mathsf{h}_f(x,y) > 0$:

$$16(9x^2y^2 - 1) > 0$$

Thus:

$$x^2y^2 > \frac{1}{9}$$

and, by taking the square root:

$$|xy| > \frac{1}{3}$$

We have that for $(1,1)$ and $(-1,-1)$, $|xy| = 1 > 1/3$, so for these two points $\mathsf{h}_f(x,y) > 0$. For $(0,0)$ instead we have that $\mathsf{h}_f(0,0) = -16 < 0$.Thus, $(0,0)$ is a saddle point for f. Finally, let's evaluate $\partial^2_{xx} f$:

$$\frac{\partial^2 f}{\partial x^2}(1,1) = 12 = \frac{\partial^2 f}{\partial x^2}(-1,-1) > 0$$

Thus we have the following result:

1. $(1,1)$ is a point of local minimum for f
2. $(0,0)$ is a saddle point for f
3. $(-1,-1)$ is a point of local minimum for f

In Picture 3.13 we plotted with Geogebra the graph of the function f. As you can see, the points $(1,1)$ and $(-1,-1)$ correspond to the points of local minimum of f and $(0,0)$ is a saddle point. ■

Example 3.19. In Example 3.15 we classified the critical points of the function:

$$f(x,y) = y^2 - x^3 - x^2$$

From Picture 3.11 we can see that $(0,0)$ is a saddle point, while $(-2/3, 0)$ is a point of local minimum for f. Now, we wanna prove this result by using the Hessian criterion. Note that the domain of the function is $\mathbb{R}^2$

so we are in two dimensions and there is no boundary, so every point of the domain is internal. Moreover, the function is of class C^2 being a polynomial function.
To apply the criterion, first let's compute the Hessian matrix of f:

$$\begin{aligned}
&\partial_x f(x,y) = -3x^2 - 2x \\
&\partial_y f(x,y) = 2y \\
&\partial^2_{xx} f(x,y) = -6x - 2 \\
&\partial^2_{xy} f(x,y) = 0 = \partial^2_{yx} f(x,y) \\
&\partial^2_{yy} f(x,y) = 2
\end{aligned}$$

So, we have:

$$\mathsf{H}f(x,y) = \begin{pmatrix} -6x-2 & 0 \\ 0 & 2 \end{pmatrix}$$

The determinant of the hessian is therefore the function:

$$\mathsf{h}_f(x,y) = -12x - 4$$

Now, in $(0,0)$ we have the following:

$$\mathsf{h}_f(0,0) = -4$$

Therefore, since $\mathsf{h}_f(0,0) < 0$, $(0,0)$ is a saddle point, as expected. Let's now evaluate $\mathsf{h}_f(-2/3,0)$:

$$\mathsf{h}_f(-2/3,0) = -12 \cdot \left(-\frac{2}{3}\right) - 4 = 8 - 4 = 4$$

Thus, $\mathsf{h}_f(-2/3,0) > 0$. To conclude, note that:

$$\partial^2_{xx} f(-2/3,0) = -6 \cdot \left(-\frac{2}{3}\right) - 2 = 4 - 2 = 2$$

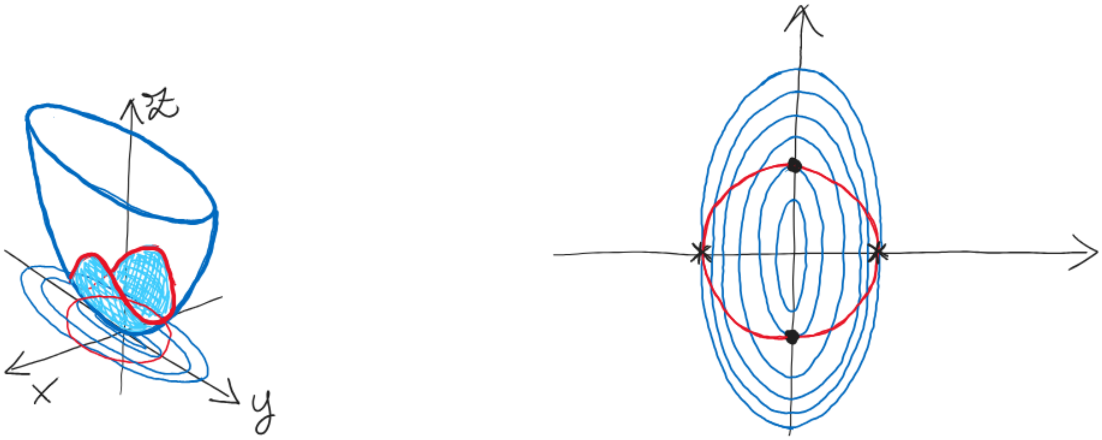
Thus, because $\mathsf{h}_f(-2/3,0) > 0$ and that $\partial^2_{xx} f(-2/3,0) > 0$, then the Hessian matrix of f in $(-2/3,0)$ is positive-definite which means that $(-2/3,0)$ is indeed a local minimum for f. ■

3.10 Lagrange multipliers

When K is a compact set and $f : K \to \mathbb{R}$ is a continuous function, by the Extreme-Value theorem (Theorem 3.1) we know there are at least two points $\vec{x}_0, \vec{x}_1 \in K$ such that $f(\vec{x}_0)$ is a global maximum for f and $f(\vec{x}_1)$ is a global minimum for f. Thanks to the Hessian criterion (Theorem 3.3) we have a way to find all points of local maximum and local minimum for f on the interior of K, however, we still don't know how to find the points of extreme value on the boundary of K.

Example 3.20. Consider the following function:

$$\begin{aligned}
&f : K \to \mathbb{R} \\
&f(x,y) := x^2 + \frac{y^2}{4}
\end{aligned}$$



(a) The graph of the elliptical paraboloid $f(x,y) = x^2 + y^2/4$

(b) The isocurves (in blue) of the elliptical paraboloid $f(x,y) = x^2 + y^2/4$ and the boundary ∂K (in red) of the domain

Figure 3.14: Graph of the elliptical paraboloid of Example 3.20

---

where $K := \overline{\mathbb{B}}(\vec{0}, 1)$ is the closed unit ball in $\mathbb{R}^2$. In Picture 3.14a there is a representation of f. From the graph, we can clearly see that $(0,0)$ is a point of global minimum for the function, but where is the global maximum? f is a function of class C^2 and $\vec{\nabla} f = \vec{0}$ only in $(0,0)$, therefore, we conclude that the global maximum cannot be an internal point. The boundary ∂K of K is the unit circle, i.e.:

$$\partial K = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\} = \mathbb{S}^1$$

In Picture 3.14b we plot the contour diagram of $f(x,y) = x^2 + y^2/4$. In blue we drew the isocurves of f and in red the boundary of K. Note that in • and in $*$, the tangent line of the isocurves of f and the tangent line of the boundary ∂K are parallel. ■

Lemma 3.1. *Suppose that $f : B \longrightarrow \mathbb{R}$ is a function of class C^1 and that B is a regular subset of $\mathbb{R}^n$, that is for every point $\vec{x} \in B$, the tangent space of B at $\vec{x}$ is well-defined. If $\vec{x}_0 \in B$ is a point of local extreme value for f in B and $\vec{\nabla} f(\vec{x}_0) \neq \vec{0}$, then $\vec{\nabla} f(\vec{x}_0)$ is orthogonal to the tangent space of B in $\vec{x}_0$.*

Proof. Consider a path $\vec{r} : [-1,1] \longrightarrow B$ so that $\vec{r}(0) = \vec{x}_0$. Because $f(\vec{x}_0) = f(\vec{r}(0))$ is a local extreme value for f, then it is also a local extreme value for the function $f \circ \vec{r} : [-1,1] \longrightarrow \mathbb{R}$. Note also that $0 \in [-1,1]$ is now an internal point, therefore the total derivative of $f \circ \vec{r}$ in 0 has to be zero:

$$\begin{aligned}
0 = \; & \frac{d(f \circ \vec{r})}{dt}(0) = \\
= \; & \frac{\partial f}{\partial x_1}(\vec{r}(0))\frac{dx_1}{dt}(0) + \cdots + \frac{\partial f}{\partial x_n}(\vec{r}(0))\frac{dx_n}{dt}(0) = \\
= \; & \frac{\partial f}{\partial x_1}(\vec{x}_0)\frac{dx_1}{dt}(0) + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}_0)\frac{dx_n}{dt}(0) = \\
= \; & \vec{\nabla} f(\vec{x}_0) \cdot \frac{d\vec{r}}{dt}(0)
\end{aligned}$$

where we used the chain rule. But $\vec{r}\,'(0)$ is a vector of the tangent space of B in $\vec{x}_0$. Moreover, any vector $\vec{v}$ of the tangent space of B in $\vec{x}_0$ is the total derivative of a path of B that passes through $\vec{x}_0$ at time $t = 0$, i.e. $\vec{v} = \vec{r}'(0)$, for some path $\vec{r}$. Therefore, the gradient $\vec{\nabla} f(\vec{x}_0)$ is orthogonal to every vector of the tangent space of B in $\vec{x}_0$, thus it is also orthogonal to the whole tangent space of B in $\vec{x}_0$. $\square$

One way of thinking about the boundary ∂K of the compact domain K of f is to regard it as the level set of another function g. So, let's assume from now on, that ∂K is as follows:

$$\partial K = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid g(x_1, \ldots, x_n) = c\}$$

for some scalar function $g : \mathbb{R}^n \to \mathbb{R}$ of class C^1 and for some number c. For the Example 3.20, such a function g could be defined as follows:

$$g(x, y) := x^2 + y^2$$

and $c = 1$, so we have that:

$$\Gamma_1 g = \mathbb{S}^1 = \partial K$$

With this assumption, thanks to Proposition 3.6, we have that if $\vec{x}_0 \in \partial K$, then $\vec{\nabla} g(\vec{x}_0)$ is orthogonal to the tangent space of ∂K in $\vec{x}_0$. But we also know, from Lemma 3.1, that if $\vec{x}_0$ is a point of local extreme value for the function f, then $\vec{\nabla} f(\vec{x}_0)$ is also orthogonal to ∂K in $\vec{x}_0$. Therefore, since both $\vec{\nabla} g(\vec{x}_0)$ and $\vec{\nabla} f(\vec{x}_0)$ are orthogonal to ∂K in $\vec{x}_0$, they have to be inevitably parallel!
This means that there exists a number $\lambda \in \mathbb{R}$ such that:

$$\vec{\nabla} f(\vec{x}_0) = \lambda \vec{\nabla} g(\vec{x}_0)$$

Such a number is called a Lagrange multiplier.

