MA271 - Mathematical Analysis 3

Instructor: Doctor Vedran Sohinger
Lecture notes written by Doctor Mario Micallef and Profesor Jose Rodrigo.

April 27, 2024

Contents

Chapter 1

Introduction

MA271, Mathematical Analysis 3 is a self-contained module. Only material lectured in class will be examined. Non-examinable will be clearly marked.

The course covers the following topics:

- Pointwise and uniform convergence (sequences and series of functions).
- Differentiation.
- Complex valued functions.

The module does not follow any specific source. There references at the end of the notes cover most of the topics in the module. I would be happy to supply a list of references that can be used to expand any of the Chapters in the notes.

The course relies heavily on material covered in first-year analysis. We recall several of the main notions that we will use.

- The triangle inequality
- Sequences and convergence
- Subsequences and The Bolzano-Weierstrass Theorem
- Cauchy sequences
- Summation
- Basic properties of power series
- The continuity of power series
- The derivative
- The diferentiability of power series
- The radius of convergence formula.
- The Riemann integral, construction and basic properties
- Uniform continuity.

Below is a brief summary of the main results.

1.1 Review limits of sequences

In first-year analysis you have studied the convergence of sequences of real numbers (Chapter 2). Here is a quick recap.

Definition 1.1. *A sequence (a_n) converges to a limit $l \in \mathbb{R}$ if for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that*

$$|a_n - l| < \varepsilon \qquad \text{for every } n \geq N.$$

In this case we write $a_n \to l$ as $n \to \infty$.

The following results were covered in first-year analysis.

- Uniqueness of limits: A sequence can have at most one limit.
- The shift rule: For any fixed $k \in N, a_n \to l$ as $n \to \infty$ if and only if $a_{n+k} \to l$ as $n \to \infty$. (This effectively means that for the question of convergence, we can disregard the first k terms of the sequence (for any fixed k we like).
- Convergent sequences are bounded: Any convergent sequence is bounded.
- If $a_n \to a$ then $|a_n| \to |a|$.
- The basic algebra of limits: Suppose $a_n \to a$ and $b_n \to b$. Then
  (i) $a_n + b_n \to a + b$;
  (ii) $a_n b_n \to ab$;
  (iii) if $b \neq 0$ then $a_n/b_n \to a/b$.
- Limits and inequalities: If $a_n \leq b_n$ for all n, $a_n \to a$ and $b_n \to b$ then $a \leq b$.
- Sandwich rule: $a_n \leq b_n \leq c_n$ with $a_n \to l$ and $c_n \to l$ implies that $b_n \to l$.

1.2 Review of continuity and differentiability

We review the notion of continuity (Chapter 5 in MA141) and *uniform continuity* (covered in Chapter 10 in MA141, and page 71 in MA139). Let $\Omega \subset \mathbb{R}$.

Definition 1.2. *Given $f : \Omega \to \mathbb{R}$, we say that f is continuous at $x \in \Omega$ if for every $\varepsilon > 0$ there exists $\delta = \delta(x, \varepsilon) > 0$ such that*

$$y \in \Omega \text{ and } |x - y| < \delta \Longrightarrow |f(y) - f(x)| < \varepsilon. \tag{1.1}$$

The key point to note from the definition above is that given a function f, $\varepsilon > 0$ and a point x there exists δ, but δ can depend on ε and x (and of course f).

Definition 1.3. *Given $f : \Omega \to \mathbb{R}$, we say that f is uniformly continuous if for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that*

$$x, y \in \Omega \text{ and } |x - y| < \delta \Longrightarrow |f(y) - f(x)| < \varepsilon. \tag{1.2}$$

The key point here is that δ can be chosen independently of x.

In the case in which $\Omega = [a, b]$ we have the following result.

Theorem 1.4. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then it is uniformly continuous.*

Before we prove the result let's consider a couple of examples in which the closed, bounded interval $[a, b]$ is replaced by an unbounded or an open domain.

Consider $f(x) = e^x$, defined in $\mathbb{R}$. Clearly this is a continuous function, but not uniformly continuous. Indeed, since f grows faster and faster for larger x it is possible to find arbitrarily small intervals in which f changes by at least ε. This example shows that the result in Theorem 1.4 is not necessarily true for unbounded domains.

We can also consider $g(x) = \frac{1}{x}$ on $(0, 1)$. Just as in the previous example, near zero, the function g grows to infinity faster and faster as we approach the origin, making it impossible to find δ independent of x that satisfies (1.2).

This result, in much more generality, not just for closed intervals on $\mathbb{R}$ will be proven in MA260 Norms, Metrics and Topologies. The key point is that the domain of f is a *compact* set (which in this case is equivalent to closed and bounded).

Proof of Theorem 1.4. We will argue by contradiction. That would mean that there exist $\varepsilon > 0$ and x_n, y_n such that $|x_n - y_n| \leq \frac{1}{n}$ but $|f(x_n) - f(y_n)| > \varepsilon$.

The sequences $\{x_n\}$ and $\{y_n\}$ are bounded, as they are in $[a, b]$, and therefore we can apply the Bolzano–Weierstrass theorem to obtain convergent subsequences $\{x_{n_k}\}_{k=1}^{\infty}$ to x and $\{y_{n_k}\}_{k=1}^{\infty}$ to y. Notice that

$$|x - y_{n_k}| \leq |x - x_{n_k}| + |x_{n_k} - y_{n_k}| \leq |x - x_{n_k}| + \frac{1}{n_k} \xrightarrow[k\to\infty]{} 0,$$

which implies that $x = y$. However we know that $|f(x_{n_k}) - f(y_{n_k})| > \varepsilon$ for all k. Since f is continuous, taking limits as k goes to infinity we obtain $0 = |f(x) - f(x)| > \varepsilon$, which is a contradiction. □

For completeness we reproduce the definition of derivative.

Definition 1.5. *Suppose $f : I \to \mathbb{R}$ is defined on an open interval I and $c \in I$. We say that f is differentiable at c if*

$$\lim_{h\to 0} \frac{f(c + h) - f(c)}{h}$$

exists. If so we call the limit $f'(c)$.

Also from that module

Lemma 1.6. *Suppose I is an open interval, $f : I \to \mathbb{R}$ and $c \in I$. Then f is differentiable at c if and only if there exists a number A and a function ε with the properties that for all x*

$$f(x) - f(c) = A(x - c) + \varepsilon(x)(x - c),$$

$\varepsilon(c) = 0$ and ε is continuous at c: ($\varepsilon(x) \to 0$ as $x \to c$). If that happens $A = f'(c)$.

1.3 Review of integration

To construct the integral on an interval $[a, b]$ we consider partitions of the interval.

In practice, a partition of the interval $[a, b]$ is determined by a collection of points $\{x_i\}_{i=0}^n$, for some n such that

$$a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b,$$

which yields the collection of intervals $I_j = [x_{j-1}, x_j]$, for $j = 1, \ldots, n$. Given a partition of $P = \{I_1, \ldots, I_n\}$ of $I = [a, b]$ we denote

$$M = \sup_I f \qquad m = \inf_I f \qquad M_k = \sup_{I_k} f \qquad m_k = \inf_{I_k} f.$$

Definition 1.7. *Given $f : [a,b] \to \mathbb{R}$ and a partition $P = \{I_1, \dots, I_n\}$ of $[a,b]$ we define the upper Riemann sum of f with respect to P as*

$$U(f,P) := \sum_{k=1}^{n} M_k |I_k|,$$

and the lower Riemann sum of f with respect to P as

$$L(f,P) := \sum_{k=1}^{n} m_k |I_k|.$$

Figure 1.1 shows the intuitive idea for calculating an integral, displaying the Lower and Upper Riemann sums, for a uniform partition with 10 intervals (for $f(x) = x^2$).

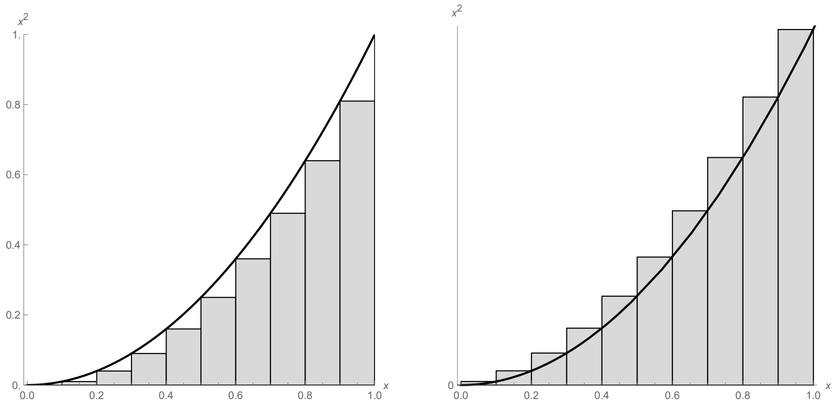

Figure 1.1: Lower (left) and Upper (right) Riemann sum of f

We will denote by $\mathscr{P}$ the set of all partitions of $[a,b]$.

Definition 1.8. *Given $f : [a,b] \to \mathbb{R}$, bounded, we define the upper Riemann integral of f by*

$$U(f) := \inf_{P \in \mathscr{P}} U(f,P).$$

We define the lower Riemann integral of f by

$$L(f) := \sup_{P \in \mathscr{P}} L(f,P).$$

Definition 1.9. *Given $f : [a,b] \to \mathbb{R}$ bounded we say that it is Riemann integrable if and only if $L(f) = U(f)$, and define its Riemann integral, denoted by $\int_a^b f(x)\mathrm{d}x$ or $\int_a^b f$, by*

$$\int_a^b f(x)\mathrm{d}x := L(f) = U(f).$$

The following result will also prove useful in showing that a function is integrable.

Theorem 1.10. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then f is integrable if and only if for every $\varepsilon > 0$ there exists a partition P of $[a,b]$ such that*

$$U(f,P) - L(f,P) < \varepsilon$$

Theorem 1.11. *Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then it is Riemann integrable.*

Theorem 1.12. *Left $f, g : [a,b] \to \mathbb{R}$ be Riemann integrable functions, and $c \in \mathbb{R}$. Then $f + g$ and cf are Riemann integrable and we have*

$$\int_a^b cf = c\int_a^b f, \qquad \int_a^b (f+g) = \int_a^b f + \int_a^b g.$$

Theorem 1.13. *Let $f, g : [a, b] \to \mathbb{R}$ be Riemann integrable functions such that $f \leq g$. Then*

$$\int_a^b f \leq \int_a^b g.$$

Theorem 1.14. *Let $f : [a, b] \to \mathbb{R}$ be an integrable function. Then $|f|$ is integrable and we have*

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

The Fundamental Theorem of Calculus explores the relationship between integration and differentiation, and how under sufficient conditions they can be understood as inverse operations. The first result we consider is when the integral of a derivative is the original function.

Theorem 1.15. *Let $F : [a, b] \to \mathbb{R}$ be a continuous function that is differentiable on (a, b) with $F' = f$. Assume that $f : [a, b] \to \mathbb{R}$ is an integrable function. Then*

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

Theorem 1.16. *Let $f : [a, b] \to \mathbb{R}$ be an integrable function and define the function $F : [a, b] \to \mathbb{R}$ by*

$$F(x) := \int_a^x f(t)\mathrm{d}t.$$

Then F is continuous of $[a, b]$. Additionally if f is continuous at $c \in [a, b]$ then $F'(c) = f(c)$, with the derivatives at a and b understood as one-sided derivatives.

Chapter 2

Sequences and Series of Functions

In this Chapter we will consider sequences and series of functions and aspects relating to pointwise and uniform convergence and its interactions with continuity, integrability and differentiability questions.

2.1 Pointwise convergence

We will consider sequences of functions $f_n : \Omega \to \mathbb{R}$ from a fixed domain Ω. Here we do not make any assumptions about Ω, i.e. being open or closed, bounded or unbounded for example. While most examples will be in one dimension, unless otherwise noted they apply to higher dimensions. We start by defining pointwise convergence.

Definition 2.1. *Let $(f_n)_{n=1}^{\infty}$ be a sequence of functions, with $f_n : \Omega \to \mathbb{R}$. We say that (f_n) or f_n converges pointwise to $f : \Omega \to \mathbb{R}$ if and only if for every $x \in \Omega$ we have $\lim_{n\to\infty} f_n(x) = f(x)$. We will denote pointwise convergence by $f_n \to f$.*

Example 2.2. *Consider the sequence (f_n) given by $f_n : [0,1] \to \mathbb{R}$, $f_n(x) = x^{1/n}$.*

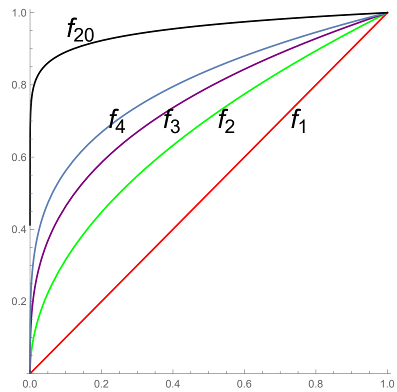

Figure 2.1: The sequence f_n for $n = 1, 2, 3, 4$ and 20.

Notice that $f_n(0) = 0$ for every n, but that for every $x \in (0,1]$ we have $\lim_{n\to\infty} x^{1/n} = 1$. As a result the limit of the sequence (f_n) is

$$f(x) = \begin{cases} 0 & x = 0, \\ 1 & x \in (0,1]. \end{cases}$$

Remark 2.3. *Notice that the above example shows that the pointwise limit of a sequence of continuous functions need not be continuous. It also produces a counterexample for the commutativity of the limits. We have*

$$\lim_{n\to\infty} \lim_{x\to 0^+} f_n(x) \neq \lim_{x\to 0^+} \lim_{n\to\infty} f_n(x),$$

as the left-hand side equals zero, while the right-hand side equals one.

Pointwise convergence clearly does not preserve continuity. It can also be very non-uniform, in the sense that while $f_n(x) \to 0$ for every x we may have $\sup_x |f_n(x) - f(x)| \to C > 0$ or even $\sup_x |f_n(x) - f(x)| \to \infty$ as n goes to infinity, as shown in the next examples.

Example 2.4. *Consider the sequences*

$$g_n(x) = \begin{cases} 2nx & x \in [0, \frac{1}{2n}) \\ -2n(x - \frac{1}{n}) & x \in [\frac{1}{2n}, \frac{1}{n}) \\ 0 & x \in [\frac{1}{n}, 1] \end{cases} \qquad h_n(x) = \begin{cases} 2n^2 x & x \in [0, \frac{1}{2n}) \\ -2n^2(x - \frac{1}{n}) & x \in [\frac{1}{2n}, \frac{1}{n}) \\ 0 & x \in [\frac{1}{n}, 1]. \end{cases}$$

It is easy to see that g_n and h_n are continuous and converge to the function $f = 0$. However, for every n we have $g_n(1/(2n)) = 1$ (with that being the maximum of g_n) and therefore

$$\sup_{x \in [0,1]} |g_n(x) - 0| = 1.$$

The situation is worse for the sequence (h_n), known as the Witch's hat. Indeed $h_n(1/(2n)) = n$, which shows that while $h_n \to 0$ we have

$$\sup_{x \in [0,1]} |h_n(x) - 0| \to \infty.$$

Pointwise convergence and integrability do not interact as one would hope. Indeed, even if we assume that the pointwise limit is integrable we may not have $\lim \int f_n = \int \lim f_n$.

Example 2.5. *Consider $f_n(x) = \chi_{[n,n+1)}(x)$, where χ_I is the indicator of the set I, i.e., takes value 1 if $x \in I$ and zero otherwise. Clearly f_n converges pointwise to $f = 0$. However,*

$$1 = \int f_n \neq \int f = 0.$$

We can think of this, as "the mass scaping to infinity" (along the x axis). In the latter calculation, we are considering the improper Riemann integral $\int_{-\infty}^{+\infty}$ on all of $\mathbb{R}$.

Another example of this phenomena, can be found by considering $g_n(x) = n\chi_{(0,1/n)}(x)$ we also have that g_n converges to 0, while having $\int g_n = 1$ for every n. We can think of this as "pointwise convergence allowing the mass to go to infinity" (along the y axis this time). The Witch's hat above also provides a similar example, in this case with continuous functions.

Example 2.6. *Another sequence that will play a role in several modules this year is $f_n(x) = \sin(n\,x)$. This sequence is connected to Fourier series and will be heavily studied in MA250 PDE for example. Notice that for $x = k\pi$ with $k \in \mathbb{Z}$ the limit exists and equals 0. If $x = p/q\,\pi$ with $p/q \notin \mathbb{Z}$ then there is no limit. Indeed $\sin(nqx) = 0$ while $\sin((2nq+1)x) = \sin(x) \neq 0$. If x is an irrational multiple of π, then the rest of the division of nx by 2π is dense in $[0, 2\pi]$ and there is no limit.*

Despite the fact that $\sin(nx)$ does not have a limit for most x, you will see in MA250 that for every integrable function f

$$\int_{-\pi}^{\pi} f(x) \sin(nx) \mathrm{d}x \to 0 \qquad \text{as } n \to \infty.$$

This result, known as the Riemann–Lebesgue Lemma, suggests that $\sin(nx)$ goes to zero in some sense (known as the weak sense, which will be covered in Measure Theory, Functional Analysis and Fourier Analysis). We can also consider the sequence $g_n(x) = \frac{\cos(nx)}{n}$. As cosine is a bounded function it is easy to see that g_n converges pointwise to 0. Since g_n are smooth we can also consider $g_n'(x) = -\sin(nx)$. This tells us that even for smooth functions, having g_n converge pointwise to g does not imply that g_n' converges to g' even if g is smooth.

The final example we consider is one of a sequence (f_n) such that $\int (f_n - f) \mathrm{d}x$ converges to zero, but where f_n does not converge pointwise to f.

Example 2.7. *We will consider functions defined on $[0,1]$. Let*

$$f_0(x) = \chi_{[0,1]}(x),$$

$$f_1(x) = \chi_{[0,1/2]}(x), \qquad f_2(x) = \chi_{[1/2,1]}(x),$$

$$f_3(x) = \chi_{[0,1/4]}(x), \qquad f_4(x) = \chi_{[1/4,1/2]}(x), \qquad f_5(x) = \chi_{[1/2,3/4]}(x), \qquad f_6(x) = \chi_{[3/4,1]}(x).$$

Notice that each function is an indicator of an interval, and that in each group above the intervals sweep $[0,1]$. When we move to the next block the length of the corresponding intervals gets divided by 2 and therefore we consider twice as many functions for each group. While the process is clear from the list writing a formula for f_n is annoying to say the least. You can check that the following works. For an index

$$n \in [\sum_{l=0}^{k-1} 2^l, \sum_{l=0}^{k} 2^l], \qquad k = 1, 2, \ldots$$

we set f_n as the indicator of the interval

$$\left[\frac{n - \sum_{l=0}^{k-1} 2^l}{2^k}, \frac{n - \sum_{l=0}^{k-1} 2^l + 1}{2^k} \right].$$

Since the length of the intervals tends to zero it is clear that $\int f_n \to 0$, but since the intervals keep sweeping the entire interval $[0,1]$ the sequence f_n does not converge to zero (or any other function for that matter). This is contrary to the intuition that if the area between f and f_n is going to zero the functions f_n must be approaching, and therefore converging to f.

2.2 Uniform convergence

We now consider the notion of uniform convergence.

Definition 2.8. *Let $f_n : \Omega \to \mathbb{R}$ be a sequence of functions. We say that (f_n) converges uniformly to $f : \Omega \to \mathbb{R}$ if and only if for every $\varepsilon > 0$ there exists $N(\varepsilon)$ such that $|f_n(x) - f(x)| < \varepsilon$ for every $x \in \Omega$ and for all $n > N(\varepsilon)$.*

The key different with pointwise convergence is that N depends only on ε and not on x. For pointwise convergence we first froze x and consider the convergence of $f_n(x)$ to $f(x)$. We will denote uniform convergence by $f_n \rightrightarrows f$.

As before we are not making any assumption on Ω. In order to simplify the presentation we introduce the notation

$$\|f\|_\infty = \sup_{x \in \Omega} |f(x)|.$$

With this notation we have

$$f_n \rightrightarrows f \iff \forall \varepsilon > 0, \exists N(\varepsilon) \text{ such that } \|f_n - f\|_\infty < \varepsilon \ \forall n > N(\varepsilon).$$

Remark 2.9. *Clearly uniform convergence implies pointwise convergence. The converse is of course false, as can be seen by considering the sequence from Remark 2.3. Namely, we note that $f_n(1/2^n) = 1/2$ and so $\|f_n - f\|_\infty \geq 1/2$. Alternatively, one can argue by contradiction and apply Theorem 2.13 below.*

Definition 2.10. *A sequence (f_n) of functions in Ω is called uniformly Cauchy if and only if for every $\varepsilon > 0$ there exists $N(\varepsilon)$ such that $\|f_n - f_m\|_\infty < \varepsilon$ for all $n, m > N(\varepsilon)$ (or alternatively $\sup_{x\in\Omega} |f_n(x) - f_m(x)| < \varepsilon$ for all $n, m > N(\varepsilon)$).*

Theorem 2.11. *A sequence (f_n) is uniformly convergent if and only if it is uniformly Cauchy.*

Proof. Assume that (f_n) is uniformly convergent to f, i.e. for every ε there exists N such that $\|f_n - f\|_\infty < \varepsilon/2$ for all $n > N$. Then, for $m, n > N$

$$\|f_n - f_m\|_\infty \leq \|f_n - f + f - f_m\|_\infty \leq \|f_n - f\|_\infty + \|f_m - f\|_\infty \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

For the converse, assume (f_n) is uniformly Cauchy. That means that for every x, $f_n(x)$ is a Cauchy sequence in $\mathbb{R}$ and therefore convergent. That means there exists $f(x)$ such that $f_n(x)$ converges to $f(x)$ at least pointwise. Now, we know that given $\varepsilon > 0$ there exists $N(\varepsilon) > 0$ such that $|f_n(x) - f_m(x)| < \varepsilon$ for every x and all $n, m > N(\varepsilon)$. That is

$$f_m(x) - \varepsilon < f_n(x) < f_m(x) + \varepsilon \text{ for all } x, \quad \text{and all } n, m > N(\varepsilon).$$

As the left-hand side holds for all $m > N(\varepsilon)$ we can take limits as m goes to infinity. We find

$$f(x) - \varepsilon \leq f_n(x) \leq f(x) + \varepsilon \text{ for all } x, \quad \text{and all } n > N(\varepsilon).$$

from which it follows that

$$|f(x) - f_n(x)| < 2\varepsilon \text{ for all } x, \quad \text{and all } n > N(\varepsilon),$$

which proves the result. □

Remark 2.12. *While this topic will be discussed in more depth in Norms, Metrics and Topologies it is worth noting that $\|\cdot\|_\infty$ is a norm in the space of bounded functions in Ω (we make no assumptions about it being open, closed, bounded or unbounded). $\|\cdot\|_\infty$ is referred to as the supremum norm or uniform norm. Recall that by norm we mean that it satisfies*

1. $\|f\|_\infty \geq 0$, *with* $\|f\|_\infty = 0$ *if and only if* $f = 0$,
2. $\|\lambda f\|_\infty = |\lambda| \|f\|_\infty$, *for all* $\lambda \in \mathbb{R}$, *and*
3. $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

Theorem 2.13. *Let (f_n) be a sequence of continuous functions in Ω that converges uniformly to $f : \Omega \to \mathbb{R}$. Then f is continuous.*

Proof. First notice that the uniform convergence implies that given any $\varepsilon > 0$ there exists $N > 0$ such that $\|f_n - f\|_\infty < \varepsilon/3$ for all $n > N$. In order to show that f is continuous at $x_0 \in \Omega$ we need to show that given ε there exists $\delta = \delta(\varepsilon)$ such that for all $x \in (x_0 - \delta, x_0 + \delta) \cap \Omega$ we have $|f(x) - f(x_0)| < \varepsilon$. With N as above, we choose $n > N$, fixed from now own. Since f_n is continuous at x_0 we know that there exists $\delta = \delta(\varepsilon)$ such that for all $x \in (x_0 - \delta, x_0 + \delta) \cap \Omega$ we have $|f_n(x) - f_n(x_0)| < \varepsilon/3$.

We estimate $|f(x) - f(x_0)|$ using the triangle inequality

$$\begin{aligned} |f(x) - f(x_0)| &= |f(x) - f_n(x) + f_n(x) - f_n(x_0) + f_n(x_0) - f(x_0)| \\ &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| \\ &\leq \|f_n - f\|_\infty + |f_n(x) - f_n(x_0)| + \|f_n - f\|_\infty \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}, \end{aligned}$$

for $n > N$ and $x \in (x_0 - \delta, x_0 + \delta) \cap \Omega$, with N and δ chosen as above. This completes the proof. □

We will denote the space of bounded, continuous functions with the uniform norm by $(C_b; \|\cdot\|_\infty)$.

Theorem 2.14. *$(C_b; \|\cdot\|_\infty)$ is a complete space, i.e. every Cauchy sequence converges to a continuous bounded function.*

Proof. We need to show that if (f_n) is Cauchy in the space, then there is a limit, and that the limit is bounded and continuous. First notice that a Cauchy sequence in $(C_b; \|\cdot\|_\infty)$ is, by definition, a uniformly Cauchy sequence. Theorem 2.11 implies that the sequence is convergent and since all the functions are continuous Theorem 2.13 implies the limit is continuous.

To see that it is bounded, notice that for every $x \in \Omega$

$$|f(x)| \le |f(x) - f_n(x)| + |f_n(x)|$$

for every n. Since f_n converges uniformly to f there exists n large enough $|f_n(x) - f(x)| < 1$. For that n, since f_n is bounded we have $|f_n| \le M$. These two inequalities lead to $|f(x)| \le M + 1$ for every $x \in \Omega$, proving the boundedness of f. □

Remark 2.15. *We could consider the interaction of uniform convergence and differentiation or integration. Consider for example $f_n(x) = \frac{\sin(n^2 x)}{n}$. The sequence (f_n) converges to $f = 0$ uniformly. Indeed*

$$\left|\frac{\sin(n^2 x)}{n} - 0\right| \le \frac{1}{n} \qquad \forall x.$$

Clearly all the functions f_n are smooth. The derivatives are given by $f_n'(x) = n\cos(n^2 x)$. It is easy to see that the sequence (f_n') does not converge uniformly (or pointwise). This example shows that while $f_n \rightrightarrows f$ we may not have $f_n' \rightrightarrows f'$ or even $f_n' \to f'$.

To explore integrability, we consider $g_n(x) = \frac{1}{2n}\chi_{[-n,n]}$. Recall that strictly speaking we have not defined Riemann integration in $\mathbb{R}$, but rather improper integration, via a limiting procedure. It is clear however that $\int g_n = 1$ for every n. The sequence g_n converges uniformly to $g = 0$ as we have $|g_n - 0| \le 1/(2n)$, and so $\lim \int g_n = 1 \neq 0 = \int g$. We reiterate that strictly speaking g_n are not Riemann integrable and we will prove that in fact, on a bounded interval $f_n \rightrightarrows f$ does imply $\int f_n \to \int f$.

