Chapter 1

Reminder of key concepts

Version 0.1 - 22 March 2025

In this chapter, we take a stroll down memory lane to remind ourselves of key concepts that we will assume that you are already familiar with. Another reason is to make it clear what notation we use. To keep the chapter brief, we avoid going into details and try to focus on the basic ideas, good notation and, to some extent, how these ideas are connected. A deep dive into these concepts can be found in the appendix.

1.1 Initial remarks on basic (and sometimes bad) notation

First, we make some initial remarks on notation. This section is meant as a reference if some of the "basic" notation confuses you. Concepts that are explicitly defined later in the text are not discussed here (such as open balls, limits, the derivative and the definite integral).

Notation for sets

Let A, B be two sets containing some types of points. The notations $A \cup B$, $A \cap B$ are standard for the union and intersection, respecitvely, of A and B, and need no explanation. However, as it sometimes causes confusion, we point out that by $A \subset B$ we mean that the set A is contained in, or equal to, B. If we want to express that A is strictly contained in B (which is rarely the case), we write $A \subsetneq B$.

If we have a sequence of sets $(A_j)_{j=0}^{\infty}$ (see below for sequence notation), we say that this sequence is increasing if $A_j \subset A_{j+1}$ and decreasing if $A_j \supset A_{j+1}$. Moreover, we define

$$\bigcup_{j=0}^{\infty} A_j = \{a : \exists j \in \mathbb{N} \text{ such that } a \in A_j\},$$

$$\bigcap_{j=0}^{\infty} A_j = \{a : \forall j \in \mathbb{N} \text{ we have } a \in A_j\}.$$

Notation for points and sequences

For points in $\mathbb{R}^n$, we write $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{x} = (x_1, \ldots, x_n)$ for $x_1, \ldots, x_n \in \mathbb{R}$. Note that we make no effort to separate between points and vectors in terms of notation[1]. If we write $\mathbf{x} \in \mathbb{R}^n$, and later consider the special case $n = 1$, we will sometimes write $\mathbf{x} = x$.

[1]Note that when we talk about points (or elements) of a set, we do not require there to be any type of structure. When we talk about *vectors*, this is always in the context of a vector space, which has a certain structure in addition to just being a set.

For sequences in $\mathbb{R}$ we use the notation $(a_j)_{j=0}^{\infty}$ or $(a_j)_{j\in\mathbb{N}}$. In some sense, this notation extends nicely to sequences of points from $\mathbb{R}^n$ in the sense that we can write, say, $(\mathbf{a_j})_{j=0}^{\infty}$. However, when we write out the term $\mathbf{a}_j$ we get into trouble with respect to indices. Indeed, we need to write something like

$$\mathbf{a}_j = (a_{j,1}, a_{j,2}, \ldots, a_{j,n}).$$

While this is perfectly fine, and people do it all the time, we prefer to avoid this as much as we can. In particular, for sequences in $\mathbb{R}^3$, we tend to write, say, $(\mathbf{w}_j)_{j=0}^{\infty}$ with $\mathbf{w}_j = (x_j, y_j, z_j)$ in order to keep the number of subscripts down. For this reason, we sometimes choose to formulate results for $\mathbb{R}^2$ or $\mathbb{R}^3$, even though they also hold for general $\mathbb{R}^n$.

We denote subsequences of a sequence $(x_j)_{j=0}^{\infty}$ by $(x_{j_k})_{k=0}^{\infty}$. To understand this notation, note that j_k is itself a sequence $(j_k)_{k=0}^{\infty}$. If we want to keep notation to a minimum, we sometimes denote the subsequence by a new symbol by putting, say, $u_k = x_{j_k}$. Another choice we can make, if we no longer care about the original sequence, is to say that we forget about the old sequence $(x_j)_{j=0}^{\infty}$ and instead use $(x_j)_{j=0}^{\infty}$ to denote its subsequence (people do this all the time).

Notation for sums and series

We denote sums by the Sigma symbol. The technical term for a sum is a series and when we sum an infinite number of terms, we call it an infinite series[2]. Please note that the following all mean the same thing:

$$\sum_{n=0}^{\infty} a_n = \sum_{n\in\mathbb{N}} a_n = \sum_{n\geq 0} a_n = \sum_{\{n\geq 0\}} a_n.$$

This flexibility of notation allows us to express sums of multiply indexed terms. E.g., the following all mean the same:

$$\sum_{i=0}^{3}\sum_{j=0}^{3} a_{i,j} = \sum_{i,j\in\{0,1,2,3\}} a_{i,j} = \sum_{0\leq i,j\leq 3} a_{i,j} = \sum_{\substack{0\leq i\leq 3\\ 0\leq j\leq 3}} a_{i,j}.$$

Notation for norms, inner products and matrices

For $x \in \mathbb{R}$, we denote its absolute value by $|x|$. For $\mathbf{x} \in \mathbb{R}^n$, the analogous concept is the Euclidean norm which we denote by $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$. Note that for $\mathbf{x} \in \mathbb{R}^n$, with $n = 1$, then $\|\mathbf{x}\| = |x|$.

We say that a subset $A \subset \mathbb{R}^n$ is bounded if there exists a constant $C > 0$ so that $\|x\| \leq C$ for all $x \in A$.

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we denote their inner product by $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$. We will not consider complex vector spaces in these lecture notes.

For matrices, we use the standard notation

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

We say that this an $m \times n$ matrix. We use the notation $\mathbf{e}_j$ to denote the j'th coordinate vector in $\mathbb{R}^n$ or $\mathbb{R}^m$. Specifically,

$$\mathbf{e}_j = (0, 0, \ldots, 0, 1, 0, \ldots, 0),$$

[2]However, people never call finite sums 'series', and thus we can safely refer to an infinite series as just a 'series' without any risk of confusion.

where the 1 appears as the j'th entry (which may well be the first or second entry).

By $|A|$, we denote the determinant of A.

For two vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$, we express the vector product by $\mathbf{v}_1 \times \mathbf{v}_2$. This is sometimes also referred to as their cross product.

Notation for defining functions

An $m \times n$ matrix is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. We sometimes express this by writing

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

More generally, if f is a function defined on a subset D of $\mathbb{R}$ and takes values in $\mathbb{R}$, we write

$$f : D \subset \mathbb{R} \to \mathbb{R}.$$

If f takes values in $\mathbb{R}^m$, we use the notation

$$\mathbf{f} = (f_1, f_2, \ldots, f_m).$$

Note that while this notation seems to imply that we consider the output of $\mathbf{f}$ to be a row vector, if we are to be consistent with matrices $m \times n$ matrices being functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, then strictly speaking, it is more correct to write

$$\mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}.$$

However, as the column-vector notation takes up a lot of space, we usually just use the former notation.

If $\mathbf{f}$ has domain $D \subset \mathbb{R}^m$ and takes values in $\mathbb{R}^n$, we can indicate this by writing

$$\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m.$$

Here, we call D the domain of $\mathbf{f}$ and $\mathbb{R}^m$ its co-domain. The values taken by $\mathbf{f}$ is called its range, and is sometimes denoted by $V_\mathbf{f}$ or $R(\mathbf{f})$. Note that it is usual to say that a function is *scalar valued* if the co-domain is $\mathbb{R}$, and *vector valued* if the co-domain is $\mathbb{R}^n$.

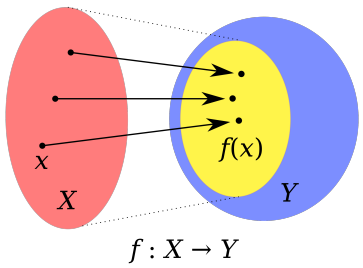

Figure 1.1: Image stolen from Wikipedia. In our case $X = D$ and $Y = \mathbb{R}^m$. The range $V_\mathbf{f}$ is indicated in yellow.

If we want to point out that the range is contained in a certain subset $B \subset \mathbb{R}^m$, we can indicate this by writing

$$\mathbf{f} : D \subset \mathbb{R}^n \to B \subset \mathbb{R}^m.$$

If it is clear that D and/or B are contained in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, (or maybe we just don't care), then we can simplify the above notation and just write, say,

$$\mathbf{f} : D \to B.$$

We add that if we want to describe a function, such as $f(x) = 2x+1$, without explicitly referring to its symbol, we can write

$$x \mapsto 2x + 1.$$

If $\mathbf{g} : D \subset \mathbb{R}^m \to B \subset \mathbb{R}^\ell$ and $\mathbf{f} : B \subset \mathbb{R}^\ell \to \mathbb{R}^n$, we can define the composition $\mathbf{f} \circ \mathbf{g}(\mathbf{x}) = \mathbf{f}(\mathbf{g}(\mathbf{x}))$ for $\mathbf{x} \in D$.. In this case, the composition is a function of the form $\mathbf{f} \circ \mathbf{g} : D \subset \mathbb{R}^n \to \mathbb{R}^m$. If we want to make the structure of the composition explicit, we sometimes write

$$\mathbf{f} \circ \mathbf{g} : D \subset \mathbb{R}^m \underset{\mathbf{g}}{\longrightarrow} B \subset \mathbb{R}^\ell \underset{\mathbf{f}}{\longrightarrow} \mathbb{R}^n.$$

We can also write something like

$$\mathbf{x} \longmapsto \mathbf{u} = \mathbf{g}(\mathbf{x}) \longmapsto \mathbf{y} = \mathbf{f}(\mathbf{u}).$$

Notation and terminology for special properties of functions

Suppose that $\mathbf{f} : D \subset \mathbb{R}^n \to B \subset \mathbb{R}^m$. Then we call $\mathbf{f}$ onto, or surjective, if, for all $\mathbf{x} \in B$, there exist $\mathbf{x} \in D$ so that $\mathbf{y} = \mathbf{f}(\mathbf{x})$. We call $\mathbf{f}$ one-to-one, or injective, if, for all $\mathbf{x}_1, \mathbf{x}_2 \in D$ we have $\mathbf{f}(\mathbf{x}_1) \neq \mathbf{f}(\mathbf{x}_2)$. If $\mathbf{f}$ is both onto B and one-to-one, we say that $\mathbf{f}$ is invertible, or bijective, from D onto B. Note that a one-to-one function is always invertible from D onto its range $V_\mathbf{f}$.

We say that a function $\mathbf{f} : D \to \mathbb{R}^m$ is bounded if there exists a constant $C > 0$ so that $\|\mathbf{f}(\mathbf{x})\| \leq C$ for all $\mathbf{x} \in D$. This is the same as saying that the range of $\mathbf{f}$ is a bounded subset of $\mathbb{R}^m$.

We say that a function $f : D \subset \mathbb{R} \to \mathbb{R}$ is increasing if, for all $x_1 \leq x_2$, we have $f(x_1) \leq f(x_2)$. This this holds with '$<$' instead of '$\leq$', we say that f is strictly increasing. If, for all $x_1 \leq x_2$, we have $f(x_1) \geq f(x_2)$ or $f(x_1) > f(x_2)$, we call f decreasing or strictly decreasing, respectively. Note that these definitions only make sense for scalar valued functions of one variable.

Some bad notation and terminology for functions (that we still use)

The notation for functions can be a bit confusing, as we will use y for different purposes. For $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ we tend to write $\mathbf{y} = \mathbf{f}(\mathbf{x})$), where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_m)$, in order to keep the notation as close as possible to the one-variable notation. However, when $D \subset \mathbb{R}^2$ or $D \subset \mathbb{R}^3$, we often write $\mathbf{f}(x, y)$ or $\mathbf{f}(x, y, z)$ in order to avoid having to use subscripts. What is worse, if $f : D \subset \mathbb{R}^2 \to \mathbb{R}$, we even sometimes write $z = f(x, y)$, also risking confusion with respect to the role of z. Hopefully, the relevant use will be clear from context.

For a function f, the symbol f^{-1} can denote many things. First, if f is scalar valued, then by f^{-1} we can mean its reciprocal function. That is $1/f$. Note that this makes no sense for vector valued functions. If $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is an invertible function, then by $\mathbf{f}^{-1}$, we can denote its inverse function, if its inverse exists. Finally, we sometimes use f^{-1} to denote the inverse image of a set with respect to the function. Indeed, if U is a subset of the co-domain of a function $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$, then we can define

$$\mathbf{f}^{-1}(U) = \{\mathbf{x} \in D : \mathbf{f}(\mathbf{x}) \in U\}.$$

Note that all three meanings of f^{-1} are different.

Last but not least: The Big-oh and little-oh notations

We will have use of both the Big-oh and Little-oh notations in this course. This notation depends on the concept of limits, that we discuss later in this chapter. However, for easy reference, we introduce the notation already now.

To make the definitions easier to formulate, we say that a property holds for all $\mathbf{x}$ "sufficiently near" a point $\mathbf{c}$ if there exists some $\delta > 0$ so that the property holds for all $\mathbf{x}$ with $\|\mathbf{x} - \mathbf{c}\| < \delta$. Sometimes we qualify this statement by saying that the property holds, say, "for all $\mathbf{x} \in D$ sufficiently near the point $\mathbf{c}$".

We say that a function $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is Big-oh of some other function $g : D \subset \mathbb{R}^n \to \mathbb{R}$, as $\mathbf{x}$ approaches $\mathbf{c}$, if there exists a constant $C > 0$, so that for all $\mathbf{x}$ sufficiently near $\mathbf{c}$, we have

$$\frac{\|\mathbf{f}(\mathbf{x})\|}{|g(\mathbf{x})|} \leq C.$$

In this case, we write

$$\mathbf{f}(\mathbf{x}) = O(g(\mathbf{x})).$$

We say that a function $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is Little-oh of some other function $g : D \subset \mathbb{R}^n \to \mathbb{R}$, as $\mathbf{x}$ approaches $\mathbf{c}$, if we have

$$\frac{\|\mathbf{f}(\mathbf{x})\|}{|g(\mathbf{x})|} \longrightarrow 0, \quad \text{as } \mathbf{x} \to \mathbf{c}.$$

In this case, we write

$$\mathbf{f}(\mathbf{x}) = o(g(\mathbf{x})).$$

Remark 1.1

Here, we only state these definitions. We will return later (in the notes or in the lectures), to give more intuition about what is going on.

1.2 Points, vectors and matrices

First, we point out that arithmetic in $\mathbb{R}^n$ is less structured than arithmetic in $\mathbb{R}$. Indeed, if we think of points in $\mathbb{R}^n$ as vectors, then we have addition in a term-by-term sense:

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

We also have a notion of multiplication with scalars. That is, for $\mathbf{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, we have

$$c\,\mathbf{x} = c(x_1, x_2, \ldots, x_n) = (cx_1, cx_2, \ldots, cx_n).$$

However, there is no natural way of multiplying a pair of vectors $\mathbb{R}^n$ to get a third vector in $\mathbb{R}^n$ as output that works for all $n \in \mathbb{N}$. Instead, the most "general" sense of multiplication we have that works for all $\mathbb{R}^n$ is the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle$ – however, it only produces output in $\mathbb{R}$, and so does not really live up to our expectation of what multiplication, perhaps, ought to do.

Exercise 1.1

(a) Does the scalar product live up to our expectation of multiplication in $\mathbb{R}^n$ when $n = 1$?

(b) For several values of n there exist "special" types of multiplication. In particular, this is true for $n = 2$ and $n = 3$. Can you figure out what these are? (In fact, for $n = 2$, there are at least two different such methods.)

Related to the inner product, we have the Euclidean norm $\|\mathbf{x}\|$ for $\mathbf{x} \in \mathbb{R}^n$. The Euclidean norm is related to the inner product on $\mathbb{R}^n$ by $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$. The Euclidean norm is an extension of the usual notion of distance in $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$. The only difference for higher dimensions is that it is harder to visualize what is going on. In particular, the triangle inequality holds in all dimensions:

$$\|\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\| + \ldots \|\mathbf{x}_n\|.$$

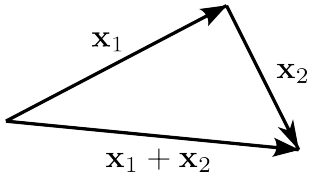

Figure 1.2: The Euclidean norm gives the length of a vector if we view it as an arrow in space. As such, the triangle inequality has the geometric interpretation as the sum of the length of two sides in a triangle is greater than the length of the third side.

Exercise 1.2 Use induction to prove the triangle inequality for all $\mathbb{R}^n$.