Theorem 3.4 (Lagrange multipliers). *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a scalar function of class C^1, let $B := \Gamma_c g$ for some number $c \in \mathbb{R}$ such that B is not empty. Suppose also that $\vec{\nabla} g(\vec{x}) \neq \vec{0}$ for every $\vec{x} \in B$. If $f : B \to \mathbb{R}$ is a scalar function of class C^1 and $\vec{x}_0$ is a point of local extreme value for f in B, then there exists a unique number $\lambda \in \mathbb{R}$, known as the Lagrange multiplier, such that:*

$$\vec{\nabla} f(\vec{x}_0) = \lambda \vec{\nabla} g(\vec{x}_0)$$

Sometimes, the equation $g(\vec{x}) = c$ that defines the c-level-set B is called a constraint for f.

Example 3.21. Suppose that we have $12m^2$ of cardboard paper and that we want to make a rectangular box out of this paper, which is as large as possible. We want to calculate the measures of such a box. So, the function we want to maximize represents the volume of the box. So, if x, y and x represent the length, the width and the depth of the box respectively, we have:

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

Note also that we can assume $x, y, z > 0$ because we don't measure lengths with negative numbers. We also know that the total amount of paper available is $12m^2$, so this means that the external surface of the box will be given by:

$$2xy + 2xz + 2yz = 12$$

We can simplify slightly this equation as follows:

$$xy + xz + yz = 6$$

This equation is our constraint, so we can decide to take:

$$g(x,y,z) := xy + xz + yz$$

and $c = 6$, so that the equation $g(x,y,z) = 6$ defines the set B. So, this means that we want to maximize the function f among the possible points on the constraint B. So, to do that, we need to use the Lagrange multipliers. Concretely, this means that we want to solve the following system of equations:

$$\begin{cases} \vec{\nabla} f(x,y,z) = \lambda \vec{\nabla} g(x,y,z) \\ g(x,y,z) = 6 \end{cases}$$

Let's start by computing the partial derivatives of f and of g:

$$\begin{aligned} \partial_x f &= yz \\ \partial_y f &= yz \\ \partial_z f &= xy \\ \partial_x g &= y + z \\ \partial_y g &= x + z \\ \partial_z g &= x + y \end{aligned}$$

So, the vectorial equation $\vec{\nabla} f = \lambda \vec{\nabla} g$ becomes:

$$\begin{cases} \partial_x f = \lambda \partial_x g \\ \partial_y f = \lambda \partial_y g \\ \partial_z f = \lambda \partial_z g \end{cases}$$

Which means:

$$\begin{cases} yz = \lambda(y+z) \\ xz = \lambda(x+z) \\ xy = \lambda(x+y) \end{cases}$$

Let's multiply the first equation by x, the second equation by y and the third equation by z, so we obtain:

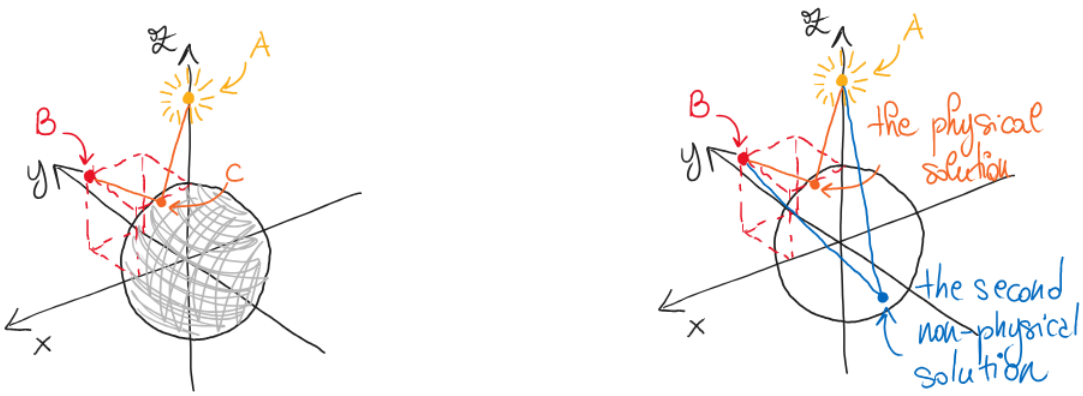
$$\begin{cases} xyz = \lambda(xy+xz) \\ xyz = \lambda(xy+yz) \\ xyz = \lambda(xz+yz) \end{cases}$$

So, we have the equations:

$$\begin{cases} \lambda(xy+xz) = \lambda(xy+yz) \\ \lambda(xy+yz) = \lambda(xz+yz) \end{cases}$$

Now, notice that if $\lambda = 0$, then $xyz = 0$, which means the volume of the box will be zero. This is definitely not the configuration that maximizes the volume of the box, so we can assume that $\lambda \neq 0$. Therefore, we can simplify by λ. Similarly, if any of x, y or z is zero, we encounter the same issue: the volume would be null. So, we can also assume that $x, y, z \neq 0$, thus we obtain that $x = y = z$. Now, let's use the second equation of the original system, i.e. $g(x,y,z) = 6$ and replace $x = y = z$:

$$6 = xy + xz + yz = x^2 + x^2 + x^2 = 3x^2$$



(a) A laser bim in point A reflects the light on a spherical mirror C and reaches the detector at point B

(b) The two solutions of Example 3.22: in orange the physical solution, in blue the non-physical one

Figure 3.15: Fermat's principle and Lagrange multipliers

So, we obtain that $x = y = z = \pm\sqrt{2}$. But remember that we assumed $x, y, z > 0$, so we have that $x = y = z = \sqrt{2}$. These are the measures for the length, width and depth that maximize the volume of the box. Finally, we can also compute the maximal volume of the box, as $xyz = 2\sqrt{2}$. ■

Example 3.22. Fermat's law establishes that light moves along the path that minimizes the time of travel. This means that if a source of light is in a point A and a light detector is at a distinct point B in space, the path that light decided to take to move from A to B is the path that takes the least amount of time. This principle has interesting consequences. In this example, we are going to explore one in particular: reflection along a spherical mirror.
Imagine that a laser pointer is in position $A = (0, 0, 2)$ and it's pointing toward a mirror shaped like the unit sphere, centred in $(0, 0, 0)$. In Picture 3.15a we represented this situation. A light detector at point $B = (1, 1, 1)$ is supposed to detect the laser bim, after that the light is reflected on the mirror. We want to find the coordinates on the spherical mirror where the light is reflected.
As assumptions, we are assuming that the air has the same density everywhere. This assumption allows us to establish that, in this scenario, the path the minimizes the time-travel, is also the shortest path from A to B that passes through one point C, the reflection point, on the spherical mirror. Note that if we have assumed that in some region the air density was different, this would have affected the speed of light. Indeed, the speed of light is not the same in every substance but it depends from the intrinsic properties of the material.

Suppose that (x, y, z) is a point on the spherical mirror. This means that:

$$x^2 + y^2 + z^2 = 1$$

This equation is the constraint of the system. So, let $g(x, y, z) := x^2 + y^2 + z^2$, so the constraint equation becomes $g(x, y, z) = 1$. Now, we know that the light, from A to C will move along a straight line. This simply because straight lines are the shortest paths in empty space. Similarly, from C to B the light bim will also move along a straight line. So, the total length of the path that the light will do is $d(A, C) + d(C, B)$, where d is the distance function, i.e. the function that measures the distance between two points. Now, it turns out that by taking f to be this function, the computations are a bit too involved. So, instead we are going to minimize the function $d^2(A, C) + d^2(C, B)$. By taking the square of the distances, we are going to find the same solution. However, the computations will be much easier. So, by calling $C = (x, y, z)$, we define our function f to be:

$$\begin{aligned} f(x,y,z) &= d^2((2,0,0),(x,y,z)) + d^2((x,y,z),(1,1,1)) = \\ &= (2-x)^2 + y^2 + z^2 + (x-1)^2 + (y-1)^2 + (z-1)^2 \end{aligned}$$

So, to recap: we are going to find the local minima of the function f among the points on the spherical mirror, i.e. the points (x,y,z) of the constraint $g(x,y,z) = 1$. This is a minimization problem on a constraint, therefore we are going to use the Lagrange multipliers technique. So we have the following system of equations:

$$\begin{cases} \vec{\nabla} f(x,y,z) = \lambda \vec{\nabla} g(x,y,z) \\ g(x,y,z) = 1 \end{cases}$$

Let's start by evaluating the partial derivatives of f and g:

$$\begin{aligned} \partial_x f &= 2x + 2(x-1) = 4x - 2 \\ \partial_y f &= 2y + 2(y-1) = 4y - 2 \\ \partial_z f &= 2(z-2) + 2(z-1) = 4z - 6 \\ \partial_x g &= 2x \\ \partial_y g &= 2y \\ \partial_z g &= 2z \end{aligned}$$

Thus we have:

$$\begin{cases} 4x - 2 = 2\lambda x \\ 4y - 2 = \lambda y \\ 4z - 6 = 2\lambda z \\ x^2 + y^2 + z^2 = 1 \end{cases}$$

So we have:

$$\begin{cases} (2-\lambda)x = 1 \\ (2-\lambda)y = 1 \\ (2-\lambda)z = 3 \\ x^2 + y^2 + z^2 = 1 \end{cases}$$

Note that $\lambda \neq 2$ because otherwise, we would have $0 = 1$. Therefore, we have that $x = y = 1/(2-\lambda)$ and that $z = 3/(2-\lambda) = 3x = 3y$. So, by using the constraint equation we find that $11x^2 = 1$, thus $x = y = \pm 1/\sqrt{11}$ and $z = \pm 3/\sqrt{11}$. So, we have two solutions: $(1/\sqrt{11}, 1/\sqrt{11}, 3/\sqrt{11})$ and $(-1/\sqrt{11}, -1/\sqrt{11}, -3/\sqrt{11})$. At this point, we need to use a little bit of our intuition to understand what is going on.