Theorem 2.16. *Lef (f_n), $f_n : [a,b] \to \mathbb{R}$ be a sequence of Riemann integrable functions that converges uniformly to $f : [a,b] \to \mathbb{R}$. Then f is Riemann integrable and $\int f_n \to \int f$.*

Proof. First we need to show that f is Riemann integrable, that is show that for every $\varepsilon > 0$ there exists a partition P of $[a,b]$ such that

$$U(f,P) - L(f,P) < \varepsilon.$$

Now, since $f_n \rightrightarrows f$ we know that for any $\varepsilon > 0$ there exists N such that $\|f_n - f\|_\infty < \varepsilon/(4(b-a))$ for $n > N$. For a fixed $n > N$ since f_n is integrable we know that given $\varepsilon > 0$ there exists a partition P of $[a,b]$ such that

$$U(f_n,P) - L(f_n,P) < \frac{\varepsilon}{2}.$$

Now, for that P

$$\begin{aligned} U(f,P) - L(f,P) &= \sum[\sup_{I_k} f - \inf_{I_k} f]|I_k| = \sum[\sup_{I_k}(f - f_n + f_n) - \inf_{I_k}(f - f_n + f_n)]|I_k| \\ &\le \sum\left[\|f - f_n\|_\infty + \sup_{I_k} f_n + \|f - f_n\|_\infty - \inf_{I_k} f_n\right]|I_k| \\ &= 2\sum\|f - f_n\|_\infty|I_k| + \sum[\sup_{I_k} f_n - \inf_{I_k} f_n]|I_k| \\ &\le 2\|f - f_n\|_\infty(b-a) + U(f_n,P) - L(f_n,P) \\ &\le 2\frac{\varepsilon}{4(b-a)}(b-a) + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

To see that $\int f_n \to \int f$, notice that

$$\left|\int_a^b f_n - \int_a^b f\right| = \left|\int_a^b (f_n - f)\right| \le \int_a^b |f_n - f| \le \int_a^b \|f - f_n\|_\infty = \|f_n - f\|_\infty(b-a).$$

Clearly the right hand side goes to zero as n goes to infinity by the uniform convergence of (f_n) to f. □

In many circumstances it is necessary to consider functions of two variables (or more) from which we construct new functions by integrating out some of the variables. We want to study several results in this direction; we start by reviewing the notions of continuity and uniform continuity in two dimensions. The definitions are analogous to Definitions 1.2 and 1.3.

Definition 2.17. *Given $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$, we say that f is continuous at x if for every $\varepsilon > 0$ there exists $\delta = \delta(x, \varepsilon) > 0$ such that*

$$y \in \Omega \text{ and } |x - y| < \delta \Longrightarrow |f(x) - f(y)| < \varepsilon. \tag{2.1}$$

Note that $|\cdot|$ has been used both to denote Euclidean distance in the plane, as in $|x - y|$, as well as for absolute value of a real number, in $|f(y) - f(x)|$.

Definition 2.18. *Given $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$, we say that it is uniformly continuous if for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that*

$$x, y \in \Omega \text{ and } |x - y| < \delta \Longrightarrow |f(x) - f(y)| < \varepsilon. \tag{2.2}$$

The key point here is that δ can be chosen independently of x. Similarly to Theorem 1.4 we have the following result.

Theorem 2.19. *Let $f : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ be a continuous function. Assume that Ω is closed and bounded. Then it is uniformly continuous.*

Proof. We will argue by contradiction. That would mean that there exists $\varepsilon > 0$ and x_n, y_n such that $|x_n - y_n| \leq \frac{1}{n}$ but $|f(x_n) - f(y_n)| > \varepsilon$.

The sequences $\{x_n\}$ and $\{y_n\}$ are bounded, as they are in Ω, which is closed and bounded, and therefore we can apply Bolzano–Weierstrass to each component to obtain convergent subsequences $\{x_{n_k}\}_{k=1}^{\infty}$ to x and $\{y_{n_k}\}_{k=1}^{\infty}$ to y. Notice that

$$|x - y_{n_k}| \leq |x - x_{n_k}| + |x_{n_k} - y_{n_k}| \leq |x - x_{n_k}| + \frac{1}{n_k} \xrightarrow[k\to\infty]{} 0,$$

which implies that $x = y$. However we know that $|f(x_{n_k}) - f(y_{n_k})| > \varepsilon$ for all k. Since f is continuous, taking limits as k goes to infinity we obtain $0 = |f(x) - f(x)| \geq \varepsilon$, which is a contradiction. □

Theorem 2.20. *Let $f : [a, b] \times [c, d] \to \mathbb{R}$ be a continuous function. Define*

$$I(t) := \int_a^b f(x, t)\mathrm{d}x$$

Then I is a continuous function on $[c, d]$.

Proof. We need to show that for every $\varepsilon > 0$ there exists δ such that $|t - t_0| < \delta$, and $t, t_0 \in [c, d]$ implies $|I(t) - I(t_0)| < \varepsilon$.

Now $I(t) - I(t_0) = \int [f(x, t) - f(x, t_0)]\mathrm{d}x$ and therefore

$$|I(t) - I(t_0)| \leq \int_a^b |f(x, t) - f(x, t_0)|\mathrm{d}x. \tag{2.3}$$

Since f is continuous on $[a, b]\times[c, d]$ it is uniformly continuous, and therefore given $\varepsilon > 0$ there exists δ such that $(x_1, t_1), (x_2, t_2) \in [a, b]\times[c, d]$ with $\sqrt{(x_1 - x_2)^2 + (t_1 - t_2)^2} < \delta$ implies that $|f(x_1, t_1)-f(x_2, t_2)| < \varepsilon/(b - a)$. Therefore if $|t - t_0| < \delta$ we have $|f(x, t) - f(x, t_0)| < \varepsilon/(b - a)$. As a result (2.3) becomes

$$|I(t) - I(t_0)| \leq \int_a^b |f(x, t) - f(x, t_0)|\mathrm{d}x < \int_a^b \frac{\varepsilon}{b - a}\mathrm{d}x = \varepsilon,$$

and we obtain the desired result. □

We can also consider differentiating I with respect to t under sufficient regularity results for f.

Theorem 2.21. *Lef $f, \frac{\partial f}{\partial t}$ be continuous functions on $[a,b]\times[c,d]$. Then, for $t \in (c,d)$*

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

Proof. Set $F(t) := \int_a^b f(x,t)\mathrm{d}x$, and $G(t) := \int_a^b \frac{\partial f}{\partial t}(x,t)\mathrm{d}x$. We want to show that $F' = G$. Note that F is differentiable by Theorem 1.16. We consider the difference between the incremental quotient that is used to define a derivative of F and the function we expect to be derivative, namely G. Let $t_0 \in (c,d)$ be given. Consider $h \in \setminus$ such that $t_0 + h \in [c,d]$. We write

$$\left|\frac{F(t_0+h)-F(t_0)}{h} - G(t)\right| = \left|\int_a^b \frac{f(x,t_0+h)-f(x,t_0)}{h} - \frac{\partial f}{\partial t}(x,t_0)\mathrm{d}x\right|,$$

which by the Mean Value Theorem, becomes, for some τ between t_0 and t_0+h

$$= \left|\int_a^b \frac{\partial f}{\partial t}(x,\tau) - \frac{\partial f}{\partial t}(x,t_0)\mathrm{d}x\right| \leq \int_a^b \left|\frac{\partial f}{\partial t}(x,\tau) - \frac{\partial f}{\partial t}(x,t_0)\right|\mathrm{d}x.$$

Now, since $\frac{\partial f}{\partial t}$ is continuous on $[a,b]\times[c,d]$ it is uniformly continuous, and therefore for every $\varepsilon > 0$ there exists δ such that for $|h| < \delta$ and τ as above we have

$$\left|\frac{\partial f}{\partial t}(x,\tau) - \frac{\partial f}{\partial t}(x,t_0)\right| < \frac{\varepsilon}{b-a}$$

This is implies that for $|h| < \delta$

$$\left|\frac{F(t_0+h)-F(t_0)}{h} - G(t_0)\right| \leq \int_a^b \frac{\varepsilon}{b-a}\mathrm{d}x = \varepsilon.$$

Taking limits as h goes to zero we have

$$|F'(t_0) - G(t_0)| \leq \varepsilon,$$

and since ε is arbitrary we are done. □

We now explore a version of Fubini's Theorem for continuous functions.

Theorem 2.22. *Let $f : [a,b]\times[c,d] \to \mathbb{R}$ be a continuous function. Then*

$$\int_a^b \left(\int_c^d f(x,y)\mathrm{d}y\right)\mathrm{d}x = \int_c^d \left(\int_a^b f(x,y)\,\mathrm{d}x\right)\mathrm{d}y$$

Proof. Since f is continuous on $[a,b]\times[c,d]$ Theorem 2.20 implies that $\int_c^d f(x,y)\mathrm{d}y$ and $\int_a^b f(x,y)\,\mathrm{d}x$ are continuous on their respective domains, and therefore Riemann integrable. Consider

$$F(t) = \int_a^t \left(\int_c^d f(x,y)\mathrm{d}y\right)\mathrm{d}x - \int_c^d \left(\int_a^t f(x,y)\,\mathrm{d}x\right)\mathrm{d}y.$$

By the FTC (Theorem 1.16), we know that F is continuous, with $F(a) = 0$. Also the first integral is differentiable with

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

We know that

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

We would now like to differentiate the second integral in F, namely

$$-\int_c^d \left(\int_a^t f(x,y)\,\mathrm{d}x \right) \mathrm{d}y$$

by differentiating inside the first integral. For that Theorem 2.21 requires that

$$\left(\int_a^t f(x,y)\,\mathrm{d}x \right)$$

be continuous in $[a,b] \times [c,d]$ as a function of t and y. Theorem 2.20 proves that it is continuous as a function of y and we actually know that it is differentiable as a function of t. However, continuity in each of the variables separately does not ensure that the function is continuous on $[a,b] \times [c,d]$. However, one can modify the proof of Theorem 2.20 to show continuity in $[a,b] \times [c,d]$. (This is left as an exercise.) Then we are allowed to differentiate inside the integral and we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_c^d \left(\int_a^t f(x,y)\,\mathrm{d}x \right) \mathrm{d}y = \int_c^d \frac{\partial}{\partial t} \left(\int_a^t f(x,y)\,\mathrm{d}x \right) \mathrm{d}y = \int_c^d f(t,y)\,\mathrm{d}y.$$

Therefore

$$F'(t) = \int_c^d f(t,y)\,\mathrm{d}y - \int_c^d f(t,y)\,\mathrm{d}y = 0,$$

Since F is continuous on $[a,b]$, $F(a) = 0$ and $F'(t) = 0$ we find $F(b) = 0$. This implies the result. □

Remark 2.23. *The continuity requirement is necessary in the previous Theorem. The following is a counterexamples to Fubini's theorem when continuity fails at just a point. Let*

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

Notice that f is not continuous at the origin. We have

$$\int_0^1 \frac{x^2 - y^2}{(x^2+y^2)^2}\mathrm{d}y = \frac{y}{x^2+y^2}\bigg|_{y=0}^{y=1} = \frac{1}{1+x^2},$$

and

$$\int_0^1 \left(\int_0^1 \frac{x^2 - y^2}{(x^2+y^2)^2}\mathrm{d}y \right) \mathrm{d}x = \int_0^1 \frac{1}{1+x^2}\mathrm{d}x = \frac{\pi}{4}.$$

In the opposite direction we get $\frac{-\pi}{4}$ by symmetry. The key here is that the function is not in L^1, i.e. $|f|$ is not integrable.

$$\int_0^1 \left(\int_0^1 \left| \frac{x^2 - y^2}{(x^2+y^2)^2} \right| \mathrm{d}y \right) \mathrm{d}x \geq \int_0^1 \left(\int_0^x \frac{x^2 - y^2}{(x^2+y^2)^2}\mathrm{d}y \right) \mathrm{d}x = \int_0^1 \frac{y}{x^2+y^2}\bigg|_{y=0}^{y=x} = \int_0^1 \frac{1}{2x}\mathrm{d}x = \infty.$$

Differentiation revisited.

We will use the notation $C^k(a,b)$ to denote functions that are k times continuously differentiable on (a,b), and $C^\infty(a,b)$ for functions that are infinitely differentiable on (a,b).

We have seen examples of sequences (f_n) that are differentiable, with (f_n) converging uniformly to f but for which f_n' does not converge to f'. In fact it is easy to construct examples of C^1 functions that converge uniformly for which f' does not exist. Consider

$$f_n(x) = \left(x^2 + 1/n\right)^{1/2}.$$

They are clearly C^1 as the $x^2 + 1/n$ never vanishes for fixed n. (f_n) converges uniformly to $f(x) = |x|$, which is not smooth at the origin. To see this notice that if

$$A := \left(x^2 + 1/n\right)^{1/2} - |x|$$

then

$$A \leq \left((x + 1/\sqrt{n})^2\right)^{1/2} - |x| \leq \frac{1}{\sqrt{n}},$$

and the uniform convergence follows.

The following result will prove rather useful.

Theorem 2.24. *Let (f_n) be a sequence of C^1 functions on $[a,b]$ (understood as a one-sided derivative). Assume $f_n \to f$ in the pointwise sense and that f_n' converges uniformly to g. Then f is C^1 and $g = f'$ or $f_n' \to f'$.*

Proof. Since $f_n' \rightrightarrows g$, Theorem 2.16 yields

$$\int_a^x g(y)\mathrm{d}y = \int_a^x \lim_{n\to\infty} f_n'(y)\mathrm{d}y = \lim_{n\to\infty} \int_a^x f_n'(y)\mathrm{d}y,$$

which by the FTC yields

$$\int_a^x g(y)\mathrm{d}y = \lim_{n\to\infty} [f_n(x) - f_n(a)] = f(x) - f(a).$$

Notice that since g is continuous this means that f is continuous. While the Theorem does not assume that g is continuous, that is a consequence of the uniform convergence of f_n' to g, since f_n are C^1. Now, the FTC implies that $\int_a^x g$ is differentiable with derivative g. Since

$$\int_a^x g = f(x) - f(a)$$

we obtain that f is differentiable and $g = f'$. □

2.3 Series of functions

In this section we consider series of functions, i.e., we study

$$\sum_{k=1}^{\infty} f_k(x),$$

with $f_k : \Omega \to \mathbb{R}$. We begin by establishing the notion of pointwise convergence and uniform convergence for a series.

Definition 2.25. *Let (f_k) be a sequence of functions $f_k : \Omega \to \mathbb{R}$. Let (S_n) be the sequence of partial sums, with $S_n : \Omega \to \mathbb{R}$ defined by*

$$S_n(x) = \sum_{k=1}^{n} f_k(x).$$

Then the series

$$\sum_{k=1}^{\infty} f_k(x)$$

converges pointwise to $S : \Omega \to \mathbb{R}$ in Ω if $S_n \to S$ pointwise on Ω and it converges uniformly to S in Ω if $S_n \rightrightarrows S$ uniformly in Ω.

Theorem 2.26. *Let (f_k), with $f_k : [a,b] \to \mathbb{R}$, be a sequence of integrable functions. Assume that $\sum_{k=1}^{\infty} f_k$ converges uniformly. Then $\sum_{k=1}^{\infty} f_k$ is Riemann integrable and*

$$\int \sum_{k=1}^{\infty} f_k = \sum_{k=1}^{\infty} \int f_k.$$

Proof. S_n is a finite sum of integrable functions and therefore integrable (by additivity). Since S_n converges uniformly Theorem 2.16 implies that S is integrable and moreover

$$\lim_{n\to\infty} \int S_n = \int \lim_{n\to\infty} S_n.$$

Since $\int S_n = \sum_{k=1}^{n} \int f_k$ and $\lim_{n\to\infty} S_n = \sum_{k=1}^{\infty} f_k$ we obtain the result. □

Theorem 2.27. *Let (f_k), with $f_k : [a,b] \to \mathbb{R}$, be a sequence of C^1 functions such that $\sum_{k=1}^{\infty} f_k$ converges pointwise. Assume that $\sum_{k=1}^{\infty} f_k'$ converges uniformly. Then*

$$\left(\sum_{k=1}^{\infty} f_k(x)\right)' = \sum_{k=1}^{\infty} f_k'(x),$$

that is, the series is differentiable and can be differentiated term-by-term.

Proof. The proof is a simple consequence of Theorem 2.24.This results says (changing the notation) that if S_n is C^1, $S_n \to S$, $S_n' \rightrightarrows g$ then $S \in C^1$ and $S' = g$ (or $S_n' \rightrightarrows S'$). If we define $S_n = \sum_{k=1}^{n} f_k$ then, it is C^1, since each f_k is C^1; it converges pointwise to $S = \sum_{k=1}^{\infty} f_k$ and finally S' converges uniformly, to g say. Then S is C^1 and $S_n' \rightrightarrows S'$. This means

$$\lim_{n\to\infty} S_n' = S' = \left(\sum_{k=1}^{\infty} f_k(x)\right)'$$

but since $S_n' = (\sum_{k=1}^{n} f_k)' = \sum_{k=1}^{n} f_k'$ we obtain the result, namely

$$\sum_{k=1}^{\infty} f_k'(x) = \left(\sum_{k=1}^{\infty} f_k(x)\right)'.$$

□

Theorem 2.28 (The Weierstrass M-test)**.** *Let (f_k) be a sequence of functions $f_k : \Omega \to \mathbb{R}$, and assume that for every k there exists $M_k > 0$ such that $|f_k(x)| \le M_k$ for every $x \in \Omega$ and $\sum_{k=1}^{\infty} M_k < \infty$. Then*

$$\sum_{k=1}^{\infty} f_k$$

converges uniformly on Ω.

Proof. Notice that it suffices to show that $S_n := \sum_{k=1}^{n} f_k$ is uniformly Cauchy (recall Theorem 2.11). Now since $\sum_{k=1}^{\infty} M_k < \infty$, given $\varepsilon > 0$ there exists N such that

$$\sum_{k=m+1}^{n} M_k < \varepsilon \qquad \text{for all } m, n > N.$$

Now

$$|S_n(x) - S_m(x)| = \left|\sum_{k=1}^{n} f_k(x) - \sum_{k=1}^{m} f_k(x)\right| = \left|\sum_{k=m+1}^{n} f_k(x)\right| \le \sum_{k=m+1}^{n} |f_k| \le \sum_{k=m+1}^{n} M_k \le \varepsilon,$$

for every x. Therefore S_n is uniformly Cauchy and the proof is complete. □

2.4 A continuous, nowhere differentiable function (THE PROOFS IN THIS SECTION ARE NOT EXAMINABLE)

In 1872 Weierstrass showed that there exist continuous functions that are nowhere differentiable. Standard examples are constructed using Fourier Series. For example

$$f(x) = \sum_{k=0}^{\infty} a^k \cos(2\pi b^k x)$$

for any $0 < a < 1 < b$ with $ab > 1$.

We will construct an example based on the sawtooth function. Consider

$$\phi(x) = \begin{cases} x - \lfloor x \rfloor & x \leq \lfloor x \rfloor + \frac{1}{2} \\ 1 - x + \lfloor x \rfloor & x > \lfloor x \rfloor + \frac{1}{2}. \end{cases}$$

The function ϕ is equal to the distance function from x to $\mathbb{Z}$.

We define, for $n = 0, 1, \ldots$

$$f_n(x) = \frac{1}{4^n}\phi(4^n x).$$

We will show that $f(x) = \sum_{n=0}^{\infty} f_n$ is continuous but nowhere differentiable. Notice that

$$0 \leq f_n \leq \frac{1}{4^n}\phi \leq \frac{1}{2}\frac{1}{4^n},$$

and that by the Weierstrass M-test we have the uniform convergence of the series. Since each f_n is continuous, and the convergence is uniform we have that f is C^0.

Given $x \in \mathbb{R}$ we will choose the sign of $h_n = \pm\frac{1}{4^{n+1}}$ in such a way that the points $4^n x$ and $4^n(x + h_n)$ both belong to the same interval of length 1/2, $[\frac{k}{2}, \frac{k+1}{2}]$ for some $k \in \mathbb{Z}$. We make this choice of sign for h_n because on each of these intervals $[\frac{k}{2}, \frac{k+1}{2}]$, the function ϕ has constant slope $+1$ or -1.

Consider the incremental quotient

$$\frac{f_n(x + h_n) - f_n(x)}{h_n} = \frac{\phi(4^n x + 4^n h_n) - \phi(4^n x)}{4^n h_n} = \pm 1.$$

Moreover, if $m < n$ the graph off_m also has slope ± 1 on the interval to which x and $x + h_n$ belong. Let us justify this last step in more detail. We argue by contradiction. Consider the case $h_n = \frac{1}{4^{n+1}}$ (the case when $h_n = -\frac{1}{4^{n+1}}$ is treated analogously). Namely, we suppose that there exists $\ell \in \mathbb{Z}$ such that

$$4^m x < \frac{\ell}{2} < 4^m x + 4^m h_n\,.$$

Multiplying the above inequalities by 4^{n-m}, we get

$$4^n x < 4^{n-m}\,\frac{\ell}{2} < 4^n x + \frac{1}{4}\,.$$

Since $4^{n-m}\,\frac{\ell}{2}$, this is a contradiction to our choice of h_n. Therefore,

$$\epsilon_m := \frac{f_m(x + h_n) - f_m(x)}{h_n} = \frac{\phi(4^m x + 4^m h_n) - \phi(4^m x)}{4^m h_n} = \pm 1.$$

However for $m \geq n + 1$ we have (since $4^m x + 4^m h_n - 4^m x = +4^m h_n = \pm 4^{m-n-1} \in \mathbb{Z}$)

$$f_m(x + h_n) - f_n(x) = \frac{1}{4^m}\left(\phi(4^m x + 4^m h_n) - \phi(4^m x)\right) = 0.$$

Therefore

$$A_n := \frac{f(x + h_n) - f(x)}{h_n} = \sum_{m=0}^{n} \frac{f_m(x + h_n) - f_m(x)}{h_n} = \sum_{m=0}^{n} \epsilon_m.$$

Therefore A_n is an even integer if n is odd and an odd integer if n is even. Hence there is no limit as n goes to infinity. Since h_n goes to zero that proves that f is not differentiable.

Chapter 3

Basic results about $\mathbb{R}^n$

3.1 Notation in $\mathbb{R}^n$

The main vector spaces that we shall consider in this module are $\mathbb{R}^n,\ n \in \mathbb{N}$. Thus, by a vector $x \in \mathbb{R}^n$ we mean the n-tuple $(x_1, \ldots, x_n),\ x_i \in \mathbb{R},\ 1 \leqslant i \leqslant n$.

For ease of writing, a vector $x \in \mathbb{R}^n$ will be written as a row vector $x = (x_1, \ldots, x_n),\ x_i \in \mathbb{R},\ 1 \leqslant i \leqslant n$. However, *in calculations* vectors will be written as column vectors

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

This is so that, if $A\colon \mathbb{R}^n \to \mathbb{R}^k$ is a linear map represented by the matrix

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kn} \end{pmatrix}$$

with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^k$, then the vector $y := Ax$ is obtained by multiplying the column vector x by the matrix A *on the left*. In index notation, if $y = (y_1, \ldots, y_k)$, then

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i \in \{1, \ldots, k\}.$$

A vector-valued function f (with values in $\mathbb{R}^k$) of the variables $x_1, \ldots, x_n$ is denoted by $f\colon U \to \mathbb{R}^k$ where $U \subset \mathbb{R}^n$ is domain of the function f, i.e., the subset in which the independent variables $x = (x_1, \ldots, x_n)$ lie.[1] Thus

$$f(x) \text{ is shorthand for } (f_1(x_1, \ldots, x_n), \ldots, f_k(x_1, \ldots, x_n))$$

and,

$$\text{for calculations, } f(x) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_k(x_1, \ldots, x_n) \end{pmatrix}.$$

We normally use a, b and c at the start of the latin alphabet to denote real numbers (i.e. scalars) and we shall use letters like x, y, p, q, u, v and w in the second half of the latin alphabet to denote vectors. Thus we shall write ax for the vector $(ax_1, \ldots, ax_n)$ without always spelling out that $a \in \mathbb{R}$ and $x \in \mathbb{R}^n$. In two and three dimensions, we will often be written $f(x, y)$ and $f(x, y, z)$. Finally, 0 will denote both the zero vector (though we shall occasionally write it as $(0, \ldots, 0)$ and the zero scalar!

[1] A real valued function $f\colon U \to \mathbb{R}$ will be referred to as a scalar function.

3.2 The Euclidean norm and inner product

The *Euclidean norm*, or *length*, or *magnitude* of $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ is denoted by $\|x\|$ and is defined by

$$\|x\| := \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}. \tag{3.1}$$

The notation is convenient given that for $n = 1,\ \|x\|$ means $|x|$.

The *direction* of a nonzero vector x is defined to be the unit vector $\frac{x}{\|x\|}$. The obvious relation

$$x = \|x\| \frac{x}{\|x\|}, \qquad x \neq 0,$$

is the mathematical statement of the informal definition of a (nonzero) vector as a quantity that has both magnitude and direction[2].

The *Euclidean distance* between x and y in $\mathbb{R}^n$ is defined as $\|x - y\|$.

The *Euclidean inner product* $x \cdot y$, also called the *dot product* and *scalar product*, of $x, y \in \mathbb{R}^n$ is defined as

$$x \cdot y := \sum_{i=1}^{n} x_i y_i.$$

Other notations for $x \cdot y$ include (x, y) and $\langle x, y \rangle$. Evidently, $\|x\| = \sqrt{x \cdot x}$.

The *Cauchy-Schwarz inequality* states that

$$|x \cdot y| \leqslant \|x\|\|y\|.$$

It follows from

$$0 \leqslant \Big\| \|y\|^2 x - (x \cdot y) y \Big\|^2 = \|y\|^4\|x\|^2 - (x \cdot y)^2\|y\|^2.$$

Definition 3.1 (Angle between two nonzero vectors)**.** *The Cauchy-Schwarz inequality implies that, if x and y are both nonzero, then there exists unique $\theta \in [0, \pi]$ such that*

$$x \cdot y = \|x\|\,\|y\| \cos\theta;$$

θ is then defined *to be the* angle *between x and y.*

Proposition 3.2 (The triangle inequality)**.** $\forall\, x, y \in \mathbb{R}^n$,

$$\|x + y\| \leqslant \|x\| + \|y\|. \tag{3.2}$$

Proof.

$$\|x + y\|^2 = (x + y) \cdot (x + y) = \|x\|^2 + 2x \cdot y + \|y\|^2 \leqslant \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

□

Replacing x by $x - z$ and y by $z - y$ we get:

$$\|x - y\| \leqslant \|x - z\| + \|z - y\|, \tag{3.3}$$

which corresponds to the familiar fact that the distance from x to y is less than or equal to the sum of the distances from x to z and z to y. Regarding $x,\ y$ and z as the vertices of a triangle, (3.3) says that the length of the edge joining x and y is less than or equal to the sum of the lengths of the edges joining x to z and z to y; this is the usual triangle inequality.

[2] Note that the zero vector does not have a direction.

Exercise 3.1. *Prove that, for all $x, y \in \mathbb{R}^n$,*

$$\Big| \|x\| - \|y\| \Big| \leqslant \|x - y\|. \tag{3.4}$$

Proposition 3.3.

(i) $\|ax\| = |a| \, \|x\| \; \forall \, a \in \mathbb{R}, \; x \in \mathbb{R}^n.$

(ii) $\|x\| \geqslant 0 \; \forall \, x \in \mathbb{R}^n$ *and* $\|x\| = 0 \; \Leftrightarrow \; x = 0.$

Warning 3.4. *It is true that $y = x$ if, and only if, $\|x - y\| = 0$. However, $\|x\| = \|y\|$ does not imply that $y = \pm x$. (Think of points on the unit circle.)*

Remark 3.5 (**$\| \cdot \|$ satisfies the definition of a norm**)**.** *A norm is a non-negative valued function $\| \cdot \| : X \to \mathbb{R}^+$ on a real vector space X which satisfies*

(i) $\|x\| \geqslant 0 \; \forall \, x \in X$ *and* $\|x\| = 0 \; \Leftrightarrow \; x = 0.$

(ii) *For every $a \in \mathbb{R}$ and $x \in X$ we have $\|ax\| = |a| \, \|x\|$.*

(iii) *The triangle inequality* (3.2)*: for every $x, y \in X$ we have $\|x + y\| \leqslant \|x\| + \|y\|$.*

3.3 Convergence in $\mathbb{R}^n$

Definition 3.6. *A sequence (x_j) of vectors in $\mathbb{R}^n$ converges to $x \in \mathbb{R}^n$ if*

$$\forall \varepsilon > 0, \; \exists \; N \in \mathbb{N} \text{ such that, } j \geq N \; \Rightarrow \|x_j - x\| < \varepsilon.$$

Proposition 3.7 (Uniqueness of limits)**.** *Let (x_j) be a sequence in $\mathbb{R}^n$. If it converges to both x and $\tilde{x}$, then $x = \tilde{x}$.*

Proof. If we assume, by contradiction that $x \neq \tilde{x}$, then $\varepsilon := \frac{1}{2}\|x - \tilde{x}\| > 0$. Since x_j converges to x, $\exists \; N_1 \in \mathbb{N}$ such that

$$j \geqslant N_1 \; \Rightarrow \|x_j - x\| < \varepsilon. \tag{3.5}$$

Similarly, since x_j also converges to $\tilde{x}$, $\exists \; N_2 \in \mathbb{N}$ such that

$$j \geqslant N_2 \; \Rightarrow \|x_j - \tilde{x}\| < \varepsilon. \tag{3.6}$$

Then, for $j \geqslant \max\{N_1, N_2\}$ we have:

$$2\varepsilon = \|x - \tilde{x}\| \leqslant \|x - x_j\| + \|x_j - \tilde{x}\| < 2\varepsilon.$$

This is of course a contradiction and therefore $x = \tilde{x}$.[3] □

The notation when we consider each of the coordinates of one of the elements x_j in the sequence can get a bit awkward. We will donte the i-th coordinate of $x_j \in \mathbb{R}^n$ by $x_{j,i}$.