Exercise 1.3 In one variable, we have the useful equivalence $|x| < R \iff -R < x < R$. Is there are a several variable counter-part to this? To keep things simple, think about this question in the context of $\mathbb{R}^2$.

Hint: This exercise is related to the title of the lecture notes :-)

Another useful/important inequality is the Cauchy-Schwarz inequality. To give it some respect, we state it as a proposition.

Proposition 1.2: The Cauchy-Schwarz inequality

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leq \Big(\sum_{i=1}^{n} x_i^2 \Big)^{1/2} \Big(\sum_{i=1}^{n} y_i^2 \Big)^{1/2}.$$

Exercise 1.4 Try to prove the Cauchy-Schwarz inequality, as it is really a nice little argument. To help you out, we split the exercise into two parts.

(a) Prove the inequality first for $\mathbb{R}^2$. That is, for all $a, b, c, d \in \mathbb{R}$, prove that

$$ab + cd \leq \sqrt{a^2 + c^2}\sqrt{b^2 + d^2}.$$

Hint: The first step is to simplify this expression so that it is easier to work with.

(b) Extend the inequality from $\mathbb{R}^2$ to $\mathbb{R}^n$ by induction.

Hint: To get the induction step, apply the $\mathbb{R}^2$ version to obtain the $\mathbb{R}^3$ version. This is very similar to how you would get the $\mathbb{R}^3$ triangle inequality from the $\mathbb{R}^2$ triangle inequality.

Expressed in terms of the inner product and Euclidean norm on $\mathbb{R}^n$, it can be rewritten as follows:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|.$$

That is, the Cauchy-Schwarz inequality is really a statement about the inner product. To get some geometric intuition about what it says, we note that in $\mathbb{R}^2$ and $\mathbb{R}^3$, where we have good geometric intuition, then we can check that the inner product gives us the angle θ between the these vectors. Specifically, in these cases, we can show that

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|x\|\|y\| \cos\theta.$$

For $\mathbb{R}^n$, with $n > 3$, we have no visual idea what an angle would look like, so we cannot really prove this statement. Instead, we turn the situation upside down and actually *define* what we mean by the angle formed by two vectors in $\mathbb{R}^n$ in terms of the inner product. That is, we make the following definition (which we probably won't use, but which is kind of fantastic).

Definition 1.3

Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we define the angle formed by these two vectors, to be the unique number $\theta \in [0, \pi]$ for which

$$\cos\theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

In particular, we see that the inner product is equal to zero if and only if $\mathbf{x}, \mathbf{y}$ are orthogonal. On the "opposite" end of the scale, the inner product of $\mathbf{x}, \mathbf{y}$ is the largest possible if they are parallell.

Exercise 1.5 Based on this discussion, what is a reasonable geometric intepretation of the Cauchy-Schwarz inequality? (There are several, so see if you can figure one out for yourself.)

Exercise 1.6 Formulate the observation on the relation between the inner product, orthogonality and parallellity as a proposition. Note that in this way, the inner product becomes a tool for detecting whether pairs of vectors are orthogonal or parallell. Also, explain why it may be an advantage to normalise these vectors, and instead work with $\mathbf{x}/\|\mathbf{x}\|$ and $\mathbf{y}/\|\mathbf{y}\|$ if we want to understand how they relate to each other in terms of angles.

Exercise 1.7 Use the Cauchy-Schwarz inequality to prove the triangle inequality.

For a pair of vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ (and only in $\mathbb{R}^3$!), we can compute their cross product

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

For us, the important thing about the cross product is that its norm $\|\mathbf{v} \times \mathbf{w}\|$ is equal to the area spanned by the parallellogram defined by $\mathbf{v}, \mathbf{w}$ when they are considered as starting from the same point.

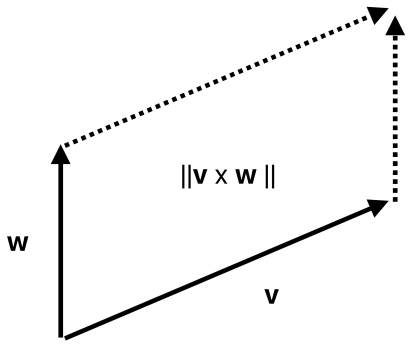

Figure 1.3: The area spanned by the parallellogram is equal to $\|\mathbf{v} \times \mathbf{w}\|$.

As noted earlier, an $m \times n$ matrix is a function of the form

$$A : \mathbb{R}^n \to \mathbb{R}^m$$

that we express as

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Such matrices are linear functions on $\mathbb{R}^n$. That is, if we define a function $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, then $\mathbf{f}$ is linear in the sense that for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ and $a_1, a_2 \in \mathbb{R}$, we have

$$\mathbf{f}(a\mathbf{x}_1 + b\mathbf{x}_2) = a_1\mathbf{f}(\mathbf{x}_1) + a_2\mathbf{f}(\mathbf{x}_2). \tag{1.1}$$

Remark 1.4: Warning on linear versus affine functions

In one variable Calculus, it makes sense to think of functions $f : \mathbb{R} \to \mathbb{R}$ as being linear if their graph is a straight line. However, the functions of the form $f(x) = kx + m$ are not what we call linear as they do not satisfy the relation (1.1). Instead, we call such functions *affine functions*.

To avoid confusion, we formulate the following definition explicitly.

Definition 1.5: Affine functions.

We say that a function $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is affine if $\mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})$ is linear.

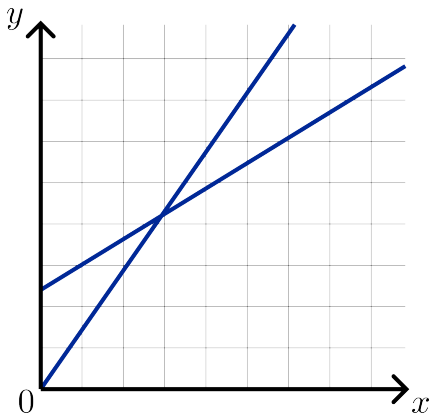

Figure 1.4: Both straight lines are graphs of affine functions, both only one is given by a linear one. Who would have thought?

Exercise 1.8

(a) Show that a straight line $f(x) = kx + m$ is linear if and only if $m = 0$.

(b) Show that $f : \mathbb{R} \to \mathbb{R}$ is linear if and only if it is of the form $f(x) = kx$.

Hint: These two exercises may seem almost identical, but the point is that you are asked to connect different definitions to the notion of being linear.

Exercise 1.9 Show that a function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there exists an $m \times n$ matrix A such that $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Hint: What is $A\mathbf{e}_j$ if $\mathbf{e}_j \in \mathbb{R}^n$ is the unit vector in the j'th coordinate direction? Can this be used to construct the matrix A from $\mathbf{f}$?

Exercise 1.10 Show that if $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is affine if and only if there exists an $m \times n$ matrix A and a vector $\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{f}(\mathbf{x}) = A\mathbf{x} + \mathbf{v}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Remark: Here, you should use the definition that $\mathbf{f}$ is affine if $\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})$ is linear. Oh, and as a hint, look to the previous exercise.

Matrices map straight lines to straight lines, but can distort angles and lengths. This means that matrices sends rectangles to parallellograms. Since matrices are linear functions, the ratio of the area of a rectangle and the parallellogram it is mapped to is always the same. In the case of $\mathbb{R}^2$, the scaling factor by which a 2×2 matrix changes the area of a rectangle in its domain is exactly equal to its determinant $|A|$. That is if, we denote the area of the rectangle R by $|R|$, and the area of the parallellogram it is mapped to by $|A(R)|$, then we have

$$|A(R)| = |A||R|.$$

The corresponding statement holds in $\mathbb{R}^m$, where concepts rectangles, parallellograms and areas should be replaced by "m-dimensional boxes, parallellepipeds and volume".

1.3 Limits of sequences

We start by recalling that a sequence of points $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ converges to a limit $L \in \mathbb{R}$ if the following ϵ-style definition holds: For all $\epsilon > 0$ there exists $N \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ the

following holds:

$$n > N \implies |a_n - L| < \epsilon.$$

Thanks to the Euclidean norm, the same definition extends painlessly to $\mathbb{R}^n$.

Definition 1.6

A sequence of points $(\mathbf{a}_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^n$ converges to a limit $\mathbf{L} \in \mathbb{R}^n$ if the following holds:

For all $\epsilon > 0$ there exists $N \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have

$$n > N \implies \|\mathbf{a}_n - \mathbf{L}\| < \epsilon.$$

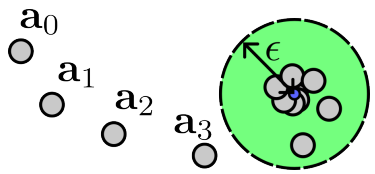

Figure 1.5: The sequence $(\mathbf{a}_n)_{n\in\mathbb{N}}$ seems to approach some point. We can see that after a certain N, all the $\mathbf{a}_n$ are in an ϵ-ball around this point.

A convenient technical result for dealing with convergence in $\mathbb{R}^n$ is that convergence can be handled on an entry-by-entry basis. To avoid annoying subscripts, we express this in $\mathbb{R}^2$.

Proposition 1.7

A sequence of points $((a_n, b_n))_{n\in\mathbb{N}}$ satisfies

$$\lim_{n\to\infty} (a_n, b_n) = (a, b)$$

if and only if

$$\lim_{n\to\infty} a_n = a \quad \text{and} \quad \lim_{n\to\infty} b_n = b.$$

Exercise 1.11 Prove Proposition 1.7

Exercise 1.12 Formulate Proposition 1.7 in the case of general $\mathbb{R}^n$ and prove it.

Remark: The only difference between the situation in this exercise, and that of $n = 2$, is that in $\mathbb{R}^n$ you also need subscripts to handle the entries of each point of the sequence. That is, you need two sets of indices.

Since convergence in $\mathbb{R}^n$ can be considered component-wise, most computational rules of limits for sequences in $\mathbb{R}^n$ follow from their counter-parts in $\mathbb{R}$. In particular, this is true for all of the following computational rules.

Proposition 1.8: Computational rules for limits of sequences in $\mathbb{R}^n$

Let $(\mathbf{a}_n)_{n\in\mathbb{N}}$ and $(\mathbf{b}_n)_{n\in\mathbb{N}}$ be two convergent sequences in $\mathbb{R}^n$. Then the following holds.

(i) $$\lim_{n\to\infty}(\mathbf{a}_n+\mathbf{b}_n)=\lim_{n\to\infty}\mathbf{a}_n+\lim_{n\to\infty}\mathbf{b}_n,$$

(ii) $$\lim_{n\to\infty}k\,\mathbf{a}_n=k\lim_{n\to\infty}\mathbf{a}_n,$$

(iii) $$\lim_{n\to\infty}(\mathbf{a}_n,\mathbf{b}_n)=(\lim_{n\to\infty}\mathbf{a}_n,\lim_{n\to\infty}\mathbf{b}_n),$$

(iv) $$\lim_{n\to\infty}\|\mathbf{a}_n\|=\Big\|\lim_{n\to\infty}\mathbf{a}_n\Big\|$$

Exercise 1.13 Prove these rules just using Proposition 1.7 and the corresponding results on $\mathbb{R}$ (that you may assume to be true).

1.4 Infinite series

In the course on one variable analysis, we define an infinite series

$$\sum_{n=1}^{\infty}a_n$$

as a limit of partial sums. More precisely, we define

$$\sum_{n=1}^{\infty}a_n=\lim_{N\to\infty}\sum_{n=1}^{N}a_n.$$

Important special cases of infinite series include the geometric series, positive series and alternating series. We now recall briefly what made these infinite series so special.

Remark 1.9: Important examples of infinite series

- **Geometric series.** Suppose $x\in\mathbb{R}$. Then for the geometric series

$$\sum_{n=0}^{\infty}x^n$$

we have a summation formula for the partial sums:

$$\sum_{n=0}^{N}x^n=\frac{1}{1-x}-\frac{x^{N+1}}{1-x}.$$

Essentially, this formula gives us complete control of the behaviour of geometric series. In particular, we see that their partial sums diverge as $N\to\infty$ for $|x|\geq 1$, while for $|x|<1$, they converge and give us the formula

$$\sum_{n=0}^{\infty}x^n=\frac{1}{1-x},\qquad |x|<1.$$

- **Alternating series.** Suppose that the sequence $a_n\geq 0$ tends monotonically to

zero. Then the alternating series

$$\sum_{n=0}^{\infty}(-1)^n a_n$$

converges to some number L. Moreover, the distance between L and the N'th partial sum satisfies

$$\left| L - \sum_{n=0}^{N}(-1)^n a_n \right| < a_{N+1}.$$

As surprisingly many infinite series turn out to be alternating, this observation sometimes gives us the easiest way to find error estimates of certain approximations.

- **Positive series.** We say that an infinite series

$$\sum_{n=0}^{\infty} a_n$$

is positive if all its terms are positive. That is, if $a_n \geq 0$ for all $n \in \mathbb{N}$. We have less control over positive series than geometric or alternating series, but the their behaviour is limited by the Balloon lemma, which says that a positive series either converges or diverges to infinity. This observation allows us to establish a bunch of convergence tests:

  - The comparison test.
  - The limit comparison test.
  - The quotient test.
  - The root test.
  - The integral test.

Exercise 1.14 Another important special case, that we do not mention above, are the telescoping series. With the word 'telescoping' in mind, try to figure out the sum of the following infinite series.

$$(a) \quad \sum_{n=1}^{\infty}\left(\sin\frac{1}{n} - \sin\frac{1}{n+1}\right) \qquad (b) \quad \sum_{n=1}^{\infty}\frac{1}{n(n+1)}.$$

Hint: In (b), think of how you would solve the correspong integral.

Exercise 1.15 Recall what each of the convergence tests for positive functions say.

Exercise 1.16 Recall the visual proof for the alternating series test.

Exercise 1.17 Recall the proof for the formula of the partial sums of the geometric series.

But what can we say about infinite series more generally? That is, if we do not know if it is geometric, alternating or positive? Well, there are some things which we can say.

Remark 1.10: Further rremarks on the convergence of infinite series

- **The divergence test.** If the terms of an infinite series do **not** go to zero, then the infinite series must diverge. Warning: Note that the converse of the divergence test is not true. The harmonic series is an example that illustrates this.
- **Absolute and conditional convergence.** If the infinite series $\sum_{n=0}^{\infty} |a_n|$ converges, then we can prove that the infinite series $\sum_{n=0}^{\infty} a_n$ also converges. Specifically, we say that it **converges absolutely**. This is a way to force an infinite series to become a positive series, so that we can apply the theory for positive series. Warning: Not all convergent series are absolutely convergent. The alternating harmonic series is an example that illustrates this. We call such series **conditionally convergent**. They tend to be "badly behaved", which we discuss further below,.

We now say some words on the tails of infinite series.

Remark 1.11: Remarks on the importance of tails

- **Approximations using tails.** If an infinite series converges, then it is possible to approximate its value using its partial sums. To determine the quality of the approximation, one strategy is to find error bounds on its tail. Specifically, if we know that
$$\sum_{n=0}^{\infty} a_n = L$$
for some value of L, then by splitting up the infinite series into a partial sum and corresponding tail, we have
$$\sum_{n=0}^{N} a_n + \sum_{n=N+1}^{\infty} a_n = L$$
which gives
$$\text{error} = L - \sum_{n=0}^{N} a_n = \sum_{n=N+1}^{\infty} a_n.$$
So any esimate on the tail will automatically translate into an estimate of the error of the approximation of L using partial sums. This is great, since partial sums can be computed numerically on a computer.
- **Even more on tails.** We also mention that since any infinite series can be written as, say,
$$\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{1000} a_n + \sum_{n=1001}^{\infty} a_n,$$
then any question about its behaviour can be reduced to a question of the behaviour of **any one of its tails**. Indeed, if you split out the first 1000 terms, then the value of their sum can easily be checked numerically, and so the question of the behaviour of the infinite series can be redirected to this (or any other) tail. In particular, if the infinite series becomes geometric, alternating or positive after some index N, then we can use this to study the infinite series.

Finally, we offer some perspective on why we call conditionally convergent series "badly

behaved" (such as the alternating harmonic series). Here is one way to explain what is going on. Note that the statement of the theorem may seem hard to understand at first glance (we do not expect you to have seen it before), so illustrate what is going on the example directly after the statement.