In Picture 3.15b we represented the two solutions. As you can see, with the first solution $(1/\sqrt{11}, 1/\sqrt{11}, 3/\sqrt{11})$ we precisely obtain what we were expecting: the reflection point on the mirror where the bim light hits, by moving along the path that minimizes the time. The second solution does not make sense physically: here the math doesn't know that the light cannot travel through the mirror, but that can only reflect. Incidentally, this solution is also not a point of minimum.

This is not a contradiction: the theorem tells us that points of local minimum or local maximum will satisfy the Lagrange multipliers equation, but it doesn't say that if a point is a solution of such equation is necessarily a point of local extreme value. ■

MODULE 4

Multiple integrals

In this chapter, we introduce integration calculus for multivariable functions.

4.1 Review on ordinary integrals

Before we go into the multivariable setting, let's recap the definition of integrals for ordinary functions. Suppose that $f : [a, b] \to \mathbb{R}$ is a function defined over a closed interval. Note that we can extend this construction to functions defined over generic subsets $\Omega \subseteq \mathbb{R}$, but this makes things a bit more complicated, so, for now, we will simply work with closed intervals.
The construction needs the following steps:

1. First, we define a special class of functions, called step functions.
2. Second, we define the integral of a step function. These functions are the simplest ones to integrate.
3. Third, we are going to define a sequence of step functions that better and better approximate the function f.
4. Fourth, finally, we define the integral of f as a limit of integrals of step functions.

Let's start with the first point. To define what a step function is, first, we introduce the notion of a partition.

Definition 4.1. *A* ***partition*** *of a non-empty set $\Omega \subseteq \mathbb{R}^n$ is a collection of open sets $\{S_i\}$ (possibly infinite) such that:*

1. *Every S_i is a subset of Ω*
2. *Every S_i is non-empty*
3. *Taken any pair of sets S_i, S_j with $i \neq j$, S_i and S_j are disjoint, i.e. $S_i \cap S_j = \emptyset$*
4. *The union of all the the closure of S_i is equal to Ω, i.e.:*

$$\bigcup_i \overline{S}_i = \Omega$$

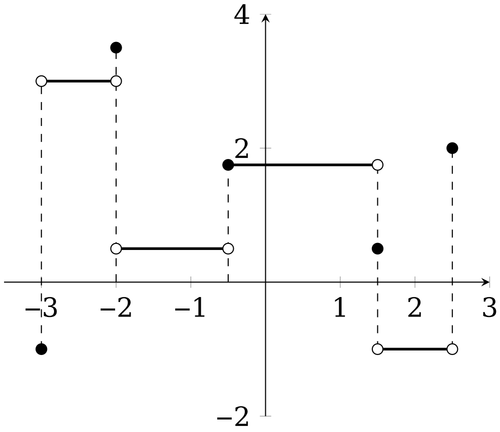


Figure 4.1: Example of a step function. By definition, there are some open intervals (x_i, x_{i+1}) where the function is constant. Notice that the definition does not specify what the function should be on the extreme points x_i.

If the partition has a finite number of sets $S_1, S_2, \ldots, S_n$, the partition is called finite.

In Picture ?????? there is a representation of a possible partition of a set Ω. Intuitively speaking, a partition is a collection of tiles that covers the whole set Ω and that they can only overlap over the boundaries. A finite partition of $[a, b]$ is given by a finite number of numbers $x_0, x_1, , \ldots, x_n \in [a, b]$ such that:

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b$$

Indeed, with such a list of $x_0, x_1, \ldots, x_n$ we can take the intervals $(x_0, x_1), (x_1, x_2), \ldots, (x_{n-1}, x_n)$ and this is a partition of $[a, b]$. To see that, note that all of these intervals are disjoint, because $x_0 < x_1 < \cdots < x_n$. Moreover, each $(x_i, x_{i+1}) \subseteq [a, b]$, because $a \leq x_i < x_{i-1} \leq b$. Finally, by taking the closure of (x_i, x_{i+1}) we have the closed interval $[x_i, x_{i+1}]$, so the union of all these closed intervals $[x_0, x_1] \cup [x_1, x_2] \cup \cdots \cup [x_{n-1}, x_n]$ is precisely equal to $[a, b]$.

With this in mind, we can define a step function as follows:

Definition 4.2. *Let $\Omega \subseteq \mathbb{R}^n$ be a non-empty set and let $P := \{S_1, S_2, \ldots, S_n\}$ be a finite partition of Ω. A* ***step function*** *for the partition P is a scalar function $s : \Omega \to \mathbb{R}$ such that on each S_i, the function s is constant. Concretely, this means that there is a list of real numbers $s_1, s_2, \ldots, s_n \in \mathbb{R}$ such that:*

1. *for every $x \in S_1$, $s(x) = s_1$*
2. *for every $x \in S_2$, $s(x) = s_2$*

$\vdots$

i. for every $x \in S_i$, $s(x) = s_i$

$\vdots$

n. for every $x \in S_n$, $s(x) = s_n$

When $\Omega = [a, b]$, we can rephrase this definition as follows: a step function is a function $s : [a, b] \to \mathbb{R}$ such that there are $x_0, x_1, \dots, x_n \in [a, b]$ and $s_1, s_2, \dots, s_n \in \mathbb{R}$ such that $a = x_0 < x_1 < \cdots < x_n = b$ and for each $x \in (x_i, x_{i+1})$, $s(x) = s_i$, for every $i = 0, \dots, n-1$. In Picture 4.1 there is an example of a step function. Note that on the points $x_0, x_1, \dots, x_n$ the function can have any value.
Now that we have defined the notion of a step function, we can introduce the integral of such a function. Intuitively, the graph of a step function can be regarded as a collection of boxes $[x_i, x_{i+1}] \times [0, s_i]$, so the whole (oriented) area of a step function is going to be the sum of the areas of the boxes. The area of each box is given by the height s_i times the length of the basis, i.e. $x_{i+1} - x_i$. This length is also called the measure of the set $[x_{i+1}, x_i]$.

Definition 4.3. *Given a step function $s : [a, b] \to \mathbb{R}$ for a partition P of $[a, b]$ given by $x_0 < x_1 < \cdots < x_n$, the* ***integral*** *of s over $[a, b]$ is the number so defined:*

$$\int_a^b s(x)\mathsf{d}x = s_1(x_1 - x_0) + s_2(x_2 - x_1) + \cdots + s_n(x_n - x_{n-1}) = \sum_{k=1}^{n-1} s_k(x_k - x_{k-1})$$

where s_i is the value that s assumes in the interval (x_i, x_{i+1}).

Now that we have a notion of integral for step functions, the idea is to extend this construction to other kinds of functions. Many times, in mathematics, we have an operation that is easy to define for a special class of functions. To extend it to a larger class of functions, use the fact that these latter can be approximated by functions of the first class. Note that if we consider a continuous function $f : [a, b] \to \mathbb{R}$ and we fix an integer $n \in \mathbb{N}$, we can construct a step function s_n that looks like f.

The idea is to split the interval $[a, b]$ into n small intervals each of the same length $\Delta x = (b - a)/n$. So, we have the following partition: $x_0 = a$, $x_1 := x_0 + \Delta x$, $x_2 := x_1 + \Delta x = x_0 + 2\Delta x$, $x_i = x_0 + i\Delta x$, $x_n = b$. Now, we pick a point from each of these intervals. It can be any point, we decide to take the right extreme point. Now, from each of these points, $x_1, x_2, \dots, x_n$ we take the value of the function at these points $f(x_1), f(x_2), \dots, f(x_n)$, so our step function $s_n : [a, b] \to \mathbb{R}$ is defined as follows:

$$s_n(x) := \begin{cases} f(x_1) & \text{if } x \in [x_0, x_1) \\ f(x_2) & \text{if } x \in [x_1, x_2) \\ \vdots \\ f(x_i) & \text{if } x \in [x_i, x_{i+1}) \\ \vdots \\ f(x_n) & \text{if } x \in [x_{n+1}, x_n] \end{cases}$$

Finally, we can take the integral of s_n over $[a, b]$. In Picture ?????? there is a representation of this construction. Notice that, by increasing n, we find a better and better approximation for the area of the function f. So, the idea is to take the limit of the numerical sequence $S_n := \int_a^b s_n(x)dx$.