Proposition 3.8 (Componentwise Convergence)**.** *A sequence (x_j) of vectors in $\mathbb{R}^n$ converges to $x_0 \in \mathbb{R}^n$ if, and only if, for each $i \in \{1, \ldots, n\}$, $\lim\limits_{j \to \infty} x_{j,i} = x_{0,i}$, where $x_j = (x_{j,1}, \ldots, x_{j,n})$ and $x_0 = (x_{0,1}, \cdots, x_{0,n})$.*

[3]Note how the triangle inequality is crucial for proving the uniqueness of the limit of a sequence.

Proof that convergence implies componentwise convergence. Note that

$$\forall\ i \in \{1, \dots n\},\ |x_{0,i} - x_{j,i}| \leqslant \|x_0 - x_j\| = \left[\sum_{k=1}^{n} (x_{0,k} - x_{j,k})^2\right]^{1/2}.$$

Now by definition of convergence, given $\varepsilon > 0,\ \exists\ N \in \mathbb{N}$ such that

$$j \geqslant N \Rightarrow\ \|x_0 - x_j\| < \varepsilon$$

and therefore

$$j \geqslant N \Rightarrow\ |x_{0,i} - x_{j,i}| < \varepsilon\ \forall i \in \{1, \dots n\},$$

i.e., $\lim_{j\to\infty} x_{j,i} = x_{0,i}$ for every $i \in \{1, \dots n\}$.

Proof that componentwise convergence implies convergence. Given $\varepsilon > 0$ and $i \in \{1, \dots, n\},\ \exists\ N_i \in \mathbb{N}$ such that $j \geqslant N_i\ \Rightarrow\ |x_{0,i} - x_{j,i}| < \varepsilon/\sqrt{n}$. Set $N := \max\{N_1, \dots, N_n\}$. Then

$$j \geqslant N\ \Rightarrow\ \|x_0 - x_j\| = \left(\sum_{k=1}^{n} (x_{0,k} - x_{j,k})^2\right)^{1/2} < \varepsilon,$$

i.e., $\lim_{j\to\infty} x_j = x_0$. □

Remark 3.9. *Proposition 3.8 allows us to reduce questions of convergence of a vector-valued sequence to the corresponding (more familiar) questions of convergence of a sequence of real numbers.*

The norm $\|\cdot\|$ that we defined in (3.1) corresponds to the standard notion of distance we are used to. However we could have defined other alternative norms.

Definition 3.10 (Max-norm $\|\cdot\|_\infty$)**.** *The* max-norm, *which is denoted by* $\|\cdot\|_\infty$, *is defined by*

$$\|x\|_\infty := \max\{|x_1|, \dots, |x_n|\}, \qquad x = (x_1, \dots, x_n). \tag{3.7}$$

The following definition provides yet another norm on $\mathbb{R}^n$.

Definition 3.11 (The 'Manhattan or taxi cab norm' $\|\cdot\|_1$)**.**

$$\|x\|_1 := |x_1| + \cdots + |x_n|, \qquad x = (x_1, \dots, x_n). \tag{3.8}$$

In fact there is a full family of norms $\|x\|_p$ for $1 \leq p < \infty$ given by

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

The Euclidean norm (3.1) corresponds to $p = 2$, and $\|\cdot\|_\infty$ corresponds to taking the limit as $p \to \infty$.

Exercise 3.2 (Comparison of the Euclidean norm with $\|\cdot\|_\infty$ and $\|\cdot\|_1$)**.** *Prove that*

$$\|x\|_\infty \leqslant \|x\| \leqslant \sqrt{n}\, \|x\|_\infty \tag{3.9}$$

and that

$$\|x\| \leqslant \|x\|_1 \leqslant \sqrt{n}\, \|x\|. \tag{3.10}$$

Furthermore, verify that $\|\cdot\|_\infty$ and $\|\cdot\|_1$ satisfy the triangle inequality and the relations stated in Remark 3.5 for a norm; indeed, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are actually norms.

This exercise shows that in Definition 3.6 we could have equivalently used $\|\cdot\|_1$ or $\|\cdot\|_\infty$ instead of $\|\cdot\|$ to define the limit of a sequence.

Since Proposition 3.8 reduces convergence to componentwise convergence the we have the following result.

Proposition 3.12 (Sequence sum rule)**.** *If x_j converges to x, y_j converges to y and $a, b \in \mathbb{R}$ then*

$$\lim_{j\to\infty}(ax_j + by_j) = ax + by.$$

Exercise 3.3 (Sequence product rules)**.** *Let (a_j) be a sequence of real numbers that converges to a and let (x_j) and (y_j) be sequences of vectors in $\mathbb{R}^n$ that converge to x and y respectively. Prove that*

(i) the sequence of vectors (a_jx_j) converges to ax and

(ii) the sequence of real numbers (not vectors!) $x_j \cdot y_j$ converges to $x \cdot y$.

Definition 3.13 (Boundedness of a sequence)**.** *A sequence (x_j) is* bounded *if there $\exists\ M > 0$ such that $\|x_j\| \leqslant M$ for every $j \in \mathbb{N}$.*

Proposition 3.14 (Boundedness of a convergent sequence)**.** *If (x_j) converges to x, then (x_j) is bounded.*

It is possible to prove this proposition by using componentwise convergence and the boundedness of real sequences. A more direct proof is based on the following lemma.

Lemma 3.15. *If (x_j) converges to x then the sequence of real numbers $\|x_j\|$ converges to $\|x\|$.*

Proof. Given $\varepsilon > 0$, $\exists\ N \in \mathbb{N}$ such that $j \geqslant N \ \Rightarrow \|x_j - x\| < \varepsilon$. It follows from the reverse triangle inequality that

$$\text{for } j \geqslant N, \quad |\,\|x_j\| - \|x\|\,| \leqslant \|x_j - x\| < \varepsilon.$$

□

Proof of boundedness of a convergent sequence. By the lemma, the convergence of (x_j) implies the convergence of $\|x_j\|$. The boundedness of a convergent sequence of real numbers then implies that $\|x_j\|$ is bounded and therefore, by definition, (x_j) is bounded. □

Remark 3.16. *Note that the converse of Lemma* (3.15) *does not hold, not even when $n = 1$. (Use a mathematical software package to plot the sequence $(\cos n, \sin n)$ in the plane for a demonstration of how badly the converse of Lemma* (3.15) *can fail.)*

Proposition 3.17 (Completeness of $\mathbb{R}^n$)**.** *Let (x_j) be a Cauchy sequence in $\mathbb{R}^n$, that is, $\forall\, \varepsilon > 0$, $\exists\ N \in \mathbb{N}$ such that $j, k \geqslant N \ \Rightarrow \ \|x_j - x_k\| < \varepsilon$. Then (x_j) converges to some $x \in \mathbb{R}^n$.*

Sketch proof. Show that each component $x_{j,i}$, $1 \leqslant i \leqslant n$, is a Cauchy sequence of real numbers. Then use the completeness of $\mathbb{R}$ and componentwise convergence of x_j. □

3.4 Subsequences and the Bolzano-Weierstrass theorem

The Bolzano-Weierstrass theorem is one of the most important theorems about sequences of real numbers. It states that every bounded sequence of real numbers has a convergent subsequence. It generalises immediately to sequences in $\mathbb{R}^n$.

Theorem 3.18 (Bolzano-Weierstrass for a bounded sequence of vectors)**.** *A bounded sequence (x_j) in $\mathbb{R}^n$ has a convergent subsequence (x_{j_ℓ}).*

Sketch of the proof. The proof of the Bolzano-Weierstrass in Chapter 3 in MA141 is done in $\mathbb{R}$. The argument below is a complete proof in two dimensions, and with Let $x_j = (x_{j,1}, \dots, x_{j,n})$ be a bounded sequence in $\mathbb{R}^n$. Then $x_{j,1}$ is a bounded sequence in $\mathbb{R}$ and therefore, by the Bolzano-Weierstrass Theorem it has a convergent subsequence $x_{j_k,1}$ which converges to $x_1^* \in \mathbb{R}$. Since we are only interested in finding a subsequence, we can consider the following sequence, indexed by k, $(x_{j_k,1}, \dots, x_{j_k,n})$. So far we have constructed a subsequence of the original for which the first coordinate is a convergent sequence.

Consider now the sequence $x_{j_k,2}$. The sequence is of course bounded and therefore, by Bolzano-Weierstrass, it has a subsequence $x_{j_{k_l},2}$ which converges to $x_2^* \in \mathbb{R}$. Notice that since $x_{j_k,1}$ is convergent, so is $x_{j_{k_l},1}$. Therefore, if we consider the sequence indexed by l, $(x_{j_{k_l},1}, \ldots, x_{j_{k_l},n})$, we now have convergent sequences in the first two components.

It is hopefully clear that, aside from running out letters (and having to resort to cleverer notation), we can repeat this procedure n times, iteratively constructing subsequences to ensure that every component is convergent.

□

3.5 Continuity

3.5.1 Definitions of continuity and continuous limit

We define continuity following the results in year 1 (see Definition 1.2). The only changes are in the dimension of the domain and the target of the function. We consider $U \subset \mathbb{R}^n$, $p \in U$ and a function $f \colon U \to \mathbb{R}^k$.

Definition 3.19 (ε-δ Definition of Continuity)**.** *Given $f \colon U \subset \mathbb{R}^n \to \mathbb{R}^k$ we say that f is continuous at p if,*

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \text{ such that, for } x, p \in U,\ \ \|x - p\| < \delta \ \Rightarrow\ \|f(x) - f(p)\| < \varepsilon.$$

Notice that the two norms $\|\cdot\|$ in the definition above corresponds to norms in different spaces, namely $\mathbb{R}^n$ and $\mathbb{R}^k$, but we dot make a distinction in the notation.

Definition 3.20 (Sequential Definition of Continuity)**.** *$f \colon U \subset \mathbb{R}^n \to \mathbb{R}^k$ is continuous at p if, for every sequence (x_j) in U which converge to p, $(f(x_j))$ converges to $f(p)$.*

Exercise 3.4. *Check that these two definitions are equivalent*

Hint: The argument is the same as that given in First Year Analysis. That the ε-δ definition implies the sequential definition is straightforward. The converse proceeds by proving the contrapositive, i.e., one assumes the failure of the ε-δ definition and then one constructs a sequence x_j in U which converges to p but for which $f(x_j)$ does not converge to $f(p)$.

We say that f is continuous, without specifying a particular point, if it is continuous at all points of its domain. If we wish to emphasize the domain U on which f is continuous, then we say that f is continuous on U.

Notation 3.21. *The space of functions continuous on U with values in $\mathbb{R}^k$ is denoted by $C(U, \mathbb{R}^k)$ or $C^0(U, \mathbb{R}^k)$.*[4] *When $k = 1$, we simply write $C(U)$ or $C^0(U)$.*

Definition 3.22 (Continuous limit)**.** *$f \colon U \to \mathbb{R}^k$ has a (continuous) limit at $p \in U$ if there exists $q \in \mathbb{R}^k$ such that*

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \text{ such that, } x \in U \text{ and } 0 < \|x - p\| < \delta \ \Rightarrow\ \|f(x) - q\| < \varepsilon.$$

We then write $\lim_{x \to p} f(x) = q$.

Just as for limits of sequences, continuous limits are unique. It is also clear that f is continuous at p if, and only if, $\lim_{x \to p} f(x) = f(p)$. Notice that the definition of continuous limit tacitly assumes that there exist points in U, different from x, which are arbitrarily close to x.

[4] The superscript will later denote the number of derivatives.

3.5.2 Separate continuity

In this subsubsection, we shall restrict ourselves to the case $n = 2$ and to functions defined on all of $\mathbb{R}^2$. The generalisation to higher dimensions is straightforward but the presentation is much easier in two dimensions using x and y as variables.

Given a real valued function $f(x, y)$, we consider two families of functions $\{g^y : \mathbb{R} \to \mathbb{R}\}_{y\in\mathbb{R}}$ and $\{h^x : \mathbb{R} \to \mathbb{R}\}_{x\in\mathbb{R}}$ defined by

$$g^y(x) := f(x, y) =: h^x(y). \tag{3.11}$$

Thus g^s is the restriction of f to the horizontal line $y = s$ and h^t is the restriction of f to the vertical line $x = t$.

Definition 3.23 (Separate continuity). *A function $f : \mathbb{R}^2 \to \mathbb{R}$ is* separately continuous *at (x_0, y_0) if g^{y_0} is continuous at x_0 as a function of x and h^{x_0} is continuous at y_0 as a function of y.*

Two natural questions arise:

(i) Does continuity imply separate continuity?

(ii) Does separate continuity imply continuity?

Exercise 3.5. *Prove that continuity implies separate continuity.*

As for question (ii), the example below shows that separate continuity does not imply continuity.

Example 3.24. *Define $f : \mathbb{R}^2 \to \mathbb{R}$ by*

$$f(x, y) := 1\,, \ \text{ if } xy \neq 0, \qquad f(x, y) := 0\,, \ \text{ if } xy = 0.$$

$g^0(x) = 0$ for every $x \in \mathbb{R}$ and $h^0(y) = 0$ for every $y \in \mathbb{R}$. In particular, both g^0 and h^0 are continuous at 0 and therefore, f is separately continuous at $(0, 0)$.

However, f is not continuous at $(0, 0)$ because $\lim_{(x,y)\to(0,0)} f(x, y)$ does not exist. We can establish this by finding two sequences (a_j, b_j) and (α_j, β_j) both of which converge to $(0, 0)$ but for which $\lim_{j\to\infty} f(a_j, b_j) \neq \lim_{j\to\infty} f(\alpha_j, \beta_j)$; we also require $(a_j, b_j) \neq (0, 0)$ and $(\alpha_j, \beta_j) \neq (0, 0)\ \forall j \in \mathbb{N}$. So take, for example, $a_j = \alpha_j = \beta_j = 1/j$ and $b_j = 0$. Then $f(a_j, b_j) = 0$ and $f(\alpha_j, \beta_j) = 1\ \forall j$ and therefore $\lim_{j\to\infty} f(a_j, b_j) = 0 \neq 1 = \lim_{j\to\infty} f(\alpha_j, \beta_j)$. By uniqueness of limits, if $\lim_{(x,y)\to(0,0)} f(x, y)$ were to exist, $\lim_{j\to\infty} f(a_j, b_j)$ and $\lim_{j\to\infty} f(\alpha_j, \beta_j)$ would have to have the same value. Since they do not, we conclude that $\lim_{(x,y)\to(0,0)} f(x, y)$ does not exist.

The following are easy to verify (left as exercises):

- f is continuous at all points (x, y) such that $xy \neq 0$,
- f is not separately continuous at points $(x, 0)$ such that $x \neq 0$ (because h^x is then not continuous at 0) and similarly,
- f is not separately continuous at points $(0, y)$ such that $y \neq 0$.

3.5.3 Basic properties of continuous functions

Throughout this subsubsection, $U \subset \mathbb{R}^n$, $p \in U$ and $a, b \in \mathbb{R}$.

The following results are basic properties of continuous functions with effectively the same proof as in one dimension.

Proposition 3.25 (The sum of continuous functions is continuous). *If $f, g : U \to \mathbb{R}^k$ are both continuous at p then, $af + bg$ is continuous at p.*

The proof of this is just an application of the sum rule for limits of sequences of vectors.

Proposition 3.26 (The product of a continuous *scalar (real) valued* function with a continuous vector-valued function is continuous)**.** *If $f\colon U \to \mathbb{R}$ and $g\colon U \to \mathbb{R}^k$ are both continuous at p then, fg is continuous at p where $(fg)(x) := (f(x))(g(x))$.*

The proof of this is just an application of the product rule for a convergent sequence of real numbers and a convergent sequence of vectors.

Exercise 3.6. *Suppose that $f\colon U \to \mathbb{R}^k$ is continuous at p. Prove that if $f(p) \neq 0$ then there exists $\delta > 0$ such that $\|f(x)\| > \frac{1}{2}\|f(p)\|\ \forall\, x \in U$ for which $\|x - p\| < \delta$.*

Exercise 3.7. *Suppose that $f\colon U \to \mathbb{R}$ is continuous at $p \in U$ and $f(x) \neq 0\ \forall\, x \in U$. Prove that $1/f$ is continuous at p.*

Corollary 3.27. *Suppose that $f\colon U \to \mathbb{R}$ and $g\colon U \to \mathbb{R}^k$ are both continuous at p and that $f(x) \neq 0\ \forall\, x \in U$. Then g/f is continuous at p.*

Proposition 3.28 (The composition of continuous functions is continuous)**.** *If $U \subset \mathbb{R}^n,\ V \subset \mathbb{R}^k,\ f\colon U \to \mathbb{R}^k$ is continuous at $p \in U,\ f(U) \subset V,\ g\colon V \to \mathbb{R}^m$ is continuous at $f(p) \in V$, then $g \circ f\colon U \to \mathbb{R}^m$ is continuous at p.*

The proof of this is just an application of the sequential definition of continuity.

Proposition 3.29 (Componentwise continuity)**.** *Recall that $f\colon U \to \mathbb{R}^k$ can be written as*

$$(x_1, \ldots, x_n) = x \mapsto f(x) = (f_1(x_1, \ldots, x_n), \ldots, f_k(x_1, \ldots, x_n)), \quad x \in U.$$

f is continuous at p if, and only if, $\forall\, i \in \{1, \ldots, k\},\ f_i\colon U \to \mathbb{R}$ is continuous at p.

Remark 3.30. *The proposition above says that f is continuous if, and only if, all of its component functions are continuous. The proof is a straightforward application of the sequential definition of continuity and the equivalence of convergence and componentwise convergence for sequences in $\mathbb{R}^n$.*

This proposition suggests that most (but not all[5]) of the features related to the continuity (and, as we shall see, differentiability) of functions $f\colon \mathbb{R}^n \to \mathbb{R}^k$ that are different from those of functions $f\colon \mathbb{R} \to \mathbb{R}$ arise when $n > 1$; whether k is greater than 1 is less significant.

3.5.4 Constructing continuous functions of several variables from continuous real valued functions of a single real variable.

A function like $f(x, y) = \dfrac{xy}{x^2 + y^2}$ looks continuous on $\mathbb{R}^2 \setminus \{(0, 0)\}$, but how do we prove it without resorting to the ε-δ definition or the sequential definition of continuity? We know $g(x) = x$ and $h(x) = x^2$ are continuous as functions of the *single* real variable x. However, what we need to know is that the functions $\gamma\colon \mathbb{R}^2 \to \mathbb{R}$ and $\eta\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\gamma(x, y) := x, \qquad \eta(x, y) := x^2$$

are continuous as functions of *two* variables. That is precisely the content of the next proposition, which follows from the following easy lemma.

Lemma 3.31. *Write $\mathbb{R}^{n+\ell}$ as $\mathbb{R}^n \oplus \mathbb{R}^\ell$, that is*

$$\mathbb{R}^{n+\ell} = \{(x, y) : x \in \mathbb{R}^n,\ y \in \mathbb{R}^\ell\}.$$

Denote by π_1 and π_2 the two projections of $\mathbb{R}^{n+\ell}$ onto $\mathbb{R}^n$ and $\mathbb{R}^\ell$ respectively:

$$\pi_1(x, y) := x, \qquad \pi_2(x, y) := y, \qquad x \in \mathbb{R}^n,\ y \in \mathbb{R}^\ell.$$

Then π_1 and π_2 are continuous.

[5] See, for instance, Corollary 4.25.

Proof. Fix $(x_0, y_0) \in \mathbb{R}^{n+\ell}$ and, given $\varepsilon > 0$, choose $\delta = \varepsilon$. Then

$$\|(x,y) - (x_0,y_0)\| < \delta \;\Rightarrow\; \|\pi_1(x,y) - \pi_1(x_0,y_0)\| = \|x - x_0\| \leqslant \|(x,y) - (x_0,y_0)\| < \varepsilon,$$

that is, π_1 is continuous. The continuity of π_2 is proved similarly. □

Proposition 3.32. *Consider $E \subset \mathbb{R}$, $a \in E$ and a function $g\colon E \to \mathbb{R}$. For $i \in \{1, \ldots, n\}$, define $\pi_i\colon \mathbb{R}^n \to \mathbb{R}$ by*

$$\pi_i(x_1, \ldots, x_i, \ldots, x_n) := x_i$$

and let $U_i := \pi_i^{-1}(E) := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_i \in E\}$. Define $f\colon U_i \to \mathbb{R}$ by $f(x_1, \ldots, x_n) := g(x_i)$, that is, $f(x) = g(\pi_i(x)) = g \circ \pi_i(x)$. Suppose that g is continuous at a. Then f is continuous at all points of $\pi_i^{-1}\{a\} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_i = a\}$.

Proof. By Lemma 3.31, π is continuous on $\mathbb{R}^n$ and therefore, by the continuity of composition of continuous functions, $f = g \circ \pi_i$ is continuous on $\pi_i^{-1}\{a\}$. □

We can use Proposition 3.32 and the results in section 3.5.3 to prove the continuity of $f(x,y) = \dfrac{xy}{x^2+y^2}$ on $\mathbb{R}^2 \setminus \{(0,0)\}$ as follows. Consider the four functions, each defined on $\mathbb{R}^2$ by

$$\gamma(x,y) := x, \qquad \eta(x,y) := x^2, \qquad \sigma(x,y) := y, \qquad \tau(x,y) := y^2.$$

Proposition 3.32 tells us that the continuity of these four functions follows from the continuity (proved in First Year Analysis) of $g(t) = t$ and $h(t) = t^2$ as functions of the *single* real variable t. Now

$$f(x,y) = \frac{(\gamma(x,y))(\sigma(x,y))}{(\eta(x,y)) + (\tau(x,y))}$$

and therefore, the continuity of f on $\mathbb{R}^2 \setminus \{(0,0)\}$ follows from the continuity of the product, sum and quotient of continuous functions at points where the denominator does not vanish.

A similar approach can be followed for most functions given by explicit formulas. However, the continuity of a function at points where the function is given special values (not by a formula) has to be investigated by separate arguments.

The following two examples are intended to clarify what is meant by 'natural domain of definition' of a function defined by an expression involving familiar continuous functions. The natural domain of

$$F(x,y) = \frac{x^2 \sin(y)}{e^x - \cosh y}$$

is $\mathbb{R}^2 \setminus \{(\log(\cosh(y)), y) : y \in \mathbb{R}\}$ and F is continuous on this set. Similarly,

$$f(x,y,z) := \left(\frac{\log(x+y)}{\sin z}\,,\, \arccos(y)\sqrt{1 + (\cos(xe^z))^2}\right)$$

is continuous on $\{(x,y,z) \in \mathbb{R}^3 : x + y > 0,\ -1 \leqslant y \leqslant 1,\ z \neq n\pi, n \in \mathbb{Z}\}$.

Proposition 3.32 is, in a sense, 'obvious' and you need not quote it explicitly when appealing to the continuity of functions given by expressions similar to the ones above.

3.5.5 Caution with taking limits in dimension $\geqslant 2$

If $a \in \mathbb{R}$ then a can be approached from only two directions, namely left and right. So, $\lim_{x\to a} f(x)$ exists if the right-hand limit $\lim_{x\to a+} f(x)$ and the left-hand limit $\lim_{x\to a-} f(x)$ both exist and are equal.

The situation is much more complicated in higher dimensions. It suffices to illustrate the issue by considering some of the many different ways of approaching $(0,0)$ in $\mathbb{R}^2$. We may, for instance, approach $(0,0)$ along any line $ax+by = 0$. We can also follow more complicated paths from a point in $\mathbb{R}^2\setminus\{(0,0)\}$ to

$(0,0)$. For example, we can proceed along (x, x^2), i.e. along the parabola $y = x^2$. Indeed, we can approach $(0,0)$ along the graph of any continuous function $\psi(x)$ for which $\psi(0) = 0$. This still does not exhaust all possibilities because, for instance, we may approach $(0,0)$ along a spiral like $(t\cos(1/t), t\sin(1/t),\ t > 0$. So, by Proposition 3.28, if $f\colon \mathbb{R}^2 \to \mathbb{R}$ is continuous at $(0,0)$ then $\lim_{t\to 0} f(\varphi(t), \psi(t))$ would have to exist for any pair of functions $\varphi, \psi\colon \mathbb{R} \to \mathbb{R}$ that are continuous at 0 and equal to 0 there. This should make it clear that continuity is much more restrictive than separate continuity and indeed, more restrictive than continuity along lines, which we shall now define.

Definition 3.33 (continuity along lines, also called linear continuity)**.** *A function* $f\colon \mathbb{R}^n \to \mathbb{R}^k$ *is* continuous along lines *(also referred to as* linearly continuous*) at* x_0 *if the restriction* f^L *of* f *to the line* L *passing through* x_0 *is continuous for every such line* L*.*

The line L through x_0 in the direction of $v \in \mathbb{R}^n$ is parameterised by $r_v(t) := x_0 + tv$. Therefore f is continuous along lines at x_0 if $f \circ r_v$ is continuous at $t = 0$ for *every choice* of $v \in \mathbb{R}^n$. In particular,

$$\lim_{t\to 0} f(x_0 + tv) = f(x_0)\ \ \forall\, v \in \mathbb{R}^n,$$

that is, $\lim_{t\to 0} f(x_0 + tv)$ is *independent* of v. We have seen above that continuity implies continuity along lines.

In the next example we will exhibit **a function which is separately continuous at all points of** $\mathbb{R}^2$ **but which is not continuous along lines through** $(0,0)$**.**

Example 3.34. *Define* $f\colon \mathbb{R}^2 \to \mathbb{R}$ *by*

$$f(x,y) = \frac{xy}{x^2+y^2}\,, \quad \text{if } (x,y) \neq (0,0), \qquad f(0,0) := 0.$$

For $y \neq 0,\ g^y$ (defined by (3.11)) is a continuous function of x and $g^0(x) = 0\ \forall\, x \in \mathbb{R}$. Therefore, g^y is continuous for any choice of $y \in \mathbb{R}$. By similar reasoning, h^x is continuous for any choice of $x \in \mathbb{R}$. This shows that f is separately continuous at all points of $\mathbb{R}^2$. However, note that

$$\left.\begin{aligned} f(t,t) &= \tfrac{1}{2},\\ f(t,2t) &= \tfrac{2}{5},\\ \text{and } f(t,-t) &= -\tfrac{1}{2} \end{aligned}\right\}\ \forall\, t \in \mathbb{R}\setminus\{0\}.$$

Therefore, $\lim_{(x,y)\to(0,0)} f(x,y)$ depends on the line in $\mathbb{R}^2$ along which we approach $(0,0)$. In particular, it is not possible to assign any value to f at $(0,0)$ that would make it continuous along lines through $(0,0)$.