Proposition 1.12: Riemann's rearrangement theorem

Suppose that the infinite series

$$\sum_{n=0}^{\infty} a_n$$

is conditionally convergent. Then, for any $L \in \mathbb{R}$, $L = \infty$ or $L = -\infty$, there exists an increasing sequence of finite subsets $U_N \subset \mathbb{N}$ so that

$$(i) \quad U_N \subset U_{N+1} \qquad (ii) \quad \bigcup_{N=1}^{\infty} U_N = \mathbb{N}$$

and for which

$$\lim_{N\to\infty} \sum_{n\in U_N} a_n = L.$$

Example 1.13

A simple example to illustrate what the above theorem says is to point out the following 'trivial' case. Indeed, choose the sets U_N so that

$$\begin{aligned} U_0 &= \{0\} \\ U_1 &= \{0,1\} \\ U_2 &= \{0,1,2\} \\ &\vdots \\ U_N &= \{0,1,2,\ldots,N\}. \end{aligned}$$

and so forth. Then for any sequence $(a_n)_{n=0}^{\infty}$, we get

$$\sum_{n\in U_N} a_n = \sum_{n=0}^{N} a_n.$$

And so,

$$\lim_{N\to\infty} \sum_{n\in U_N} a_n$$

is just the standard definition of what we mean by an infinite series.

The point of the above theorem, that we hoped the example illustrated, is that by choosing the sets U_N in a different way, we can change the value of the conditionally convergent infinite series. This is **not** the case for absolutely convergent series. Their value will remain the same, no matter how you choose the U_N (as long as they satisfy the condition of the theorem).

Below, there are some exercises you can do if you find this interesting. To make things easier to talk about, we formulate the following definition.

Definition 1.14: Summation method

We call a sequence $U = (U_N)_{N\in\mathbb{N}}$ of finite subsets $U_N \subset \mathbb{N}$ a summation method, if it satisfies (i) and (ii) of the above proposition. Moreover, if

$$\lim_{N\to\infty} \sum_{n\in U_N} a_n = L,$$

then we write

$$\sum_{n=0}^{\infty} a_n \overset{(U)}{=} L.$$

Exercise 1.18 Replace the 'standard' choice of the U_N, used in the above example, by $U_N = \{0, 1, \ldots, 2N\}$. What is the value of the so-called Grandi series

$$\sum_{n=0}^{\infty} (-1)^n$$

using this summation method? What is the value, if any, of the Grandi series in the 'usual' summation method? What does this say about the divergence test?

Exercise 1.19 Find a summation method U for which

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$$

Remark: You need to formulate the summation method in terms of an algorithm (no other way is possible, I think). Try to implement this algorithm in, say, Python, to see if your algorithm works.

Exercise 1.20 (Project idea) Using the language of summation methods, you can also think about how to define infinite series of multiply indexed sequences. In particular, suppose you have a sequence $(a_{i,j})_{i,j\in\mathbb{N}}$, how should we define

$$\sum_{i,j\in\mathbb{N}} a_{i,j} \ ?$$

As it turns out, there is no "obvious" or "canonical" choice - instead there are several reasonable ways to do this. Extend the definition of summation methods to sequnces indexed by $\mathbb{N}^2$, and see if you can come up with some reasonable methods. Moreover, try to come up with "reasonable" examples of multiply indexed sequences that behave differently with respect to these summation methods.

1.5 Representations of functions in one and several variables

In the next couple of sections, we are going to review limits, derivatives and integrals of functions in one and several variables. To prepare for this, we briefly consider some ways in which we can think of and visualise such functions.

Representations of functions $f : D \subset \mathbb{R} \to \mathbb{R}$

The most familiar should be the functions $f : D \subset \mathbb{R} \to \mathbb{R}$. We do not need to say much, but point out that you need to know exactly what the graph of such a function represents.

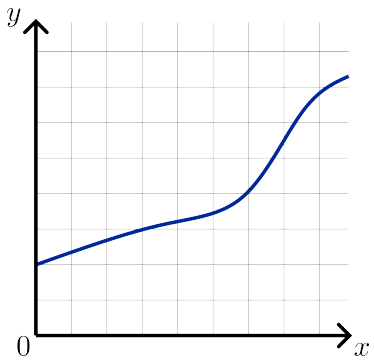

Figure 1.6: Here we can see an example of a function $f : D \subset \mathbb{R} \to \mathbb{R}$. We can also see that f is increasing since the graph goes up to the right.

> **Exercise 1.21** The above graph is also an illustration of the inverse function of f, which we denote by f^{-1}.

Representations of functions $\mathbf{f} : D \subset \mathbb{R} \to \mathbb{R}^m$

Next, we consider functions $\mathbf{f} : D \subset \mathbb{R} \to \mathbb{R}^m$. We often think of such functions as representing curves in $\mathbb{R}^m$, or, even better, to describe the path of a particle moving through $\mathbb{R}^m$. For this reason, we often use t as the variable (imagining it to represent 'time') and write

$$\mathbf{f}(t) = \big(f_1(t), f_2(t), \ldots, f_d(t)\big).$$

When we think of the variable as time, the analogy with the moving particle, helps us to take into account that while such a function may trace out a curve, it may do so at different 'speeds'. Moreover, it can also happen that the particle, stops and reverses direction, or that its path overlaps. We also mention that since the word 'curve' begins with the letter 'c', people even tend to favour writing $\mathbf{c}(t)$ for such functions.

Example 1.15

Consider the function $\mathbf{f} : [0, \pi] \to \mathbb{R}^2$ given by

$$\mathbf{f}(t) = (\cos t, t \sin t).$$

There are two possible ways for us to visualize this: either we draw the set $\{(\cos t, t \sin t) : t \in [0, \pi]\}$ or we draw the graphs of the component functions $f_1(t) = \cos t$ and $f_2(t) = t \sin t$.

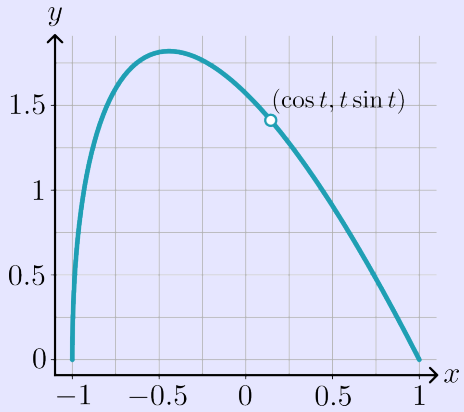

Figure 1.7: We can visualize the curve $\mathbf{f}(t) = (\cos t, t \sin t)$ as its graph in the plane.

While the path analogy is tempting and solid, it makes the most sense for continuous curves. Moreover, for any notion of speed to make sense, we also need for the function to be differentiable. While continuity and differentiability are both terms that will be discussed further below, we already now make this definition.

Definition 1.16

We call a continuous function $\mathbf{f} : [a, b] \to \mathbb{R}^m$ a continuous parameterised curve. If, in addition, $\mathbf{f}$ is differentiable, we call it a differentiable parameterised curve. Moreover, by a *curve*, we mean the image of a continuous or differentiable parameterised curve.

Remark 1.17: More bad terminology

We sometimes refer to continuous parameterised curves and differentiable parameterised curves simply as 'parametereised curves' or even as 'curves'. However, while we need to accept this, we should be aware that this is a so-called 'abuse of notation'.

Exercise 1.22 Show that a curve may have different parameterisations.

Hint: Why not consider Example 1.15?

Exercise 1.23 In one variable analysis, we tend to think of the graphs of functions $f : D \subset \mathbb{R} \to \mathbb{R}$ as curves in $\mathbb{R}^2$. Can you strengthen this connection by expressing the graph of such a function f is a curve, in the sense discussed above? (Please feel free to assume that f is continuous if you want to avoid the abuse of terminology.)

Representations of functions $f : D \subset \mathbb{R}^n \to \mathbb{R}$

We now turn to another special case, namely functions $f : D \subset \mathbb{R}^n \to \mathbb{R}$. These are functions of the form

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

These functions are harder to visualise, but in the special case $n = 2$, i.e., when $f : D \subset \mathbb{R}^2 \to \mathbb{R}$, we tend to think of them as some type of surface traced out by the graph of f. This is an exact analogue to to how we think of the graph of $f : D \subset \mathbb{R} \to \mathbb{R}$ as a curve in $\mathbb{R}^2$ (see exercise 1.22, above).

Example 1.18

An example of such a function is the function $f : [-\pi, \pi]^2 \subset \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = \sin(x + y).$$

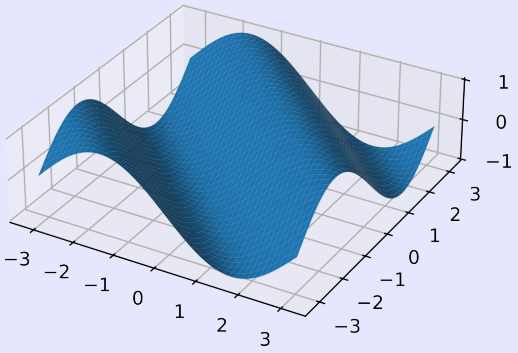

Figure 1.8: We can visualize $f(x, y) = \sin(x + y)$ as its graph, i.e. the set $\{(x, y, \sin(x + y)) : (x, y) \in [-\pi, \pi]\}$. This set is seen here, and looks very much like some form of surface. The advantage here is that one easily gets an idea of what the function "looks like".

Another way, which is essentially the same, is to think of the function in terms of its 'contour' lines.

Example 1.19

Let us this time consider the function $f : [-2\pi, 2\pi]^2 \subset \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = \cos x + \sin y.$$

We can visualize this function by drawing so-called contour lines. The idea is to look at different values the function takes, and draw the "level sets" of those values, i.e. draw the sets $f^{-1}(\{z\}) \subset \mathbb{R}^2$ for these values of z. This allows one to get an idea of where the values of the function changes. One can also go one step further and color the levels sets differently depending on the value z the function takes there.

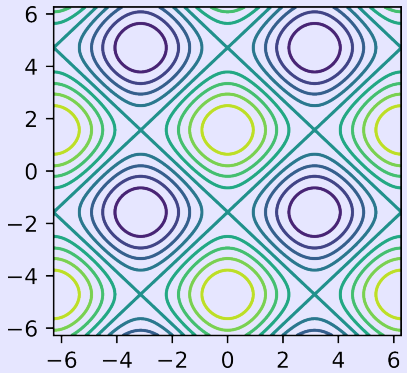

Figure 1.9: Contour lines of the function $f(x, y) = \cos x + \sin y$. Darker colors represent lesser values of $f(x, y)$ and brighter colors represent greater values of $f(x, y)$. This figure has the advantage that it is easier to guess where the maxima and minima of the function are, and one can also use it to easily identify saddle points.

Representations of functions $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$

Next, we turn to functions $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$. A very pragmatic way of thinking of such functions is to write

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))$$

and then consider the functions $f_j : D \subset \mathbb{R}^n \to \mathbb{R}$ separately. However, while this may be a good strategy for some computational and technical tasks, we risk missing an overall understanding of what it is that the function $\mathbf{f}$ is up to. In fact, when n and m are large, it may be really hard to get a good intuition. However, for small n and m, there are options. Let us consider some.

The cases $\mathbf{f} : D \subset \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbf{f} : D \subset \mathbb{R}^3 \to \mathbb{R}^3$ offer a couple of reasonable ways of representing the functions, that we only discuss in the former case (since it is easier to make pretty pictures). The first is to think of such functions in terms of how they deform $\mathbb{R}^2$.

Example 1.20

To motivate why the following function is interesting, let us start by looking at the complex plane. Recall that the complex numbers $\mathbb{C}$ consist of the numbers of the form $z = x + iy$ where $x, y \in \mathbb{R}$ and $i^2 = -1$. A simple function one might consider on the complex plane is then the function $g : \mathbb{C} \to \mathbb{C}$ given by $g(z) = z^2$. If we look at z in the form $x + iy$, then this gives

$$g(x + iy) = (x + iy)^2 = x^2 + 2xiy + (iy)^2 = (x^2 - y^2) + i(2xy).$$

If we now think of complex numbers as points in $\mathbb{R}^2$, then this then corresponds to the function $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathbf{g}(x, y) = (x^2 - y^2, 2xy).$$

We will consider *almost* this function. In particular, we will consider $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\mathbf{f}(x, y) = \frac{1}{4}(x^2 - y^2, 2xy),$$

as this will give a nicer visualization.

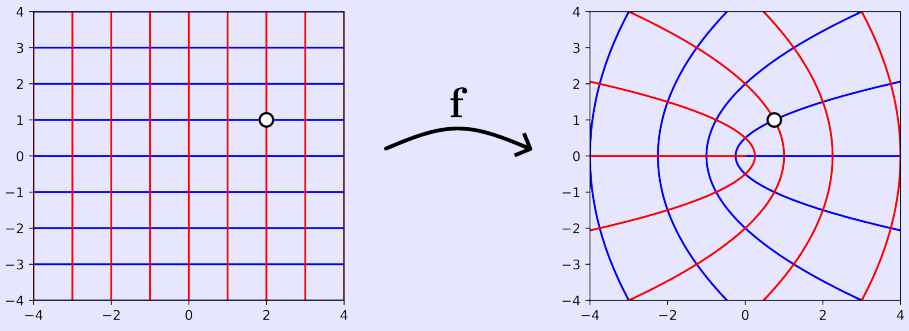

Figure 1.10: We can visualize $\mathbf{f}(x, y) = \frac{1}{4}(x^2 - y^2, 2xy)$ by looking at how it transforms the grid lines in the plane. To make it clear how this work, we have also marked a point in the original plane and the point it maps to. The blue lines on the left map to blue lines on the right. Can you see any interesting properties of $\mathbf{f}$ from this?

Another way of representing such functions is to think of them as vector fields.

Example 1.21

Let us look at the function $\mathbf{f} : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}^2$ given by

$$\mathbf{f}(x,y) = \frac{1}{\|(x,y)\|}(-y,x).$$

A way to visualize this function is to think of it like this: for each point (x,y) in the plane, the function gives us a vector $\mathbf{f}(x,y)$, which we can think of as an arrow based at (x,y). This gives us a visualization of the function as in the following figure.

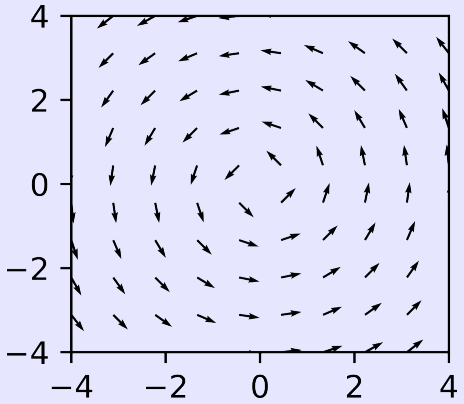

Figure 1.11: We select some grid of points (x,y) in the plane, and at each of them compute the vector $(u,v) = \mathbf{f}(x,y)$. Then we draw each vector (u,v) as an arrow based at the point (x,y). This is a very common visualization technique when dealing with ordinary differential equations.

As you can imagine, similar representations can be achieved in $\mathbb{R}^3$, but please don't make us do this.

We now consider the case $\mathbf{f} : D \subset \mathbb{R}^2 \to \mathbb{R}^3$. This case sort of touches on both the case of functions $\mathbb{R}^2 \to \mathbb{R}$ (whose graphs looked like surfaces) and $\mathbb{R} \to \mathbb{R}^3$ (that traced out curves).

Example 1.22

As we saw in example 1.18, functions $\mathbf{f} : D \subset \mathbb{R}^2 \to \mathbb{R}$ give rise to surfaces (whatever that means) in $\mathbb{R}^3$. But we would just as well think of surfaces which are not given by graphs of such functions. For example, what about the surface of the unit ball? Surely this should be a surface, but it cannot be the graph of a function (each (x,y) would have two distinct z on the surface associated with it). So how can we describe this surface? The idea is to try to put some sort of coordinate system on it. The way we do this is by finding some function which maps some domain $D \subset \mathbb{R}^2$ to out surface. For the surface of the unit ball, an example of this is given by the spherical coordinate function $\mathbf{f} : [0,\pi] \times [0,2\pi) \subset \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$\mathbf{f}(\theta,\varphi) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta).$$

It turns out that the range of this function, i.e. the set $\{f(\theta,\varphi) : \theta \in [0,\pi],\ \varphi \in [0,2\pi)\}$, is precisely then the surface of the unit ball. We visualize this in the following figure.