Definition 4.4. *Given a function $f : [a, b] \to \mathbb{R}$ we say that f is* ***integrable*** *if the sequence:*

$$S_n := \int_a^b s_n(x)\mathsf{d}x$$

converges, i.e. if the limit:

$$\lim_{n\to\infty} \int_a^b s_n(x)\mathsf{d}x$$

exists. When such limit exists, we call it the ***integral*** *of f on $[a, b]$ and we call $[a, b]$ the* ***integration domain****. Finally, we will use the following notation:*

$$\int_a^b f(x)\mathsf{d}x := \lim_{n\to\infty}\int_a^b s_n(x)\mathsf{d}x$$

4.2 Integrals for vector functions and arclength

In this section, we will focus on extending the notion of integration to multivariable functions. We are going to study only two cases:

1. The integral of a vector function $\vec{f} : [a, b] \to \mathbb{R}^n$
2. The integral of a scalar function $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}^2$ is a bounded region of the plane

Let's start by considering vector functions $\vec{f} : [a, b] \to \mathbb{R}^n$. For each $t \in [a, b]$ $\vec{f}(t) = \langle x_1(t), x_2(t), \ldots, x_n(t)\rangle$, and each x_i is a function $x_i : [a, b] \to \mathbb{R}$. So, $\vec{f}$ is integrable if and only if all $x_1, x_2, \ldots, x_n$ are integrable as well and in that case we have:

$$\int_a^b \vec{f}(t)\mathsf{d}t = \left\langle \int_a^b x_1(t)\mathsf{d}t, \int_a^b x_2(t)\mathsf{d}t, \ldots, \int_a^b x_n(t)\mathsf{d}t \right\rangle$$

Example 4.1. Consider the following function:

$$\vec{f} : [0, \pi] \to \mathbb{R}$$
$$\vec{f}(t) := \langle \cos t, \sin t, t\rangle$$

Let's find the integral of this function from 0 to π:

$$\begin{aligned}
&\int_0^\pi \vec{f}(t)\mathsf{d}t = \\
= \quad & \left\langle \int_0^\pi \cos t\mathsf{d}t, \int_0^\pi \sin t\mathsf{d}t, \int_0^\pi t\mathsf{d}t \right\rangle = \\
= \quad & \left\langle \sin t|_0^\pi, -\cos t|_0^\pi, \left.\frac{t^2}{2}\right|_0^\pi \right\rangle = \\
= \quad & \left\langle \sin\pi - \sin 0, -\cos\pi + \cos 0, \frac{\pi}{2} \right\rangle = \\
= \quad & \left\langle 0, 2, \frac{\pi}{2} \right\rangle
\end{aligned}$$

This concludes the example. ■

Since we are talking about vector functions and integration, we dedicate some time to discussing the notion of arclength of a curve. Given a vector function $\vec{f} : [a, b] \to \mathbb{R}^n$, we can think of $\vec{f}$ as a path that starts at time $t = a$ in $\vec{f}(a)$ and ends at time $t = b$ in $\vec{f}(b)$. A good question is to ask what is the total length of this path. Such length is usually called arclength of the path. To understand how to compute such value, let's first understand what is the infinitesimal length $\mathsf{d}s$ of the path.

In Picture ??????? you can see that if we fix a point along the path $\vec{f}(t_0)$ and we imagine moving slightly further at $\vec{f}(t + t_0)$ for a very small $t > 0$, we can approximate the function $\vec{f}$ at $t + t_0$ using the linear approximation, which reads as follows:

$$\vec{f}(t + t_0) \simeq \vec{f}(t_0) + (t - t_0)\vec{f}'(t_0)$$

So, the distance between $\vec{f}(t_0)$ and $\vec{f}(t+t_0)$ is approximately equal to the distance between $\vec{f}(t_0)$ and $\vec{f}(t_0)+(t-t_0)\vec{f}'(t_0)$, which gives:

$$\mathrm{d}s = \left|\vec{f}(t+t_0)-\vec{f}(t_0)\right| \cong \left|\vec{f}(t_0)+(t-t_0)\vec{f}'(t_0)-\vec{f}(t_0)\right| = \left|(t-t_0)\vec{f}'(t_0)\right| = \left|\vec{f}'(t_0)\right|\mathrm{d}t$$

where we indicated with $\mathrm{d}t$ the infinitesimal interval of time $t-t_0$. So, to find the arclength, we need to sum all of these infinitesimal contributes $\mathrm{d}s$:

$$\mathsf{L}\vec{f}_a^b := \int_a^b \left|\vec{f}'(t)\right|\mathrm{d}t = \int_a^b \sqrt{x_1'^2(t)+x_2'^2(t)+\cdots+x_n'^2(t)}\mathrm{d}t$$

Proposition 4.1. *The arclength of a path $\vec{f}:[a,b]\to\mathbb{R}?n$ is given by the following formula:*

$$\mathsf{L}\vec{f}_a^b := \int_a^b \left|\vec{f}'(t)\right|\mathrm{d}t = \int_a^b \sqrt{x_1'^2(t)+x_2'^2(t)+\cdots+x_n'^2(t)}\mathrm{d}t$$

Example 4.2. Consider the function:

$$\vec{f}:[0,\pi]\to\mathbb{R}^3 \qquad\qquad \vec{f}(t):=\langle\cos t,\sin t,t\rangle$$

Let's calculate the arclength of this function. First, let's evaluate $\vec{f}'(t)$ and $\left|\vec{f}'(t)\right|$:

$$\vec{f}'(t)=\langle-\sin t,\cos t,1\rangle$$
$$\left|\vec{f}'(t)\right|=\sqrt{\sin t^2 t+\cos^2 t+1}=\sqrt{2}$$

Therefore:

$$\mathsf{L}\vec{f}_0^\pi = \int_0^\pi \left|\vec{f}'(t)\right|\mathrm{d}t = \int_0^\pi \sqrt{2}\mathrm{d}t = \sqrt{2}\,t|_0^\pi = \pi\sqrt{2}$$

So, the arclength of the path is $\pi\sqrt{2}$. ■

4.3 Definition of multivariable integrals

Let's now focus on the definition of the integral of functions $f:\Omega\to\mathbb{R}$ where $\Omega\subseteq\mathbb{R}^2$. We want to generalize the construction that we developed before for ordinary functions. So, the steps are as follows:

1. Define step functions in this context
2. Define the integral of step functions
3. Construct a sequence of step functions that better and better approximate the function $f:\Omega\to\mathbb{R}$
4. Define the integral for f over Ω

To do that, we first decide to simplify the choice of f: instead of taking f defined over any possible region Ω of the plane, we decided instead to take a simpler situation: we work with rectangles.

Definition 4.5. *An* ***open rectangle*** *is a subset R of $\mathbb{R}^2$ defined as the Cartesian product between two open intervals, i.e. $R = (a, b) \times (c, d)$, for $a < b$ and $c < d$ four real numbers. Equivalently:*

$$R = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d\}$$

A ***closed rectangle*** *$\overline{R}$ is the closure of an open rectangle R. Concretely, this means that $\overline{R} = [a, b] \times [c, d]$.*

So, in the following f is going to be a real-valued function defined over a closed rectangle $\overline{R} \subseteq \mathbb{R}^2$, i.e. $f : \overline{R} \to \mathbb{R}$. In Definition 4.2 we already gave a definition of a step function. However, this concept is too generic to be applicable in this context. The reason is that a partition for a subset or $\mathbb{R}^2$ can be very complicated to work with. So, instead, we take partitions that contain only open rectangles.

So, the idea is to cover the domain of the function f with a finite number of rectangular tiles $R_1, R_2, \ldots, R_n$ that do not overlap but along the borders and define a step function over such partition. In Picture ????? you can see an example of such a step function. A good representation of a step function is a collection of rectangular boxes.

To define the integral of such a step function, we just need to sum the (oriented) volumes of all of these boxes. Each box has a volume of $s_i \mathsf{m}(R_i)$, where s_i is the height and $\mathsf{m}(R_i)$ is the area of the base rectangle R_i. Sometimes, $\mathsf{m}(R_i)$ is called the measure of R_i.

Definition 4.6. *Let $s : \overline{R} \to \mathbb{R}$ be a step function for the partition $\{R_1, R_2, \ldots, R_n\}$, where each $R_i = (a_i, b_i) \times (c_i, d_i)$ is an open rectangle. The* ***integral*** *of s over R is defined as follows:*

$$\int_R s(x, y) \mathsf{d}A := s_1 \mathsf{m}(R_1) + s_2 \mathsf{m}(R_2) + \cdots + s_n \mathsf{m}(R_n) = \sum_{k=1}^{n} s_i \mathsf{m}(R_i)$$

where s_i is the value of the step function over the rectangle R_i and:

$$\mathsf{m}(R_i) := (b_i - a_i) \cdot (d_i - c_i)$$

is called the ***measure*** *of R_i.*

Now that we have a notion of integral of a step function we want to extend the notion of integral to more generic functions. So, in order to do that, we need to build a sequence of step functions $\{s_n\}$ that better and better approximates f over R.
So, let's fix an integer n and suppose that $R = [A, B] \times [C, D]$ where A, B, C, D are real numbers. Let's divide the intervals $[A, B]$ and $[C, D]$ into n subintervals of the same length $\Delta x := (B - A)/n$ and $\Delta y := (D - C)/n$, respectively. So, if we call $x_0 = A, x_1 = x_0 + \Delta x, \ldots, x_i = x_0$

MODULE 5

Appendix

5.1 Matrix theory*

In this appendix, we just give a brief introduction to matrix theory. There's a lot to say about this topic and our brief exposition is definitely not exhaustive. So, intuitively, a matrix $m \times n$ is a table of numbers with m rows and n columns. Similarly, matrices can also contain functions or other stuff. One of the first things to know about matrices is that they are in one-to-one with linears function. Let's recall this definition here:

Definition 5.1. *Given two vector spaces $\mathbb{V}$ and $\mathbb{V}'$, a **linear function** is a function $f : \mathbb{V} \to \mathbb{V}'$ that satisfies the following condition, known as linearity:*

$$f(a\vec{u} + b\vec{v}) = af(\vec{u}) + bf(\vec{v})$$

for every real number a, b and every vector $\vec{u}, \vec{v} \in \mathbb{V}$.