Remark 3.35. *(This is not examinable) One may be tempted to think that a separately continuous function fails to be continuous only at isolated points. This is not the case. For a fixed* $(a,b) \in \mathbb{R}^2$ *define* $f_{(a,b)}\colon \mathbb{R}^2 \to \mathbb{R}$ *by* $f_{(a,b)}(x,y) := f(x-a, y-b)$ *where* f *is as in Example 3.34. Let* (a_n, b_n) *be an enumeration of* $\mathbb{Q}\times\mathbb{Q}$*, i.e., of all points in* $\mathbb{R}^2$ *both of whose coordinates are rational. Then define* $F\colon \mathbb{R}^2 \to \mathbb{R}$ *by*

$$F(x,y) := \sum_{n=0}^{\infty} 2^{-n} f_{(a_n,b_n)}(x,y).$$

It can be shown that F *is separately continuous, but discontinuous precisely on* $\mathbb{Q}\times\mathbb{Q}$*. Conceptually, the sum in the definition of* F *disperses the discontinuity of* f *to all the rational points in* $\mathbb{R}^2$*.*

The next example exhibits **a function** $f\colon \mathbb{R}^2 \to \mathbb{R}$ **which is continuous along lines through** $(0,0)$ **but which is not continuous at** $(0,0)$**.**

Example 3.36. *Define* $f\colon \mathbb{R}^2 \to \mathbb{R}$ *by* $f(x,y) = 1$ *if* $0 < y < x^2$ *and* $f(x,y) = 0$ *otherwise. Show that* $\lim_{t\to 0} f(tv) = 0 = f(0,0)\ \forall\, v \in \mathbb{R}^2$*. However, show also that* f *is discontinuous at* $(0,0)$*.*

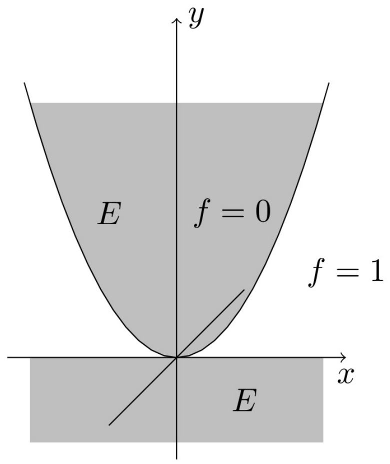

Figure 3.1: Diagram of the function

In the diagram below, the grey shaded region E is the set on which $f = 0$, i.e.,

$$E := \{(x,y) : f(x,y) = 0\} = \{(x,y) : y \leqslant 0 \text{ or } y \geqslant x^2\}.$$

Given $v \in \mathbb{R}^2$, $\exists\, \tau > 0$ such that, $|t| < \tau \;\Rightarrow\; tv \in E$. (If $v = (a,b)$, take $\tau = b/a^2$ if $ab \neq 0$ and $\tau = +\infty$ if $ab = 0$.) In other words, $f(tv) = 0\ \forall\, t \in (-\tau,\tau)$. It follows that $\lim_{t\to 0} f(tv) = 0 = f(0,0)\ \forall\, v \in \mathbb{R}^2$, as claimed.

Finally, $\lim_{x\to 0} f(x, \frac{1}{2}x^2) = 1 \neq 0 = f(0,0)$ which shows that f is discontinuous at $(0,0)$.

In the next example, we consider the following question: Suppose we demand that $f\colon \mathbb{R}^2 \to \mathbb{R}$ be continuous along *any* line in $\mathbb{R}^2$ and not just the ones that pass through a chosen point. Would f then have to be continuous? Remarkably, this is still not the case, as demonstrated by the following example.

Example 3.37. *Define* $f\colon \mathbb{R}^2 \to \mathbb{R}$ *by*

$$f(x,y) = \frac{x^2 y}{x^4 + y^2}, \quad \text{if } (x,y) \neq (0,0), \qquad f(0,0) := 0.$$

For each $k \in \mathbb{R}$, define $g^k\colon \mathbb{R} \to \mathbb{R}$ to be the restriction of f to the line $y = kx$ through the origin in $\mathbb{R}^2$, i.e.,

$$g^k(x) := f(x,kx) = \frac{kx}{x^2 + k^2}\ \forall\, x \in \mathbb{R}.$$

Also define $g^\infty\colon \mathbb{R} \to \mathbb{R}$ to be the restriction of f to the y-axis, i.e.,

$$g^\infty(x) := f(0,y) = 0\ \forall\, y \in \mathbb{R}.$$

We see that, for any choice of $k \in \mathbb{R} \cup \infty$, g^k is continuous.

We now consider the restriction of f to the parabola $y = x^2$. This is given by $\varphi\colon \mathbb{R} \to \mathbb{R}$ where

$$\varphi(x) := f(x,x^2) = \frac{x^2 x^2}{x^4 + (x^2)^2} = \frac{1}{2}, \quad \text{if } x \neq 0, \qquad \varphi(0) = f(0,0) = 0.$$

Thus φ is not continuous. It follows from Propositon 3.28 that, since $g(x) := (x,x^2)$ is continuous, f cannot be continuous at $(0,0)$ because, if it were, then $\varphi = f \circ g$ would also have to be continuous, which it is not. Indeed, we have also shown that $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist. In particular, the function f is continuous away from $(0,0)$ (we will not show this). It is also continuous along lines at $(0,0)$, but it is not continuous at $(0,0)$.

Chapter 4

Rudiments of topology of $\mathbb{R}^n$ and Continuity

4.1 Closed and open subsets of $\mathbb{R}^n$

Definition 4.1 (Closed set). *$X \subset \mathbb{R}^n$ is defined to be* closed *if, whenever (x_j) is a sequence of points in X which converges to $x \in \mathbb{R}^n$ then the limit x also belongs to X.*

Definition 4.2 (Open set). *$U \subset \mathbb{R}^n$ is defined to be* open *if, $\forall\, x \in U,\ \exists\ \varepsilon > 0$ such that $y \in \mathbb{R}^n$ and $\|y - x\| < \varepsilon \ \Rightarrow\ y \in U$.*

By convention, *the empty set is defined to be both open and closed.*

Proposition 4.3. *A set is open if, and only if, its complement is closed.*

Proof. Suppose that $U \subset \mathbb{R}^n$ is open and that U^c is not empty (if U^c is empty, then it is closed by definition). In order to show that U^c is closed, we consider a sequence (x_j) in U^c which converges to $x \in \mathbb{R}^n$. We have to show that x lies in U^c. If it does not, then, since U is open, $\exists\ \varepsilon > 0$ such that $\|y - x\| < \varepsilon \ \Rightarrow\ y \in U$. But $\lim_{j\to\infty} x_j = x$ and therefore, $\exists\ N \in \mathbb{N}$ such that

$$j \geqslant N \ \Rightarrow\ \|x_j - x\| < \varepsilon \ \Rightarrow\ x_j \in U$$

which contradicts the assumption that $x_j \notin U\ \forall\, j \in \mathbb{N}$.

For the converse, let X be a closed subset of $\mathbb{R}^n$ whose complement X^c is nonempty. To prove that X^c (which we assume to be nonempty) is open, we have to show that, given $y \in X^c\ \exists\ \varepsilon > 0$ such that $\|x - y\| \geqslant \varepsilon\ \forall\, x \in X$. ($\varepsilon$ is allowed to depend on y but not on $x \in X$). If this were not the case, then, we could find $y \in X^c$ and a sequence (x_j) in X such that $\|x_j - y\| \leqslant 1/j$. But then $y = \lim_{j\to\infty} x_j$ and, since X is closed, y must belong to X, contrary to the assumption that $y \notin X$. □

Remark 4.4. *Most textbooks first define an open set and then define a closed set to be the complement of an open set. Of course, these textbooks then have to prove that a closed set satisfies Definition 4.1.*

The definition of open set motivates the following definition.

Definition 4.5 (Open (Euclidean) ball). *The* open ball *of radius $r > 0$ centred at $a \in \mathbb{R}^n$ is denoted by $\mathbb{B}(a, r)$ or $\mathbb{B}_r(a)$ and is defined by*

$$\mathbb{B}_r(a) \equiv \mathbb{B}(a, r) := \{x \in \mathbb{R}^n : \|x - a\| < r\}.$$

We abbreviate $\mathbb{B}_r(0)$ to $\mathbb{B}_r$ and $\mathbb{B}_1$ to just $\mathbb{B}$.

The definition of an open set can now be rephrased as

$$U \subset \mathbb{R}^n \text{ is } \textit{open} \text{ if, } \forall\, x \in U,\ \exists\ \varepsilon > 0 \text{ such that } \mathbb{B}_\varepsilon(x) \subset U.$$

This is the definition of an open set that is given in most textbooks. The following proposition justifies the use of the adjective 'open' in the definition of an open ball.

Proposition 4.6. *An open ball is open, i.e., it satisfies the definition of an open set.*

Proof. For each $y \in \mathbb{B}(a,r)$ we need to find $\rho_y > 0$ so that the open ball $\mathbb{B}(y,\rho_y) \subset \mathbb{B}(a,r)$.

To this end, set $\rho_y = r - \|y-a\|$. (This value of ρ_y is suggested by a picture which you should draw.) Then, since $\|y-a\| < r$, we have $\rho_y > 0$ and, for $x \in \mathbb{B}(y,\rho_y)$,

$$\|x-a\| \leqslant \|x-y\| + \|y-a\| < \rho_y + \|y-a\| = r,$$

i.e., the open ball $\mathbb{B}(y, r-\|y-a\|) \subset \mathbb{B}(a,r)$ as required. □

Definition 4.7 (Closed ball)**.** *The* closed ball *of radius $r > 0$ centred at $a \in \mathbb{R}^n$ is denoted by $\overline{\mathbb{B}(a,r)}$ or $\overline{\mathbb{B}_r(a)}$ and is defined by*

$$\overline{\mathbb{B}_r(a)} \equiv \overline{\mathbb{B}(a,r)} := \{x \in \mathbb{R}^n : \|x-a\| \leqslant r\}.$$

We abbreviate $\overline{\mathbb{B}_r(0)}$ to $\overline{\mathbb{B}_r}$ and $\overline{\mathbb{B}_1}$ to just $\overline{\mathbb{B}}$.

The following proposition justifies the use of the adjective 'closed' in the definition of a closed ball.

Proposition 4.8. *A closed ball is closed, i.e., it satisfies the definition of a closed set.*

Sketch proof. This proposition can be proved in at least two ways. One way is to prove that the complement of a closed ball is open. Another way is to prove that, if x_j is a sequence in $\overline{\mathbb{B}(a,r)}$ which converges to x, then $\|x-a\| \leqslant r$, i.e., $x \in \overline{\mathbb{B}(a,r)}$. □

Proposition 4.9 (An arbitrary union of open sets is open)**.** *If U_λ is open for all $\lambda \in \Lambda$, where Λ is an indexing set (which could be uncountable), then $\bigcup_{\lambda\in\Lambda} U_\lambda$ is open.*

Proof. If $p \in \bigcup_{\lambda\in\Lambda} U_\lambda$ then $\exists\ \lambda^* \in \Lambda$ such that $p \in U_{\lambda^*}$. But U_{λ^*} is open and therefore, $\exists\ \varepsilon > 0$ such that $\mathbb{B}(p,\varepsilon) \subset U_{\lambda^*}$. In particular, $\mathbb{B}(p,\varepsilon) \subset \bigcup_{\lambda\in\Lambda} U_\lambda$, i.e., $\bigcup_{\lambda\in\Lambda} U_\lambda$ is open. □

Definition 4.10 (ε-neighbourhood)**.** *Let E be* any *subset of $\mathbb{R}^n$. Given $\varepsilon > 0$, the* ε-neighbourhood *$\mathscr{N}(E,\varepsilon)$ of E is defined by*

$$\mathscr{N}(E,\varepsilon) := \bigcup_{x\in E} \mathbb{B}(x,\varepsilon).$$

By the previous proposition, $\mathscr{N}(E,\varepsilon)$ is open.

Example 4.11. *Let $E := \mathbb{Q} \cap [0,1] \subset \mathbb{R}$. Enumerate the rational numbers in E by a sequence $x_1, x_2, \ldots$. Given $\varepsilon > 0$, let*

$$\mathscr{O} := \bigcup_{j=1}^{\infty} (x_j - 2^{-j}\varepsilon\,,\, x_j + 2^{-j}\varepsilon).$$

Then $\mathscr{O}$ is open. The sum of the lengths of the intervals that make up $\mathscr{O}$ is $\varepsilon\sum_{j=1}^{\infty} 2^{1-j} = 2\varepsilon$. Therefore, if $\varepsilon < \frac{1}{2}$, $\mathscr{O}$ cannot contain all the irrationals between zero and 1. This example shows how complicated open sets can be.

Proposition 4.12 (The finite intersection of open sets is open)**.**
If $U_1, U_2, \ldots, U_m$ are all open, then $\bigcap_{j=1}^m U_j$ is also open.

Proof. If $p \in \bigcap_{j=1}^m U_j$ then

$$\exists\ \varepsilon_1 > 0 \text{ such that } \mathbb{B}(p,\varepsilon_1) \subset U_1,$$
$$\ldots\,,$$
$$\exists\ \varepsilon_m > 0 \text{ such that } \mathbb{B}(p,\varepsilon_m) \subset U_m.$$

Set $\varepsilon := \min\{\varepsilon_1, \ldots, \varepsilon_m\} > 0$. Then $\mathbb{B}(p,\varepsilon) \subset \bigcap_{j=1}^m U_j$, i.e., $\bigcap_{j=1}^m U_j$ is open. □

Corollary 4.13. *An arbitrary intersection of closed sets is closed and the finite union of closed sets is closed.*

Sketch proof. Consider the complements of the relevant closed sets and apply the preceding propositions together with de Morgan's laws on complements, unions and intersections. □

Remark 4.14. *Note that a subset of $\mathbb{R}^n$ may be neither open, nor closed. For example $[0,1)$ in $\mathbb{R}$ or*

$$\{(x,y)|\, x^2+y^2 \leq 1\} \cap \{(x,y)\,|y>0\}.$$

Note, too, that $\emptyset$ and $\mathbb{R}^n$ are both open and closed.

Terminology. The collection of open subsets of $\mathbb{R}^n$ is called a *topology* of $\mathbb{R}^n$.

4.2 Continuity and topology

4.2.1 Continuity in terms of open sets

The ε-δ definition of continuity at p for a function $f\colon U \to \mathbb{R}^k,\ p \in U \subset \mathbb{R}^n$ can be phrased as

$$\forall\, \varepsilon > 0\ \exists\ \delta > 0 \text{ such that } f(\mathbb{B}(p,\delta) \cap U) \subset \mathbb{B}(f(p),\varepsilon). \tag{4.1}$$

Equivalently,

$$\forall\, \varepsilon > 0\ \exists\ \delta > 0 \text{ such that } \mathbb{B}(p,\delta) \cap U \subset f^{-1}(\mathbb{B}(f(p),\varepsilon)). \tag{4.2}$$

Informally and pictorially, f is continuous at p if $f(x)$ can be guaranteed to stay near $f(p)$ (ε-near) by requiring x to stay sufficiently close (δ-close) to p in U.

Theorem 4.15 (Continuity via open sets and closed sets)**.** *The following statements are equivalent.*

(i) $f\colon \mathbb{R}^n \to \mathbb{R}^k$ is continuous at all points of $\mathbb{R}^n$.

(ii) for all open subsets V of $\mathbb{R}^k,\ f^{-1}(V)$ is open.

(iii) for all closed subsets $\mathscr{F}$ of $\mathbb{R}^k,\ f^{-1}(\mathscr{F})$ is closed.

Proof. Suppose that f is continuous at all points of U and let $V \subset \mathbb{R}^k$ be open. Then, for each $p \in f^{-1}(V)$, $\exists\ \varepsilon > 0$ such that $\mathbb{B}(f(p),\varepsilon) \subset V$. By continuity of f at p as stated in (4.2), $\exists\ \delta(p) > 0$ such that $\mathbb{B}(p,\delta(p)) \subset f^{-1}(\mathbb{B}(f(p),\varepsilon)) \subset f^{-1}(V)$, which shows that $f^{-1}(V)$ is open. (We have assumed that $f^{-1}(V)$ is not empty; if it is, it would still be an open set.)

Conversely, given $p \in \mathbb{R}^n$ and $\varepsilon > 0$, the ball $\mathbb{B}(f(p),\varepsilon)$ is open and therefore, it follows (by assumption) that $f^{-1}(\mathbb{B}(f(p),\varepsilon))$ is open. In particular, $\exists\ \delta > 0$ such that $\mathbb{B}(p,\delta) \subset f^{-1}(\mathbb{B}(f(p),\varepsilon))$, which is precisely the statement of the ε-δ definition of continuity as expressed by (4.2).

Finally, the equivalence of (ii) and (iii) follows from $f^{-1}(V^c) = (f^{-1}(V))^c$ and the fact that a set is closed if, and only if, its complement is open. □

Remark 4.16. *If $f\colon \mathbb{R}^n \to \mathbb{R}^k$ is continuous, then the image of an open set need not be open. For instance, consider the constant map $f(x) = 0\ \forall\, x \in \mathbb{R}^n$.*

If $f\colon \mathbb{R}^n \to \mathbb{R}^k$ is continuous, then the image of a closed set need not be closed. For instance, consider the map $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = \dfrac{x^2}{x^2+1}\ \forall\, x \in \mathbb{R}$. Then $f(\mathbb{R}) = [0,1)$ which is neither a closed subset nor an open subset of $\mathbb{R}$.

Example 4.17 (Open and closed sets via Theorem 4.15)**.** *Show that*

(i) the unit sphere $S^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$ is a closed subset of $\mathbb{R}^n$.

(ii) The set $E := \{(x,y) \in \mathbb{R}^2 : xy\big(\sin(1/x)\big)\big(\cos(1/y)\big) > -1\}$ *is an open subset of* $\mathbb{R}^2$.

Solution.

(i) Define $f \in C(\mathbb{R}^n)$ by $f(x) = \|x\|$. Then $S^{n-1} = f^{-1}(\{1\})$ which, by part (iii) of Theorem 4.15, is a closed subset of $\mathbb{R}^n$ because $\{1\}$ is a closed subset of $\mathbb{R}$.

(ii) Define $f \colon \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x,y) := xy\big(\sin(1/x)\big)\big(\cos(1/y)\big) \text{ if } xy \neq 0, \qquad f(x,y) = 0 \text{ if } xy = 0.$$

Then f is continuous and $E = f^{-1}((-1,\infty))$ which, by part (ii) of Theorem 4.15, is an open subset of $\mathbb{R}^2$ because $(-1,\infty)$ is an open subset of $\mathbb{R}$.

Remark 4.18. *Close inspection of the proof of Theorem 4.15 will reveal that if* $U \subset \mathbb{R}^n$ *is open then* $f \colon U \to \mathbb{R}^k$ *is continuous at all points of* $\mathbb{R}^n$ *if, and only if, for all open subsets* V *of* $\mathbb{R}^k$, $f^{-1}(V)$ *is open. However, it is no longer true that the preimage of a closed set is necessarily closed, that is, statement (iii) of Theorem 4.15 no longer applies.*

Similarly, if $U \subset \mathbb{R}^n$ *is closed then* $f \colon U \to \mathbb{R}^k$ *is continuous at all points of* $\mathbb{R}^n$ *if, and only if, for all closed subsets* $\mathscr{F}$ *of* $\mathbb{R}^k$, $f^{-1}(\mathscr{F})$ *is closed. However, it is no longer true that the preimage of an open set is necessarily open, that is, statement (ii) of Theorem 4.15 no longer applies.*

The extension of Theorem 4.15 to functions $f \colon U \to \mathbb{R}^k$*, where* U *is an arbitrary subset of* $\mathbb{R}^n$*, requires the notion of sets that are* open/closed relative to U. *We will not explore these notions in this module.*

4.2.2 Continuity and sequential compactness

Definition 4.19 (Sequentially compact subset)**.** $K \subset \mathbb{R}^n$ *is* sequentially compact *if every sequence* x_j *in* K *has a convergent subsequence* x_{j_ℓ} *whose limit is in* K.

Definition 4.20. $X \subset \mathbb{R}^n$ *is* bounded *if* $\exists\, M > 0$ *such that* $\|x\| \leqslant M \;\forall\, x \in X$.

Theorem 4.21. $K \subset \mathbb{R}^n$ *is sequentially compact if, and only if,* K *is closed and bounded.*

Proof. Suppose that K is sequentially compact. To prove that K is closed, we consider a sequence x_j in K which converges to $x \in \mathbb{R}^n$ and then we have to show that $x \in K$. By the sequential compactness of K, x_j has a subsequence x_{j_ℓ} whose limit is in K. But $x = \lim_{j\to\infty} x_j = \lim_{\ell\to\infty} x_{j_\ell} \in K$. The proof that K is closed is complete.

To prove that K is bounded, assume, for a contradiction, that it is unbounded. Then there exists a sequence x_j in K such that $\|x_j\| \geqslant j \;\forall\, j \in \mathbb{N}$. By the sequential compactness of K, x_j has a subsequence x_{j_ℓ} whose limit is in K. In particular, x_{j_ℓ} is bounded, i.e., $\exists\, M > 0$ such that $\|x_{j_\ell}\| \leqslant M \;\forall\, \ell \in \mathbb{N}$. But by definition of subsequence, $j_\ell \geqslant \ell$ and, by the way the sequence x_j was chosen, $\|x_{j_\ell}\| \geqslant j_\ell$. Therefore,

$$M \geqslant \|x_{j_\ell}\| \geqslant j_\ell \geqslant \ell \;\forall\, \ell \in \mathbb{N}.$$

This clearly cannot hold and we conclude that K must be bounded.

We now assume that K is closed and bounded and prove that it is sequentially compact. So consider an arbitrary sequence x_j in K. Since K is bounded, x_j must be bounded and, by the Bolzano-Weierstrass theorem, it has a convergent subsequence x_{j_ℓ} whose limit x must be in K, since K is closed. The proof that K is sequentially compact is complete. □

Theorem 4.21 is important because it enables us to determine easily whether a set is sequentially compact. For instance, the theorem asserts that a closed ball $\overline{\mathbb{B}(a,r)}$ is sequentially compact without having to check whether all its sequences contain a convergent subsequence! Similarly, we can assert that the sphere $\mathbb{S}^{n-1}(a,r) := \{x \in \mathbb{R}^n : \|x - a\| = r\}$ is sequentially compact; it is clearly bounded and we showed above (for $a = 0$ and $r = 1$, but the proof is virtually identical) that it is closed.

Theorem 4.22 (Continuity preserves sequential compactness). *If $f\colon K \to \mathbb{R}^k$ is continuous and K is sequentially compact then $f(K)$ is also sequentially compact.*

Proof. Let (y_j) be a sequence in $f(K)$. Then, for each $j \in \mathbb{N}$, $\exists\ x_j \in K$ such that $f(x_j) = y_j$. By the sequential compactness of K, there exists a convergent subsequence (x_{j_ℓ}) of (x_j) such that $\lim_{\ell\to\infty} x_{j_\ell} = x \in K$. By continuity of f at x, $\lim_{\ell\to\infty} y_{j_\ell} = \lim_{\ell\to\infty} f(x_{j_\ell}) = f(x) \in f(K)$, i.e., $f(K)$ is sequentially compact. □

Theorem 4.23 (Extreme Value Theorem). *Let $K \subset \mathbb{R}^n$ be sequentially compact and let $f\colon K \to \mathbb{R}$ be continuous. Then $\exists\ x^*, x_* \in K$ such that*

$$f(x_*) \leqslant f(x) \leqslant f(x^*)\ \forall\, x \in K.$$

This theorem asserts that a continuous real valued function on a sequentially compact space *attains* its extreme values, i.e., max and min. This theorem was proved in First Year Analysis in the case that K is a closed, bounded interval. It is one of the most important theorems of elementary mathematical analysis because, for instance, it is used in the proof of Rolle's Theorem which, in turn, is used in the proof of Taylor's theorem.

Proof of Extreme Value Theorem. By the previous theorem and Theorem 4.21, $f(K) \subset \mathbb{R}$ must be closed and bounded. Therefore, $M := \sup f(K)$ and $m := \inf f(K)$ are both finite because $f(K)$ is bounded. By definition of sup and inf, there exist sequences $a_j, b_j \in f(K)$ such that $\lim_{j\to\infty} a_j = m$ and $\lim_{j\to\infty} b_j = M$. But $f(K)$ is closed and therefore, $m, M \in f(K)$, i.e., $\exists\ x^*, x_* \in K$ such that

$$f(x_*) = m \leqslant f(x) \leqslant M = f(x^*)\ \forall\, x \in K.$$

□

Remark 4.24. *The notion of supremum cannot be extended from $\mathbb{R}$ to $\mathbb{R}^k,\ k \geqslant 2$. That is why we had to restrict ourselves to scalar functions in the preceding theorem. The best we can do for vector-valued functions is stated in the following corollary.*

Corollary 4.25. *Let $K \subset \mathbb{R}^n$ be sequentially compact and let $f\colon K \to \mathbb{R}^k$ be continuous. Then $\exists\ x^*, x_* \in K$ such that*

$$\|f(x_*)\| \leqslant \|f(x)\| \leqslant \|f(x^*)\|\ \forall\, x \in K.$$

Proof. Let us note that the function $g : K \to \mathbb{R}$ given by $g(x) := \|f(x)\|$ is continuous. In order to see this, we use the triangle inequality to obtain that for all $x, y \in K$, we have

$$|g(x) - g(y)| = |\|f(x)\| - \|f(y)\|| \leq \|f(x) - f(y)\|\,. \tag{4.3}$$

More precisely, given $x \in K$ and $\varepsilon > 0$, by continuity of f at x, it follows that there exists $\delta > 0$ such that $\|f(x) - f(y)\| < \varepsilon$ for all $y \in K$ with $\|x - y\| < \delta$. Substituting this into (4.3), we obtain that $|g(x) - g(y)| < \varepsilon$ for all such y. The result now follows from Theorem 4.23. □

Chapter 5

The space of linear maps and matrices

Notation 5.1.

(i) The space of linear maps, *i.e.,* $\{A : \mathbb{R}^n \to \mathbb{R}^k \mid A \text{ is linear}\}$, *shall be denoted by* $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ *and* $L(\mathbb{R}^n, \mathbb{R}^n)$ *shall be abbreviated to* $L(\mathbb{R}^n)$.[1]

(ii) The space of $k \times n$ matrices with real entries *shall be denoted by* $\mathbb{R}^{k,n}$.[2]

To a matrix

$$(a_{ij}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kn} \end{pmatrix} \in \mathbb{R}^{k,n}$$

we associate (subconsciously?!) $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ defined by

$$\mathbb{R}^n \ni x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto Ax := \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^k. \tag{5.1}$$

Let $\{v_1, \dots, v_n\}$ and $\{w_1, \dots, w_k\}$ be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively, i.e.,

$$v_j = (0, \dots, 0, \underset{\substack{\uparrow \\ j^{\text{th}}\text{position among } n \text{ entries}}}{1}, 0, \dots, 0) \in \mathbb{R}^n, \qquad w_i = (0, \dots, 0, \underset{\substack{\uparrow \\ i^{\text{th}}\text{position among } k \text{ entries}}}{1}, 0, \dots, 0) \in \mathbb{R}^k.$$

Then,

$$Av_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{kj} \end{pmatrix} = \sum_{i=1}^{k} a_{ij} w_i, \quad j \in \{1, \dots, n\}, \tag{5.2}$$

and therefore, (a_{ij}) is the matrix representation of A with respect to the standard bases on $\mathbb{R}^n$ and $\mathbb{R}^k$.

On a few occasions it shall be useful to express this association of (a_{ij}) with A as defined above more formally as a map $\mu \colon \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k) \to \mathbb{R}^{k,n}$, i.e.,

$$\mu(A) := (a_{ij}), \quad \text{where } A \text{ and } (a_{ij}) \text{ are related by (5.2).} \tag{5.3}$$

It is easy to verify that μ is a linear isomorphism. Moreover, since we shall be using standard bases on $\mathbb{R}^n$ and $\mathbb{R}^k$ throughout (unless otherwise eexampleicitly stated), we shall switch between the linear map A and the associated matrix $\mu(A) = (a_{ij})$ without warning.