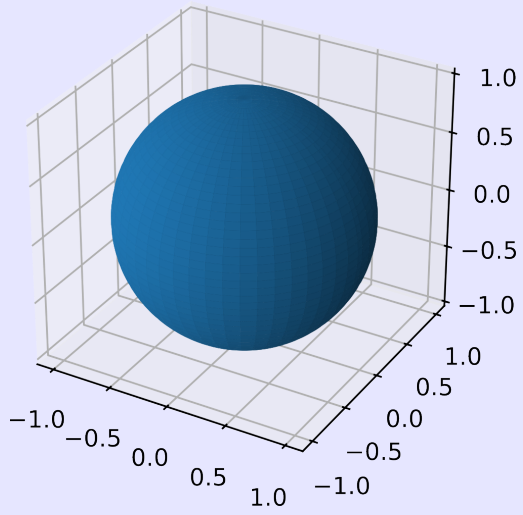

Figure 1.12: The surface of the unit ball can be given as the image of a continuous function $\mathbf{f} : D \subset \mathbb{R}^2 \to \mathbb{R}^3$.

Indeed, the analogy to the situations with curves is strong enough that we make the following definition. Again, we remark that while we use the terms continuity and differentiability here, they will only be discussed later in this chapter.

Definition 1.23: Paremeterised surfaces in $\mathbb{R}^3$

We call a continuous function $\mathbf{f} : [a, b] \times [c, d] \to \mathbb{R}^3$ a parameterised surface. If, in addition, $\mathbf{f}$ is differentiable, we call it a differentiable parameterised surface. Moreover, by a surface in $\mathbb{R}^3$, we mean the image of a continuous or differentiable parameterised surface.

Remark 1.24: Even more bad terminology

The same remark as we did for curves applies here (see Remark 1.17).

Exercise 1.24 Repeat exercise 1.23, except this time show that it makes sense, in terms of the definitions made here, to think of the graphs of functions $\mathbf{f} : D \subset \mathbb{R}^2 \to \mathbb{R}$ as surfaces (again, please think of f as being continuous if this helps you sleep at night - also, assume whatever you want about D).

1.6 Limits of functions and continuity

Definition of limits and continuity for $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$

To define limits and continuity in the general case, we start by recalling the epsilon-type definition of the limit of a function $f : D \subset \mathbb{R} \to \mathbb{R}$. Indeed, we say that $f(x)$ converges to $L \in \mathbb{R}$ as $x \to c$ if the following holds. For every $\epsilon > 0$, there exists a $\delta > 0$ so that for all $x \in D$, we have

$$\begin{cases} |x - c| < \delta \\ x \neq c \end{cases} \implies |f(x) - L| < \epsilon.$$

We sometimes express this as

$$\lim_{x\to c} f(x) = L \quad \text{or} \quad f(x) \underset{x\to c}{\longrightarrow} L.$$

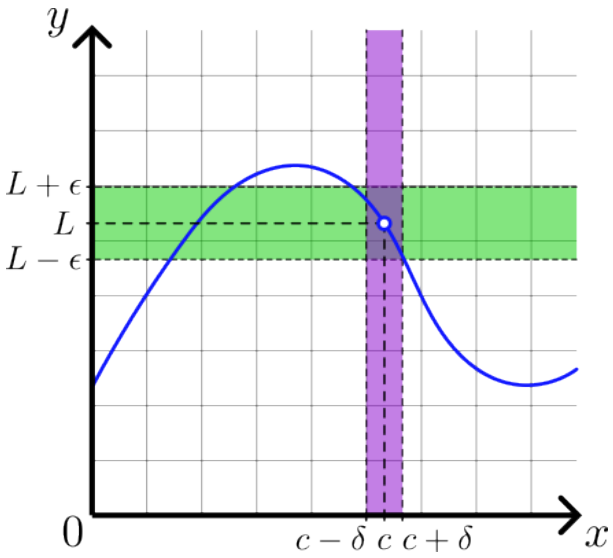

Figure 1.13: For any given ϵ, we can find some δ so that when x is δ-close to c, $f(x)$ is ϵ-close to L.

There is a piece of fine print here. In order to be able to consider the limit as $x \to c$, we need to be able to approach c in the domain of f other than c itself. We call points $c \in D$ for which this is possible limit points of D. Here is a proper definition.

Definition 1.25: Limit points of a set in $\mathbb{R}^n$.

Let $D \subset \mathbb{R}^n$ be a set.

(i) We call $\mathbf{c} \in \mathbb{R}^n$ a limit point of D if, for all $\delta > 0$, there exist $\mathbf{x} \in D\backslash\{\mathbf{c}\}$ so that $\|\mathbf{x} - \mathbf{c}\| < \delta$.

(ii) Any $\mathbf{c} \in D$ that is not a limit point of D is called an *isolated point*.

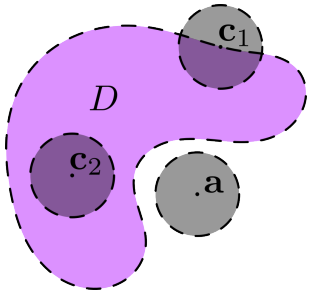

Figure 1.14: The points $\mathbf{c}_1$ and $\mathbf{c}_2$ are limit points of D since every ball around them contains some point of D distinct from them, but as we can also see, $\mathbf{a}$ is not a limit point of D, since there is some ball around it completely disjoint from D. $\mathbf{a}$ is not even a limit point of the set $D \cup \{\mathbf{a}\}$. As we can see, being a limit point is independent of being an element of a set, as $\mathbf{c}_2$ is an element of D, while $\mathbf{c}_1$ and $\mathbf{a}$ are not.

Example 1.26

Let us consider the set $D = [0,1) \cup \{2\}$. Then any $c \in [0,1]$ is a limit point of D, since for all $\delta > 0$, we can find $x \in [0,1)$ that are δ close to c. However, $c = 2$ is an isolated point as we cannot find points in $D \backslash \{2\} = [0,1)$ arbitrarily close to c.

With this terminology, we formulate the definition of the limit for functions in several variables.

Definition 1.27

Let $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ be a function and $\mathbf{c}$ a limit point of D. Then, we say that $\mathbf{f}$ converges to $\mathbf{L}$ as $\mathbf{x} \to \mathbf{c}$ if, for every $\epsilon > 0$, there exists $\delta > 0$ such that for all $\mathbf{x} \in D$ we have

$$\begin{cases} \|\mathbf{x} - \mathbf{c}\| < \delta \\ \mathbf{x} \neq \mathbf{c} \end{cases} \implies \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{c})\| < \epsilon.$$

If this holds, we write

$$\lim_{\mathbf{x} \to \mathbf{c}} \mathbf{f}(\mathbf{x}) = \mathbf{L}.$$

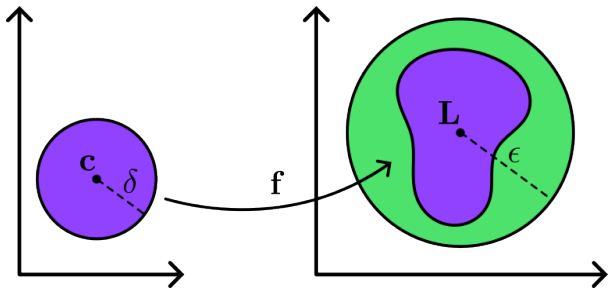

Figure 1.15: For any ϵ-ball around $\mathbf{L}$, we can find some δ-ball around $\mathbf{c}$ which maps into the ϵ-ball. It might even be easier to understand this definition in $\mathbb{R}^2$.

Having defined the limit of a function, we can proceed to define continuity. Essentially, we just want to say that $\mathbf{f}$ is continuous at a point $\mathbf{x} \in D$ if

$$\lim_{\mathbf{x} \to \mathbf{c}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{c})$$

However, to make sure that we go for a definition which will match a topological definition we will see later in the course, we go for the following, slightly more technical looking, definition.

Definition 1.28: Continuity

Let $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ be a function and $\mathbf{c} \in D$. Then, we say that $\mathbf{f}$ is continuous at $\mathbf{c}$ if the following holds.

For every $\epsilon > 0$, there exists $\delta > 0$ such that for all $\mathbf{x} \in D$ we have

$$\|\mathbf{x} - \mathbf{c}\| < \delta \implies \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{c})\| < \epsilon.$$

We say that $\mathbf{f}$ is a continuous function if this holds for all $\mathbf{c} \in D$.

Remark 1.29: Technical remark

If $\mathbf{c}$ is a limit point of D, then the above definition says exactly that

$$\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{c}).$$

However, at isolated points there is a difference. On the one hand, if $\mathbf{c} \in D$ is isolated, then definition of the limit does not apply. On the other hand, the condition of the definition of continuity still makes sense (and trivially holds, as $\mathbf{x} = \mathbf{x}$ is the only point satisfying $\|\mathbf{x} - \mathbf{c}\| < \delta$ when δ is small enough). In particular, it follows that a function os always continuous on its isolated points.

Limits and continuity in the special case $\mathbf{f} : D \subset \mathbb{R} \to \mathbb{R}^m$

We now take a quick tour of some important special cases, in order to get a better sense of how to treat limits and continuity in higher dimensions. We basically follow in the footsteps of Section 1.5. The first stop we make are functions $\mathbf{f} : D \subset \mathbb{R} \to \mathbb{R}^n$. These functions are of the form

$$\mathbf{f}(x) = \big(f_1(x), f_2(x), \ldots, f_d(x)\big).$$

This case is particularly friendly, as we can consider the limit on a entry-by-entry basis (recall Proposition 1.7).

Example 1.30

Consider the function $\mathbf{f}(x) = (3x, 2x)$. Then

$$\begin{aligned}\lim_{x\to 3} \mathbf{f}(x) &= \lim_{x\to 3} (3x+1, 2x+2) \\ &= \Big(\lim_{x\to 3} 3x, \lim_{x\to 3} 2x\Big) = (9, 6).\end{aligned}$$

Exercise 1.25 Suppose that $\mathbf{f} : [0,1] \subset \mathbb{R} \to \mathbb{R}^2$ and let $c \in [0,1]$ and $\mathbf{L} = (L_1, L_2) \in \mathbb{R}^2$. Show that

$$\lim_{t\to c} \mathbf{f}(t) = \mathbf{L}$$

if and only

$$\lim_{t\to c} f_1(t) = L_1 \quad \text{and} \quad \lim_{t\to c} f_2(t) = L_2.$$

Exercise 1.26 Show that $\mathbf{f}(t) = \big(f_1(t), f_2(t)\big)$ is continuous at a point c if and only if f_1 and f_2 are continuous at c.

Exercise 1.27 Determine which of the following functions define paths:

(i) $\mathbf{f} : [-\pi, \pi] \to \mathbb{R}^3$, defined by $\mathbf{f}(t) = (\cos t \sin t, e^{t^2}, t+1)$.

(ii) $\mathbf{f} : [0,1]^2 \to \mathbb{R}^2$, defined by $\mathbf{f}(x, y) = (\cos x, \sin y)$.

(iii) $\mathbf{f} : [-2, 2] \to \mathbb{R}^2$, defined by

$$\mathbf{f}(t) = \begin{cases} \left(\frac{\sin t}{t}, \cos t\right), & t \neq 0, \\ (1,1), & t = 0. \end{cases}$$

(iv) $f : [0, 1] \to [0, 1]$ defined by $f(t) = t$.

(v) $f : \mathbb{R} \to \mathbb{R}$, defined by $f(t) = t$.

Limits and continuity in the special case $f : D \subset \mathbb{R}^n \to \mathbb{R}$

We now turn to another special case, namely functions $f : D \subset \mathbb{R}^n \to \mathbb{R}$. How do we check limits and establish continuity for such functions? As sequences have much more "freedom of movement" in $\mathbb{R}^n$ than in $\mathbb{R}$, there are some additional technical issues to deal with when considering concrete examples.

Example 1.31

Consider the function

$$f(x, y) = \frac{x^2 y}{x^4 + y^2}.$$

Let us check whether it has a limit at the origin. We can do this in several ways.
First, let us check what happens when we let (x, y) tend to $(0, 0)$ along the y-axis. This amounts to the following computation:

$$\lim_{t\to 0} f(0, t) = \lim_{t\to 0} \frac{0 \cdot t}{0 + t^2} = \lim_{t\to 0^+} 0 = 0.$$

What about along the straight line $y = kx$? This amounts to computing:

$$\lim_{t\to 0} f(t, kt) = \lim_{t\to 0} \frac{t^2 \cdot kt}{t^4 + (kt)^2} = \lim_{t\to 0} \frac{kt}{t^2 + k^2} = 0.$$

Taken together, we have now checked the limit along every line toward the origin. The limit is always 0, so surely the function has limit 0 whenever (x, y) tends toward zero? Well, actually no. Here is what happens when we tend to 0 along the path $t \mapsto (t, t^2)$:

$$\lim_{t\to 0} f(t, t^2) = \lim_{t\to 0} \frac{t^4}{t^4 + t^4} = \frac{1}{2}.$$

Crazy stuff, no? We now pause the example for a second, to discuss how to figure out what is going on.

The problem in the above example is that we say that by $(x, y) \to (0, 0)$, we mean that the Euclidean distance between (x, y) and $(0, 0)$ tends to zero. What this means is that we want to know how the function value behaves as we consider what happens on *all* points in the open Euclidean balls

$$B(0, t) = \{(x, y) : \|(x, y)\| < t\}$$

as $t \to 0$. Another way to approach this problem is to describe these open balls using polar coordinates. That is, if we write

$$x = r \cos\theta$$
$$y = r \sin\theta$$

then

$$B(0, t) = \{(r\cos\theta, r\sin\theta) : r \in [0, t), \theta \in [0, 2\pi)\}.$$

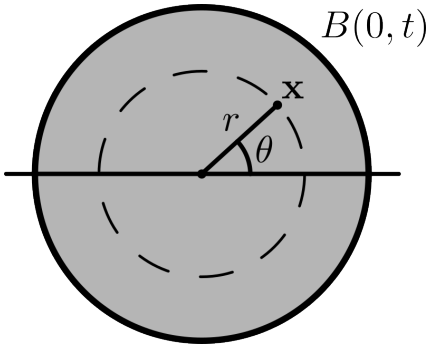

Figure 1.16: The point $\mathbf{x}$ can be described uniquely by its distance from the origin, r, and its angle with the x-axis, θ. With this we can write $\mathbf{x} = (r\cos\theta, r\sin\theta)$.

We now revisit the example.

Example 1.32

Again, we consider the mysterious function

$$f(x,y) = \frac{x^2 y}{x^4 + y^2}.$$

However, we now use polar coordinates to check the limit. First, we rewrite the formula for f in terms of polar coordinates:

$$f(x,y) = f(r\cos\theta, r\sin\theta) = \frac{r^3\cos^2\theta\sin\theta}{r^4\cos^4\theta + r^2\sin^2\theta} = \frac{r\cos^2\theta\sin\theta}{r^2\cos^4\theta + \sin^2\theta}$$

If this expression tends to 0 as $r \to 0$ *independent* of θ, then the limit of $f(x,y)$ as (x,y) tends to $(0,0)$ is equal to 0. If the limit depends on θ, then f does not have a limit as (x,y) tends to $(0,0)$.

So, to see what is the case here, we can now, for fixed r, check what the maximum value, with respect to θ is.