Linear functions are cool for many different reasons. One of them is that they are pretty easy to work with, once we introduce two bases of the vector spaces $\mathbb{V}$ and $\mathbb{V}'$. Now, let's simplify a little bit the problem and suppose that $\mathbb{V}$ and $\mathbb{V}'$ has finite dimension. This means that there are two bases $B := \{\vec{v}_1, \dots, \vec{v}_n\}$ and $B' := \{\vec{v}'_1, \dots, v'_m\}$ of $\mathbb{V}$ and $\mathbb{V}'$, respectively. By definition of basis, every vector $\vec{u} \in \mathbb{V}$ can now be written in a unique way as follows:

$$\vec{u} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$$

where $x_1, \dots, x_n$ are real numbers. So, if f is a linear function, we can write:

$$f(\vec{u}) = f(x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) = x_1 f(\vec{v}_1) + \cdots + x_n f(\vec{v}_n)$$

So, once we know what are $f(\vec{v}_1), \dots, f(\vec{v}_n)$ we know, for every $\vec{u} \in \mathbb{V}$ what will be $f(\vec{u})$. Another way to say this is saying that to define a linear function $f : \mathbb{V} \to \mathbb{V}'$ we only need to know what vectors are $f(\vec{v}_1), \dots, f(\vec{v}_n) \in \mathbb{V}'$. However, having also a basis B' of $\mathbb{V}'$, and knowing that $f(\vec{v}_1), \dots, f(\vec{v}_n)$ are vectors in $\mathbb{V}'$, we can write each of them in a unique way as follows:

$$\begin{aligned} f(\vec{v}_1) &= a_{1,1}\vec{v}'_1 + a_{2,1}\vec{v}'2 + \cdots + a_{m,1}\vec{v}'_m \\ f(\vec{v}_2) &= a_{1,2}\vec{v}'_1 + a_{2,2}\vec{v}'2 + \cdots + a_{m,2}\vec{v}'_m \\ &\vdots \\ f(\vec{v}_n) &= a_{1,n}\vec{v}'_1 + a_{2,n}\vec{v}'2 + \cdots + a_{m,n}\vec{v}'_m \end{aligned}$$

So, to specify a linear function f, given two bases B and B', we just need the numbers $a_{i,j}$, with $1 \leq i \leq n$ and $1 \leq j \leq m$. A nice representation of these data is given by a matrix.

Definition 5.2. *Let $f : \mathbb{V} \to \mathbb{V}'$ be a linear function between two vector spaces and $B := \{\vec{v}_1, \dots, \vec{v}_m\}$, $B' := \{v'_1, \dots, \vec{v}'_n\}$ two bases for $\mathbb{V}$ and $\mathbb{V}'$, respectively. The* ***associated matrix*** *of f w.r.t. B and B', is the matrix:*

$$A := \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}$$

Concretely, the columns of the matrix A are the components of the vectors $f(\vec{v}_1), \dots, f(\vec{v}_n)$ w.r.t. to B'.

So, this shows that every linear function $f : \mathbb{V} \to \mathbb{V}'$ between two finitely-dimensional vector spaces $\mathbb{V}$ and $\mathbb{V}'$ defines a matrix, whenever two bases are chosen for $\mathbb{V}$ and $\mathbb{V}'$. The opposite is also true: every matrix defines a linear function. To understand this, we first need to introduce some important operations of matrices. In the following, we will simplify the notation of matrices and we will write $A = \left(a_{i,j}\right)_{m \times n}$ to indicate an $m \times n$ matrix A, whose entries are the numbers $a_{i,j}$, with $1 \leq i \leq m$ and $1 \leq j \leq n$.

Definition 5.3. *Suppose that $A = \left(a_{i,j}\right)_{m \times n}$ and $C = \left(c_{i,j}\right)_{m \times n}$ are two matrices with same width and height and $r, s \in \mathbb{R}$ be two real numbers. Then we define:*

$$rA + sC := \left(ra_{i,j} + sc_{i,j}\right)_{m \times n}$$

Now, suppose that $A = \left(a_{i,j}\right)_{m \times n}$ and $C = \left(c_{l,m}\right)_{x \times p}$ are two matrices, this time A is large n and C is tall n. We define the multiplication of A with C as follows:

$$AC := \left(a_{i,1}c_{1,j} + a_{i,2}c_{2,j} + \dots + a_{i,p}c_{k,j}\right)_{n \times p}$$

To better understand matrix multiplication, let's think of the rows of the matrix A as horizontal vectors $\vec{u}_1, \dots, \vec{u}_m$, so:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix} = \begin{pmatrix} \vec{u}_1 \\ \vec{u}_2 \\ \vdots \\ \vec{u}_m \end{pmatrix}$$

To be clear, this means the following:

$$\begin{aligned} \vec{u}_1 &:= \langle a_{1,1}, a_{1,2}, \dots, a_{1,n} \rangle \\ \vec{u}_2 &:= \langle a_{2,1}, a_{2,2}, \dots, a_{2,n} \rangle \\ &\vdots \\ \vec{u}_m &:= \langle a_{m,1}, a_{m,2} \dots, a_{m,n} \rangle \end{aligned}$$

Now, let's think of the columns of the matrix C as vertical vectors, so:

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,p} \\ a_{2,1} & a_{2,2} & \dots & a_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,p} \end{pmatrix} = \begin{pmatrix} \vec{u}'_1 & \vec{u}'_2 & \dots & \vec{u}'_p \end{pmatrix}$$

So:

$$\begin{aligned} \vec{u}'_1 &:= \langle a_{1,1}, a_{1,2}, \dots, a_{1,p} \rangle \\ \vec{u}'_2 &:= \langle a_{2,1}, a_{2,2}, \dots, a_{2,p} \rangle \\ &\vdots \\ \vec{u}'_p &:= \langle a_{n,1}, a_{n,2} \dots, a_{n,p} \rangle \end{aligned}$$

So, the multiplication matrix AC can be defined as follows:

$$AC = \begin{pmatrix} \vec{u}_1 \\ \vec{u}_2 \\ \vdots \\ \vec{u}_m \end{pmatrix} \begin{pmatrix} \vec{u}'_1 & \vec{u}'_2 & \dots & \vec{u}'_p \end{pmatrix} = \begin{pmatrix} \vec{u}_1 \cdot \vec{u}'_1 & \vec{u}_1 \cdot \vec{u}'_2 & \dots & \vec{u}_1 \cdot \vec{u}'_p \\ \vec{u}_2 \cdot \vec{u}'_1 & \vec{u}_2 \cdot \vec{u}'_2 & \dots & \vec{u}_2 \cdot \vec{u}'_p \\ \vdots & \vdots & \ddots & \vdots \\ \vec{u}_m \cdot \vec{u}'_1 & \vec{u}_m \cdot \vec{u}'_2 & \dots & \vec{u}_m \cdot \vec{u}'_p \end{pmatrix}$$

Let's do an example. Let:

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \\ 2 & -1 \end{pmatrix} \qquad C = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 4 & 1 & -3 & 0 \end{pmatrix}$$

Then:

$$AC = \begin{pmatrix} 1 & 2 \\ 0 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 0 \\ 4 & 1 & -3 & 0 \end{pmatrix} = \begin{pmatrix} 8 & 3 & -7 & 0 \\ 12 & 3 & -9 & 0 \\ -4 & 1 & 1 & 0 \end{pmatrix}$$

So far we showed that linear functions define matrices, now let's employ matrix multiplication to show that every matrix is associated to a linear function. The idea is to think of vectors in a vector space $\mathbb{V}$ as matrices with one single column, so if $\vec{v} = \langle x_1, \dots, x_n \rangle$, then, given a matrix $A = (a_{i,j})_{m\times n}$ we define:

$$f(\vec{v}) := A\vec{v} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Now, this gives a matrix $m \times 1$, which we interpret as a vector in $\mathbb{R}^m$. It turns out that f so defined is a linear function from $\mathbb{V}$ to $\mathbb{R}^m$.

Proposition 5.1. *Suppose that $f : \mathbb{V} \to \mathbb{V}'$ and $g : \mathbb{V}' \to \mathbb{V}''$ are two linears function between finitely-dimensional vector spaces $\mathbb{V}, \mathbb{V}'$ and $\mathbb{V}''$. Moreover, suppose that C and A are the associated matrices of f and g, respectively. Then, the associated matrix to the composition of f with g, i.e. the linear function $g \circ f : \mathbb{V} \to \mathbb{V}''$ is the multiplication of A with C, i.e. AC.*

There's a lot to say about matrices, but for our purposes, we just need another topic: determinants. For every square matrix, i.e. a matrix $n \times n$ A, we can define a special number, called determinant and often denoted by $\det A$, which contains some important information of A. Defining this number for any square matrix however, is not an easy task and requires too much time and concepts. So, here we just give the definition of this concept in two special cases: when $n = 2$ and $n = 3$.