[1] Other notations in use are $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R}^k)$ and $\mathrm{End}(\mathbb{R}^n)$.

[2] Other notations in use are $\mathbb{R}^{k \times n}$, $M(k \times n, \mathbb{R})$, $M_{k \times n}(\mathbb{R})$ and $M_{kn}(\mathbb{R})$. $M(n, \mathbb{R})$ is sometimes used as an abbreviation of $M(n \times n, \mathbb{R})$.

It is easy to check that the identification

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kn} \end{pmatrix} \longleftrightarrow (a_{11}, \dots, a_{1n}, a_{21}, \dots, a_{2n}, \dots, a_{k1}, \dots, a_{kn}) \tag{5.4}$$

between $\mathbb{R}^{k,n}$ and $\mathbb{R}^{nk}$ is a linear isomorphism. It follows that

$$\dim(\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)) = \dim(\mathbb{R}^{k,n}) = nk.$$

5.1 Two norms on the space of linear maps and matrices

We shall be discussing the continuity of maps like

$$f\colon \mathbb{R}^n \to L(\mathbb{R}^k, \mathbb{R}^\ell), \quad f\colon L(\mathbb{R}^n) \to \mathbb{R} \quad \text{and} \quad f\colon L(\mathbb{R}^n) \to L(\mathbb{R}^n).$$

As we have seen, this necessitates defining a notion of distance (norm) on $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$, or equivalently, a norm on $\mathbb{R}^{k,n}$. The first such notion that comes to mind is to use the identification (5.4) and define the so-called *Frobenius norm* $\|\cdot\|_F$ by

$$\|(a_{ij})\|_F := \left(\sum_{i=1}^{k} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}. \tag{5.5}$$

This is fine, but we shall make more use of the operator norm (defined below) as it turns out to be more convenient.[3]

The operator norm arises from studying how large $|Ax|$ can get relative to $|x|$ as x ranges over $\mathbb{R}^n$. We can do this using (5.1) and the Cauchy-Schwarz inequality:

$$\|Ax\|^2 = \sum_{i=1}^{k} \Big(\sum_{j=1}^{n} a_{ij} x_j \Big)^2 \leqslant \sum_{i=1}^{k} \left(\Big(\sum_{j=1}^{n} a_{ij}^2 \Big) \Big(\sum_{j=1}^{n} x_j^2 \Big) \right) = \Big(\sum_{i=1}^{k} \sum_{j=1}^{n} a_{ij}^2 \Big) \|x\|^2 = \|(a_{ij})\|_F^2 \, \|x\|^2.$$

In particular, for $x \neq 0$ we have

$$\sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|^2}{\|x\|^2} \leqslant \|(a_{ij})\|_F^2. \tag{5.6}$$

This makes possible the following definition.

Definition 5.2. *The* operator norm $|||A|||$ *of* $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ *is defined by*

$$|||A||| := \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}. \tag{5.7}$$

In practice, (5.7) *is often used in the form*

$$\|Ax\| \leqslant |||A||| \, \|x\| \ \forall\, x \in \mathbb{R}^n. \tag{5.8}$$

Also in practice, one may somehow establish that $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ *satisfies*

$$\|Ax\| \leqslant M \, \|x\| \ \forall\, x \in \mathbb{R}^n \ \text{ for some } M > 0.$$

Then (5.7) *implies that* $|||A||| \leqslant M$.

The expression $\dfrac{\|Ax\|}{\|x\|}$ *can be equivalently written as* $\left\| \dfrac{1}{\|x\|} Ax \right\|$ *and, by the linearity of* A, $\left\| \dfrac{1}{\|x\|} Ax \right\| = \left\| A \Big(\dfrac{x}{\|x\|} \Big) \right\|$. *Since* $\left\| \Big(\dfrac{x}{|x|} \Big) \right\| = 1$, *we have the following equivalent definition of* $|||A|||$*:*

$$|||A||| := \sup_{\|x\|=1} \|Ax\|. \tag{5.9}$$

[3]Observe that $\|\mu(A)\|_F^2 = \operatorname{trace}\big((\mu(A))^T (\mu(A))\big) = \operatorname{trace}\big((\mu(A))(\mu(A))^T\big)$, where the superscript T denotes transpose and the trace of a matrix is the sum of its diagonal entries.

Observe that the supremum in (5.9) is being taken over a sequentially compact set (namely, the unit sphere S^{n-1} in $\mathbb{R}^n$). We shall see in Proposition 5.5 below that this can be an advantage of (5.9) over (5.7).

5.1.1 Comparison of the two norms

Recalling from (5.3) that $\mu(A) = (a_{ij})$ we can rewrite (5.6) as

$$|||A|||^2 \leqslant \|\mu(A)\|_F^2. \tag{5.10}$$

From (5.2) we see that

$$\|\mu(A)\|_F^2 = \|(a_{ij})\|_F^2 := \sum_{j=1}^n \sum_{i=1}^k a_{ij}^2 = \sum_{j=1}^n |Av_j|^2 \leqslant \|A\|^2 \sum_{j=1}^n |v_j|^2 = n\|A\|^2. \tag{5.11}$$

Combining (5.11) and (5.10) gives the comparison

$$\frac{1}{\sqrt{n}}\|\mu(A)\|_F \leqslant |||A||| \leqslant \|\mu(A)\|_F. \tag{5.12}$$

5.1.2 Properties of the operator norm

The properties of $|||\cdot|||$ in the next proposition justify calling it a norm.

Proposition 5.3. *In (i), (ii) and (iii) below,* $A, B \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ *and* $a \in \mathbb{R}$.

(i) $|||A||| = 0 \;\Leftrightarrow\; A = 0.$

(ii) $|||aA||| = |a|\,|||A|||.$

(iii) Triangle inequality. $|||A + B||| \leqslant |||A||| + |||B|||.$

Proof. The first two items are elementary and the proofs are left to the reader. For the third item,

$$\|(A+B)x\| = \|Ax + Bx\| \leqslant \|Ax\| + \|Bx\| \leqslant (|||A||| + |||B|||)\|x\|$$

and therefore

$$|||A + B||| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|(A+B)x\|}{\|x\|} \leqslant |||A||| + |||B|||.$$

□

Proposition 5.4 (Composition bound)**.** $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ *and* $B \in L(\mathbb{R}^k, \mathbb{R}^m) \;\Rightarrow\; BA \in L(\mathbb{R}^n, \mathbb{R}^m)$ *and* $\|BA\| \leqslant \|B\|\,\|A\|$.

Proof. $\|(BA)(x)\| = \|B(Ax)\| \leqslant |||B|||\|Ax\| \leqslant |||B|||\,|||A|||\,|x|$ and therefore, $|||BA||| \leqslant |||B|||\,|||A|||$. □

Proposition 5.5. $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ *is injective if, and only if,* $\exists\, \alpha > 0$ *such that* $\|Ax\| \geqslant \alpha|x|\ \forall\, x \in \mathbb{R}^n$. *(Note that* k *does not have to be equal to* n*.)*

Proof. If $Ax = 0$ and $\|Ax\| \geqslant \alpha\|x\|$ for some $\alpha > 0$ then $x = 0$, i.e., A is injective.

The converse is proved by establishing the contrapositive, i.e., suppose that there is a sequence x_j in $\mathbb{R}^n \setminus \{0\}$ such that $\|Ax_j\|/\|x_j\| \to 0$ as $j \to \infty$. Set $u_j := x_j/\|x_j\|$. Then $|u_j| = 1\ \forall\, j \in \mathbb{N}$ and $Au_j \to 0$ as $j \to \infty$. Since S^{n-1} is sequentially compact, there exists a subsequence u_{j_ℓ} which converges to $u \in S^{n-1}$. Let us note that the map $x \mapsto \|Ax\|$ is continuous. Namely, by the triangle inequality and linearity, we have for all $x, y \in \mathbb{R}^n$ that

$$|\|Ax\| - \|Ay\|| \leq \|Ax - Ay\| = \|A(x-y)\| \leq |||A|||\,\|x - y\|,$$

from where we deduce the continuity of $x \mapsto \|Ax\|$. Therefore, $\|Au\| = \lim_{j\to\infty} \|Au_j\| = 0$, i.e., $u \in \ker(A)$. It follows that A is not injective. □

Remark 5.6. *Proposition 5.5 can be regarded as a* quantitative measure of injectivity. *An injective linear map has to keep a nonzero vector x away from zero. The larger the value of α in the inequality $\|Ax\| \geqslant \alpha\|x\|$, the more the linear map A pushes x away from zero.*

A better way of saying this is to consider a perturbation of A by a matrix B to get the matrix $A+B$. We then have the following

Proposition 5.7. *Suppose that $A, B \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ satisfy*

$$\|Ax\| \geqslant \alpha\|x\| \text{ for some } \alpha > 0 \tag{5.13}$$

and $|||B||| < \alpha$. Then $A+B$ is still injective.

Proof.

$$\|(A+B)x\| \geqslant \|Ax\| - \|Bx\| \geqslant \alpha\|x\| - |||B|||\|x\| = \delta\|x\|,$$

where $\delta := \alpha - |||B||| > 0$. Therefore $(A+B)x = 0 \;\Rightarrow\; x = 0$, which proves that $A+B$ is injective. $\square$

This proposition can be interpreted as saying that if (5.13) *holds then the open ball $\mathbb{B}(A,\alpha) \subset \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$*[4] *is contained in the set of injective linear transformations in $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$.*

So, a larger value of α in (5.13) *indicates that A is able to withstand perturbations by 'larger' (as measured by the operator norm) linear transformations while maintaining injectivity.*

5.2 Convergence and continuity in $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$

These are defined in exactly the same way as for sequences in $\mathbb{R}^n$ and functions $f\colon U \to \mathbb{R}^k,\ U \subset \mathbb{R}^n$. For instance, a sequence $(A_j)_{j\in\mathbb{N}}$ of linear transformations in $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ converges to $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ if, $\forall\,\varepsilon > 0,\ \exists\, N \in \mathbb{N}$ such that $j \geqslant N \;\Rightarrow |||A_j - A||| < \varepsilon$.

Similarly, for $r > 0,\ \mathbb{B}(A, r) := \{B \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k) : |||B - A||| < r\}$.

A moment's thought will reveal that, because of (5.12), we could also use $\|\cdot\|_F$ instead of the operator norm to define these notions. Recall that $\|\cdot\|_F$ on $\mathbb{R}^{k,n}$ is the same as $|\cdot|$ on $\mathbb{R}^{nk}$ via (5.4) and (5.5). Therefore, the completeness[5] of $\mathbb{R}^{nk}$ that was established in Proposition 3.17, immediately implies the completeness of $\mathbb{R}^{k,n}$ with respect to $\|\cdot\|_F$. The completeness of $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ with respect to $\|\cdot\|$ then follows from (5.12).

Property (i) in Proposition (5.3) and the triangle inequality are precisely the properties that are needed to prove the usual properties of limits of sequences and continuous limits, such as uniqueness of limit, sum rule, boundedness of convergent sequences, etc..

5.2.1 Continuity of functions involving matrices or linear maps

A function $f\colon U \to \mathbb{R}^{k,n}$ is continuous at $x \in U$ if $\forall\,\varepsilon > 0,\ \exists\,\delta > 0$ such that $|y - x| < \delta \;\Rightarrow \|f(y) - f(x)\|_F < \varepsilon$. As above, since $\|\cdot\|_F$ on $\mathbb{R}^{k,n}$ is the same as $\|\cdot\|$ on $\mathbb{R}^{nk}$ via (5.4) and (5.5), we can use Proposition 3.29 to assert that

$$x \overset{f}{\mapsto} \begin{pmatrix} a_{11}(x) & \dots & a_{1n}(x) \\ \vdots & & \vdots \\ a_{k1}(x) & \dots & a_{kn}(x) \end{pmatrix} : U \to \mathbb{R}^{k,n}$$

is continuous at x if, and only if, $\forall\, i \in \{1, \dots, k\},\ j \in \{1, \dots, n\},\ x \mapsto a_{ij}(x)$ is continuous at x.

A function $F\colon U \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ is continuous at $x \in U$ if $\forall\,\varepsilon > 0,\ \exists\,\delta > 0$ such that $\|y - x\| < \delta \;\Rightarrow |||F(y) - F(x)||| < \varepsilon$.

[4] This ball is taken with respect to the operator norm on $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$; see §5.2.

[5] A space X is complete if every Cauchy sequence in X converges to an element of X.

Remark 5.8. *Because of* (5.12), *we see that* $F\colon U \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ *is continuous at* $x \in U$ *if, and only if,* $\mu(F)\colon U \to \mathbb{R}^{k,n}$ *is continuous at* $x \in U$.

This remark is very useful because it provides a practical way of checking the continuity of $F\colon U \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$. Namely, we simply have to check whether all the matrix entries of the matrix representation $\mu(F)$ (with respect to the standard bases on $\mathbb{R}^n$ and $\mathbb{R}^k$) are continuous.

The continuity of $f\colon \mathbb{R}^{k,n} \to \mathbb{R}^\ell$ and of $F\colon \mathbb{R}^{k,n} \to \mathbb{R}^{\ell,m}$ is defined similarly by identifying $(\mathbb{R}^{k,n}, \|\cdot\|_F)$ with $(\mathbb{R}^{nk}, \|\cdot\|)$, as in the next proposition and the example below it.

Proposition 5.9 (Continuity of the determinant function)**.** *The map* $\Delta\colon \mathbb{R}^{n,n} \to \mathbb{R}$ *defined by* $\Delta(a_{ij}) := \det(a_{ij})$ *is continuous with respect to the norm* $\|\cdot\|_F$ *on* $\mathbb{R}^{n,n}$.

Proof. The determinant is simply a polynomial of degree n in its n^2 variables

$$a_{11}, \dots, a_{1n}, a_{21}, \dots, a_{2n}, \dots, a_{n1}, \dots, a_{nn}. \tag{5.14}$$

Therefore, its continuity follows from the identifications (5.4) and (5.5) of $(\mathbb{R}^{n,n}, \|\cdot\|_F)$ with $(\mathbb{R}^{n^2}, \|\cdot\|)$ and the usual continuity of polynomials[6] on $\mathbb{R}^{n^2}$. □

Example 5.10. *Define* $F\colon \mathbb{R}^{2,2} \to \mathbb{R}^{2,2}$ *by* $F(A) = A^2$. *If*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \ \textit{then}\ A^2 = \begin{pmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{pmatrix}$$

and therefore, F *can be viewed as* $\varphi\colon \mathbb{R}^4 \to \mathbb{R}^4$ *defined by*

$$\varphi(a, b, c, d) := (a^2 + bc, ab + bd, ac + cd, bc + d^2),$$

which is clearly continuous and therefore F *is continuous.*

[6] See Proposition 3.32 and use the algebraic properties of continuous functions. Each term of $\Delta(a_{ij})$ is, in fact, linear in each of the n^2 variables (5.14) and this is what makes Δ an example of a *multilinear map*. The special case of the determinant of a 2×2 matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ may help to clarify matters. This determinant is then just the function on $\mathbb{R}^4$ defined by $\Delta(a, b, c, d) := ad - bc$, which is easily seen to be continuous by Proposition 3.32 and the algebraic properties of continuous functions.

Chapter 6

The Derivative

From now on, the domain $U \subset \mathbb{R}^n$ of a function $f\colon U \to \mathbb{R}^k$ shall be an *open* subset of $\mathbb{R}^n$, unless otherwise stated. In particular, this means that when $p \in U$ and a limit like $\lim_{x\to p}$ is considered, x is allowed to approach p from any direction.

6.1 Directional derivative

The rate of change of a function of two or more variables depends on the direction in which that change is measured. For example, $f(x, y) = x$ increases as we move to the right along a line parallel to the x-axis but it does not change at all when we move vertically along a line parallel to the y-axis. So we introduce the notion of directional derivative in order to take this dependence on direction into account.

Given $v \in \mathbb{R}^n$ with $v \neq 0$, the line $L_{x,v}$ passing through $x \in \mathbb{R}^n$ in the direction of v is parameterised by $r(t) = x + tv,\ t \in \mathbb{R}$. When $v = 0$, we have that $r(t) = x$ for all t. Since U is open, $\exists\, \tau > 0$ such that $x + tv \in U\ \forall\, t \in (-\tau, \tau)$. The restriction $g_{x,v}$ of f to this segment of $L_{x,v}$ is defined by

$$g_{x,v}(t) := f(x + tv) \in \mathbb{R}^k, \quad |t| < \tau.$$

Since $g_{x,v}(t) = (g_1(t), \ldots, g_k(t))$ is a function of a single real real variable, we can differentiate it component by component in the usual way.[1]

Definition 6.1. *The* directional derivative $\partial_v f(x)$ *is defined by*

$$\begin{aligned}
\partial_v f(x) &:= \frac{d}{dt} g_{x,v}(t)\Big|_{t=0} \\
&= \frac{d}{dt} f(x + tv)\Big|_{t=0} && (6.1)\\
&= \lim_{t\to 0} \frac{f(x + tv) - f(x)}{t}. && (6.2)
\end{aligned}$$

Example 6.2. *Calculate $\partial_v f(x, y)$ for the function $f(x, y) := x^2 - y^2$ in the direction of $v = (a, b)$.*

Solution. $f((x, y) + t(a, b)) = (x + ta)^2 - (y + tb)^2 = x^2 + 2tax + t^2a^2 - y^2 - 2tby - t^2b^2$. Therefore

$$\frac{d}{dt} f(x + tv) = 2ax + 2ta^2 - 2by - 2tb^2,$$

$$\partial_v f(x, y) = \frac{d}{dt} f(x + tv)\Big|_{t=0} = 2ax - 2by.$$

[1] In the definition 3.33 of linear continuity, $g_{x,v}$ was denoted by f^L.

6.1.1 Directional derivative and continuity

A function of one variable that is differentiable must, in particular, be continuous. So it is reasonable to ask the following question.

Suppose that $\partial_v f(x)$ exists for all $v \in \mathbb{R}^n$, does it follow that f is continuous at x?

Somewhat surprisingly, the answer is no!

Example 6.3. *As in Example 3.36, let $f(x, y) = 1$ if $0 < y < x^2$ and $f(x, y) = 0$ otherwise. Show that $\partial_v f(0, 0)$ exists for all $v \in \mathbb{R}^2$ even though f is not continuous at $(0, 0)$!*

Solution. As in Example 3.36, given $v \in \mathbb{R}^2$, $\exists\ \tau > 0$ such that $f(tv) = 0\ \forall\, t \in (-\tau, \tau)$. It follows that $\partial_v f(0, 0) = \lim\limits_{t\to 0} \frac{f(tv)-f(0,0)}{t} = 0\ \forall\, v \in \mathbb{R}^2$.

So, we need a definition of derivative that is more restrictive than partial and directional derivatives just as continuity is more restrictive than separate and linear continuity. The key idea, which we are about to describe, is to regard the derivative as providing a 'best' affine linear approximation of a function. In this process we move away from the 'kinematic' notion of derivative as rate of change and adopt, instead, a 'mapping' viewpoint of functions in which a nonlinear map is approximated by an affine linear one. Hence the need to fully understand linear maps as in a module on Linear Algebra.

6.2 The (Fréchet) Derivative as an affine linear approximation

6.2.1 Affine linear approximation in the 1-variable case

Given $x \in (a, b) \subset \mathbb{R}$, the derivative at x, $f'(x)$, of a function $f\colon (a, b) \to \mathbb{R}$ is defined by

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}. \tag{6.3}$$

This definition cannot be readily extended to functions $f\colon \mathbb{R}^n \to \mathbb{R}^k$ because it is not possible to divide vectors in $\mathbb{R}^k$ by vectors in $\mathbb{R}^n$, even when $n = k$. We can get around this difficulty by rewriting (6.3) as

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - f'(x)\,h\|}{\|h\|} = 0. \tag{6.4}$$

Then, rewriting $f(x+h) - f(x) - f'(x)\,h$ as $f(x+h) - (f(x) + f'(x)\,h)$, we can interpret (6.4) as saying that, for 'small' h, the (nonlinear) mapping $h \mapsto f(x+h)$, x *fixed*, is optimally approximated by the affine linear map $h \mapsto f(x) + f'(x)\,h$. Observe that this is a *mapping* viewpoint of the derivative, which is conceptually different from the hitherto held *kinematic* viewpoint of rate of change. It is this mapping viewpoint which we are able to generalise to the notion of derivative of functions of several variables.

6.2.2 The (Fréchet) Derivative

By analogy with (6.4), we make the following definition.

Definition 6.4. *$f\colon U \to \mathbb{R}^k$ is differentiable at $x \in U$ if $\exists\ A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ such that*

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0, \tag{6.5}$$

As above, rewriting $f(x+h) - f(x) - Ah$ as $f(x+h) - (f(x) + Ah)$, we can interpret this definition as saying that, for 'small' h, the (nonlinear) map $h \mapsto f(x+h)$ is optimally approximated by the affine linear map $h \mapsto f(x) + Ah$.

Exercise 6.1. *Show that the linear map A in* (6.5)*, if it exists, is unique. This justifies saying that the Fréchet derivative provides the optimal linear approximation of f.*

Remark 6.5. *A real number a can also be viewed as the linear map $A\colon \mathbb{R} \to \mathbb{R}$ defined by $Ah = ah$. Indeed a is the 1×1 matrix representation of A with respect to the standard basis 1 of $\mathbb{R}$. Similarly, the real number $f'(x)$ in* (6.3) *is the 1×1 matrix representation of the linear map $h \mapsto f'(x)\, h$ in* (6.4)*.*

Notation 6.6. *If a linear map A that satisfies* (6.5) *exists, it is called the (Fréchet)* derivative *of f at x and it is denoted by $Df(x)$.*

Thus (6.5) can be rewritten as

$$\lim_{h\to 0} \frac{\|f(x+h) - (f(x) + Df(x)\, h)\|}{\|h\|} = 0. \tag{6.6}$$

Exercise 6.1 justifies calling the affine linear map $h \mapsto f(x)+Df(x)h$ the best affine linear approximation of the map $h \mapsto f(x+h)$. The ε-δ formulation of (6.6) provides us with a way of quantifying how good this approximation is. Namely, by definition of continuous limit we have that

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \text{ so that } 0 < \|h\| < \delta \ \Rightarrow\ \frac{\|f(x+h) - (f(x) + Df(x)\, h)\|}{\|h\|} < \varepsilon.$$

Multiplying both sides by $\|h\|$ and then allowing $h = 0$ we deduce that

$$\forall\, \varepsilon > 0,\ \exists\, \delta > 0 \text{ so that } \|h\| < \delta \ \Rightarrow\ \|f(x+h) - (f(x) + Df(x)\, h)\| \leqslant \varepsilon\|h\|. \tag{6.7}$$

We shall refer to (6.7) and (6.6) also as the definition of the derivative $Df(x)$ of f at x. Note that equality has to be allowed in (6.7) because we have allowed the possibility that $h = 0$, which can be convenient in many situations.

Proposition 6.7 (Differentiability implies continuity). *If $f\colon U \to \mathbb{R}^k$ is differentiable at $x \in U$ then f is continuous at x.*

Proof. By (6.7) we have that $\forall\, \varepsilon > 0\ \exists\, \delta > 0$ such that

$$\begin{aligned} \|h\| < \delta \ &\Rightarrow\ \|f(x+h) - (f(x) + Df(x)\, h)\| \leqslant \varepsilon\|h\| \\ &\Rightarrow\ \|f(x+h) - f(x)\| \leqslant \|Df(x)\, h\| + \varepsilon\|h\| \\ &\Rightarrow\ \|f(x+h) - f(x)\| < (\|Df(x)\|_{op} + \varepsilon)\|h\|. \end{aligned}$$

Set $\delta_* := \min\{\delta, \varepsilon/(\|Df(x)\|_{op} + \varepsilon)\}$. Then $|h| < \delta_* \ \Rightarrow\ \|h\| < \delta$ and therefore we can use the above chain of implications to conclude that

$$\|h\| < \delta_* \ \Rightarrow\ \|f(x+h) - f(x)\| < (\|Df(x)\|_{op} + \varepsilon)\delta_* < \varepsilon.$$

We have proved the claim that f is continuous at x. □

6.2.3 Differentiability of components of vector-valued functions

Exercise 6.2. *Given $f\colon U \to \mathbb{R}^k$, $f(x) = (f_1(x), \dots, f_k(x))$, prove that f is differentiable at $x \in U$ if, and only if, for each $i \in \{1, \dots, k\}$, $f_i\colon U \to \mathbb{R}$ is differentiable at x.*

Remark 6.8. *Compare Exercise 6.2 with Proposition 3.29 on componentwise continuity.*

6.2.4 Relation between the derivative and directional derivative

Proposition 6.9. *If $Df(x)$ exists then $\partial_v f(x)$ exists for all $v \in \mathbb{R}^n$ and $\partial_v f(x) = Df(x)v$. In particular,* if f is differentiable at x, *then $\partial_v f(x)$ is linear in v, i.e.,*

$$\partial_{av+bw} f(x) = a\, \partial_v f(x) + b\, \partial_w f(x)\ \forall\, a, b \in \mathbb{R} \text{ and } \forall\, v, w \in \mathbb{R}^n. \tag{6.8}$$

Proof. If $v = 0$ there is nothing to prove. So, we assume that $v \neq 0$. Then, replacing h in (6.6) by tv and removing $\|\cdot\|$ where that is allowed, we get

$$\lim_{t\to 0} \frac{f(x+tv) - f(x) - Df(x)(tv)}{t\|v\|} = 0. \tag{6.9}$$

Multiply both sides of (6.9) by $\|v\|$ and use $Df(x)(tv) = tDf(x)v$ by the linearity of $Df(x)$ so as to get

$$\lim_{t\to 0} \frac{f(x+tv) - f(x)}{t} = Df(x)v, \quad \text{i.e.,} \partial_v f(x) = Df(x)v.$$

Finally, by linearity of $Df(x)$,

$$\partial_{av+bw} f(x) = Df(x)(av+bw) = aDf(x)v + bDf(x)w = a\,\partial_v f(x) + b\,\partial_w f(x).$$

□

Example 6.3 shows that the converse of Proposition 6.9 does not hold.

Remark 6.10. *We give another proof of Proposition 6.9 after we have proved the chain rule; see §6.6.3.*

6.3 Partial derivatives, gradient and Jacobian matrix

Let $\{v_1, \ldots, v_n\}$ be the standard basis of $\mathbb{R}^n$, i.e.,

$$v_i = (0, \ldots, 0, \underset{\substack{\uparrow \\ i^{\text{th}} \text{position among } n \text{ entries}}}{1}, 0, \ldots, 0) \in \mathbb{R}^n.$$

Definition 6.11. *For $1 \leqslant i \leqslant n$, $\partial_{v_i} f(x)$ is called the i^{th}-*partial derivative *of $f\colon U \to \mathbb{R}^k$ at $x \in U$. It is more simply denoted by $\partial_i f(x)$.*

Since

$$\begin{aligned}\partial_{v_i} f(x) &= \lim_{t\to 0} \frac{f(x+tv_i) - f(x)}{t} \\ &= \lim_{t\to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots x_n) - f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots x_n)}{t}\end{aligned}$$

$\partial_i f(x)$ is calculated by differentiating $f(x_1, \ldots, x_n)$ with respect to the i^{th} variable, treating all the other variables as constant[2]. It is therefore also common to write $\frac{\partial f}{\partial x_i}(x)$ or $\frac{\partial}{\partial x_i} f(x_1, \ldots, x_n)$ instead of $\partial_i f(x)$.