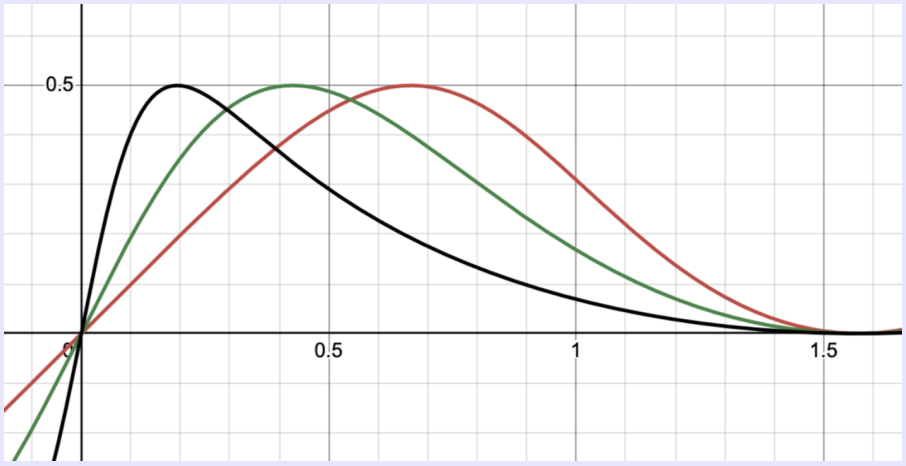

Figure 1.17: The graph of $\theta \mapsto f(r\cos\theta, r\sin\theta)$ for $r = 1$ (red), $r = 0.5$ (green) and $r = 0.2$ (black). As we see, for each value of r, there is an angle at which the function takes the value $1/2$. As we saw in Example 1.31, this happens exactly along the path $t \mapsto (t, t^2)$.

Exercise 1.28 Check whether $f(x,y) = x^2y^2/(x^4+y^2)$ has a limit as (x,y) approaches the origin.

Remark 1.33

Note that when we consider limits for functions defined on subsets of $\mathbb{R}^n$ for $n \geq 2$, there are infinitely many directions that can be considered. Moreover, it is not enough to check the limit along all straight lines. In particular, there is no hope of splitting the consideration of limits into, say, "left-sided" and "right-sided" limits as in the one variable case.

Some theorems on limits and continuity in the general case

We now return to the general case of functions $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$. As with the case of functions $f : D \subset \mathbb{R} \to \mathbb{R}^m$, we can reduce technical computations by working component-by-component. This is the content of the following result (which is a generalisation of exercise **??**, above).

Lemma 1.34

Let $\mathbf{f} = (f_1, \ldots, f_m) : D \subset \mathbb{R}^n \to \mathbb{R}^m$ and let $\mathbf{c}$ be a limit point of D. Then

$$\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x})$$

exists if and only if

$$\lim_{\mathbf{x}\to\mathbf{c}} f_j(\mathbf{x})$$

exists for all $j = 1, \ldots, m$. Furthermore, in that case,

$$\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x}) = \left(\lim_{\mathbf{x}\to\mathbf{c}} f_1(\mathbf{x}), \ldots, \lim_{\mathbf{x}\to\mathbf{c}} f_m(\mathbf{x}) \right).$$

With the lemma in hand, the following result follows in a fairly straight-forward manner from the one variable theory.

Proposition 1.35

Let $\mathbf{f}, \mathbf{g} : D \subset \mathbb{R}^n \to \mathbb{R}^m$, and let $\mathbf{c}$ be a limit point of D. Then we have the following:

(i) If $\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x})$ and $\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{g}(\mathbf{x})$ exist, then so does $\lim_{\mathbf{x}\to\mathbf{c}} (f(\mathbf{x}) + g(\mathbf{x}))$, and

$$\lim_{\mathbf{x}\to\mathbf{c}} (f(\mathbf{x}) + g(\mathbf{x})) = \lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x}) + \lim_{\mathbf{x}\to\mathbf{c}} \mathbf{g}(\mathbf{x}).$$

(ii) If $\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x})$ exists, then so does $\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{k}f(\mathbf{x})$ for all $k \in \mathbb{R}$, and

$$\lim_{\mathbf{x}\to\mathbf{c}} k\mathbf{f}(\mathbf{x}) = k \lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x}).$$

If, in addition, $\mathbf{h} : U \subset \mathbb{R}^m \to \mathbb{R}^d$, then the following also holds:

(iii) If $\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{f}(\mathbf{x}) = \mathbf{d}$ exists and is a limit point of U, and $\lim_{\mathbf{y}\to\mathbf{d}} h(\mathbf{y})$ exists, then so does $\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{h}(\mathbf{f}(\mathbf{x}))$, and we have that

$$\lim_{\mathbf{x}\to\mathbf{c}} \mathbf{h}(\mathbf{f}(\mathbf{x})) = \lim_{\mathbf{y}\to\mathbf{d}} \mathbf{h}(\mathbf{y}).$$

Exercise 1.29 Use Lemma 1.34 and the one variable theory to prove the properties listed in Proposition 1.35.

Exercise 1.30 Are there any other limit laws for functions you can think of? If so, we just forgot them, so it is best that you formulate and prove them so that you can use them later in the course.

We wrap this up with this section with some additional exercises on limits for the benefit of enthusiastic students.

Exercise 1.31 Let A be an $m \times n$ matrix, and let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ given by $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ be its associated function. Show that $\mathbf{f}$ is continuous.

Exercise 1.32

(i) Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Show that $\mathbf{f}$ is continuous.

(ii) Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be an affine function. Show that $\mathbf{f}$ is continuous.

Exercise 1.33 A function $f : \mathbb{R}^2 \to \mathbb{R}$ is called a polynomial in two variables if it is a finite sum of terms of the form $ax^n y^m$, where $a \in \mathbb{R}$ and $n, m \in \mathbb{N}$.

(i) Show that $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = 2xy + (xy^2 - x^5y^3)^2 xy$$

is a polynomial in two variables.

(ii) Show that if f, g are polynomials in two variables, then so are $f + g$, $f - g$ and fg. Is f/g ever a polynomial in two variables?

(iii) Show that a polynomial in two variables is continuous.

(iv) Give a definition of a polynomial in n variables and show that all polynomials in n variables are continuous.

Exercise 1.34 Give an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is not continuous at $(0, 0)$, such that, for each $\lambda \in \mathbb{R}$, the function $f_\lambda : \mathbb{R} \to \mathbb{R}$ given by $f_\lambda(x) = f(x, \lambda x)$ is continuous.

Exercise 1.35 Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is such that

$$\lim_{t \to 0} f(\mathbf{g}(t)) = 0$$

for all continuous functions $\mathbf{g} : \mathbb{R} \to \mathbb{R}^n$ with $\mathbf{g}(t) \neq \mathbf{0}$ for $t \neq 0$, and $\mathbf{g}(0) = \mathbf{0}$. Is it necessarily the case that

$$\lim_{\mathbf{x} \to \mathbf{0}} f(\mathbf{x}) = 0?$$

1.7 Differentiability

Definition of the differentiability and the derivative of $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$

To define what we mean by the derivative of a function $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$, we first recall the definition of the derivative of a function $f : D \subset \mathbb{R} \to \mathbb{R}$ at a point $c \in \mathbb{R}$:

$$f'(c) = \lim_{x \to a} \frac{f(x) - f(c)}{x - c}.$$

If this limit exists, we say that f is differentiable at c, and call $f'(c)$ its derivative at that point.

In higher dimensions, the situation is more complicated. To avoid some (potentially unnecessary) complications (but who knows?), we will only define the derivative at points that are especially easy to reach in terms of limits. To make this precise, we make the following definition.

Definition 1.36: Inner points of a set.

We say that a point $c \in D \subset \mathbb{R}^n$ is an inner point of D if there exists a $\delta > 0$ so that for all points $x \in \mathbb{R}^n$, the following holds:

$$\|\mathbf{x} - \mathbf{c}\| < \delta \implies x \in D.$$

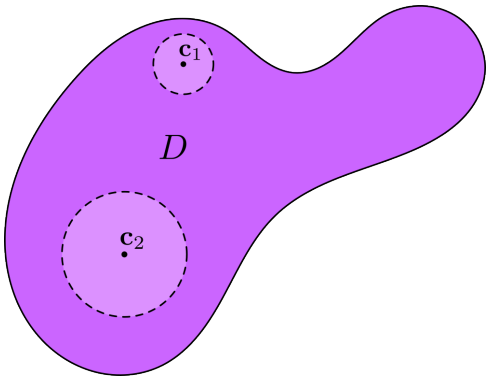

Figure 1.18: The points $\mathbf{c}_1$ and $\mathbf{c}_2$ are interior points of D, since we can fit balls around them in D.

Example 1.37

Let us consider the set $D = [0, 1) \cup \{2\}$. Then any $c \in (0, 1)$ is an inner point of D. Indeed, for $\delta = \min\{1 - c, c\}$, it holds that

$$|x - c| < \delta \implies x \in [0, 1).$$

The points $c = 0$ and $c = 2$ are not inner points of D, since no matter what $\delta > 0$ we choose, there will be points outside of D for which $|x - c|$ holds. Note that $c = 1$ cannot be an inner point of D simply because it does not belong to D (so no need to consider if there is a δ that would work).

Exercise 1.36 Show that if $\mathbf{c}$ is an inner point of $D \subset \mathbb{R}^n$ then it is also a limit point of D.

With this terminology in hand, we are able to formulate a general definition of the derivative for vector valued functions of several variables.

Definition 1.38: Differentiability

Let $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ be a function and $\mathbf{c}$ an inner point of D. Then, we say that $\mathbf{f}$ is differentiable at $\mathbf{c}$ if there exists an $m \times n$ matrix, that we denote by $\mathbf{Df}(\mathbf{c})$, so that

$$\frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{c}) - \mathbf{Df}(\mathbf{c})(\mathbf{x} - \mathbf{c})\|}{\|\mathbf{x} - \mathbf{c}\|} \xrightarrow[\mathbf{x}\to\mathbf{c}]{} 0.$$

We refer to the matrix $\mathbf{Df}(\mathbf{c})$ as the derivative of $\mathbf{f}$ at $\mathbf{c}$.

Here, we immediately rewrite this definition in terms of the little-oh notation.

Remark 1.39: Little-oh formulation of differentiability

In terms of the little-oh notation, a function $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at an inner point $\mathbf{c} \in D$ if there exists an $m \times n$ matrix $\mathbf{Df}(\mathbf{c})$, so that, for $\mathbf{x}$ close to $\mathbf{c}$, we have

$$\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{c}) - \mathbf{Df}(\mathbf{c})(\mathbf{x} - \mathbf{c}) = o\Big(\|\mathbf{x} - \mathbf{c}\|\Big).$$

Note that this says that the affine function

$$\mathbf{x} \mapsto \mathbf{f}(\mathbf{c}) + \mathbf{Df}(\mathbf{c})(\mathbf{x} - \mathbf{c})$$

is a good approximation of $\mathbf{f}$ for $\mathbf{x}$ near $\mathbf{c}$.

Remark 1.40: Warning.

Be aware that the derivative of functions in the context of calculus of several variables has many names and symbols. Two other popular names for exactly the same object that we define above seem to be the Jacobian and the total derivative. Other symbols include

$$\frac{\partial(f_1, f_2, \ldots, f_m)}{\partial(x_1, x_2, \ldots, x_m)} \quad \text{and} \quad \mathbf{J_f}(\mathbf{u}).$$

An issue with Definition 1.38 is that it does not explicitly tell us what the matrix $\mathbf{Df}(\mathbf{c})$ looks like. It does not even tell us that the derivative is unique (it is!). Our next goal is to formulate a proposition that reveals the formula for the derivative (see Appendix **??** for details of how to prove this). To this end, we first need to define what we mean by a partial derivative.

Definition 1.41

Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a function and $\mathbf{c}$ an inner point of D. By the j'th partial derivative of f at $\mathbf{c}$, we define

$$\partial_j f(\mathbf{c}) = \lim_{x_j \to c_j} \frac{f(\mathbf{c} + x_j \mathbf{e}_j) - f(\mathbf{c})}{x_j - c_j}$$

where $\mathbf{c} = (c_1, c_2, \ldots, c_n)$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{e}_j$ is the j'th unit vector in $\mathbb{R}^m$.

Example 1.42

Computing partial derivatives is just like computing derivatives in one variable calculus. Indeed, let

$$f(x, y) = 4xy + 3x + 2y.$$

To compute the partial deriviative with respect to the first variable (i.e., x), we just need to pretend that all other variables are constants:

$$\partial_1 f(x, y) = 4y + 3.$$

To compute the partial deriviative with respect to the second variable (i.e., y), we just need to pretend that all other variables are constants:

$$\partial_2 f(x, y) = 4x + 2.$$

That's it.

We are now ready to provide the formula for the derivative matrix.

Theorem 1.43

Let $\mathbf{f} = (f_1, f_2, \ldots, f_m) : D \subset \mathbb{R}^n \to \mathbb{R}^m$ and let $\mathbf{c}$ be an inner point of D. Then, if $\mathbf{f}$ is differentiable at $\mathbf{c}$ it follows that

$$\mathbf{Df}(\mathbf{c}) = \begin{pmatrix} \partial_1 f_1(\mathbf{c}) & \partial_2 f_1(\mathbf{c}) & \cdots & \partial_n f_1(\mathbf{c}) \\ \partial_1 f_2(\mathbf{c}) & \partial_2 f_2(\mathbf{c}) & \cdots & \partial_n f_2(\mathbf{c}) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_m(\mathbf{c}) & \partial_2 f_m(\mathbf{c}) & \cdots & \partial_n f_m \mathbf{c}) \end{pmatrix}.$$

Exercise 1.37 Does this theorem say anything about the uniqueness of the derivative?

Exercise 1.38 Can this matrix ever exist without $\mathbf{f}$ being differentiable at $\mathbf{c}$?

The astute reader may notice that we are not completely out of the woods yet, in terms of the theory. Sure, if there is an affine function that approximates $\mathbf{f}$ well, then we call the function differentiable. And if $\mathbf{f}$ differentiable, we know that the derivative is a matrix consisting of partial derivatives. But it does not follow that if the derivative matrix exists, then $\mathbf{f}$ is differentiable. Indeed, there are examples where this is false. This means we need the following theorem to be on steady ground.

Theorem 1.44

Let $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ and let $\mathbf{c}$ be an inner point of D. Then the following are equivalent.

(i) f is continuously differentiable near $\mathbf{c}$ (that is, $\mathbf{x} \mapsto \mathbf{Df}(\mathbf{c})\mathbf{x}$ is continuous).

(ii) f has all partial derivatives near $\mathbf{c}$, and the partial derivatives are all continuous.

Tour of special cases of the derivative: $f : D \subset \mathbb{R} \to \mathbb{R}^m$

We now take a quick tour of some important special cases of differentiability, in order to get a better sense of how to work with the derivative in higher dimensions. The first stop we make

are functions $\mathbf{f} : D \subset \mathbb{R} \to \mathbb{R}^m$. Recall that these are functions of the form

$$\mathbf{f}(x) = \Big(f_1(x), f_2(x), \ldots, f_n(x)\Big),$$

and that we think of them as curves, or better yet, trajectories, of particles in $\mathbb{R}^n$. According to Theorem 1.43, their derivatives are of the form

$$\mathbf{Df}(x) = \begin{pmatrix} f_1'(x) \\ f_2'(x) \\ \vdots \\ f_m'(x) \end{pmatrix}.$$

Example 1.45: Geometric interpretation of the derivative of paths

Let us consider the path $\mathbf{f}(t) = \big(\cos t, \sin t\big)$ in $\mathbb{R}^2$. The curve traced by this path is exactly the unit circle. The derivative at time t is given by

$$\mathbf{Df}(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}.$$

This can be visualised as follows.

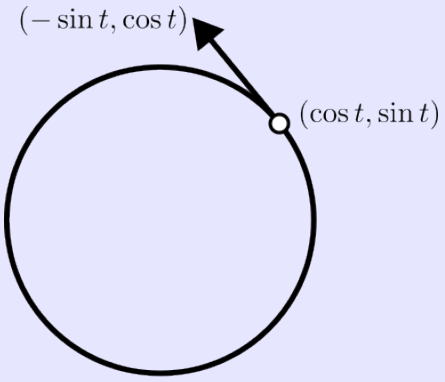

Figure 1.19: As we see, the derivative is always pointing in the direction of travel.

Exercise 1.39 Consider the above example.

(a) Interpret geometrically how the derivative affected if we instead consider

$$\mathbf{f}(t) = \big(\cos(2t), \sin(2t)\big).$$

(b) Interpret geometrically how the derivative affected if we instead consider

$$\mathbf{f}(t) = \big(\cos(\phi(t)), \sin(\phi(t))\big)$$

for some differentiable function $\phi : \mathbb{R} \to \mathbb{R}$.

Exercise 1.40 Prove the formula for the derivative of $f : D \subset \mathbb{R} \to \mathbb{R}^2$ directly from the definition of the derivative. That is, without using Theorem **??**.