Definition 5.4. *For a 2×2 matrix A as follows:*

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$$

*the **determinant** of A is the number so defined:*

$$\det A := \begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} := a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$$

For a 3×3 matrix A as follows:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$$

*the **determinant** of A is the number so defined:*

$$\det A := \begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} := a_{1,1} \begin{vmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{vmatrix} - a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} + a_{1,3} \begin{vmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{vmatrix} =$$

$$= a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{2,1}a_{3,2}a_{1,3} - a_{1,3}a_{2,2}a_{3,1} - a_{1,2}a_{2,1}a_{3,3} - a_{2,3}a_{3,2}a_{1,1}$$

Glossary

angle between The angle between two non-zero non-parallel arrows $\vec{u}$ and $\vec{v}$ is the smallest angle formed on the plane generated by $\vec{u}$ and $\vec{v}$ that stays between $\vec{u}$ and $\vec{v}$. In case $\vec{u}$ and $\vec{v}$ are parallel, we simply say that the angle between $\vec{u}$ and $\vec{v}$ is zero. See Definition 1.1. 6

angular velocity The angular velocity $\vec{\omega}$ is defined as the vector $(\vec{r} \times \vec{v})/r^2$, where $\vec{r}$ is the position, $r := |\vec{r}|$ and $\vec{v}$ is the velocity. 23

arclength The arclength of a path from a point $\vec{x}_0$ to a point $\vec{x}_1$ is the length of this path between these two points. 103, 104

arrow An arrow is a segment equipped with an orientation. See Definition 1.1. 6

associated matrix The associated matrix of a linear function $f : \mathbb{V} \to \mathbb{V}'$ is the matrix defined by $(f(\vec{v}_1) \ldots f(\vec{v}_n))$, where $B := \{\vec{v}_1, \ldots, \vec{v}_n\}$ is a basis for $\mathbb{V}$. See Definition 5.2. 107, 108

associativity A binary operation μ over a set G is associative if for every $g, h, i \in G$, $\mu(\mu(g, h), i) = \mu(g, \mu(h, i))$. See Definition 1.3. 6, 7

barycentre The point of a three-dimensional body which represents the centre of mass of the body. Intuitively, this is the centre of the body, considering the possibility that the mass is not uniformly distributed. 5

basis A basis of a vector space $\mathbb{V}$ is a family $B := \{\vec{v}_i\}$ of vectors of $\mathbb{V}$ so that every other vector $\vec{u} \in \mathbb{V}$ can be written in a unique way as a finite linear combination of vectors of B. See Definition 1.7. 10–13, 16, 20, 62, 106, 107

binormal vector The binormal vector of a vector function $\vec{f}$ at t_0 is the versor $\vec{\mathsf{B}}(t_0) := \vec{\mathsf{T}}(t_0) \times \vec{\mathsf{N}}(t_0)$, where $\vec{\mathsf{T}}$ is the tangent vector of $\vec{f}$ and $\vec{\mathsf{N}}(t_0)$ is the normal vector. See Definition 2.13. 57

bivariate Gaussian The bivariate Gaussian is the two-dimensional bell curve. See Example 3.1. 60, 63, 68, 74, 77, 89, 90

boundary The boundary $\partial\Omega$, or also called the border, of a subset $\Omega \subseteq \mathbb{R}^n$ is the difference between closure and the interior of Ω. See Definition 2.3. 44, 45, 84

bounded A bounded subset of $\mathbb{R}^n$ is a subset that can be entirely included in a closed ball. See Definition 3.7. 80–82, 110

Cartesian plane The Cartesian plane is the vector space $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. See Example 1.2. 9

chirality Chirality, from the Greek word "hand", is a property of some objects and molecules. Something is chiral if it exists in two different species, each the reflection of the other. For example, dice are chiral, spirals are chiral and lots of molecules have chiral properties. 21

class C^0 A class C^0 function is a continuous function. See Definition 3.1. 66

class C^1 A class C^1 function is a function that has all partial derivatives and these are continuous functions. See Definition 3.1. 66, 95, 96

class C^2 A class C^2 function is a function that has all partial derivatives and these are class C^1 functions. See Definition 3.1. 87, 88, 90–92, 94, 95

class C^k A class C^k description=A class C^2 function is a function that has all partial derivatives and these are class C^{k-1} functions. See Definition 3.1. See Definition 3.1. 66

closed A closed set $C \subseteq \mathbb{R}^n$ is a subset of $\mathbb{R}^n$, which is the complementary of an open set, i.e. $\mathbb{R}^n \setminus C$ is open. See Definition 2.3. 44, 80–82, 110

closed ball An closed ball centred in $\vec{x}_0$ of radius $r > 0$ is the set of points $\vec{x}$ that are within r distance from $\vec{x}_0$, including the boundary. See Definition 2.2. 43–45, 80

closed rectangle A closed rectangle is the Cartesian product between two closed intervals $[a, b] \times [c, d]$. 105, 113

closure The closure $\overline{\Omega}$ of a subset $\Omega \subseteq \mathbb{R}^n$ is the smallest closed set that includes Ω. See Definition 2.3. 44–47, 84, 100, 101, 105, 109

commutative group A commutative group is a group where the binary operation is commutative. See Definition 1.3. 6, 7, 9

commutativity A group (G, μ) is commutative whenever for every $g, h \in G$, $\mu(g, h) = \mu(h, g)$. See Definition 1.3. 6, 7

compact A compact subset of $\mathbb{R}^n$ is a closed and bounded subset of $\mathbb{R}^n$. See Definition 3.8. 80–82

complement Given a set U, the complement of a subset $A \subseteq U$ is the set of all the elements of U that are not in A and is denoted by $U \setminus A$. 80

component Given a basis $B = \{\vec{v}_i\}$ of a vector space $\mathbb{V}$, the components of a vector $\vec{u}$ w.r.t. B are the unique real numbers $x_1, \ldots, x_{n^*}$ such that $\vec{u} = x_1\vec{v}_1 + \cdots + x_{n^*}\vec{v}_{n^*}$. See Definition 1.9 and 1.7. 13, 16

constraint A constraint for the variables x, y and z is an equation of the form $F(x, y, z) = 0$, for some scalar function $F : \mathbb{R}^3 \to \mathbb{R}$. 72, 96

continuity Continuity is the property of a continuous function. See Definition 2.5. 48

continuous A continuous function is a function that has no jump. See Definition 2.5. 47, 110

contour diagram The contour diagram of a scalar function is the collection of all the level-sets of the function. See Definition 3.5. 76, 77

counterexample A counterexample is an example of a construction that contradicts an assumption. For example, it is true that every isosceles triangle has two internal angles that are the same. So, one could suspect that this is true for every triangle, i.e. one could think that every triangle has two internal angles that are the same. However, a counterexample of this statement is given by a triangle with sides long 1, 2 and 3. 82, 86

critical A point $\vec{x}_0$ is critical for a scalar function f if f is not differentiable in $\vec{x}_0$ or if $\vec{x}_0$ is a stationary point for f. See Definition 3.10. 84–86

cross product Same as vector product. 23

curvature The curvature of a vector function $\vec{f}$ at time t is defined as follows: $\kappa(t) := \left|\vec{\mathsf{T}}'(t)\right| / \left|\vec{f}'(t)\right|$, where $\vec{\mathsf{T}}(t)$ is the tangent vector of $\vec{f}$ at time t. 56

cylinder A cylinder is a two-dimensional surface obtained by translation of a plane curve along a fixed direction. 35, 36

degrees of freedom The degrees of freedom of a system represents the minimum number of independent parameters we need in order to completely describe every possible state of the system. For example, the dynamic of a bumblebee can be entirely described using 6 degrees of freedom: 3 parameters for the position in space, 3 for the velocity the bumblebee has. 5, 6

determinant The determinant of a square matrix is a special number that captures some properties of the matrix. See Definition 5.4. 29, 108

differentiability Differentiability is the property of a function to be differentiable. See Definition 2.7. 48, 50

differentiable A function $\vec{f} : \Omega \to \mathbb{R}^m$, with $\Omega \in \mathbb{R}^n$, is differentiable in a point $\vec{x}_0$ of it's domain if there is a linear function $\vec{L}_{\vec{x}_0}$ such that $\vec{f}(\vec{x}) \simeq \vec{f}(\vec{x}_0) + \vec{L}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$ for $\vec{x}$ close enough to $\vec{x}_0$. See Definition 2.7. 49, 50, 54, 59, 65, 66, 77

differential The differential of a function $\vec{f} : \Omega \to \mathbb{R}^m$ at $\vec{x}_0 \in \Omega$ is the linear function $\vec{L}_{\vec{x}_0}$ so that $\vec{f}(\vec{x}) \simeq \vec{f}(\vec{x}_0) + \vec{L}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$ for $\vec{x}$ close enough to $\vec{x}_0$. The differential is often denoted by $\mathsf{d}\vec{f}_{\vec{x}_0}$. See Definition 2.7. 49, 59, 61, 62, 73

dimension The dimension of a vector space is the number of vectors in a basis of the vector space. See Definition 1.1. 11

direction Whenever $\vec{u}$ is a non-zero arrow, The direction of $\vec{u}$, is the line where $\vec{u}$ lies on. See Definition 1.1. 6

directional derivative The directional derivative of a function $\vec{f}$ at $\vec{x}_0$ along the direction of the versor $\hat{u}$ is given by $\partial_u \vec{f}(\vec{x}_0) := \mathsf{d}\vec{f}_{\vec{x}_0}(\hat{u})$. 59–63, 66

distance The distance between two points $\vec{x}_1, \vec{x}_2$, in an inner product space, or more generally in a normed space, is given by $|\vec{x}_1 - \vec{x}_2| = \sqrt{(\vec{x}_1 - \vec{x}_2) \cdot (\vec{x}_1 - \vec{x}_2)}$. See Definition 2.2. 43, 44, 98, 112

domain The domain of a function is the set of points where the function is well-defined. 45

dot product See scalar product. 15

Euclidean space The Euclidean space is the vector space $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^3$. See Example 1.2. 9

extreme value An extreme value of a scalar function f is a global maximum or a global minimum of f. See Definition 3.6. 79, 85

global maximum A global maximum of a scalar function f is a value $f(\vec{x}_0)$ so that, for any other $\vec{x}$, $f(\vec{x}) \leq f(\vec{x}_0)$. See Definition 3.6. 79, 80, 82, 85, 95

global minimum A global minimum of a scalar function f is a value $f(\vec{x}_0)$ so that, for any other $\vec{x}$, $f(\vec{x}) \geq f(\vec{x}_0)$. See Definition 3.6. 79, 82, 95

gradient The gradient of a scalar function f is the vector whose components are the partial derivatives of f. See Definition 3.4. 74, 75, 77

group A group is a set G equipped with a binary operation μ which is unital, associative and it has inverses. See Definition 1.3. 6