Bearing in mind that $f(x) = (f_1(x), \ldots, f_k(x))$ we have

$$\partial_i f(x) = (\partial_i f_1(x), \ldots, \partial_i f_k(x)), \text{ and so, } \partial_i f(x) \text{ is } \textit{a vector in } \mathbb{R}^k.$$

If f is a function of a few variables, say two, it is common to write $f(x, y)$ instead of $f(x_1, x_2)$, and to write f_x instead of $\frac{\partial f}{\partial x}$ or $\partial_1 f$. Similarly, f_y is shorthand for $\frac{\partial f}{\partial y}$. Similar shorthand applies to functions of 3 or 4 variables. For 5 variables or more, it is usually more convenient to number the variables, rather than choose distinct letters!

6.3.1 Algebraic rules for partial derivatives.

Since partial differentiation involves differentiating with respect to a single variable, the usual algebraic rules of differentiation apply. For instance,

$$\text{if } f, g\colon U \to \mathbb{R}^k, \text{ then } \partial_i(f+g) = \partial_i f + \partial_i g.$$

Similarly,

$$\text{if } f\colon U \to \mathbb{R} \text{ and } g\colon U \to \mathbb{R}^k \text{ then } \partial_i(fg) = (\partial_i f)g + f\partial_i g.$$

[2]See Example 6.17 further down.

6.3.2 Gradient and Jacobian matrix

Definition 6.12. *The* Jacobian matrix at x, $\partial f(x)$, *of* $f\colon U \to \mathbb{R}^k$, $f(x) = (f_1(x), \dots, f_k(x))$ but written as a column vector, *is defined by*

$$\partial f(x) = \begin{pmatrix} \partial_1 f_1(x) & \dots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_k(x) & \dots & \partial_n f_k(x) \end{pmatrix} = \begin{pmatrix} \partial_1 f(x) & \dots & \partial_n f(x) \\ \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} \partial f_1(x) & \dots \\ \vdots & \\ \partial f_k(x) & \dots \end{pmatrix}$$

where $\partial_1 f(x), \dots, \partial_n f(x)$ *are the* vector-valued *partial derivatives of the* vector-valued *function* f *(the values of* f *are vectors in* $\mathbb{R}^k$*) and* $\partial f_1, \dots, \partial f_k$ *are the Jacobian* $1 \times n$ *matrices (row vectors) of the* scalar-valued *functions* $f_1, \dots, f_k$.

Definition 6.13. *The* gradient at x, $\nabla f(x)$, *of a* scalar valued function $f\colon U \to \mathbb{R}$ *is defined to be the* column vector

$$\nabla f(x) := \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix}.$$

Thus $\nabla f(x)$ *is the vector in* $\mathbb{R}^n$ *which is the* transpose *of the row vector* $\partial f(x)$, $\nabla f(x) = \big(\partial f(x)\big)^T$.

Remark 6.14. *For a* scalar valued function $f\colon U \to \mathbb{R}$, $\partial f(x)$ *represents a* linear functional *on* $\mathbb{R}^n$ *defined by*

$$\mathbb{R}^n \ni v \mapsto (\partial f(x))(v) := (\partial_1 f(x))v_1 + \dots + (\partial_n f(x))v_n \in \mathbb{R}, \qquad v = (v_1, \dots, v_n).$$

Using the Euclidean inner product, this linear functional $\partial f(x)$ *is identified with the vector* $\nabla f(x)$*:*

$$(\partial f(x))(v) = (\nabla f(x)) \cdot v.$$

However, be warned that the distinction between ∇f *and* ∂f *is often suppressed, even in these notes!*

Proposition 6.15. *If* $f\colon U \to \mathbb{R}^k$ *is differentiable at* $x \in U$ *and* $h \in \mathbb{R}^n$ *then*

$$Df(x)h = \partial f(x)h. \tag{6.10}$$

Remark 6.16. *It is important to appreciate the difference between the two sides of* (6.10)*. On the left hand side we have* the linear map $Df(x)$ *acting on the vector* h *whereas on the right hand side we have* the matrix $\partial f(x)$ *multiplying the vector* h*. In other words,* $\partial f(x)$ is the matrix representation of $Df(x)$ with respect to the standard bases on $\mathbb{R}^n$ and $\mathbb{R}^k$. *More formally,* $\partial f(x) = \mu(Df(x))$ *where* $\mu\colon \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k) \to \mathbb{R}^{k,n}$ *is defined by* (5.3).

Proof of Proposition 6.15. $h = h_1 v_1 + \dots + h_n v_n$ and therefore, by linearity of $Df(x)$,

$$Df(x)h = \sum_{i=1}^{n} h_i Df(x) v_i = \sum_{i=1}^{n} h_i \partial_i f(x) = \partial f(x) h,$$

where, in the second equality, we have used Proposition 6.9. □

Example 6.17. *Calculate the Jacobian matrix* ∂f *of* $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ *defined by*

$$f(x, y, z) := \left(e^{x^3+2y}, \frac{\sin x}{\sqrt{1 + y^2 z^4}} \right).$$

Solution.

$$\partial f(x, y, z) = \begin{pmatrix} 3x^2 e^{x^3+2y} & 2e^{x^3+2y} & 0 \\ \frac{\cos x}{\sqrt{1+y^2z^4}} & -\frac{yz^4 \sin x}{(1+y^2z^4)^{3/2}} & -\frac{2y^2z^3 \sin x}{(1+y^2z^4)^{3/2}} \end{pmatrix}.$$

All the entries of the Jacobian matrix ∂f are continuous functions and we shall see in §6.7 that this implies the existence of the derivative $Df(x, y, z) \in L(\mathbb{R}^3, \mathbb{R}^2)$, which is the linear map defined by

$$Df(x, y, z)(r, s, t) = \begin{pmatrix} 3x^2 e^{x^3+2y} & 2e^{x^3+2y} & 0 \\ \frac{\cos x}{\sqrt{1+y^2z^4}} & -\frac{yz^4 \sin x}{(1+y^2z^4)^{3/2}} & -\frac{2y^2z^3 \sin x}{(1+y^2z^4)^{3/2}} \end{pmatrix} \begin{pmatrix} r \\ s \\ t \end{pmatrix} = \partial f(x, y, z) \cdot \begin{pmatrix} r \\ s \\ t \end{pmatrix}.$$

6.3.3 Why so many different notations for the same thing?!

We have seen in Propositions 6.9 and 6.15 that,

when f is differentiable at x, $Df(x)h = \partial_h f(x)$ and $Df(x)h = \partial f(x)h$.

So why bother with three different ways of writing the same thing?

The reason is that even when $Df(x)$ does not exist, $\partial_h f(x)$ and $\partial f(x)h$ may both still exist but it may happen that they are not equal! (See Example 6.19 below.) In other words, the existence of the Jacobian matrix $\partial f(x)$ does not guarantee that the linear map it defines is the derivative $Df(x)$, *unless $Df(x)$ is known to exist.*

The reader would be right to wonder at this stage how to go about calculating the derivative Df if this cannot be simply done by computing the Jacobian matrix ∂f. Fortunately, as has already been pointed out in Example 6.17, it suffices to verify further that all the entries of ∂f are continuous for then, by Theorem 6.30, Df exists and its matrix representation with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^k$ is given by ∂f.

However, there are situations where it may be necessary to to calculate Df directly from the definition (6.6). This would be the case, for example, at points outside the natural domain of definition of f where it is assigned a special value. There are also situations where it is more convenient, to calculate $\partial_v f$ directly from the definition (6.1), as in §6.5.1.

Remark 6.18. *The notions of partial derivative, directional derivative and (Fréchet) derivative are the differentiable analogues of separate continuity, continuity along lines and continuity.*

Example 6.19. *Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by*

$$f(x,y) = \frac{x^3}{x^2+y^2} \text{ if } (x,y) \neq (0,0), \quad f(0,0) = 0.$$

(i) Show that $\partial_v f(0,0)$ exists for all $v \in \mathbb{R}^2$. In particular, calculate $\partial f(0,0)$.

(ii) Show that $\partial_v f(0,0) \neq \partial f(0,0)\, v$ and that $\partial_v f(0,0)$ is not linear in v.

(iii) Calculate $f_x(x,y)$ and $f_y(x,y)$ for $(x,y) \neq (0,0)$ and show that f_x and f_y are not continuous at $(0,0)$.

(iv) Explain why $Df\,(0,0)$ does not exist.

Solution. We shall let $v = (a,b) \in \mathbb{R}^2$ throughout.

(i) As always, $\partial_{(0,0)} f(0,0) = 0$. If $(a,b) \neq (0,0)$ we have $f(ta,tb) = \dfrac{ta^3}{a^2+b^2}$ $\forall\, t \in \mathbb{R}$ and therefore,

$$\partial_v f(0,0) = \frac{d}{dt} f(ta,tb)\Big|_{t=0} = \frac{d}{dt}\frac{ta^3}{a^2+b^2}\Big|_{t=0} = \frac{a^3}{a^2+b^2}.$$

So, $\partial_v f\,(0,0)$ exists for all $v \in \mathbb{R}^2$ and, in particular,

the Jacobian matrix $\partial f\,(0,0) = (\partial_{(1,0)} f\,(0,0)\,,\, \partial_{(0,1)} f\,(0,0)) = (1,0).$

(ii) $\partial f(0,0)\,(a,b) = a \neq \partial_{(a,b)} f(0,0)$, unless $b = 0$. $\partial_{(a,b)} f(0,0)$ is not linear in (a,b) because $\frac{a^3}{a^2+b^2}$ is not a linear function of a and b.

(iii)
$$f_x(x,y) = \frac{x^2(x^2+3y^2)}{(x^2+y^2)^2}, \qquad f_y(x,y) = \frac{-2x^3y}{(x^2+y^2)^2}.$$
We have seen that $f_x(0,0) = \partial_{(1,0)}f(0,0) = 1$ but $\lim_{y\to 0} f_x(0,y) = 0$. Therefore, f_x is not continuous at $(0,0)$. Similarly $f_y(0,0) = \partial_{(0,1)}f(0,0) = 0$ but $\lim_{x\to 0} f_y(x,x) = -\frac{1}{2}$. Therefore, f_y is also not continuous at $(0,0)$.

(iv) $\partial_v f(0,0)$ is not linear in v and therefore, by Proposition 6.9, $Df\,(0,0)$ does not exist. The situation is similar to that in Example 3.37 of a real valued function of two variables which fails to be continuous even though its restriction to any line in $\mathbb{R}^2$ is continuous. We shall see below that the lack of differentiability of f at $(0,0)$ can be understood geometrically as the failure of the graph of f (which lies in $\mathbb{R}^3$) to have a tangent plane at $(0,0,0)$.

6.4 Geometric approximation and approximation of functions (NOT COVERED IN CLASS AND NOT EXAMINABLE)

6.4.1 Tangent to a curve

Let $r\colon [a,b] \to \mathbb{R}^k$, $r(t) = (x_1(t),\ldots,x_k(t))$, be a continuously differentiable parameterisation of a curve $C = r([a,b]) \subset \mathbb{R}^k$. By this we mean that the functions $\frac{dx_1}{dt},\ldots,\frac{dx_n}{dt}$ are all continuous. Assume that $r'(t) = (\frac{dx_1}{dt},\ldots,\frac{dx_k}{dt}) \neq 0\ \forall\, t \in [a,b]$, i.e., the parameterisation r is *regular*. Using the rate of change definition of derivative given by (6.3), we can then interpret $r'(t)$ as the vector tangent to C at $r(t)$.[3] The line $L_{r(t)}$ tangent to C at $r(t)$ is parameterised by
$$\ell(h) = r(t) + r'(t)h.$$
But $r'(t) = \partial r(t)$ and therefore, the affine linear approximation of $h \mapsto r(t+h)$ by $h \mapsto r(t)+\partial r(t)h = \ell(h)$ is a parameterisation of the tangent line $L_{r(t)}$. In other words, the affine linear approximation of $h \mapsto r(t+h)$ by $h \mapsto r(t) + \partial r(t)h$ for small h corresponds to the geometric approximation of C by $L_{r(t)}$ near $r(t)$.

In the special case that C is itself a line, then $L_{r(t)}$ is the same as C. This is the geometric manifestation of the fact that, as discussed in Example 6.20, the best affine linear approximation of an affine linear map is itself!

6.4.2 Tangent plane of a surface

Let U be an open subset of $\mathbb{R}^2$ and let $r\colon U \to \mathbb{R}^3$ be a continuously differentiable parameterisation of a surface $S = r(U) \subset \mathbb{R}^3$. By this we mean that if $r(u,v) = (x(u,v), y(u,v), z(u,v))$ then all six partial derivatives x_u, y_u, z_u, x_v, y_v and z_v are continuous. Assume that ∂r is of rank 2, the maximal rank that it can have, at all points of U, i.e., the parameterisation r is *regular*. Since
$$r_u = (x_u, y_u, z_u),\ r_v = (x_v, y_v, z_v) \text{ and } \partial r = \begin{pmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{pmatrix}$$
we see that ∂r is of rank 2 if, and only if, r_u and r_v are linearly independent.[4] As in the preceding discussion for a curve C, the affine linear approximation of $(h,k) \mapsto r(u+h, v+k)$ by
$$(h,k) \mapsto r(u,v) + \partial r(u,v)(h,k) = r(u,v) + hr_u(u,v) + kr_v(u,v)$$

[3] We can also view $r(t)$ as the position of a particle at time t and then $r'(t)$, also denoted $\dot{r}(t)$, is the velocity of the particle.

[4] For example,
$$r(u,v) := ((\cos v)(\sin u), (\sin v)(\sin u), \cos u), \quad 0 < v < 2\pi,\ 0 < u < \pi,$$
is a regular parameterisation of the unit sphere minus the prime meridian, i.e., the semicircle running from the North Pole $(0,0,1)$ to the South Pole $(0,0,-1)$ via $(1,0,0)$.

is then a parameterisation of the plane $T_{r(u,v)}S$ tangent to S at $r(u,v)$. Once again, the affine linear approximation of $(h,k) \mapsto r(u+h, v+k)$ for small h and k corresponds to the geometric approximation of S by $T_{r(u,v)}S$ near $r(u,v)$.

6.4.3 Graph of a scalar function of 2 variables

Given $f\colon U \to \mathbb{R},\ U \subset \mathbb{R}^2$, the graph $\mathscr{G}_f$ of f is the surface parameterised by

$$r(x,y) = (x, y, f(x,y)).$$

For example, if $f(x,y) = \sqrt{1-x^2-y^2},\ x^2+y^2 < 1$, then $r(x,y) = (x, y, \sqrt{1-x^2-y^2})$ is another parameterisation of the upper hemisphere.

Note that $r_x = (1,0,f_x)$ and $r_y = (0,1,f_y)$ are linearly independent for any function f. A parameterisation of the plane tangent to G_f at $(x,y,f(x,y))$ is given by

$$\begin{aligned}(h,k) \mapsto r(x,y) + (Dr(x,y))(h,k) &= (x,y,f(x,y)) + h(1,0,f_x) + k(0,1,f_y)\\ &= (x+h, y+k, f(x,y) + hf_x + kf_y)\\ &= (x+h, y+k, f(x,y) + (h,k)\cdot(\nabla f(x,y))).\end{aligned}$$

Thus we see that f is not differentiable at $(x_0,y_0) \in U$ if, and only if, $\mathscr{G}_f$ does not have a tangent plane at $(x_0, y_0, f(x_0,y_0))$. For example, $(x,y) \to |(x,y)| = \sqrt{x^2+y^2}$ is not differentiable at $(0,0)$ because none of its partial derivatives exist at 0. We see this geometrically by noting that the graph of $(x,y) \to |(x,y)|$ on $\mathbb{R}^2$ is a circular cone about the z-axis with an apex at the origin where the cone does not have a tangent plane.

6.4.4 Orders of approximation of a function

For arbitrary values of n and k, it is not possible to provide simple geometric interpretations of $Df(x) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ similar to those presented above; consider, for example, $n = 3$ and $k = 2$. Therefore we have to change our viewpoint when defining the derivative from that of rate of change or tangent line and tangent plane to that of best approximation by a linear map. Linear maps are the simplest maps, after constant maps, and they are fully understood (rank, eigenvalues, etc.) by the methods of linear algebra. We can then transfer this knowledge of linear maps to differentiable maps up to an error that can be quantified by (6.7).

Recalling Taylor's theorem, we see that

(i) a function $h \mapsto f(x+h)$ which is continuous at $h = 0$ admits an approximation by the constant $f(x)$. The error of the approximation is measured by $\varepsilon = \varepsilon\|h\|^0$ and therefore, this approximation is said to be of zeroth order in h.

(ii) a function $h \mapsto f(x+h)$ which is differentiable at $h = 0$ can be approximated by the affine linear map $h \mapsto f(x) + Df(x)\,h$. According to (6.7), the error of the approximation is now measured by $\varepsilon\|h\|$ and therefore, this approximation is said to be of first order (equivalently, linear) in h. Furthermore, for small h, $\varepsilon\|h\| \ll \varepsilon$, i.e., this first order approximation is much better (i.e., the error is smaller) than that demanded by continuity, or even Lipschitz continuity.

(iii) Later on in this module, we shall show that if $h \mapsto f(x+h)$ is twice differentiable at $h = 0$ then it admits an approximation of the form $h \mapsto f(x) + Df(x)\,h + ($*quadratic polynomial* in $h)$. The error of the approximation is now measured by $\varepsilon\|h\|^2$ and therefore, this approximation is said to be of second order (equivalently, quadratic) in h. Quadratic polynomials are also studied in linear algebra under the topic of symmetric bilinear forms.

The above discussion should make clear that, when discussing derivatives of functions of several variables, the significance of derivative moves away from that of rate of change to that of approximation by polynomials which are 'simple' enough to be amenable to detailed study.

6.5 Examples of direct calculation of the derivative from its definition (NOT COVERED IN CLASS AND NOT EXAMINABLE)

Example 6.20. *Show that the derivative of the affine linear map $f\colon \mathbb{R}^n \to \mathbb{R}^k$ defined by*

$$f(x) = Ax + y_0,\ A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k),\ y_0 \in \mathbb{R}^k,$$

is given by $Df(x) = A\ \forall\, x \in \mathbb{R}^n$.

Solution. To calculate $Df(x)$ we need to consider $f(x+h) - f(x)$ and look for the term linear in h:

$$f(x+h) - f(x) = \big(A(x+h) + y_0\big) - (Ax + y_0) = Ah,$$

where, by linearity, $A(x+h) = Ax + Ah$. So, $f(x+h) - f(x) - Ah = 0\ \forall\, h \in \mathbb{R}^n$. It follows from the definition of derivative by (6.6) that $Df(x) = A\ \forall\, x \in \mathbb{R}^n$. We have also shown that, as expected, the best affine linear approximation of the affine linear map f is f itself!

In particular, if $n = 1$ and $f \in L(\mathbb{R}, \mathbb{R}^k)$ is defined by

$$f(t) = tv + y_0, \quad t \in \mathbb{R},\ v, y_0 \in \mathbb{R}^k,$$

then $Df(t)$ is the linear map $h \mapsto hv\colon \mathbb{R} \to \mathbb{R}^k$ and $\partial f(t) = v\ \forall\, t \in \mathbb{R}$. This is an extension of Remark 6.5 and a special case of Remark 6.16. Namely, a vector $v \in \mathbb{R}^k$ can also be viewed as the linear map $A\colon \mathbb{R} \to \mathbb{R}^k$ defined by $Ah = hv$. Indeed v is the $k \times 1$ matrix representation of A with respect to the standard basis $w_1, \ldots, w_k$ of $\mathbb{R}^k$.

6.5.1 Differentiation of matrix-valued functions

The spaces $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ and $\mathbb{R}^{k,n}$ are both vector spaces that can be identified with $\mathbb{R}^{nk}$ and therefore, the definition of derivative given by (6.6) can also be applied to functions with domain and/or range in these spaces. The only change that is needed is the replacement of $\|\cdot\|$ in (6.6) by the operator norm $\|\cdot\|_{op}$ or $\|\cdot\|_F$.

Example 6.21. *Show directly from the definition 6.6 that the derivative of the quadratic map*

$f\colon L(\mathbb{R}^n) \to L(\mathbb{R}^n)$ defined by $f(A) := A^2$

is given by $(Df(A))(H) = AH + HA\ \forall\, H \in L(\mathbb{R}^n)$. (Note that $AH + HA = 2AH$ only if A and H commute, i.e., $AH = HA$.)

Solution. As in Example 6.20, we consider $f(A+H) - f(A)$ and look for the term linear in H:

$$f(A+H) - f(A) = (A+H)(A+H) - A^2 = AH + HA + H^2.$$

The term linear in H is $AH + HA$ and so, we define a linear map $\Lambda_A\colon L(\mathbb{R}^n) \to L(\mathbb{R}^n)$ by $\Lambda_A(H) := AH + HA$. Then, $f(A+H) - f(A) - \Lambda_A(H) = H^2$ and therefore,

$$\lim_{H\to 0} \frac{\|f(A+H) - f(A) - \Lambda_A(H)\|_{op}}{\|H\|_{op}} = \lim_{H\to 0} \frac{\|H^2\|_{op}}{\|H\|_{op}} \leqslant \lim_{H\to 0} \frac{\|H\|_{op}^2}{\|H\|_{op}} = \lim_{H\to 0} \|H\|_{op} = 0.$$

We have shown that $(Df(A))(H) = \Lambda_A(H) = AH + HA$.

Remark 6.22. *We can also calculate the directional derivative $\partial_H f(A)$ directly from its definition:*

$$\begin{aligned}
\partial_H f(A) &= \frac{d}{dt} f(A + tH)\Big|_{t=0} \\
&= \frac{d}{dt}(A^2 + tAH + tHA + t^2H^2)\Big|_{t=0} \\
&= (AH + HA + 2tH^2)\Big|_{t=0} \\
&= AH + HA.
\end{aligned}$$

This is in agreement with Proposition 6.9 according to which $(Df(A))(H) = \partial_H f(A)$.

Indeed, since the entries of $f(A)$ *are quadratic expressions in the entries of* A[5] *then we know that the partial derivatives of* f *with respect to the variables of the entries of* A *are continuous and therefore, as we shall see in §4.7,* f *is differentiable and the calculation of* $(Df(A))(H)$ *can be reduced to that of* $\partial_H f(A)$*, which is often much simpler.*

6.6 The Chain Rule (NOT COVERED IN CLASS AND NOT EXAMINABLE)

Theorem 6.23. *Let* U *and* V *be open subsets of* $\mathbb{R}^n$ *and* $\mathbb{R}^k$ *respectively. Suppose that* $f\colon U \to \mathbb{R}^k$ *is differentiable at* $x \in U$ *and that* $f(x) \in V$*. Suppose further that* $g\colon V \to \mathbb{R}^m$ *is differentiable at* $f(x)$*. Then* $g \circ f\colon \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* x *and*

$$D(g \circ f)(x) = Dg(f(x)) \circ Df(x). \tag{6.11}$$

The following two lemmas will be useful in the proof of the Chain Rule.

Lemma 6.24. *Given* $f\colon U \to \mathbb{R}^k$, $x \in U$, $r > 0$ *such that* $\mathbb{B}(x, r) \subset U$ *and* $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$*, define* $\Delta_{x,A} f\colon \mathbb{B}(0, r) \to \mathbb{R}^k$ *by*

$$\Delta_{x,A} f(h) = \begin{cases} \frac{f(x+h)-f(x)-Ah}{\|h\|}, & \text{if } h \neq 0, \\ 0, & \text{if } h = 0. \end{cases} \tag{6.12}$$

Then f *is differentiable at* x *with* $Df(x) = A$ *if, and only if,* $\Delta_{x,A} f$ *is continuous at* 0.

Proof. If $\Delta_{x,A} f$ is continuous at 0 then $\lim_{h\to 0} \|\Delta_{x,A} f(h)\| = \|\lim_{h\to 0} \Delta_{x,A} f(h)\| = \|\Delta_{x,A} f(0)\| = 0$, Therefore, (6.5) holds and f is differentiable at x with $Df(x) = A$.

Conversely, if f is differentiable at x and we set $A = Df(x)$ in (6.12) then, (6.6) asserts that $\lim_{h\to 0} \|\Delta_{x,A} f(h)\| = 0$. But then $\lim_{h\to 0} \Delta_{x,A} f(h) = 0 = \Delta_{x,A} f(0)$, which is precisely the statement that $\Delta_{x,A} f$ is continuous at 0. □

Notation 6.25. *If* f *is differentiable at* x*, then we let* $\Delta_x f(h)$ *denote* $\Delta_{x,Df(x)} f(h)$.

Lemma 6.26. *Let* $\tau > 0$ *and consider a function* δ *from the open ball* $\mathbb{B}_\tau \subset \mathbb{R}^n$ *to* $\mathbb{R}^k$ *defined by*

$$\delta(h) := \xi(h)\,\eta(h), \; 0 < \|h\| < \tau, \quad \delta(0) := 0,$$

where, $\xi\colon \big(\mathbb{B}_\tau \setminus \{0\}\big) \to \mathbb{R}$ *is bounded and* $\eta\colon \mathbb{B}_\tau \to \mathbb{R}^k$ *is continuous at* $0 \in \mathbb{B}_\tau$ *and* $\eta(0) = 0$*. Then* δ *is continuous at* $0 \in \mathbb{B}_\tau$.

Proof. By continuity of η at 0, given $\varepsilon > 0$, $\exists\, \sigma \in (0, \tau)$ such that $\|h\| < \sigma \;\Rightarrow\; \|\eta(h)\| < \varepsilon$.

By boundedness of ξ, $\exists\, M > 0$ such that $\|\xi(h)\| < M \;\forall\, h \in \mathbb{B}_\tau \setminus \{0\}$.

Therefore, $0 < \|h\| < \sigma \;\Rightarrow\; \|\delta(h)\| < M\varepsilon$, i.e., $\lim_{h\to 0} \delta(h) = 0 = \delta(0)$ and this completes the proof of the lemma. □

Proof of Chain Rule. As in the proof of Proposition 6.7 we have

$$f(x + h) = f(x) + Df(x)h + \Delta_x f(h)\|h\|$$

and

$$g(f(x) + k) = g(f(x)) + Dg(f(x))k + \Delta_{f(x)} g(k)\|k\| \tag{6.13}$$

[5] For instance, if $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ then $f(A) = \left(\begin{smallmatrix} a^2+bc & ab+bd \\ ca+dc & cb+d^2 \end{smallmatrix}\right)$.

where

$$\Delta_x f(h) := \begin{cases} \frac{f(x+h)-f(x)-Df(x)h}{\|h\|}, & \text{if } h \neq 0, \\ 0, & \text{if } h = 0. \end{cases}$$

and

$$\Delta_{f(x)} g(k) := \begin{cases} \frac{g(f(x)+k)-g(f(x))-Dg(f(x))k}{\|k\|}, & \text{if } k \neq 0, \\ 0, & \text{if } k = 0. \end{cases}$$

Set $k(h) := Df(x)h + \Delta_x f(h)\|h\|$ in (6.13). Then, by linearity of $Dg(f(x))$,

$$\begin{aligned} g(f(x+h)) = g(f(x)) &+ Dg(f(x))(Df(x)h) \\ &+ \|h\| Dg(f(x))(\Delta_x f(h)) + \|k(h)\| \, \Delta_{f(x)} g(k(h)). \end{aligned}$$

Therefore,

$$g(f(x+h)) - g(f(x)) - Dg(f(x)) \circ Df(x)h = \|h\| \big(\delta_1(h) + \delta_2(h)\big)$$

where,

$$\delta_1(h) := Dg(f(x))(\Delta_x f(h)),$$

$$\text{and } \delta_2(h) := \frac{\|k(h)\|}{\|h\|} \Delta_{f(x)} g(k(h)), h \neq 0, \quad \delta_2(0) := 0.$$

The proof of the Chain Rule will be complete once we prove that

$$\lim_{h \to 0} \|\delta_1(h)\| = 0 \qquad \text{and} \qquad \lim_{h \to 0} \|\delta_2(h)\| = 0.$$

We start with $\delta_1(h)$.

$$\|\delta_1(h)\| \leqslant \|Dg(f(x))\|_{op} \, \|\Delta_x f(h)\|$$

and, by differentiability of f at x, we have $\lim_{h \to 0} \|\Delta_x f(h)\| = 0$. It follows immediately that $\lim_{h \to 0} \|\delta_1(h)\| = 0$.