Exercise 1.41 In Remark 1.39, it was pointed out that the affine function $\mathbf{x} \mapsto \mathbf{f}(\mathbf{c}) + \mathbf{Df}(\mathbf{c})(\mathbf{x} - \mathbf{c})$ is a good approximation of $\mathbf{f}$ for $\mathbf{x}$ close to $\mathbf{c}$. Write out this affine function in more detail when $f : D \subset \mathbb{R} \to \mathbb{R}^m$.

Tour of special cases of the derivative: $f : D \subset \mathbb{R}^n \to \mathbb{R}$

We now move on to consider the special case $f : D \subset \mathbb{R}^n \to \mathbb{R}$. In this case, the function is on the form

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

From the formula of the derivative, we obtain

$$\mathbf{D}f(\mathbf{x}) = \Big(\partial_1 f(\mathbf{x}), \partial_2 f(\mathbf{x}), \ldots, \partial_n f(\mathbf{x})\Big).$$

This is exactly what we usually call the gradient of f, and denote by

$$\nabla f(\mathbf{x}) = \Big(\partial_1 f(\mathbf{x}), \partial_2 f(\mathbf{x}), \ldots, \partial_n f(\mathbf{x})\Big).$$

The choice of the terminology 'gradient' for this derivative hints at its geometric interpretation.

Example 1.46

Let us consider the example $f(x, y) = x^2 + 2y^2$. The gradient of this function is

$$\nabla f(x, y) = (2x, 4y)\,.$$

Notice that the gradient is a function $\mathbb{R}^2 \to \mathbb{R}^2$, and we can therefore think of it as a vector field. If we visualise it together with some level curves $f(x, y) = c$ (which are ellipses) then we will see that the gradient vector field will always be orthogonal to these level curves. Since a function is constant along its level curves, then if it is differentiable, the rate of change will the largest when we move orthogonally with respect to these curves. That is, if we move in the direction of the gradient.

picture of level curves and gradient field.. or picture of people walking on a mountain?

We point out that both the geometric interpretation of the gradient and the derivative of a path boil down to vectors related to the shape of curves. On the one hand, the derivative of a function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^n$ gives us information about the direction and speed of a particle whose travel is descibed by the path parameterised by $\mathbf{f}$. On the other hand, the gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ tells us the normal vector of the countour lines on a map of the landscape (or surface) parameterised by f.

Tour of special cases of the derivative: $f : D \subset \mathbb{R}^2 \to \mathbb{R}^2$

We now consider a final special case, namely $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$. In this case, the derivative is given by

$$\mathbf{Df}(\mathbf{c}) = \begin{pmatrix} \partial_1 f_1(\mathbf{c}) & \partial_2 f_1(\mathbf{c}) \\ \partial_1 f_2(\mathbf{c}) & \partial_2 f_2(\mathbf{c}) \end{pmatrix}.$$

Note that we can rewrite the derivative using the gradient notation as follows:

$$\mathbf{Df}(\mathbf{c}) = \begin{pmatrix} \nabla f_1(\mathbf{c}) \\ \nabla f_2(\mathbf{c}) \end{pmatrix}.$$

Moreover, the linear approximation of $\mathbf{f}$ can be expressed as

$$\begin{aligned}\mathbf{f}(\mathbf{c}+\mathbf{h}) &= \mathbf{f}(\mathbf{c}) + \mathbf{Df}(\mathbf{c})\mathbf{h} + o(\|\mathbf{h}\|) \\ &= \mathbf{f}(\mathbf{c}) + \begin{pmatrix} \partial_1 f_1(\mathbf{c}) & \partial_2 f_1(\mathbf{c}) \\ \partial_1 f_2(\mathbf{c}) & \partial_2 f_2(\mathbf{c}) \end{pmatrix} \mathbf{h} + o(\|\mathbf{h}\|) \\ &= \begin{pmatrix} f_1(\mathbf{c}) + \partial_1 f_1(\mathbf{c})h_1 + \partial_2 f_1(\mathbf{c})h_2 \\ f_2(\mathbf{c}) + \partial_1 f_2(\mathbf{c})h_1 + \partial_2 f_2(\mathbf{c})h_2 \end{pmatrix} + o(\|vbh\|).\end{aligned}$$

We now investigate two ways to geometrically understand what this tells us.

Example 1.47

Let us consider the function $\mathbf{f}(x, y) = (xy, x + y)$. Its derivative at a point $\mathbf{c} = (a, b)$ is given by

$$\mathbf{Df}(\mathbf{c}) = \begin{pmatrix} b & a \\ 1 & 1 \end{pmatrix}$$

We now try to interpret this derivative in two ways. First, in terms of how $\mathbf{f}$ distorts areas in $\mathbb{R}^2$ and second with respect to $\mathbf{f}$ as a vector field on $\mathbb{R}^2$. No matter the perspective, however, the key point is that for (x, y) close to (a, b), we know that

$$\begin{aligned}\mathbf{f}(x, y) &= \mathbf{f}(a, b) + \mathbf{Df}(a, b) \begin{pmatrix} x - a \\ y - b \end{pmatrix} + o\left(\|(x, y) - (a, b)\|\right) \\ &= \begin{pmatrix} ab \\ a + b \end{pmatrix} + \begin{pmatrix} b & a \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x - a \\ y - b \end{pmatrix} + o\left(\|(x, y) - (a, b)\|\right)\end{aligned}$$

So, let us take the deformation perspective. The above formula tells us that for (x, y) close to (a, b), then $\mathbf{f}$ will act more or less as a matrix for points (x, y) near (a, b). This tells us several things. In particular, the determinant of $\mathbf{Df}(a, b)$ tells us whether this matrix shrinks or expands areas. Close to the point (a, b) this will then also be true for $\mathbf{f}$. Moreover, any matrix will deform circles to ellipses. So, for very small circles centered at (a, b) this will also be approximately true for $\mathbf{f}$. We leave checking the details as an exercise.

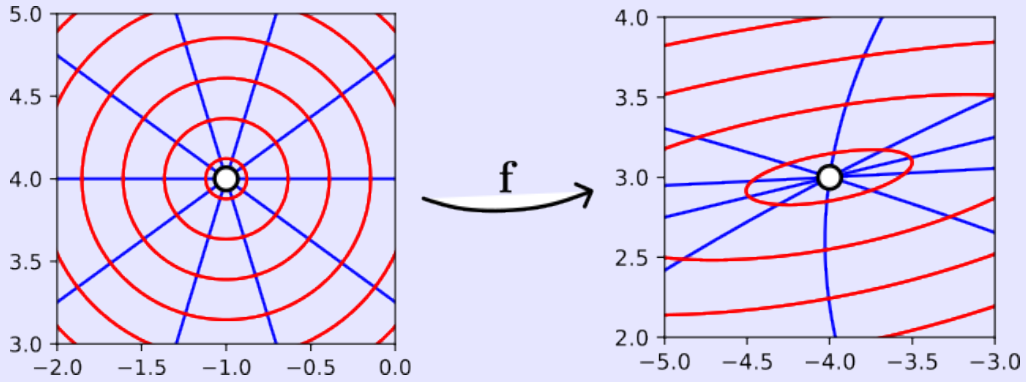

Figure 1.20: Close to (a, b), circles are approximately deformed into ellipses.

We now turn to the vector field perspective. In this perspective, we think of $\mathbf{f}$ as assigning, for each point (x, y), a vector $(u, v) = \mathbf{f}(x, y)$. What does the derivative matrix tell us about this? Well, similar to the deformation perspective above, it tells us how this vector field changes as we consider points close to (a, b). If we think of the vector field as some type of flow of water, or as an electric or magnetic field, then it should be intuitively

plausible that this information can tell us if the field twists, turns, or becomes stronger in any direction seen from the point (a, b). Specifically, the derivative tells us what the divergence and curl of the vector field is locally. (Which is not so surprising, when you start learning about these things.)

Exercise 1.42 Figure out how the derivative matrix from the above example distorts a small circle $h^2 + k^2 = \epsilon^2$. Here, you can try to figure out the following.

(a) How are the two main axes of the circle transformed? What is their new orientation, what is the new angle between them, and what are their new lengths?

(b) What is the proportion of the area of the circle and the ellipse it is sent to?

Hint: Question (a) can be solved by naive geometric considerations. Question (b) can be settled by considering the determinant.

Computational rules for the derivative

We now discuss computational rules for the derivative. In a sense, this part of the section is a bit of an anti-climax, as the only result of any importance we will deal with here is the chain rule (most other computational rules are small potatoes in comparison and are dealt with in the exercises). The two other "big" computational theorems on the derivative in several variables - namely the inverse function theorem and the implicit function theorem - will appear later in a later chapter.

We state the one variable and several variable versions back-to-back for easy comparison.

Proposition 1.48: Chain rule in one variable

Suppose that the functions $g : A \subset \mathbb{R} \to \mathbb{R}$ and $f : B \subset \mathbb{R} \to \mathbb{R}$ are differentiable at inner points $a \in A$ and $g(a) \in B$, respectively.

Then $f \circ g$ is differentiable at a, and

$$\frac{\mathrm{d}}{\mathrm{d}x} f\Big(g(x)\Big) = f'\Big(g(x)\Big) g'(x).$$

Proposition 1.49: Chain rule in several variables

Suppose that the functions $\mathbf{g} : A \subset \mathbb{R}^n \to \mathbb{R}^\ell$ and $\mathbf{f} : B \subset \mathbb{R}^\ell \to \mathbb{R}^m$ are differentiable at inner points $\mathbf{a} \in A$ and $\mathbf{g}(\mathbf{a}) \in B$, respectively.

Then $\mathbf{f} \circ \mathbf{g}$ is differentiable at $\mathbf{a}$, and

$$\mathbf{D}(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = \mathbf{D}\mathbf{f}\Big(\mathbf{g}(\mathbf{a})\Big)\mathbf{D}\mathbf{g}(\mathbf{a})$$

We illustrate the several variable version of the chain rule with an example.

Example 1.50

We consider the functions $\mathbf{g}(x, y) = (2xy, x + y)$ and $\mathbf{f}(u, v) = (\sin(v), \cos(u))$. From the formula

$$\mathbf{f} \circ \mathbf{g}(x, y) = \big(\sin(x + y), \cos(2xy)\big)$$

we can directly compute the derivative, without any use of the chain rule, at a point $(a, b) \in \mathbb{R}^2$:

$$\mathbf{D}(\mathbf{f} \circ \mathbf{g})(a, b) = \begin{pmatrix} \cos(a+b) & \cos(a+b) \\ -2b\sin(ab) & -2a\sin(ab) \end{pmatrix}.$$

Let us now use the chain rule, to see if we get the same result. First, we compute the derivative of $\mathbf{f}$ at a point (u, v), without making the connection $(u, v) = \mathbf{g}(a, b)$ just yet.

$$\mathbf{Df}(u, v) = \begin{pmatrix} 0 & \cos(v) \\ -\sin(u) & 0 \end{pmatrix}.$$

Next, we compute the derivative of $\mathbf{g}$ at a point (a, b).

$$\mathbf{Dg}(a, b) = \begin{pmatrix} 2b & 2a \\ 1 & 1 \end{pmatrix}.$$

We now compute the product of these matrices:

$$\mathbf{Df}(u, v)\mathbf{Dg}(a, b) = \begin{pmatrix} 0 & \cos(v) \\ -\sin(u) & 0 \end{pmatrix} \begin{pmatrix} 2b & 2a \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \cos(v) & \cos(v) \\ -2b\sin(u) & -2a\sin(u) \end{pmatrix}.$$

Finally, letting $(u, v) = (2ab, a + b)$, we see that the result obtained through the chain rule is the same as the one obtained through the direct computation.

Exercise 1.43 State and prove the summation rule for the derivative in several variables.

Exercise 1.44 State and prove the product rule for the derivative in several variables. Here the 'state' part requires a little thinking.

Exercise 1.45 State and prove the product rule for the derivative in several variables. Here the 'state' part requires a little thinking. In the prove part, you can actually use the chain rule (but it is not necessary).

Exercise 1.46 State and prove the reciprocal rule for the derivative in several variables. Again, the 'state' part requires a little thinking. Also again, in the prove part, you can actually use the chain rule (but it is not necessary). Deduce the quotient rule as a consequence.

Exercise 1.47 Suppose that a function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is invertible. Based on this, state and prove a version of the inverse function theorem.

Hint: Take a look at the one variable version of this result as inspiration. Also, note that you are allowed to assume that f is invertible. In the standard formulation of the inverse function theorem, it is not assumed that f is invertible, instead this must also be proven - and this is the hard part, which we will return to later in the course. Here, you are allowed to assume that this is part is known, which leaves the "easiest" part. That is, to deduce a formula for the derivative of the inverse function of $\mathbf{f}$.

1.8 Definite integrals

A quick run-through of the Riemann integral in one variable

The main point of definite integrals in one variable is to allow us to capture the notion of an area under a graph. To this end, we need to define what we mean by partitions, mesh size and what we mean by Riemann sums.

First, we recall that by a partition P of an interval $[a, b]$, we mean a set $P = \{x_0, x_1, \ldots, x_n\}$ so that

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b.$$

The main job of the partition is to split the inteval $[a, b]$ into subintervals $[x_{k-1}, x_k]$ that together cover $[a, b]$. We use the notation $\Delta x_k = x_k - x_{k-1}$ to indicate the width of each subinterval defined by the partition.

For a function $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ we can use the partition P to approximate the area under its graph. To do this, we also need to choose points $c_k \in [x_{k-1}, x_k]$ so that we can compute heights corresponding to each subinterval defined by the partition. Based on the partition and choice of points c_k, we can define the following Riemann sum of f:

$$R(f, P, (c_k)_{k=1}^n) = \sum_{k=1}^{n} f(c_k)\Delta x_k.$$

pictar

Note that if we do not care about the choice of the points c_k, then we ignore these in the notation and just write $R(f, P)$ for the Riemann sum.

As a measure of how well we would expect such a Riemann sum to approximate the area under the graph of a function, we use the mesh size. The mesh size is defined to be the biggest gap between consecutive elements of a partition, and we denote it by $\text{mesh}(P)$. That is, we define

$$\text{mesh}(P) = \max\{\Delta x_1, \Delta x_2, \ldots, \Delta x_n\}.$$

We can now state Riemann's definition.

Definition 1.51: Riemann's definition

We say that a function $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ is Riemann integrable on $[a, b]$ if the following holds. There exists a unique number A so that for all sequences $R_n = R(f, P_n)$ of Riemann sums for f on $[a, b]$ with $\text{mesh}(P_n) \to 0$, we have

$$\lim_{n\to\infty} R(f, P_n) = A.$$

We define this unique number A to be the integral of f over $[a, b]$ and denote it by the symbol

$$\int_a^b f(x)\,\mathrm{d}x.$$

Given this definition, the question becomes: which types of functions are Riemann integrable? Riemann's definition is often considered to unwieldy to approach directly. Therefore, it is common to approach the theory via what we choose to call Darboux' criterion for Riemann integrability. To formulate Darboux' criterion, we need to single out two extreme cases related to Riemann sums.

To this end, we assume that f is a bounded function on $[a, b]$ (this means that the range of f is contained in a finite interval). Given a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$, the

completeness axiom now tells us that the following quantities exist:

$$m_k = \inf\{f(x) : x \in [x_{k-1}, x_k]\} \qquad \text{and} \qquad M_k = \sup\{f(x) : x \in [x_{k-1}, x_k]\}.$$

This allows us to define the lower and upper Riemann sums of f with respect to the partition P, respectively, as follows:

$$L(f, P) = \sum_{k=1}^{n} m_k \Delta x_k \qquad \text{and} \qquad M_k = \sum_{k=1}^{n} M_k \Delta x_k.$$

Darboux' criterion is as follows.

Definition 1.52: Darboux' integrability criterion

Let $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ be a bounded function. Then we say that f satisfies Darboux' integrability criterion if, for all $\epsilon > 0$, there exists $\delta > 0$ such that for all partitions P of $[a, b]$ the following holds.

$$\text{mesh}(P) < \delta \implies U(f, P) - L(f, P) < \epsilon.$$

The main point is that for concrete examples, it is much easier to check Darboux' integrability criterion than to verify that Riemann's definition holds. This means that working with the theory for definite integrals becomes much easier once one has put down the required effort to obtain the following result.