Hessian matrix The Hessian matrix of a function of class C^2 is the matrix of all the second-order partial derivatives of a function. See Definition 3.12. 87–91, 94

infinitesimal An infinitesimal quantity is usually intended to be a positive and very small number. 103, 104

inner product See scalar product. 15

inner product space An inner product space is a vector space $\mathbb{V}$ equipped with a scalar product $\cdot : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$. See Definition 1.11. 15, 42, 43

integrable An integrable function is a function that can be integrated. See Definition **??**. 102, 103

integral The integral is an operation over functions that returns the oriented area or volume of hyper-volume underneath the function. See Definition **??**. 102, 103, 105

integration domain The integration domain of an integral is the region in which the function is integrated. See Definition **??**. 103

interior The interior $\mathring{\Omega}$ of a subset $\Omega \subseteq \mathbb{R}^n$ is the largest open set that is included in Ω. See Definition 2.3. 44, 45, 109

internal A point $\vec{x}_0$ is internal of a set Ω if there is at least an open ball $\mathbb{B}(\vec{x}_0, r)$ that is entirely included within Ω. See Definition 3.11. 84–86, 91, 92, 94

inverse We say that an element $g \in G$ is invertible with respect to the binary unital operation μ, if there is another element $h \in G$ such that $\mu(g, h) = 1_G = \mu(h, g)$, where 1_G is the unit of μ. See Definition 1.3. 6, 7, 9

inverse arrow The inverse arrow of an arrow $\vec{u}$ is the unique arrow with the same length and direction of $\vec{u}$ but opposite orientation. It is denoted by $-\vec{u}$. 7

inversion An inversion, or reflection, is a transformation that changes an object into one of its reflections in a flat mirror. 21, 113

isocurve Cf. level-set. 77, 95

isomorphism An isomorphism of two mathematical objects is an invertible transformation between them that preserves their structures. So, for example, an isomorphism of vector spaces is a linear bijection. 20

isosurface Cf. level-set. 77

Lagrange multiplier The Lagrange multipliers technique is used to find local extreme values of a function in the presence of a constraint. See Theorem 3.4. 96

left-handed A triple of vectors $(\vec{u}, \vec{v}, \vec{w})$ so that $\vec{w} \perp \vec{u}$ and $\vec{w} \perp \vec{v}$ is left-handed if it is not right-handed. 25

level-curve Cf. level-set. 77

level-set The c-level set of a scalar function f is the set $\Gamma_c(f) := \{\vec{x} \mid f(\vec{x}) = c\}$. See Definition 3.5. 76, 77, 96, 110, 111

level-surface Cf. level-set. 77

line generated The line generated by a non-zero arrow or an arrow is the direction of the arrow or vector. See Definition 1.1. 6

linear A linear function, sometimes also called linear morphism, is a function $f : \mathbb{V} \to \mathbb{V}'$ between two vector spaces $\mathbb{V}$ and $\mathbb{V}'$ which satisfies linearity, i.e. for every $a, b \in \mathbb{R}$ and $\vec{u}, \vec{v} \in \mathbb{V}$, $f(a\vec{u} + b\vec{v}) = af(\vec{u}) + bf(\vec{v})$. 49, 59, 62, 89, 106–108, 110

linearization The linearization of a function $\vec{f}$ in a point $\vec{x}_0$ for which $\vec{f}$ is differentiable, is the function $\vec{\lambda}(\vec{x}) := \vec{f}(\vec{x}_0) + \mathsf{d}\vec{f}_{\vec{x}_0}(\vec{x} - \vec{x}_0)$. See Definition 3.3. 67

local extreme value $\vec{x}_0$ is a point of local extreme value for a scalar function f if it is a point of local maximum or of local minimum for f. See Definition 3.9. 83–86, 91, 92, 95, 96

local maximum $\vec{x}_0$ is a point of local maximum for a scalar function f if there is an open ball $\mathbb{B}(\vec{x}_0, r)$ such that for every $\vec{x} \in \mathbb{B}(\vec{x}_0, r)$, $f(\vec{x}) \leq f(\vec{x}_0)$. See Definition 3.9. 83, 84, 86, 91

local minimum $\vec{x}_0$ is a point of local minimum for a scalar function f if there is an open ball $\mathbb{B}(\vec{x}_0, r)$ such that for every $\vec{x} \in \mathbb{B}(\vec{x}_0, r)$, $f(\vec{x}) \geq f(\vec{x}_0)$. See Definition 3.9. 83, 86, 91, 93

measure The measure of a bounded set $\Omega \subseteq \mathbb{R}^n$ is the integral of the characteristic function $\chi_\Omega : \Omega \to \mathbb{R}$, $\chi_\Omega(\vec{x}) = 1$ over Ω. In 1 dimension, the measure corresponds to the length, in 2 dimensions, to the area and in 3 dimensions to the volume of Ω. See Definition **??**. 102, 105

metric space A metric space is a set A equipped with a distance $d : A \times A \to \mathbb{R}$. See Cool Example 2.1. 43

modulus The modulus of an arrow, or of a vector, is the length of the vector or arrow. Sometimes, it is also called magnitude or norm. See Definition 1.1. 6, 7, 12, 13, 54

multivariable function A multivariable function is a function of the type $f : \mathbb{R}^n \to \mathbb{R}^m$. See Definition 2.1. 40, 42

negative-definite An $n \times n$ square matrix is negative-definite is its quadratic-form is 0. See Definition 3.14. 90, 91

negative-semidefinite An $n \times n$ square matrix is negative-semidefinite is its quadratic-form is ≤ 0. See Definition 3.14. 90, 91

Newton Newton was a person, you know the apple guy, but it's also a unit of measure for forces and it corresponds to $1N = 1kg \cdot m/s^2$, so $1N$ is the same force of the weight of $1kg$ of any substance on the surface of our planet, or of about $2.204lb$. 19

norm A norm is function $\|-\| : \mathbb{V} \to \mathbb{R}$ which measures the length of vectors. See Cool-Stuff 1.2. 16, 20, 43, 112

normal vector The normal vector of a vector function $\vec{f}$ at t_0 is the versor $\vec{\mathsf{N}}(t_0) := \vec{\mathsf{T}}'(t_0)/\left|\vec{\mathsf{T}}'(t_0)\right|$, where $\vec{\mathsf{T}}$ is the tangent vector of $\vec{f}$. See Definition 2.10. 54–57, 109, 112

normed space A normed space is a vector space $\mathbb{V}$ equipped with a norm $\|-\|$. 16, 20, 43

not aligned Three points A, B and C in space are not aligned if they are distinct and they do not lie on the same line. 33

open An open set $A \subseteq \mathbb{R}^n$ is a subset or $\mathbb{R}^n$ so that every point $\vec{x}_0 \in A$ admits at least an open ball $\mathbb{B}(\vec{x}_0, \delta)$, centred in $\vec{x}_0$ and of some radius $\delta > 0$ which is totally included in A, i.e. $\mathbb{B}(\vec{x}_0, \delta) \subseteq A$. See Definition 2.3. 44, 80, 100

open ball An open ball centred in $\vec{x}_0$ of radius $r > 0$ is the set of points $\vec{x}$ that are within r distance from $\vec{x}_0$, excluding the boundary. See Definition 2.2. 42–45, 80

open rectangle An open rectangle is the Cartesian product between two open intervals $(a, b) \times (c, d)$. 105, 113

orientation The orientation of an arrow, indicates in which direction the arrow is pointing. See Definition 1.1. 6–8, 11, 12, 17, 18, 21–23, 30, 37, 39, 54, 76, 79

oriented angle The oriented angle from $\vec{u}$ to $\vec{v}$ is a real number θ so defined: $|\theta|$ is the angle between $\vec{u}$ and $\vec{v}$ and has positive sign when interpreted as the angle from $\vec{u}$ to $\vec{v}$ and negative sign when interpreted as the angle from $\vec{v}$ to $\vec{u}$. 23

origin The origin is the starting point of the orthonormal unit vectors is a system of coordinates. 5, 6, 11, 21, 27

orthogonal Two non-zero arrows $\vec{u}$ and $\vec{v}$ are orthogonal if the angle between $\vec{u}$ and $\vec{v}$ is $90°$. We use the symbol $\vec{u} \perp \vec{v}$ to indicate that $\vec{u}$ and $\vec{v}$ are orthogonal. See Definition 1.1. 6, 11–13, 18, 23, 34–36, 55, 77

orthonormal A basis $B = \{\vec{v}_i\}$ of a vector space $\mathbb{V}$ equipped with a scalar product $\cdot$ is orthonormal whenever for every i, $|\vec{v}_i| = 1$ and for every i, j, so that $i \neq j$, $\vec{v}_i \cdot \vec{v}_j = 0$. 13, 16

osculating circle The osculating circle of a vector function $\vec{f}$ is the circle that lies on the osculating plane, has radius $1/\kappa(t)$ and centre $\vec{f}(t) + \kappa(t)\vec{\mathsf{N}}(t)$. See Definition 2.12. 56, 57

osculating plane The osculating plane of a vector function $\vec{f}$ is the plane generated by the tangent vector $\vec{\mathsf{T}}(t)$ and the normal vector $\vec{\mathsf{N}}(t)$. See Definition 2.12. 56, 112