We move on to $\delta_2(h)$. For $h \neq 0$, set

$$\xi(h) := \frac{\|k(h)\|}{\|h\|} \leqslant \frac{\|Df(x)h\|}{\|h\|} + \|\Delta_x f(h)\| \leqslant \|Df(x)\|_{op} + \|\Delta_x f(h)\|.$$

The continuity of $\Delta_x f$ at 0 implies that $\xi(h)$ is bounded on $\mathbb{B}_\tau \setminus \{0\}$ for some $\tau > 0$. Next set

$$\eta(h) := \Delta_{f(x)} g(k(h)).$$

$k(h)$ is a continuous function of h and $k(0) = 0$. Therefore, by differentiability of g at $f(x)$ and Proposition 6.7, $\eta(h)$ is a continuous function of h and $\eta(0) = 0$. We may therefore apply Lemma 6.26 to $\delta_2(h) = \xi(h)\,\eta(h)$ to conclude that $\lim_{h \to 0} \|\delta_2(h)\| = 0$.

The proof that $g \circ f$ is differentiable at x and (6.11) holds is complete. □

6.6.1 Jacobian form of chain rule

The linear isomorphism $\mu \colon \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k) \to \mathbb{R}^{k,n}$ defined by (5.3) takes composition of linear transformations to matrix multiplication. Therefore, under the same hypotheses as for Theorem 6.23 we have

$$\partial\, g \circ f(x) = \partial g(f(x)) \cdot \partial f(x) \quad \text{where } \cdot \text{ stands for matrix multiplication.} \tag{6.14}$$

More eexampleicitly,

$$\begin{pmatrix} \partial_1 \, g_1 \circ f(x) & \dots & \partial_n \, g_1 \circ f(x) \\ \vdots & & \vdots \\ \partial_1 \, g_m \circ f(x) & \dots & \partial_n \, g_m \circ f(x) \end{pmatrix} =$$

$$\begin{pmatrix} \partial_1 g_1(f(x)) & \dots & \partial_k g_1(f(x)) \\ \vdots & & \vdots \\ \partial_1 g_m(f(x)) & \dots & \partial_k g_m(f(x)) \end{pmatrix} \begin{pmatrix} \partial_1 f_1(x) & \dots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_k(x) & \dots & \partial_n f_k(x) \end{pmatrix}. \quad (6.15)$$

The entry in the j^{th} row and i^{th} column of $\partial \, g \circ f(x)$ can be written as $\dfrac{\partial}{\partial x_i}(g_j(f(x))$. If we set $y = f(x)$ and we see g as a function of $y = (y_1, \dots, y_k)$ then the entry in the j^{th} row and r^{th} column of $\partial g(f(x))$ can be written as $\dfrac{\partial g_j}{\partial y_r}(f(x))$. Then, (6.15) can be written as

$$\frac{\partial}{\partial x_i}(g_j(f(x)) = \sum_{r=1}^{k} \frac{\partial g_j}{\partial y_r}(f(x)) \, \frac{\partial y_r}{\partial x_i}(x) \quad \text{where, by } \frac{\partial y_r}{\partial x_i}(x) \text{ we mean } \frac{\partial f_r}{\partial x_i}(x). \quad (6.16)$$

(6.16) is perhaps more memorable than (6.15) because we can imagine cancelling ∂y_r from the denominator and numerator in the terms of the sum in (6.16). However, it is important to appreciate the difference between $\dfrac{\partial}{\partial x_i}(g_j(f(x))$ and $\dfrac{\partial g_j}{\partial y_i}(f(x))$. In the first expression, the function $g_j \circ f$ is being differentiated with respect to its i^{th} variable $(1 \leqslant i \leqslant n)$ and evaluated at x whereas in the second expression it is the function g_j that is being differentiated with respect to its i^{th} variable $(1 \leqslant i \leqslant k)$ and then evaluated at $y = f(x)$.

6.6.2 Calculating with the chain rule and gradient

Given $f \colon \mathbb{R}^n \to \mathbb{R}$ and $g \colon \mathbb{R} \to \mathbb{R}$ the i^{th} partial derivative of $g \circ f$ can be computed using (6.16) with $k = 1$ and $m = 1$:

$$\partial_i \, g \circ f(x) = g'(f(x))(\partial_i f(x))$$

and therefore,

$$\nabla \, g \circ f(x) = g'(f(x)) \nabla f(x).$$

In the example that follows, the gradient of a function will be written as a row vector (to make it easier to write)!

Example 6.27. *Calculate* $\nabla \|x\|, \ x \in \mathbb{R}^n \setminus \{0\}$, *by applying the chain rule to* $g \circ f$ *where* $f(x) := \|x\|^2$ *and* $g(t) := \sqrt{t}, \ t > 0$.

Solution. $g'(t) = \frac{1}{2\sqrt{t}}$ and, since $f(x) = x_1^2 + \dots + x_n^2$, we have that $\partial_i f(x) = 2x_i$, i.e., $\nabla f(x) = (\partial_1 f(x), \dots, \partial_n f(x)) = (2x_1, \dots, 2x_n) = 2x$. Therefore,

$$\nabla \|x\| = \nabla \, g \circ f(x) = g'(f(x)) \nabla f(x) = \frac{1}{2\sqrt{\|x\|^2}} \, 2x = \frac{x}{\|x\|}. \quad (6.17)$$

The component form of (6.17) is

$$\frac{\partial}{\partial x_i} \|x\| = \frac{x_i}{\|x\|}.$$

Another common application of the chain rule occurs in the calculation of the derivative of $f \circ r$ where $r \colon \mathbb{R} \to \mathbb{R}^n$ is a parameterisation of a path in $\mathbb{R}^n$ and $f \colon \mathbb{R}^n \to \mathbb{R}$ is a scalar function. In this case, if

$r(t) = (x_1(t), \ldots, x_n(t))$ then

$$\begin{aligned}(f \circ r)'(t) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(r(t)) \frac{dx_i}{dt} \\ &= \nabla f(r(t)) \cdot r'(t). \end{aligned} \tag{6.18}$$

Example 6.28. *Fix $x \in \mathbb{R}^n$ and define $r \colon \mathbb{R} \to \mathbb{R}^n$ by $r(t) := tx$, i.e., if $x \neq 0$ then r is a parameterisation of the line through 0 and x. Given $f \colon \mathbb{R}^n \to \mathbb{R}$ calculate $(f \circ r)'(t)$ in terms of ∇f.*

Solution. $r'(t) = x$ and therefore, by (6.18), $(f \circ r)'(t) = x \cdot (\nabla f(tx))$. Equivalently,

$$\frac{d}{dt} f(r(t)) = x_1 \frac{\partial f}{\partial x_1}(tx) + \cdots + x_n \frac{\partial f}{\partial x_n}(tx).$$

6.6.3 Another proof of Proposition 6.9

Given $v \in \mathbb{R}^n$, $\exists\, \delta > 0$ such that $x + tv \in U \;\forall\, t \in (-\delta, \delta)$. As in Example 6.28 define $r \colon (-\delta, \delta) \to \mathbb{R}^n$ by $r(t) := x + tv$, i.e., r is a parameterisation of a line segment through x in the direction of v. Then $\partial_v f(x) := \frac{d}{dt} f(r(t))\Big|_{t=0}$ and therefore,

$$\begin{aligned}\partial_v f(x) &= Df(x)\, \frac{d}{dt}(x + tv)\Big|_{t=0} \qquad \text{(chain rule)} \\ &= Df(x)v. \end{aligned}$$

6.6.4 Application of chain rule to the verification of a PDE satisfied by a function

Example 6.29. *Suppose that $f \colon \mathbb{R} \to \mathbb{R}$ is differentiable and define $u \colon \mathbb{R}^2 \to \mathbb{R}$ by $u(x,y) := f(x^2 e^{-y})$. Show that u satisfies the PDE $xu_x + 2u_y = 0$.*

Solution.

$$\begin{aligned} u_x(x,y) &= f'(x^2 e^{-y}) \frac{\partial}{\partial x}(x^2 e^{-y}) = 2x e^{-y} f'(x^2 e^{-y}), \\ u_y(x,y) &= f'(x^2 e^{-y}) \frac{\partial}{\partial y}(x^2 e^{-y}) = -x^2 e^{-y} f'(x^2 e^{-y}). \end{aligned}$$

Therefore $xu_x + 2u_y = 2x^2 e^{-y} f'(x^2 e^{-y}) + (-2x^2 e^{-y} f'(x^2 e^{-y})) = 0$.

6.7 Continuity of partial derivatives implies differentiability (NOT COVERED IN CLASS AND NOT EXAMINABLE)

As noted in Example 6.17, partial derivatives are easily computed using the familiar rules of differentiation. Unfortunately, as Example 6.19 shows, the existence of all partial derivatives of a function at all points does not guarantee its differentiability. However, it is worth noting that in Example 6.19 Df was shown to not exist at a point where the partial derivatives are discontinuous. The partial derivatives of f in Example 6.3 do not even exist at points of the form (x, x^2), $x \neq 0$ and f_y does not exist at points on the real axis different from $(0,0)$. These examples raise the possibility that continuity of the partial derivatives may guarantee differentiability. That is the content of the next theorem.

Theorem 6.30. *Consider $f \colon U \to \mathbb{R}^k$ and suppose there exists $\mathbb{B}(x,r) \subset U$ such that the Jacobian matrix $\partial f(y)$ exists at all points of $\mathbb{B}(x,r)$ and that ∂f is continuous at x. Then f is differentiable at x and $Df(x)\,h = \partial f(x)\,h \;\forall\, h \in \mathbb{R}^n$.*

Remark 6.31. *Recall from §5.2.1 that, writing f as $(f_1, \dots, f_k)$, ∂f is continuous at x if, and only if, $\partial_1 f_1,\ \dots,\ \partial_n f_1,\ \partial_1 f_2,\ \dots,\ \partial_n f_2,\ \dots,\ \partial_1 f_k,\ \dots,\ \partial_n f_k$ are all continuous at x.*

Note, too, that we need to make an assumption on the behaviour of the partial derivatives of f at all *points y sufficiently near x in order to conclude the existence of Df at just x.*

Proof. We shall only give the proof in the simplest case $n = 2,\ k = 1$.

For $0 < \|(h_1, h_2)\| < r$ define,

$$\Delta f(h_1, h_2) := f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) - h_1\,\partial_1 f(x_1, x_2) - h_2\,\partial_2 f(x_1, x_2). \tag{6.19}$$

We need to show that

$$\lim_{(h_1,h_2)\to(0,0)} \frac{\Delta f(h_1, h_2)}{\|(h_1, h_2)\|} = 0. \tag{6.20}$$

If we succeed, then we would have proved that $Df(x_1, x_2)$ is the linear map from $\mathbb{R}^2$ to $\mathbb{R}$ defined by

$$Df(x_1, x_2)(h_1, h_2) := h_1\,\partial_1 f(x_1, x_2) + h_2\,\partial_2 f(x_1, x_2).$$

Partial derivatives only provide information along lines parallel to the axes. Therefore, we have to break $f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2)$ into differences along the axes as follows:

$$\begin{aligned} f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) &= \big(f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2)\big) + \big(f(x_1 + h_1, x_2) - f(x_1, x_2)\big) \\ &= \qquad\qquad II \qquad\qquad + \qquad\qquad I. \end{aligned}$$

The second term I can be written in terms of $\partial_1 f$ by applying the mean value theorem to $f(\cdot, x_2)$, x_2 *fixed*. Namely,

$$\exists\ \theta_1 \in (0,1) \text{ such that } f(x_1 + h_1, x_2) - f(x_1, x_2) = h_1 \partial_1 f(x_1 + \theta_1 h_1, x_2). \tag{6.21}$$

Similarly, for the first term II,

$$\exists\ \theta_2 \in (0,1) \text{ such that } f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) = h_2 \partial_2 f(x_1 + h_1, x_2 + \theta_2 h_2). \tag{6.22}$$

Substituting (6.21) and (6.22) in the definition (6.19) of $\Delta f(h_1, h_2)$ we get

$$\begin{aligned} \Delta f(h_1, h_2) = h_1\big(&\partial_1 f(x_1 + \theta_1 h_1, x_2) - \partial_1 f(x_1, x_2)\big) \\ &+ h_2\big(\partial_2 f(x_1 + h_1, x_2 + \theta_2 h_2) - \partial_2 f(x_1, x_2)\big). \end{aligned} \tag{6.23}$$

By continuity of $\partial_1 f$ and $\partial_2 f$ at (x_1, x_2), given $\varepsilon > 0,\ \exists\ \delta > 0$ (which we may, and shall, assume to be less than r) such that

$$\begin{aligned} \|(\tilde h_1, \tilde h_2)\| < \delta\ \Rightarrow\ &\|\partial_1 f(x_1 + \tilde h_1, x_2 + \tilde h_2) - \partial_1 f(x_1, x_2)\| < \varepsilon \\ \text{and}\quad &\|\partial_2 f(x_1 + \tilde h_1, x_2 + \tilde h_2) - \partial_2 f(x_1, x_2)\| < \varepsilon. \end{aligned} \tag{6.24}$$

Now $\|(h_1, \theta_2 h_2)\| < \|(h_1, h_2)\|$ and $\|(\theta_1 h_1, 0)\| < \|(h_1, h_2)\|$ and therefore, if $\|(h_1, h_2)\| < \delta$ then, using (6.24) in (6.23) with $(\tilde h_1, \tilde h_2) = (\theta_1 h_1, 0)$ in the first term and $(\tilde h_1, \tilde h_2) = (h_1, \theta_2 h_2)$ in the second term, yields

$$\|\Delta f(h_1, h_2)\| < \varepsilon(|h_1| + |h_2|) \leqslant \varepsilon\sqrt{2}\,\|(h_1, h_2)\|.$$

In other words, we have established (6.20) and completed the proof of Theorem 6.30. ☐

6.7.1 The space of continuously differentiable functions

Definition 6.32. *Suppose that* $f: U \to \mathbb{R}^k$ *is differentiable on* U*. Then* f *is said to be* continuously differentiable at $p \in U$ *if the map* $x \mapsto Df(x): U \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ *is continuous at* p*. More eexampleicitly,*

$$\forall \varepsilon > 0,\ \exists\, \delta > 0 \text{ such that } \|x - p\| < \delta \ \Rightarrow\ \|Df(x) - Df(p)\|_{op} < \varepsilon.$$

Proposition 6.33. $f: U \to \mathbb{R}^k$ *is continuously differentiable on* U *if, and only if,* $\partial f: U \to \mathbb{R}^{k,n}$ *is continuous on* U*.*

Proof. Recall that $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ and $\mathbb{R}^{k,n}$ are identified via the map $\mu: \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k) \to \mathbb{R}^{k,n}$ defined by (5.3) and that, by Proposition 6.15, if $Df(x)$ exists, then $\partial f(x) = \mu(Df(x))$. Furthermore, by the discussion in §5.2.1, the continuity at p of $x \mapsto Df(x): U \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^k)$ implies the continuity at p of $x \mapsto \partial f(x): U \to \mathbb{R}^{k,n}$.

Conversely, if ∂f is continuous on U then Theorem 6.30 assures us that Df exists at all points of U and that $\partial f = \mu \circ Df$. We can then appeal again to the discussion in §5.2.1 to assert that the continuity at p of ∂f implies the continuity at p of Df. □

This proposition is useful because it provides us with a practical way of checking continuous differentiability, namely, we simply have to compute all the first order partial derivatives $\partial_i f_j$ of $f = (f_1, \ldots, f_k)$ and verify that they are all continuous. This means that most functions that one can eexampleicitly write down in terms of polynomials, exponential, logarithm, etc. are continuously differentiable on their 'natural' domain of definition.

Notation 6.34.

$$C^1(U, \mathbb{R}^k) := \{f: U \to \mathbb{R}^k \mid \partial f: U \to \mathbb{R}^{k,n} \text{ is continuous}\}.$$

$C^1(U) := C^1(U, \mathbb{R})$.

Chapter 7

Complex Analysis

This part of the course is an introduction to complex analysis. The main topics will be complex differentiability, power series and contour integrals. Basic notions and properties for complex numbers were introduction in Year 1 and we only provide a quick review here.

7.1 Review of basic facts about $\mathbb{C}$

The field of complex numbers is given by

$$\mathbb{C} = \{z = x + \mathrm{i}y, \quad x, y \in \mathbb{R}\},$$

with $\mathrm{i}^2 = -1$. For $z = x + \mathrm{i}y$ as above we say that x is the real part of z, denoted by $x = \mathbf{Re}\, z$ and that y is the imaginary part of z, denoted by $y = \mathbf{Im}\, z$. By $|z|$ we denote the modulus (or norm) of z, given by $\sqrt{x^2 + y^2}$. We denote by $\bar{z}$ the complex conjugate of z. That is, if $z = x + \mathrm{i}y$ then $\bar{z} = x - \mathrm{i}y$. It is easy to see that

1. $\bar{\bar{z}} = z$,
2. $\overline{z + w} = \bar{z} + \bar{w}$,
3. $\overline{zw} = \bar{z}\bar{w}$,
4. $|z|^2 = z\bar{z}$ and $|\bar{z}| = |z|$.

Notice that we can identify $\mathbb{C}$ with $\mathbb{R}^2$, simply by identifying $z = x + \mathrm{i}y$ with (x, y). In this way $|z|$ corresponds to the Euclidean norm in $\mathbb{R}^2$. We will not use $\|\cdot\|$, the notation that we used for the norm in $\mathbb{R}^2$.

The notions of convergence, open and closed for $\mathbb{C}$ are identical to those in the plane (see results in Definition 4.1.

Definition 7.1. *We say that $(z_n)_{n=1}^{\infty} \subset \mathbb{C}$ converges to z if and only if $|z_n - z|$ tends to zero as n goes to ∞. That is, if for every $\varepsilon > 0$ there exists $N > 0$ such that $|z_n - z| < \varepsilon$ for all $n > N$.*

Definition 7.2. *We say that $\Omega \subset \mathbb{C}$ is open if and only for every $x \in \Omega$ there exits $r > 0$ such that $B_r(x) = \{z \in \mathbb{C} |\, |z - x| < r\} \subset \Omega$. We say that Ω is closed if and only if Ω^c is open.*

Definition 7.3. *A set $K \subset \mathbb{C}$ is sequentially compact if and only if for every sequence $(x_j)_{j\in\mathbb{N}} \subset K$ has a convergent subsequence $(x_{j(l)})_{l\in\mathbb{N}}$ whose limit is in K.*

Now, maps in $f : \Omega \subset \mathbb{C} \to \mathbb{C}$, are given by a pair of real-valued functions $f(z) = u(z) + \mathrm{i}v(z)$, the real part u of f and the imaginary part v of f. We can think of those two functions as functions of z or as functions in $\mathbb{R}^2$ of x and y, the real and imaginary part of z. This means we can also think of f as a function from $\Omega \subset \mathbb{R}^2$ to $\mathbb{R}^2$.

Definition 7.4. *Given $f : \Omega \subset \mathbb{C} \to \mathbb{C}$ we say that it is continuous at $z_0 \in \Omega$ if and only if for every $\varepsilon > 0$ there exists δ such that $|z - z_0| < \delta$, with $z \in \Omega$ implies that $|f(z) - f(z_0)| < \varepsilon$.*

Notice that the notion of continuity coincides with the one defined for maps from $\mathbb{R}^2$ to $\mathbb{R}^2$. We will now consider the notion of differentiability, where the two notions differ very significantly.

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at a point p if and only if there exists a linear map $Df(p) \in L(\mathbb{R}^n; \mathbb{R}^k)$ such that

$$\lim_{h\to 0} \frac{\|f(p+h) - f(p) - Df(p)h\|}{\|h\|} = 0. \tag{7.1}$$

The reason for introducing that definition arose from the fact that when $k > 1$ we have no notion of division for the quantity we would like to study

$$\lim_{h\to 0} \frac{f(p+h) - f(p)}{h}$$

as division by $h \in \mathbb{R}^n, n > 1$ is not well defined. However, in $\mathbb{C}$ we do have a notion of multiplication and therefore we can use that quotient to define differentiability.

Definition 7.5. *Let $\Omega \subset \mathbb{C}$ be an open set and $z \in \Omega$. We say that f is complex differentiable at z if and only if the limit*

$$\lim_{h\to 0} \frac{f(z+h) - f(z)}{h} \tag{7.2}$$

exists. We denote the limit by $f'(z)$.

In contrast to what happened in the real valued case, where the derivative was a linear map from $\mathbb{R}^n$ to $\mathbb{R}^k$, which in our case would mean from $\mathbb{R}^2$ to $\mathbb{R}^2$, corresponding to a 2 by 2 matrix, in the complex case we obtain a complex number. Before studying how to reconcile this difference, we look at the consequences of the definition for the real and imaginary part of f. Let's write $h = \Delta x + \mathrm{i}\Delta y$, and $f(z) = u(z) + \mathrm{i}v(z)$, which we can also think of as $f(x, y) = u(x, y) + \mathrm{i}v(x, y)$. Then the quotient in the definition of complex derivative can be rewritten as

$$\frac{f(z+h) - f(z)}{h} = \frac{u(x+\Delta x, y+\Delta y) - u(x,y) + \mathrm{i}[v(x+\Delta x, y+\Delta y) - v(x,y)]}{\Delta x + \mathrm{i}\Delta y}.$$

We could consider multiple ways of sending $\Delta x + \mathrm{i}\Delta y$ to zero, obtaining the same answer if the limit exists. We will consider the two obvious options, sending Δx first to zero followed by Δy, and the reverse, Δy first followed by Δx. We find

$$\begin{aligned}
&\lim_{\Delta y\to 0}\lim_{\Delta x\to 0} \frac{u(x+\Delta x, y+\Delta y) - u(x,y) + \mathrm{i}[v(x+\Delta x, y+\Delta y) - v(x,y)]}{\Delta x + \mathrm{i}\Delta y} \\
&\qquad = \lim_{\Delta y\to 0} \frac{u(x, y+\Delta y) - u(x,y) + \mathrm{i}[v(x, y+\Delta y) - v(x,y)]}{\mathrm{i}\Delta y} \\
&\qquad = \frac{1}{\mathrm{i}} \lim_{\Delta y\to 0} \frac{u(x, y+\Delta y) - u(x,y)}{\Delta y} + \mathrm{i}\frac{v(x, y+\Delta y) - v(x,y)}{\Delta y} \\
&\qquad = \frac{1}{\mathrm{i}} \left[\frac{\partial u}{\partial y}(x,y) + \mathrm{i}\frac{\partial v}{\partial y}(x,y)\right] = v_y(x,y) - \mathrm{i}u_y(x,y),
\end{aligned}$$

while

$$\begin{aligned}
&\lim_{\Delta x\to 0}\lim_{\Delta y\to 0} \frac{u(x+\Delta x, y+\Delta y) - u(x,y) + \mathrm{i}[v(x+\Delta x, y+\Delta y) - v(x,y)]}{\Delta x + \mathrm{i}\Delta y} \\
&\qquad = \lim_{\Delta x\to 0} \frac{u(x+\Delta x, y) - u(x,y) + \mathrm{i}[v(x+\Delta x, y) - v(x,y)]}{\Delta x}
\end{aligned}$$

$$= \lim_{\Delta x \to 0} \frac{u(x+\Delta x, y) - u(x,y)}{\Delta x} + \mathrm{i}\frac{v(x+\Delta x, y) - v(x,y)}{\Delta x}$$
$$= \left[\frac{\partial u}{\partial x}(x,y) + \mathrm{i}\frac{\partial v}{\partial x}(x,y)\right] = u_x(x,y) + \mathrm{i}v_x(x,y).$$

This immediately means that at the very least we need to demand some relationships between the partial derivatives of u and v to hold in order to have a complex derivative. Namely

$$u_x = v_y \qquad u_y = -v_x \tag{7.3}$$

These equations are known as the Cauchy–Riemann equations. These are clearly necessary conditions, but at this point in no way guarantee that a complex derivative would exists if (7.3) is satisfied.

By considering two simple examples, it is easy to see that the notion of complex derivative is highly restrictive as functions that are obviously smooth when considered as a map from $\mathbb{R}^2$ to $\mathbb{R}^2$ are not actually complex differentiable. First we consider $f(z) = z$. Notice that $f'(z)$ exists and equals 1. Indeed

$$f'(z) = \lim_{h\to 0} \frac{z+h-z}{h} = 1.$$

However if we consider $g(z) = \bar{z}$, we obtain a function that is not complex differentiable. We have

$$\lim_{h\to 0} \frac{g(z+h) - g(z)}{h} = \lim_{h\to 0} \frac{\bar{z} + \bar{h} - \bar{z}}{h} = \lim_{h\to 0} \frac{\bar{h}}{h},$$

a limit that does not exist. (Consider for example the limits obtained by taking h along the real or the imaginary axis.) The function g does not satisfy the Cauchy–Riemann equations. We have $g(z) = x - \mathrm{i}y$, and therefore

$$u_x = 1, \quad v_y = -1, \quad u_y = 0, \quad v_x = 0.$$

When considering g as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ we have $g(x,y) = (x,-y)$ we clearly have a differentiable function, as all components are smooth functions. (The existence of continuous partial derivatives suffices to obtain differentiability, see Theorem 6.30.)

Definition 7.6. *We say that $f : \Omega \to \mathbb{C}$ is analytic (or holomorphic) in a neighbourhood U of z if it is complex differentiable everywhere in U. We say that f is entire if it is analytic in the whole of $\mathbb{C}$.*

A function can be differentiable at one point, but not necessarily analytic. Consider as an example the function $f(z) = |z|^2$. We will show that the function is complex differentiable at 0, but that it is not analytic, as it is not complex differentiable outside the origin. Notice that $f(z) = x^2 + y^2$, and $u = x^2 + y^2$ and $v = 0$. When computing the Cauchy–Riemann equations we find

$$u_x = 2x \qquad u_y = 2y, \qquad v_x = v_y = 0.$$

The Cauchy–Riemann equations mean $2x = 0$ and $2y = 0$, which is only satisfied at the origin. Now, to check that f is complex differentiable at the origin

$$\left.\frac{|z+h|^2 - |z|^2}{h}\right|_{z=0} = \frac{|h|^2}{h} = \bar{h} \xrightarrow[h\to 0]{} 0,$$

proving that f is complex differentiable at the origin with derivative 0.

We will now revisit the Cauchy–Riemann equations and connect complex differentiability with the dependence of the function on $\bar{z}$. Consider $f(z)$ as given by $u(x,y) + \mathrm{i}v(x,y)$. Using the fact that $x = \frac{z+\bar{z}}{2}$ and $y = \frac{z-\bar{z}}{2\mathrm{i}}$ we can rewrite the function back in terms of z and $\bar{z}$. Now, we could consider the derivative of f with respect to $\bar{z}$. Applying the chain rule we would obtain

$$\frac{\partial u}{\partial \bar{z}} = u_x \frac{1}{2} - u_y \frac{1}{2\mathrm{i}} \qquad \frac{\partial v}{\partial \bar{z}} = v_x \frac{1}{2} - v_y \frac{1}{2\mathrm{i}}$$

Therefore

$$\frac{\partial f}{\partial \bar z} = u_{\bar z} + \mathrm{i} v_{\bar z} = u_x \frac{1}{2} - u_y \frac{1}{2\mathrm{i}} + \mathrm{i}\left[v_x \frac{1}{2} - v_y \frac{1}{2\mathrm{i}}\right],$$

which we can simplify to

$$\frac{\partial f}{\partial \bar z} = \frac{1}{2}\left[u_x - v_y\right] + \mathrm{i}\frac{1}{2}\left[v_x + u_y\right].$$

Notice that if the function is complex differentiable, it satisfies the Cauchy–Riemann equations and therefore the expression above is identically zero. In this sense we say that if a function is complex differentiable, then

$$\frac{\partial f}{\partial \bar z} = 0.$$

This illustrates why $f(z) = \bar z$ or $g(z) = |z|^2 = z\bar z$ were not complex differentiable.