Theorem 1.53: Darboux' theorem

Let $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$. Then f is satisfies Darboux' integrability criterion on $[a, b]$ if and only if f is Riemann integrable on $[a, b]$.

should we prove this somewhere? (not here, though)

The most important class of such functions are the continuous functions on intervals $[a, b]$ (and, in a mostly trivial extension, the finitely piece wise continuous functions).

Theorem 1.54: Continuous functions are Riemann integrable

All continuous functions $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ are Riemann integrable. In particular, this means that for such f, we have

$$\lim_{n\to\infty} \sum_{k=1}^{n} f(c_k)\Delta x_k = \int_a^b f(x)\,\mathrm{d}x$$

whenever the Riemann sums inside the limit on the left-hand are based on a sequence of partitions $P_n = \{x_0, \ldots, x_n\}$ with $\text{mesh}(P_n) \to 0$.

Proof. We do not include the proof, but indicate some highlights. The difficult part of this proof is to show that if f is continuous on an interval $[a, b]$ then f is also uniformly continuous there. We will address uniform continuity in a later chapter (reference location). Once this is done, it is enough to show that uniform continuous functions satisfy Darboux' criterion, which is on the level of a fairly standard epsilon-delta type proof. □

Exercise 1.48 Look up a definition of piece-wise continuous functions somewhere (or suggest a definition yourself). Use this to prove that piece wise continuous functions are Riemann integrable.

Remark: If you come up with a definition and realise that the functions you call piece wise continuous are not integrable, you should modify your definition or add some suitable condition.

How to extend the Riemann integral to rectangles in several variables

We now turn to how the theory of integration is established for functions $f : R \subset \mathbb{R}^n \to \mathbb{R}$ where $R = [a_1, b_1] \times [a_n, b_n]$ is what we call an n-dimensional rectangle. To extend what we did above to this case, we basically just need to follow in the same footsteps, and make sure we come up with good notation along the way (no new ideas are required as long as we only consider such bounded rectangles).

Definition 1.55: Partitions and mesh size in $\mathbb{R}^n$

We say that a set $\mathbf{P} = P_1 \times \cdots \times P_n$ is a partition of $R = [a_1, b_1] \times \cdots \times [a_n, b_n]$ if each P_j is a partition of $[a_j, b_j]$. We define the mesh size of $\mathbf{P}$ to be

$$\text{mesh}(\mathbf{P}) = \text{mesh}(P_1) \cdot \text{mesh}(P_2) \cdots \text{mesh}(P_n).$$

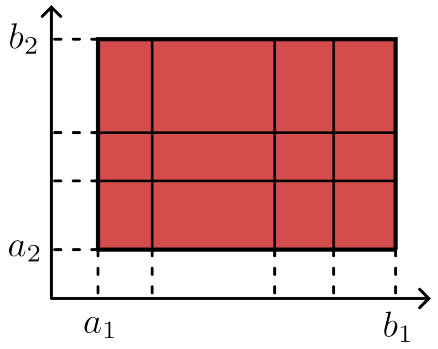

Figure 1.21: In $\mathbb{R}^2$ the mesh size is really a mesh size.

We now define Riemann sums. To keep the notation transparent, we only give this definition for $\mathbb{R}^2$. The extension to $\mathbb{R}^n$ should be conceptually clear, but it requires a bit of notational ugliness that we prefer to avoid.

Definition 1.56

Let $f : R \subset \mathbb{R}^2 \to \mathbb{R}$ be a function, $\mathbf{P} = \{x_0, \ldots, x_n\} \times \{y_0, \ldots, y_m\}$ a partition of R, and $(\mathbf{c}_{i,j})$ points so that $\mathbf{c}_{i,j} \in [x_{i-1}, x_i] \times [y_{j-1}, y_j]$.

Then by the Riemann sum of f with respect to $\mathbf{P}$ and the points $(\mathbf{c}_{i,j})$, we mean the sum

$$R(f, \mathbf{P}, (\mathbf{c}_{i,j})) = \sum_{i=1}^{n} \sum_{j=1}^{m} f(\mathbf{c}_{i,j}) \Delta x_i \Delta y_j.$$

If we do not care about the choice of the points $\mathbf{c}_{i,j}$, we supress these from the notation, and only write $R(f, \mathbf{P})$.

If we assume f to be bounded, we can also define the lower and upper Riemann sum of f with respect to $\mathbf{P}$. These are the sums

$$L(f,\mathbf{P}) = \sum_{i=1}^{n}\sum_{j=1}^{m} m_{i,j}\Delta x_i \Delta x_j$$

$$U(f,\mathbf{P}) = \sum_{i=1}^{n}\sum_{j=1}^{m} M_{i,j}\Delta x_i \Delta x_j,$$

where

$$m_{i,j} = \inf\left\{f(x,y) : (x,y) \in [x_{i-1},x_i]\times[y_{j-1},y_j]\right\}$$
$$M_{i,j} = \sup\left\{f(x,y) : (x,y) \in [x_{i-1},x_i]\times[y_{j-1},y_j]\right\}.$$

picture - things not so easy to visualise now

Exercise 1.49 Find a nice way to express this directly in $\mathbb{R}^n$ (and let me know how).

We now give a slightly more slimmed, but still correct, version of the story that we told for $\mathbb{R}$. Here are the two definitions of integrability.

Definition 1.57: Darboux' integrability criterion for $\mathbb{R}^n$

We say that a bounded function $f : R \subset \mathbb{R}^n \to \mathbb{R}$ satisfies Darboux' integrability criterion if the following holds. For all $\epsilon > 0$, there exist $\delta > 0$ so that for all partitions $\mathbf{P}$ of R, we have

$$\text{mesh}(\mathbf{P}) < \delta \implies U(f,\mathbf{P}) - L(f,\mathbf{P}) < \epsilon.$$

Definition 1.58: Riemann integrability in $\mathbb{R}^n$

We say that a function $f : R \subset \mathbb{R}^n \to \mathbb{R}$ is Riemann integrable if the following holds. There exists a unique number A so that for all sequences $R_n(f,\mathbf{P}_n)$ of Riemann sums of f on R with $\text{mesh}(\mathbf{P}_n) \to 0$, we have

$$\lim_{n\to\infty} R_n(f,\mathbf{P}_n) = A.$$

We define this unique number A to be the integral of f over $[a,b]\times[c,d]$ and denote it by the symbol

$$\int_R f(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

As before, their connection is provided by the following result, which is obtained in the same way.

Theorem 1.59: Darboux' theorem

Let $f : R \subset \mathbb{R}^n \to \mathbb{R}$. Then f is satisfies Darboux' integrability criterion on R if and only if f is Riemann integrable on R.

Theorem 1.60

All continuous functions $f : R \subset \mathbb{R}^n \to \mathbb{R}$ are Riemann integrable. In particular, this means that for such f, we have

$$\lim_{n\to\infty} R(f, \mathbf{P}_n) = \int_R f(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

whenever the Riemann sums inside the limit on the left-hand are based on a sequence of partitions $\mathbf{P}_n$ with $\mathrm{mesh}(\mathbf{P}_n) \to 0$.

How to extend the Riemann integral to Jordan sets in several variables

A difference between integration in $\mathbb{R}$ and $\mathbb{R}^n$ is that we want to integrate functions over a lot of reasonable subsets in $\mathbb{R}^n$ that cannot be immediately expressed in terms of rectangles.

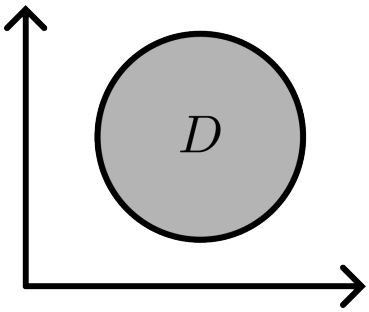

Figure 1.22: This is a disc in $\mathbb{R}^2$, and not a rectangle. Note that in $\mathbb{R}$, both discs and rectangles reduce to intervals.

To integrate over other types of sets in $\mathbb{R}^n$, the trick is to consider these sets as part of a larger rectangle. For this reason, and to make this approach work, we need to assume that these more weird sets that we want to integrate over are bounded.

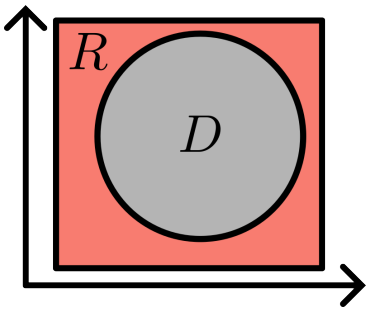

Figure 1.23: By this account, a disc is a weird set – but at least it is contained in a rectangle.

The trick is then to pose the question of what it means to integrate over the disc as a question of whether the so-called indicator function of the disc, considered as a function "living" on this larger rectangle, can be integrated. We make this more precise.

Definition 1.61: Jordan sets

We say that a bounded set $A \subset \mathbb{R}^n$ is a Jordan set if the so-called indicator function

$$\mathbf{1}_A(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in A \\ 0 & \mathbf{x} \notin A \end{cases}$$

is Riemann integrable on some rectangle R containing A.

indicate what is going on via a picture

This motivates the following definition.

Definition 1.62

Let $A \subset \mathbb{R}^n$ be a bounded set and $f : A \subset \mathbb{R}^n \to \mathbb{R}$ a function. Then we say that f is Riemann integrable if the function defined by

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \mathbf{x} \in A \\ 0 & \mathbf{x} \notin A \end{cases}$$

is Riemann integrable on any bounded rectangle R containing A. If this is the case, we define the Riemann integral of f over A by

$$\int_A f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_R \tilde{f}(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

We can now prove the following.

Theorem 1.63

Suppose that $A \subset \mathbb{R}^n$ is a closed and bounded Jordan set and that $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is a continuous function. Then f is Riemann integrable on A.

Proof. We only indicate the proof. The point is to first show that if f is continous on a closed and bounded subset of $\mathbb{R}^2$ then f is also uniformly continuous on $\mathbb{R}^2$. This is actually something we will prove in the course, so forget that part for now. After this, the point is to combine this with the property that A is a Jordan set. This works like a charm, since the boundary of a Jordan set is very small from the point of view of the partitions of a Riemann sum. But this implies exactly that the part of the upper and lower Riemann sum where uniform continuity does not help us is very very small, which allows us to complete the proof. □

Exercise 1.50 Turn the last part of the sketch - where the uniform continuity is combined with the Jordan set property - into a proof.

Exercise 1.51 Could we use the Jordan set approach to integrate functions over more complicated sets in $\mathbb{R}$ than intervals? See if you can find some.

How to extend the Riemann integral to vector valued functions

We have now considered how to integrate functions $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ and functions $f : R \subset \mathbb{R}^n \to \mathbb{R}$. But what about an extension to the case $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$? Well, there is basically

nothing to discuss from the point of view of what functions we can integrate. Indeed, once we have established the theory for subsets $A \subset \mathbb{R}^n$, then the extension to functions whose range are in $\mathbb{R}^m$ is just a matter of component-wise application. To keep notation simple, let us consider the case $f : [a, b] \subset \mathbb{R} \to \mathbb{R}^2$. That is, we have

$$\mathbf{f}(x) = (f_1(x), f_2(x)),$$

Then we just define

$$\int_a^b \mathbf{f}(x) = \Big(\int_a^b f_1(x)\,\mathrm{d}x, \int_a^b f_2(x)\,\mathrm{d}x \Big).$$

That is, the question of integrating $\mathbf{f}$ reduces to the more classical question of integrating the functions $f_1, f_2 : [a, b] \subset\to \mathbb{R}$.

Computational rules for the Riemann integral in several variables

We now formulate some "basic" computational rules that apply to integrals of functions $f : R \subset \mathbb{R}^n \to \mathbb{R}$.

Proposition 1.64: Basic computational rules

Suppose that $f, g : A \subset \mathbb{R}^n \to \mathbb{R}$ are Riemann integrable functions and $\lambda \in \mathbb{R}$ a constant. Then

$$\int_A f(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_A g(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_A \Big(f(\mathbf{x}) + g(\mathbf{x}) \Big)\,\mathrm{d}\mathbf{x}$$

$$\int_A \lambda\, f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \lambda \int_A f(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

Moreover, if $f \leq g$ on A, then

$$\int_A f(\mathbf{x})\,\mathrm{d}\mathbf{x} \leq \int_A g(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

The triangle inequality holds:

$$\Big| \int_A f(\mathbf{x})\,\mathrm{d}\mathbf{x} \Big| \leq \int_A |f(\mathbf{x})|\,\mathrm{d}\mathbf{x}.$$

Suppose that $B_1, B_2 \subset A$ are disjoint and closed Jordan sets. Then

$$\int_{B_1} f(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{B_2} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{B_1 \cup B_2} f(\mathbf{x})\,\mathrm{d}x.$$

Suppose that J is the boundary of a closed and bounded Jordan set. Then

$$\int_J f(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0.$$

When we consider 'how' to compute integrals, the details depend more on whether or not you are in the several or single variable setting. In the single variable setting, we have the following familiar rules.

Proposition 1.65

Suppose that $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ is continuous. Then

(i) Mean value theorem: There exist $c \in [a, b]$ such that

$$\int_a^b f(x)\,\mathrm{d}x = f(c)(b-a).$$

(ii) Fundamental theorem of Calculus: Suppose that $c, x \in [a, b]$, then

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_c^x f(t)\,\mathrm{d}t = f(x).$$

(iii) Evaluation formula: Suppose that F is any primitive function of f, then

$$\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a).$$

(iv) Change of variables formula: Let φ be a differentiable function, then

$$\int_a^b f(\varphi(x))\varphi'(x)\,\mathrm{d}x = \int_{\varphi(a)}^{\varphi(b)} f(u)\,\mathrm{d}u.$$

(v) Integration by parts formula: Suppose that $f, g : [a, b] \subset \mathbb{R} \to \mathbb{R}$ are continuously differentiable functions, then

$$\int_a^b f(x)g'(x)\,\mathrm{d}x = \Big[f(x)g(x)\Big]_a^b - \int_a^b f'(x)g(x)\,\mathrm{d}x.$$

But what formulas hold for $f : A \subset \mathbb{R}^n \to \mathbb{R}$? This question depends heavily on how much of the several variables theory you have had the time to encounter. Here, we assume that you have not seen results such as Green's theorem, Stokes theorem and Gauss' theorem, that extend, somehow, the notion of integration by parts and the fundamental theorem of calculus to functions on $\mathbb{R}^n$. Instead, we focus on more "basic" parts of the several variables theory.

First, we formulate Fubini's theorem, which is useful for reducing the question of computign integrals over rectangles, to the question of computing them on intervals. We formulate it for continuous functions, in order to avoid some technicalities.

Proposition 1.66: Fubini's theorem for continuous functions

Suppose that $f : [a, b] \times [c, d] \subset \mathbb{R}^2 \to \mathbb{R}$ is continuous. Then

$$\iint_{[a,b]\times[c,d]} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int_a^b \Big(\int_c^d f(x, y)\,\mathrm{d}y\Big)\,\mathrm{d}x = \int_c^d \Big(\int_a^b f(x, y)\,\mathrm{d}x\Big)\,\mathrm{d}y.$$

Example 1.67

Consider the function $f(x, y) = xy^2$ on the rectangle $[0, 1]^2$. To compute its integral, we

use Fubini:

$$\begin{aligned}\iint_{[0,1]^2} xy^2 \,\mathrm{d}x\,\mathrm{d}y &= \int_0^1 \Big(\int_0^1 xy^2 \,\mathrm{d}y\Big)\,\mathrm{d}x \\ &= \int_0^1 \Big(\int_0^1 xy^2 \,\mathrm{d}y\Big)\,\mathrm{d}x \\ &= \int_0^1 \Big[\frac{xy^3}{3}\Big]_{y=0}^{y=1} \mathrm{d}x \\ &= \int_0^1 \frac{x}{3}\,\mathrm{d}x = \Big[\frac{x^2}{6}\Big]_{x=0}^{x=1} = \frac{1}{6}.\end{aligned}$$

But how to compute integrals that are not defined on rectangles. This can be approached in essentially two ways, if the goals is to compute the integral exactly. The first is to attempt to describe the area of integration in terms of inequalities. Instead of attempting a general description of this approach (not sure if any useful such description is possible), we illustrate this by an example.