parallel Two arrows $\vec{u}$ and $\vec{v}$ are parallel when they generate parallel lines. In case one of the two has zero length, we immediately say that they are parallel. We use $\vec{u} \parallel \vec{v}$ to denote that $\vec{u}$ and $\vec{v}$ are parallel and $\vec{u} \nparallel \vec{v}$ that $\vec{u}$ and $\vec{v}$ are not parallel. See Definition 1.1. 6, 11, 12

partial derivative The partial derivatives of a differentiable function $\vec{f}$ in a point $\vec{x}_0$ are the vectors $\mathsf{d}\vec{f}_{\vec{x}_0}(\hat{\mathbf{i}})$, $\mathsf{d}\vec{f}_{\vec{x}_0}(\hat{\mathbf{j}})$ and $\mathsf{d}\vec{f}_{\vec{x}_0}(\hat{\mathbf{k}})$. See Definition 3.2. 50, 62, 64–66

partition A partition of a non-empty set is a collection of subsets that covers the whole set so that there is no overlap between these subsets except for the boundaries. See Definition 4.1. 100, 101, 105

path A path is a continuous function of the type $\Omega \to \mathbb{R}^n$, where $\Omega \subseteq \mathbb{R}$ is an interval or the whole $\mathbb{R}$. 41

plane curve A plane curve is a curve that lies on a plane. 110

plane generated Given two non-parallel (non-zero) arrows $\vec{u}$ and $\vec{v}$, the plane generated by $\vec{u}$ and $\vec{v}$ is the unique plane where both the vectors lie. See Definition 1.1. 6, 11, 12, 56, 112

polynomial A polynomial is an algebraic expression obtained by multiplying and summing letters and numbers together. The letters are often called the variables and the numbers are often called coefficients. An example of a polynomial in the variable x is the expression $x^2 + 2x + 1$. 9, 11

positive-definite An $n \times n$ square matrix is positive-definite is its quadratic-form is > 0. See Definition 3.14. 90, 91, 94

positive-semidefinite An $n \times n$ square matrix is positive-semidefinite is its quadratic-form is ≥ 0. See Definition 3.14. 90, 91

quadratic-form The quadratic form of an $n \times n$ square matrix A is the function $q_A : \mathbb{R}^n \to \mathbb{R}$ so defined $q_A(\vec{u}) = \vec{u} \cdot A\vec{u}$. See Definition 3.13. 90, 91, 112, 113

quadric surface A quadric surface is a two-dimensional surface described by a second-order polynomial equation in three variables. 36

r.h.r. Right-hand rule. See Definition 1.8. 11–13, 21, 23, 25, 27, 113

rectangle See open rectangle and closed rectangle. 104

reflection See inversion. 21

regular Intuitively, a regular function is a function that doesn't make sudden variations. Regularity is a term that is not technical, and that refers to different kinds of regularity. Regular could mean, continuous, differentiable, or differentiable with continuity and so on. This is just a vague term for any of these technical properties. 42, 113

regularity Regularity is the property that regards regular functions. 48

right-hand rule The right-hand rule is a convention to decide the orientation of three arrows, one of which is orthogonal to both of the other two. See Definition 1.8. 11, 21–23

right-handed A triple of vectors $(\vec{u}, \vec{v}, \vec{w})$ so that $\vec{w} \perp \vec{u}$ and $\vec{w} \perp \vec{v}$ is right-handed if it satisfies the r.h.r.. 25, 111

rigidly Moving rigidly something, means translating it without any stretch, change of direction, or orientation. 6, 7

rotation A rotation is a movement that changes the direction of an object in space. 21

saddle A saddle point $\vec{x}_0$ for a function f is a stationary point which is not a point of local extreme value. See Definition 3.15. 85, 91–94

scalar action The scalar action is the function that defines how scalar, i.e. real numbers, act on vectors. In the case of arrows, the scalar action changes the modulus of the arrows of a factor equal to the number multiplied by the arrow. See Definition 1.5. 8, 9, 14, 21, 50, 59

scalar function A scalar function is a function of the type $f : \mathbb{R}^n \to \mathbb{R}$. See Definition 2.1. 40, 41, 50, 64, 72, 74, 77

scalar product The scalar product is an operation $\cdot$ that takes two vectors $\vec{u}$ and $\vec{v}$ and gives a real number $\vec{u} \cdot \vec{v}$. The standard Euclidean scalar product is defined as $\vec{u} \cdot \vec{v} := |\vec{u}||\vec{v}| \cos\theta$, where θ is the angle between $\vec{u}$ and $\vec{v}$. 14–17, 19, 21, 41–43, 74, 110, 111

scalar projection The scalar projection of a vector $\vec{v} \in \mathbb{V}$ onto a non-zero vector $\vec{u}$ is the number $\mathrm{comp}_{\vec{u}}\vec{v} := (\vec{u} \cdot \vec{v})/|\vec{u}|$. See Definition 1.12. 17, 19

scalar triple product The scalar triple product of three vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ is the real number $\vec{u} \cdot (\vec{v} \times \vec{w})$. See Definition 1.14. 29

singleton A singleton is a set that contains only a point $\{x_0\}$. 46

speed We mean for speed the amount of space that a body covers, usually measured in meters, in a unit of time, usually in seconds. The relationship with the velocity is that the speed is the modulus of the velocity. 22, 23, 39, 114

sphere A sphere centred in $\vec{x}_0$ of radius $r > 0$ is the set of points $\vec{x}$ that have exactly distance r from $\vec{x}_0$. See Definition 2.2. 43, 44

square matrix A square matrix is a matrix $n \times n$. 108, 110

stationary A point $\vec{x}_0$ is stationary for a scalar function f if f is differentiable in $\vec{x}_0$ and $\vec{\nabla} f(\vec{x}_0) = \vec{0}$. See Definition 3.10. 83, 85, 91, 92

step function A step function over a non-empty set Ω is a function that is constant over each subset S_i of a finite partition of Ω. See Definition 4.2. 101, 102

sum of arrows The sum of two arrows $\vec{u}$ and $\vec{v}$, both that start from the same point P, is the arrow $\vec{u} + \vec{v}$ that starts in P and ends where ends $\vec{v}$ once $\vec{v}$ is rigidly moved in front of $\vec{u}$. See Definition 1.2. 6, 7

system of coordinates A system of coordinates is a basis of orthonormal unit vectors oriented according to the right-hand rule, which is used to describe in a unique way every point of the three-dimensional space. 5, 12, 22

tangent plane The tangent plane of a two-dimensional surface on a point is the plane that locally touches the surface only in that point. 37, 67, 78, 83

tangent space The tangent space of a geometrical space on a point is the space that locally touches the space only in that point. 37, 67, 95

tangent vector The tangent vector of a vector function $\vec{f}$ at t_0 is the versor $\vec{\mathsf{T}}(t_0) := \vec{f}'(t_0)/\left|\vec{f}'(t_0)\right|$. See Definition 2.9. 54–57, 109, 110, 112

total derivative The total derivative of a vector function $\vec{f}$ in a point t_0 is the limit for $h \to 0$ of $(\vec{f}(t_0 + h) - \vec{f}(t_0))/h$. See Definition 2.8. 50, 59, 61, 65, 95

translation A translation is a rigid movement that does not invert or change the orientation of an object in space. 21

twice-differentiable A twice-differentiable function is a function $\vec{f} : \Omega \to \mathbb{R}^m$, with $\Omega \subseteq \mathbb{R}^n$, such that it is differentiable and all of its partial derivatives are differentiable too. For vector functions (i.e. when $n = 1$), this means that the function is differentiable and its total derivative is differentiable as well. 55, 56

unbounded A unbounded subset of $\mathbb{R}^n$ is a subset that is not bounded. See Definition 3.7. 80

unitality A binary operation μ over a set G is unital, if there is a distinct element 1_G, sometimes called the unit or the zero, such that for every element $g \in G$, $\mu(g, 1_G) = g$ and $\mu(1_G, g) = g$. See Definition 1.3. 6, 7

vector A vector is an element in a vector space. See Definition 1.6. 9–11

vector function A vector function is a function of the type $f : \mathbb{R} \to \mathbb{R}^m$. See Definition 2.1. 40, 41, 47, 50, 65

vector product The vector product of two vectors in introduced in Definition 1.13. 7, 23, 24, 26–28, 34, 41, 110

vector projection The vector projection of a vector $\vec{v} \in \mathbb{V}$ onto a non-zero vector $\vec{u}$ is the vector $\mathsf{proj}_{\vec{u}}\vec{v} := (\mathsf{comp}_{\vec{u}}\vec{v})\hat{u}$. See Definition 1.12. 17, 19, 35

vector space A vector space is a commutative group $(\mathbb{V}, +)$ equipped with a scalar action $\cdot$ which is compatible with $+$. See Definition 1.6. 9–13, 15, 16, 19, 40, 43, 62, 106, 107

velocity The velocity of a body is the vector with modulus equal to the speed and with same direction and orientation of the direction and orientation of the motion of the body. 22, 23, 39, 51, 109

versor Given a non-zero vector $\vec{u} \in \mathbb{V}$, the versor of $\vec{u}$ is the vector $\hat{u} := \vec{u}/|\vec{u}|$. For arrows, the versor of an arrow $\vec{u}$ is the arrow with the same direction and orientation of $\vec{u}$ but modulus 1. See Definition 1.12. 17, 30, 31, 54, 59, 61–63, 76, 86, 90, 109, 110, 112, 113

vertical component The vertical component of an arrow $\vec{u}$ (or of a vector in the three-dimensional space) is the arrow parallel to $\hat{\mathbf{k}}$ that summed with the projection of $\vec{u}$ along the $\hat{\mathbf{i}}\hat{\mathbf{j}}$ plane gives back $\vec{u}$. 12

w.r.t. With respect to. 11, 37, 56, 65, 72, 73, 107

zero arrow The zero arrow is the arrow with modulus equal to 0 and its denoted by $\vec{0}$. 6, 11