Example 1.68

Let us compute the integral of $f(x,y) = xy^2$ again, but this time over the first qudrant of the unit disk. There are several ways to do this, but one is to recognise that this set, which we denote by S, can be described by the inequalities

$$\begin{aligned}0 \leq& y \leq 1 \\ 0 \leq& x \leq \sqrt{1-y^2}\end{aligned}$$

This means that we can compute as follows:

$$\begin{aligned}\iint_S f(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \int_0^1 \left(\int_0^{\sqrt{1-y^2}} xy^2 \,\mathrm{d}x\right)\,\mathrm{d}y \\ &= \int_0^1 \Big[\frac{x^2y^2}{2}\Big]_{x=0}^{x=\sqrt{1-y^2}}\,\mathrm{d}y \\ &= \int_0^1 \frac{(1-y^2)y^2}{2}\,\mathrm{d}y \\ &= \int_0^1 \frac{(y^2-y^4)}{2}\,\mathrm{d}y = \Big[\frac{y^3}{6} - \frac{y^5}{10}\Big]_{y=0}^{y=1} = \frac{1}{6} - \frac{1}{10} = \frac{1}{15}.\end{aligned}$$

A second approach to integration over non-rectangular areas (or curves) is to parameterise them in terms of some other function, and then make a change of variables. We set the stage for this by a simple example.

Example 1.69

Let us again compute $f(x,y) = xy^2$ over the first quadrant of the unit disk, which we denote by S. But this time, we let us parameterise it using polar coordinates. That is,

we introduce the function

$$\mathbf{\Phi}(r,\theta) = (r\cos\theta, r\sin\theta).$$

Note that in polar coordinates, the first quadrant of the unit circle is described by $r \in [0,1]$ and $\theta \in [0,\pi/2]$. But this means that $\mathbf{\Phi}$ maps the square $[0,1]\times[0,\pi/2]$ onto S. This means that the function

$$\mathbf{f}\circ\mathbf{\Phi}$$

is a version of $\mathbf{f}$ defined on a rectangle. But how to connect the integral of the composition $\mathbf{f}\circ\mathbf{\Phi}$ over a this rectangle to the integral of $\mathbf{f}$ on the first quadrant of the unit circle? We pause the example to address this question.

Well, the point is that we need a change of variables formula, and here it is.

Proposition 1.70: Change of variables formula

Let $f : A \subset \mathbb{R}^n \to \mathbb{R}$ be a Riemann integrable function, and $\mathbf{\Phi} : R \subset \mathbb{R}^n \to A \subset \mathbb{R}^n$ an invertible and continuously differentiable function on a bounded rectangle R so that $|\mathbf{D\Phi}(\mathbf{u})| \neq 0$. Then

$$\int_A f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_R \Big(f\circ\mathbf{\Phi}(\mathbf{u})\Big)|\mathbf{D\Phi}(\mathbf{u})|\,\mathrm{d}\mathbf{u},$$

where $|\mathbf{D\Phi}(\mathbf{u})|$ denotes the determinant of derivative of $\mathbf{\Phi}$.

With this result in hand, we continue the example.

Example 1.71

We now return to the example we considered above. To use the change of variables formula, we first compute the derivative of $\mathbf{\Phi}(r,\theta) = (r\cos\theta, r\sin\theta)$:

$$\mathbf{D\Phi}(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}.$$

Next, we compute the determinant of the derivative (indeed, the determinant measures how much this function shrinks or expands area):

$$|\mathbf{D\Phi}(r,\theta)| = r\cos^2\theta + r\sin^2\theta = r.$$

The change of variables formula now gives us:

$$\begin{aligned}
\iint_S f(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \iint_{[0,1]\times[0,\frac{\pi}{2}]} f(r\cos\theta, r\sin\theta) r\,\mathrm{d}r\,\mathrm{d}\theta \\
&= \int_0^{\frac{\pi}{2}} \left(\int_0^1 r^4\cos\theta\sin^2\theta\,\mathrm{d}r\right)\mathrm{d}\theta \\
&= \int_0^{\frac{\pi}{2}} \left(\left[\frac{r^5\cos\theta\sin^2\theta}{5}\right]_{r=0}^{r=1}\right)\mathrm{d}\theta \\
&= \int_0^{\frac{\pi}{2}} \frac{\cos\theta\sin^2\theta}{5}\,\mathrm{d}\theta.
\end{aligned}$$

At this point, we make the u substitution $u = \sin\theta$ with $\mathrm{d}u = \cos\theta\,\mathrm{d}\theta$, to get

$$\int_0^{\frac{\pi}{2}} \frac{\cos\theta \sin^2\theta}{5}\,\mathrm{d}\theta = \int_0^1 \frac{u^2}{5}\,\mathrm{d}u = \left[\frac{u^3}{15}\right]_{u=0}^{u=1} = \frac{1}{15}.$$

Done!

Remark 1.72

The astute reader may have noticed that the situation in the above example does not seem to be covered by the change of variables result since Φ is not invertible when defined on $[0,1]\times[0,\pi/2]$, nor is its determinant non-zero on this rectangle (r is allowed to be zero). However, what saves the day is that we do not need to work on the closed rectangle, it is enough to work on the open rectangle $(0,1)\times(0,\pi/2)$. Indeed, the theorem does not require the rectangle R to be formed by closed intervals.

exercises?

We end this section with some remarks on curve and surface integrals. These are very helpful questions to reflect on as they bring out some important features of the theory discussed above, and how the analysis part combines nicely with the linear algebra part.

First, recall that in one variable analysis we cannot use the logic of Riemann sums directly to compute the so-called arclength (or just length) of graphs. Indeed, Riemann sums are about area and length is about, well length. The most naive way to compute the length of a graph is to approximate it by the "top lines" of Riemann sum rectangles, but this ends up in disaster.

picture
Caption: Left v right. Approximating with flat lines v diagonal lines.

This also happens in several variables. Indeed, it may seem tempting to try to compute, say, the length of a parameterised curve $\mathbf{f}:[0,1]\to\mathbb{R}^2$, using its parameterisation and the change of variables formula. Or, for that matter, to use the change of variables formula in combination with a parameterisation to compute the area of a parameterised two-dimensional surface $\mathbf{f}:[0,1]^2\subset\mathbb{R}^3$. However, this is problematic because we are dealing with three 'completely' different situations:

1. The change of variables formula is about clevely using how the determinant of $\mathbf{Df}(\mathbf{u})$ captures distortions of length, area or volume in $\mathbb{R}^n$. That is $\mathbf{f}$ maps $\mathbb{R}^n$ into itself, and the question is how this map, locally (since we need to make this work with Riemann sums) distorts n-dimensional space.

2. If we want to use this to make a change of variables formula based on $\mathbf{Df}(t)$ that we can apply to parameterised paths $\mathbf{f}:[0,1]\to\mathbb{R}^n$, then we are faced with the following fundamental problem. The derivative $\mathbf{Df}(t)$ at a point t is a 2×1 matrix. Such matrices do not have determinants. The problem is not to figure out how length is distorted when mapped from $\mathbb{R}$ to $\mathbb{R}$, but rather how length is distorted when mapped from $\mathbb{R}$ to some, potentially, complicated geometric object inside of $\mathbb{R}^2$.

3. Similarly, any change of variables formula that hopes to leverage $\mathbf{Df}(\mathbf{u})$ to compute integrals of parameterised surfaces, say, $\mathbf{f}:[0,1]^2\to\mathbb{R}^3$ faces the same problems. First of all, the determinant makes no sense on 3×2 matrices. Second, this is not about how area is distorted by an $\mathbb{R}^2\to\mathbb{R}^2$, map but rather how area is distorted when mapped from $\mathbb{R}^2$ into a potentially complicated portion of $\mathbb{R}^3$.

The above discussion reflects the discussino we started out with, above, of how to compute the length of a curve in one variable analysis. There are "naive" approximations that do not work, and then there are less naive ones that actually do. So let us briefly consider how we can find changes of variable formulas that work for parameterised curves and surfaces. As the solutions to both situations are just common sense, we explain them in two examples, followed by the general statements.

Example 1.73

Let us consider how to compute the length of the parameterised curve $\mathbf{f}(t) = (\cos \pi t, \sin \pi t)$ for $t \in [0, 1]$. Let us use the "travelling particle" analogy for all its worth to figure out what is going on. First of all, $\mathbf{f}$ parameterises a particle travelling along the top half of the unit circle. Let us compute the derivative of this parameterised curve:

$$\mathbf{Df}(t) = (-\pi \sin \pi t, \pi \cos \pi t) = \pi(-\sin \pi t, \cos \pi t).$$

This is a vector that tells us the direction of travel and velocity[a] of the particle at any given time t. So, if the analogy is good, then on a very small time interval $[t, t + \Delta t]$, we expect the distance travelled by the particle, which is traditionally denoted by $\Delta \mathbf{s}$, to be close to $\Delta t \cdot \|\mathbf{Df}(t)\|$. Summing up these quantities for every subinterval of some partition of $[0, 1]$, and then letting the mesh size tend to 0, we obtain that the length should be given by

$$\int_\gamma \mathrm{d}\mathbf{s} = \int_0^1 \|\mathbf{Df}(t)\| \, \mathrm{d}t.$$

We interpret this integral as an integral 'along the curve'. Here, we use γ to denote the curve 'itself' (that is, the image of $\mathbf{f}$).

[a] Speed is usually defined as the direction and length of the derivative vector of a parameterised path. The length of the derivative, that is, $\|\mathbf{Df}(t)\|$ is usually referred to as the velocity of the particle travelling along the path described by $\mathbf{f}$ at time t.

Exercise 1.52 Finish the above example by computing the norm and integral.

Exercise 1.53 Let us extend the above example as follows. Instead of the path representing the path of a travelling particle, let us think of it as describing the placement of a fence built of wood. We denote the curve describing this placement by γ (this is the range of $\mathbf{f}$). For $(x, y) \in \gamma$, let $h(x, y)$ denote the height of the fence at this point. Since the height is a bit uneven, we are asked to compute an integral to determine its area (so that we can buy paint, I guess). By modifying the above example slightly, figure out how much area of the fence corresponds to the "time interval" Δ. Use this to find a formula for the integral of h along the curve. That is, find a formula for the curve integral

$$\int_\gamma h \, \mathrm{d}\mathbf{s}.$$

Picture.

Exercise 1.54 It is possible to prove, under certain conditions, that such curve integrals $\int_\gamma h \, \mathrm{d}\mathbf{s}$ are independent of the parameterised curve used to compute it. Apparently the textbook Adams and Essex botches this statement. If you have the book, you can try to figure out what the problem is. Formulate a correct statement and then find a proof for

it.

Remark: It may be that this has been corrected in the newest edition. We need to check this :-)

We now turn to the example of parameterised surfaces.

Example 1.74

Let us now consider how to compute the area of the graph of, say, the surface of the graph $\mathbf{f}(x,y) = x^2 y$ for $x, y \in [0,1]$. Let us denote this surface by S. It can be parameterised by defining a function $\mathbf{\Phi} : [0,1]^2 \to \mathbb{R}^3$ by $\Phi(u,v) = (u,v,f(u,v))$. Note that Φ is one-to-one and continously differentiable, with derivative

$$\mathbf{D\Phi}(u,v) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \partial_1 f(u,v) & \partial_2 f(u,v) \end{pmatrix}.$$

So, what is going on? The point is basically this, if we draw a grid in $[0,1]^2$, parallell with the coordinate axes of $\mathbb{R}^2$, then this grid will be mapped, by $\mathbf{\Phi}$, to a grid on the surface S. If we can figure out how the squares formed by the grid on $[0,1]^2$ is deformed by Φ when mapped onto S, we should be in good shape.

But how to do this? Well, consider a point $(u,v) \in [0,1]^2$, and assume that the gridlines are so close together, that $\mathbf{D\Phi}(u,v)$ approximates $\mathbf{f}$ well close to (u,v). Next, we can consider these gridlines as orthogonal linear paths that meet at (u,v). That is, as the paths

$$t \mapsto (t,v) \qquad \text{and} \qquad t \mapsto (u,t).$$

This is not really necessary, but we can now use our intuition from the example of how to deal with curves, to notice that if we differentiate $\mathbf{\Phi}$ in the direction of the first path at (u,v), then we get the distortion of the portion of the grid line between (u,v) and $(u+\Delta t, v)$. And if we differentiate $\mathbf{\Phi}$ in the direction of the second path at (u,v), we will get the distortion of the grid line between (u,v) and $(u, v+\Delta t)$.

image of this

The nice thing is that from $\mathbf{\Phi}$ we do not only get the 'size' of the distortion in each of the direction, but we can also quite easily get the distortion in terms of a vector with both direction and length. To do this, we let $\mathbf{e}_1 = (1,0)$ and $\mathbf{e}_2 = (0,1)$ denote the coordinate unit vectors, and then compute

$$\mathbf{Df}(u,b)\mathbf{e}_1 = \mathbf{Df}(u,b)\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \partial_1 f(u,v) \end{pmatrix}$$

and

$$\mathbf{Df}(u,b)\mathbf{e}_2 = \mathbf{Df}(u,b)\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ \partial_2 f(u,v) \end{pmatrix}.$$

picture of this

And now comes the super nice thing. So, the sides of this little square of the grid on $[0,1]^2$ are distorted to the vectors we just computed. But how on earth can we compute

and compare the area of the little square of the grid on $[0,1]^2$ to the area spanned by the two vectors we just computed (which should be very close to describing the square that this little squre is mapped to on S)? Well, fortunately, we can shut off our brains, and just rely on linear algebra, because the area spanned by to vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is exactly equal to the norm of their cross product. That is, the size of the a small square at (u,v) gets the size:

$$\Delta\mathbf{S} = \|(1,0,\partial_1 f(u,v)) \times (0,1,\partial_2 f(u,v))\|\Delta u \Delta v.$$

This means that the surface area is given by

$$\int_S \mathrm{d}\mathbf{S} = \int_S \|\mathbf{Df}(u,v)\mathbf{e}_1 \times \mathbf{Df}(u,v)\mathbf{e}_2\| \,\mathrm{d}u\,\mathrm{d}v.$$

We leave the computation of the right-hand expression as an exercise.

The above examples lead us to making the following definitions (not theorems!).

Definition 1.75: Curve integrals

Let $\mathbf{c} : I \to \mathbb{R}^n$ be a one-to-one mapping that is continuously differentiable on a bounded interval I, and let γ be the curve traced out by $\mathbf{c}$. If $f : \gamma \to \mathbb{R}$ is continuous, then the curve integral of f over γ is defined to be

$$\int_\gamma f \,\mathrm{d}\mathbf{s} = \int_0^1 \Big(f \circ \mathbf{c}(t)\Big) \|\mathbf{Dc}(t)\| \,\mathrm{d}t.$$

Definition 1.76: Surface integrals

Let $\mathbf{\Phi} : R \subset \mathbb{R}^2 \to \mathbb{R}^3$ be a one-to-one mapping that is continuously differentiable on a bounded rectangle R, and let S be the surface traced out by $\mathbf{\Phi}$. If $f : S \to \mathbb{R}$ is continuous, then the surface integral of f is given by

$$\iint_S f \,\mathrm{d}\mathbf{S} = \iint_D \Big(f \circ \mathbf{\Phi}(u,v)\Big) \|\mathbf{Df}(u,v)\mathbf{e}_1 \times \mathbf{Df}(u,v)\mathbf{e}_2\| \,\mathrm{d}u\,\mathrm{d}v$$

Exercise 1.55 Finish up the computations of Example 1.74

Exercise 1.56 Come up with a geometric object that you already know the length of. Parameterise it, and the use the definition of curve integrals to compute its length area to verify that the definition makes sense.

Exercise 1.57 Come up with a geometric object that you already know the surface area of. Parameterise it, and the use the definition of surface integrals to compute its surface area to verify that the definition makes sense.

Exercise 1.58 Use the definition of surface integrals to deduce the integration formula for the surface of a body of revolution that is sometimes covered in one variable calculus courses.

Exercise 1.59 Exercise 1.53, above, was meant to motivate how to move from thinking about curve integrals as computing the length of a curve, to thinking about the integral of a function along the curve. Try to come up with a motivational example that motivates the same transition from thinking of surface integrals as computing the area of a surface, to thinking of surface integrals as integrating a function over the surface.