Now that we know that the Cauchy–Riemann equations need to be satisfied for a function to be complex differentiable we can identify the complex plane with a subspace of 2×2 matrices. This identification will allow us to connect directly complex differentiability with the standard notion of differentiability discussed in Chapter 6. We have already identified $a + \mathrm{i}b$ with the point in $\mathbb{R}^2$ given by (a, b). We can also identify it with the matrix

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Note that which factor of b contains a minus sign is just a convention. Notice that the determinant of that matrix equals $|a + \mathrm{i}b|^2$, and that therefore the matrix is invertible unless $a + \mathrm{i}b = 0$. This identification preserves the basic operations we have for complex numbers, for example summation and multiplication. That is it is possible to perform the operation $(a + \mathrm{i}b) + (c + \mathrm{i}d)$ as complex numbers or as the sum of the two corresponding matrices, with the results agreeing (modulo the identification). For the product we have

$$(a + \mathrm{i}b)(c + \mathrm{i}d) = (ac - bd) + \mathrm{i}(bc + ad)$$

and

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} ac - bd & -(bc + ad) \\ bc + ad & ac - bd \end{pmatrix},$$

proving the result. Sometimes it is useful to consider a hybrid of both identification, the one as a matrix, and the one as a point (or vector) in $\mathbb{R}^2$. For example, for the product of two complex numbers that we have just considered, we could identify it with

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix}.$$

The answer is the vector

$$\begin{pmatrix} ac - bd \\ bc + ad \end{pmatrix},$$

which corresponds to the right complex number $(ac-bd)+\mathrm{i}(bc+ad)$ and to the matrix $\begin{pmatrix} ac - bd & -(bc + ad) \\ bc + ad & ac - bd \end{pmatrix}$.

We are now ready to connect complex differentiation with Cauchy–Riemann and differentiation for functions in $\mathbb{R}^2$.

Theorem 7.7. *Let $f : \Omega \subset \mathbb{C} \to \mathbb{C}$ with Ω open. f is complex differentiable at $z = a + \mathrm{i}b \in \Omega$ if and only if f, when considered as map from $\Omega \subset \mathbb{R}^2$ to $\mathbb{R}^2$ has a derivative at the point (a, b) that satisfies the Cauchy–Riemann equations.*

Before we prove this result, we emphasize that some books will replace the right-hand side by asking that the Cauchy–Riemann equations are satisfied and that all partial derivatives are continuous. Notice that this last condition implies the existence of a derivative.

Proof. Assume that f is complex differentiable at $z = a + \mathrm{i}b$. Then we have

$$\lim_{h\to 0} \frac{f(z+h) - f(z)}{h} = f'(z),$$

which we can rewrite as

$$\lim_{h\to 0} \left| \frac{f(z+h) - f(z) - f'(z)h}{h} \right| = 0. \tag{7.4}$$

In order to prove that f is differentiable as a map in $\mathbb{R}^2$ we need to find a linear map Df that satisfies (7.1), which translates in finding a 2×2 matrix. Notice that (7.4) suggest that $f'(z) \in \mathbb{C}$ should be the map. Indeed if we identify $f'(z)$ with the corresponding matrix, and think of $f'(z)h$ not as a product of two complex numbers but as a matrix acting on the vector h then we have in fact proven that f has a derivative. Since we already know that all complex differentiable functions satisfy the Cauchy–Riemann equations we have completed that implication.

For the reverse, assuming that we have a derivative, that means that we have a 2×2 matrix which is given by

$$Df((a,b)) = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}$$

and that satisfies

$$\lim_{h\to 0} \frac{|f((a,b)+h) - f((a,b)) - Df((a,b))h|}{|h|} = 0.$$

Since the Cauchy–Riemann equations are satisfied we know that this matrix does in fact have the form

$$\begin{pmatrix} u_x & -v_x \\ v_x & u_x \end{pmatrix},$$

meaning that we could identify it with a complex number as before. We could therefore identify $Df\,h$ with the product of the complex numbers $f'(z) = u_x + \mathrm{i}v_x$ and h. Identifying (a,b) with z we obtain

$$\lim_{h\to 0} \frac{|f(z+h) - f(z) - f'(z)h|}{|h|} = 0$$

which implies that

$$\lim_{h\to 0} \frac{f(z+h) - f(z)}{h}$$

exists and equals $f'(z)$, completing the proof. □

As a consequence of the above result, since we can connect complex derivatives with derivatives of maps from $\mathbb{R}^2$ to $\mathbb{R}^2$, we have the following results:

Theorem 7.8. *Lef $f, g : \Omega \subset \mathbb{C} \to \mathbb{C}$ be complex differentiable functions. Then (asumming $g \neq 0$ in the third expression) we have that the familiar expressions*

$$(f+g)' = f' + g' \qquad (fg)' = f'g + fg' \qquad \left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2} \qquad (f(g))' = f'(g)g'$$

apply to the complex-valued case as well. For the final expression one needs to assume that the composition makes sense, i.e. the range of g is contained in the domain of f.

We conclude this section by proving that $f(z) = z^n$ is complex differentiable for every $n \in \mathbb{N}$. Using Theorem 7.7 it suffices to show that it has a derivative at every point and that it satisifies the Cauchy–Riemann equations. Notice that since it is a polynomial (once expanded in terms of x and y and considered as map from $\mathbb{R}^2$ to $\mathbb{R}^2$ we trivially have that it has a derivative). To see that it satisfies the Cauchy–Riemann equations, notice that (and similarly for v)

$$u_x = (\mathbf{Re}\, f)_x = \mathbf{Re}(f_x).$$

Therefore

$$f_x = u_x + \mathrm{i}v_x = n(x+\mathrm{i}y)^{n-1} \qquad\qquad f_y = u_y + \mathrm{i}v_y = n(x+\mathrm{i}y)^{n-1}\mathrm{i}.$$

Without computing what u_x, v_x, u_y, v_y are, notice that it follows from the expression above that

$$u_y + \mathrm{i}v_y = \mathrm{i}(u_x + \mathrm{i}v_x),$$

which implies that $u_x = v_y$ and $u_y = -v_x$, which are the Cauchy–Riemann equations.

7.2 Power Series

We want to focus on the study of power series, i.e. expressions of the form $\sum_{n=0}^{\infty} a_n z^n$. We begin by reviewing (in a very utilitarian way) some basic ideas of series for complex numbers covered in year 1.

Definition 7.9. *The series $\sum_{n=0}^{\infty} a_n$, with $a_n \in \mathbb{C}$ is convergent if and only if the sequence $S_N = \sum_{n=0}^{N} a_n$ is convergent in $\mathbb{C}$.*

Definition 7.10. *The series $\sum_{n=0}^{\infty} a_n$, with $a_n \in \mathbb{C}$ is absolutely convergent if and only if the series $\sum_{n=0}^{\infty} |a_n|$ is convergent.*

The geometric series $\sum_{n=0}^{\infty} z^n$ is convergent if and only if $|z| < 1$, and sums up to $1/(1-z)$ (with partial sums $S_N = (1-z^{N+1})/(1-z)$). We review a couple of the convergence tests from year 1.

Theorem 7.11 (Ratio Test)**.** *Consider $\sum_{n=0}^{\infty} a_n$ and assume that $a_n \neq 0$ for all n. Then*

1. *If $\limsup \frac{|a_{n+1}|}{|a_n|} < 1$ then $\sum_{n=0}^{\infty} a_n$ is convergent.*

2. *If $\frac{|a_{n+1}|}{|a_n|} \geq 1$ for all $n > N$ then $\sum_{n=0}^{\infty} a_n$ is divergent.*

In particular if $\lim \frac{|a_{n+1}|}{|a_n|}$ exists, and equals L we have convergence for $L < 1$ and divergence for $L > 1$. (The test is inconclusive if $L = 1$.)

Theorem 7.12 (Root Test)**.** *Consider $\sum_{n=0}^{\infty} a_n$. Then*

1. *If $\limsup |a_n|^{1/n} < 1$ then $\sum_{n=0}^{\infty} a_n$ converges.*

2. *If $\limsup |a_n|^{1/n} > 1$ then $\sum_{n=0}^{\infty} a_n$ diverges.*

The proofs of these results are obtained by comparison with the geometric series and will not be covered in these notes.

We will focus on studying expressions of the form

$$\sum_{n=0}^{\infty} a_n z^n \qquad \text{or} \qquad \sum_{n=0}^{\infty} a_n (z-z_0)^n,$$

with $a_n, z \in \mathbb{C}$.

Theorem 7.13. *Given $(a_n)_{n=0}^{\infty}$ there exists $R \in [0, \infty]$ such that*

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

*converges for all $|z| < R$ and diverges for $|z| > R$. (As we will see in the proof $R = \frac{1}{\limsup |a_n|^{1/n}}$.) The quantity R is called the **radius of convergence of the series**.*

Proof. We consider z given, but fixed, and apply the root test to the series $\sum_{n=0}^{\infty} a_n z^n$. This series is convergent if

$$\limsup |a_n z^n|^{1/n}$$

is less than 1 and divergent if it is greater than 1. But that translates to convergence if

$$|z| < \frac{1}{\limsup |a_n|^{1/n}}$$

and divergence when

$$|z| > \frac{1}{\limsup |a_n|^{1/n}},$$

proving the result. □

A simple application of the ratio tests yields the following result:

Theorem 7.14. *Let $a_n \neq 0$ for all $n \geq N$, and assume that $\lim \frac{|a_{n+1}|}{|a_n|}$ exists. Then $\sum_{n=0}^{\infty} a_n z^n$ has radius of convergence $R = \lim \frac{|a_n|}{|a_{n+1}|}$.*

Next we will show that within the radius of convergence a power series is actually differentiable, and that we can in fact compute the derivative term-by-term. More precisely:

Theorem 7.15. *Assume $\sum_{n=0}^{\infty} a_n z^n$ has radius of convergence R. Then for $|z| < R$ the function $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is differentiable and*

$$f'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

Proof. First we will show that the power series for $f'(z)$ does have the same radius of convergence. Notice that the radius of convergence of $\sum_{n=1}^{\infty} n a_n z^{n-1}$ and $\sum_{n=1}^{\infty} n a_n z^n$ (i.e. where we have multiplied the expression by z) is the same. To see this notice that if $\sum_{n=1}^{\infty} n a_n z^{n-1}$ is convergent for $|z| < R$ and divergent for $|z| > R$ then the same will apply to the second series. Therefore we just need to consider

$$\limsup |n a_n|^{1/n} = \lim n^{1/n} \limsup |a_n|^{1/n} = \limsup |a_n|^{1/n},$$

which shows that the radius of convergence is the same as for $\sum_{n=0}^{\infty} a_n z^n$. Notice that the series $\sum_{n=2}^{\infty} n(n-1) a_n z^{n-2}$ also has the same radius of convergence (we will need this result in our estimate below, even though we never formally compute second derivatives).

Next, notice that for $k \in \mathbb{N}$ we have

$$\frac{w^k - z^k}{w - z} = w^{k-1} + w^{k-2} z + \cdots + w z^{k-2} + z^{k-1}. \tag{7.5}$$

Now, in order to prove that f is complex differentiable and compute its derivative we study

$$\frac{f(z+h) - f(z)}{h} - \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

We denote by $w = z + h$ (and so $h = w - z$), and substitute the expression for f in terms of a series to find

$$\sum_{n=0}^{\infty} a_n \left(\frac{w^n - z^n}{w - z} - n z^{n-1} \right).$$

We look more carefully at the term in brackets. Using (7.5) we find (taking $k = n$)

$$\frac{w^n - z^n}{w - z} - n z^{n-1} = w^{n-1} + w^{n-2} z + \cdots + w z^{n-2} + z^{n-1} - n z^{n-1}$$

$$= w^{n-1} - z^{n-1} + [w^{n-2} - z^{n-2}]z + \cdots + (w-z)z^{n-2}$$

$$= (w-z)\left[\frac{w^{n-1} - z^{n-1}}{w-z} + \frac{w^{n-2} - z^{n-2}}{w-z}z + \cdots + \frac{w-z}{w-z}z^{n-2}\right]. \qquad (7.6)$$

Now, for $|z| < r < R$ and $|w| < r < R$ we have

$$\left|\frac{w^k - z^k}{w-z}\right| = |w^{k-1} + w^{k-2}z + \cdots + wz^{k-2} + z^{k-1}| < kr^{k-1}$$

and therefore

$$\left|\frac{w^k - z^k}{w-z}z^{n-k-1}\right| \le kr^{k-1}r^{n-k-1} \le kr^{n-2},$$

which substituted in (7.6) yields

$$\le |w-z|\left[|(n-1)r^{n-2} + (n-2)r^{n-2} + \cdots 2r^{n-2} + r^{n-2}|\right]$$

$$\le |w-z|r^{n-2}\frac{1}{2}n(n-1).$$

We have shown that

$$\left|\frac{f(z+h) - f(z)}{h} - \sum_{n=1}^{\infty} na_n z^{n-1}\right| \le |w-z|\frac{1}{2}\sum_{n=0}^{\infty} n(n-1)|a_n|r^{n-2} \le M|h|,$$

which goes to zero as h goes to zero. Notice that in the last inequality we have used that the series $\sum_{n=0}^{\infty} n(n-1)|a_n|r^{n-2}$ is finite, since we observed that the radius of convergence of the corresponding power series was also R. □

We have the following simple consequence of the Theorem above, which allows us to compute the coefficients a_n in terms of derivatives of f.

Corollary 7.16. *Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series with radius of convergence $R > 0$. Then $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is infinitely differentiable and moreover*

$$f^{(n)}(0) = a_n n!, \qquad n = 0, 1, 2, \ldots$$

Proof. The result is trivial for $f(0)$, as it clearly equals a_0. A simple induction argument using the formula for the derivative of f in the previous Theorem yields the desired result. □

Theorem 7.17. *Let $\sum_{n=0}^{\infty} a_n z^n$ be a power series with radius of convergence $R > 0$. Then for every $r\ in(0, R)$ the sequence of functions*

$$f_k := \sum_{n=0}^{k} a_n z^n$$

converges uniformly in $|z| \le r$.

Proof. We show the result by proving that (f_k) is uniformly Cauchy in $|z| \le r$. We have (assuming that $j \le k$)

$$|f_k(z) - f_j(z)| = |\sum_{n=j+1}^{k} a_n z^n| \le \sum_{n=j+1}^{k} |a_n|r^n \le \sum_{n=j+1}^{\infty} |a_n|r^n.$$

Since by assumption $\sum_{n=0}^{\infty} |a_n|r^n$ is finite, given any $\varepsilon > 0$ we can choose N large enough to make $|f_k(z) - f_j(z)| < \varepsilon$ for all $j, k > N$, concluding the proof. (This proof is essentially an application of the Weierstrass M-test that we covered a few weeks ago.) □

7.2.1 The exponential and the circular functions

Many of these functions should have appeared in year I, though perhaps only in the real-valued case.

Definition 7.18. *We define the following power series for $z \in \mathbb{C}$.*

$$\mathrm{e}^z := \sum_{n=0}^{\infty} \frac{1}{n!} z^n, \tag{7.7}$$

$$\cos(z) := \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n}, \tag{7.8}$$

$$\cosh(z) := \sum_{n=0}^{\infty} \frac{1}{(2n)!} z^{2n}, \tag{7.9}$$

$$\sin(z) := \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1}, \tag{7.10}$$

$$\sinh(z) := \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} z^{2n+1}. \tag{7.11}$$

The ratio test shows (Exercise) that the radius of convergence of all of the series above is $R = \infty$. Notice that using Theorem 7.15 we can prove well known identities like $(\mathrm{e}^z)' = \mathrm{e}^z$. Indeed

$$(\mathrm{e}^z)' = \left(\sum_{n=0}^{\infty} \frac{1}{n!} z^n\right)' = \sum_{n=1}^{\infty} \frac{n}{n!} z^{n-1} = \sum_{n=0}^{\infty} \frac{1}{n!} z^n = \mathrm{e}^z.$$

In fact, we can easily relate all the circular functions to the exponential.

Proposition 7.19. *The following identities hold for all $z \in \mathbb{C}$:*

$$\cos z = \frac{\mathrm{e}^{\mathrm{i}z} + e^{-\mathrm{i}z}}{2}, \qquad \sin z = \frac{\mathrm{e}^{\mathrm{i}z} - e^{-\mathrm{i}z}}{2i},$$

$$\cosh z = \frac{\mathrm{e}^{z} + e^{-z}}{2}, \qquad \sinh z = \frac{\mathrm{e}^{z} - e^{-z}}{2}.$$

Proof. We only prove the first one. The others are very similar and are left as an Exercise.

$$\begin{aligned} \frac{\mathrm{e}^{\mathrm{i}z} + e^{-\mathrm{i}z}}{2} &= \frac{1}{2}\left[\sum_{n=0}^{\infty} \frac{1}{n!}(\mathrm{i}z)^n + \sum_{n=0}^{\infty} \frac{1}{n!}(-\mathrm{i}z)^n\right] \\ &= \frac{1}{2}\left[\sum_{n=0}^{\infty} \frac{\mathrm{i}^n + (-\mathrm{i})^n}{n!} z^n\right] = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} z^{2k}, \end{aligned}$$

where we have used that

$$\mathrm{i}^n + (-\mathrm{i})^n = \begin{cases} 2(\mathrm{i})^n = 2(-1)^{n/2} & n \text{ even} \\ 0 & n \text{ odd} \end{cases}.$$

□

There are additional relationships between sine and cosine and their hyperbolic counterparts. Notice that we have

$$\cos(\mathrm{i}z) = \cosh(z) \qquad \cosh(\mathrm{i}z) = \cos(z) \qquad \sin(\mathrm{i}z) = \mathrm{i}\sinh(z) \qquad \sinh(\mathrm{i}z) = \mathrm{i}\sin(z),$$

which shows that sine and cosine are unbounded functions in the complex plane. Just consider $z = \mathrm{i}y$ for $y \in \mathbb{R}$ together with the fact that the real valued $\sinh$ and $\cosh$ grow exponentially at infinity.

Theorem 7.20. *The exponential function* e^z *satisfies the following properties*

1. $\mathrm{e}^{z+w} = \mathrm{e}^z\mathrm{e}^w$ *for all* $z, w \in \mathbb{C}$.
2. $e^z \neq 0$ *for all* $z \in \mathbb{C}$.
3. $\mathrm{e}^z = 1$ *if and only if* $z = 2k\pi\mathrm{i}$ *for* $k \in \mathbb{Z}$, *and as a result* $\mathrm{e}^{z+w} = \mathrm{e}^z$ *if and only if* $w = 2k\pi\mathrm{i}$, $k \in \mathbb{Z}$. *Notice that in particular we have shown* $\mathrm{e}^{z+2k\pi\mathrm{i}} = \mathrm{e}^z$ *for all* $k \in \mathbb{Z}$, *so in this sense the exponential is periodic in the imaginary variable.*
4. $\mathrm{e}^z = -1$ *if and only if* $z = (2k+1)\pi\mathrm{i}$ *for* $k \in \mathbb{Z}$.

Proof. We present the direct proof of part 1, without using more advanced tools from complex analysis that would reduce the heavy computational nature.

$$\mathrm{e}^z\mathrm{e}^w = \left(\sum_{n=0}^{\infty} \frac{1}{n!} z^n\right)\left(\sum_{k=0}^{\infty} \frac{1}{k!} w^k\right) = \sum_{l=0}^{\infty} \sum_{\substack{n,k \\ n+k=l}} \frac{z^n}{n!}\frac{w^k}{k!}$$
$$= \sum_{l=0}^{\infty}\sum_{j=0}^{l} \frac{1}{l!}\binom{l}{j} z^j w^{l-j} = \sum_{l=0}^{\infty} \frac{1}{l!}(z+w)^l = \mathrm{e}^{z+w}.$$

For part 2, notice that $\mathrm{e}^z e^{-z} = 1$, proving that $e^z \neq 0$. For part 3, denoting $z = x + \mathrm{i}y$ we find

$$\mathrm{e}^z = \mathrm{e}^x\mathrm{e}^{\mathrm{i}y} = \mathrm{e}^x(\cos y + \mathrm{i}\sin y),$$

which equals 1 if and only if $\mathrm{e}^x = 1$ and $\cos y + \mathrm{i}\sin y = 1$. These only happen if $x = 0$ and $y = 2\pi k$, $k \in \mathbb{Z}$. Similarly for part 4. ☐

7.2.2 Argument and Log

Every complex number $z \in \mathbb{C}\backslash\{0\}$ can be written in the form $z = |z|\mathrm{e}^{\mathrm{i}\theta}$, where θ is the angle that the vector z forms with the x axis, measured counter-clockwise. Of course that angle is not unique (but rather up to factors of 2π. Notice that for $z = 0$ there is no natural way to choose an angle.

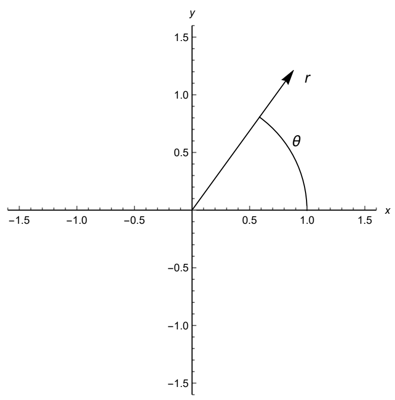

Figure 7.1: Polar representation of a complex number

We can define the (multivalued) function, for $z \neq 0$,

$$\arg(z) = \{\theta \in \mathbb{R} : z = |z|\mathrm{e}^{\mathrm{i}\theta}\}. \tag{7.12}$$

It is not a function as such, as the image is not uniquely defined, and if $\theta \in \arg(z)$ then so is $\theta + 2k\pi$. The following are easily verified properties of $\arg$.

Proposition 7.21.

1. $\arg(\alpha z) = \arg(z)$ *for all* $\alpha > 0$.
2. $\arg(\alpha z) = \arg(z) + \pi = \{\theta + \pi,$ *for* $\theta \in \arg(z)\}$ *for all* $\alpha < 0$,
3. $\arg(\bar{z}) = -\arg(z) = \{-\theta,$ *for* $\theta \in \arg(z)\}$,
4. $\arg(1/z) = -\arg(z)$,
5. $\arg(zw) = \arg(z) + \arg(w) = \{\theta + \phi,$ *with* $\theta \in \arg(z), \phi \in \arg(w)\}$.

The ambiguity of the argument function can be solved by defining the principal value Arg of the $\arg$ function to take values in $(-\pi, \pi]$. That is for any $z \in \mathbb{C}$ we have $\mathrm{Arg}(z) \in (-\pi, \pi]$.

Notice that it is impossible to define the Arg function continuously in the entire plane. In particular as we approach any point in the negative real axis, if we do it from above the Arg function will yield π, while if we do it from below it will $-\pi$. Observe that if we had made any other choice for the range of Arg there would always be a half-line where we have the same issue, the difference between the values of the argument when approaching from opposite sides is always 2π.

We want to define the logarithm by analogy of what happens in $\mathbb{R}$. In the real valued case we say (here $w, z \in \mathbb{R}$)

$$w = \log(z) \quad \text{if and only if} \quad \mathrm{e}^w = z.$$

If we could extend this for $w, z \in \mathbb{C}$, since we know that $\mathrm{e}^w = \mathrm{e}^{w+2\pi \mathrm{i} k}$ for any $k \in \mathbb{Z}$ we would have that if $w = \log(z)$ the so is $w + 2\pi \mathrm{i} k$. Therefore we would have that $\log(z)$ is a multivalued function (just like it happened before with $\arg(z)$, the argument function).

Let's write $z = |z|\mathrm{e}^{\mathrm{i}\arg(z)}$ and $w = \log(z) = u + \mathrm{i}v$. We have

$$\mathrm{e}^{u+\mathrm{i}v} = \mathrm{e}^u \mathrm{e}^{\mathrm{i}v} = z = |z|\mathrm{e}^{\mathrm{i}\arg(z)},$$

and therefore comparing the two expressions in polar form we must have

$$\mathrm{e}^u = |z| \qquad \text{and} \qquad \mathrm{e}^{\mathrm{i}v} = \mathrm{e}^{\mathrm{i}\arg(z)}.$$

That means that $u = \log|z|$, with this logarithm being the *real* logarithm. We will denote by Log the logarithm in $\mathbb{R}$ to distinguish it from the complex valued we want to define. We define the multivalued function

$$\log(z) = \mathrm{Log}|z| + \mathrm{i}\arg(z). \tag{7.13}$$

In terms of the Arg function we have

$$\log(z) = \mathrm{Log}|z| + \mathrm{i}\mathrm{Arg}(z) + 2\pi \mathrm{i} k \qquad \text{for } k \in \mathbb{Z}.$$

For example if we compute the complex logarithm of 1 we have

$$\log(1) = \mathrm{Log}|1| + \mathrm{i}\mathrm{Arg}(1) + 2\pi \mathrm{i} k = 2\pi \mathrm{i} k \qquad \text{for } k \in \mathbb{Z}.$$

Notice that the definition above makes sense provided that $z \neq 0$, where the *real* logarithm is not defined. We can now compute logarithms of negative numbers.

$$\log(-1) = \mathrm{Log}|-1| + \mathrm{i}\mathrm{Arg}(-1) + 2\pi \mathrm{i} k = \mathrm{i}\pi + 2\pi \mathrm{i} k \qquad \text{for } k \in \mathbb{Z}.$$

The complex logarithm we have just defined obeys many of the properties that we know for the real logarithm, with the caveat that we have to take care of the multi-valuedness of the function. For example

$$\log(zw) = \log z + \log w.$$

To prove this result, notice that since $\mathrm{Log}|zw| = \mathrm{Log}(|z||w|) = \mathrm{Log}|z| + \mathrm{Log}|w|$ and that $\arg(zw) = \arg(z) + \arg(w)$ we have

$$\log(zw) = \mathrm{Log}|zw| + \mathrm{i}\arg(zw) = \mathrm{Log}|z| + \mathrm{Log}|w| + \mathrm{i}\arg(z) + \mathrm{i}\arg(w) = \log z + \log w.$$

This equality needs to be understood *modulo* $2\pi\mathrm{i}$, that is there exists $k \in \mathbb{Z}$ such that

$$\log(zw) - \log(z) - \log(w) = 2\pi\mathrm{i}k.$$

Similarly we have

$$\log(z/w) = \log(z) - \log(w).$$

If we want to consider the (complex) differentiability of the $\log$ we have to deal with the multi-valuedness of the $\arg$ function. Indeed if we consider the incremental quotient

$$\frac{\log(z + \Delta z) - \log z}{\Delta z}$$

we need to make sure that as we approach z both $\log$s approach the same value. We know that this cannot be done continuously in the entire plane, and that we need to remove a semi-line arising from the origin. For example if we consider $\mathbb{C}\backslash\{x \leq 0\}$ we can consider the **principal branch** of the logarithm, which by an abuse of notation we denote by Log, just like the real logarithm, by

$$\mathrm{Log}(z) = \mathrm{Log}|z| + \mathrm{i}\mathrm{Arg}(z).$$

This function, defined on $\mathbb{C}\backslash\{x \leq 0\}$ is single valued. If we consider points of the form $z = x \pm \mathrm{i}\varepsilon$, for $x < 0$ and $\varepsilon > 0$ small, we find

$$\lim_{\varepsilon\to 0}\mathrm{Log}(x \pm \mathrm{i}\varepsilon) = \mathrm{Log}(-x) \pm \mathrm{i}\pi,$$

showing that the function could not be extended continuously along $\{x < 0\}$. This half-line is called a **branch cut**. It is possible to compute the derivative of Log directly from the definition, or in terms of its inverse. However, for practical purposes, once we know it is differentiable, from the identity

$$\mathrm{e}^{\mathrm{Log}z} = z$$

we find

$$e^{\mathrm{Log}z}(\mathrm{Log}z)' = 1$$

from which it follows that $(\mathrm{Log}z)' = 1/z$.

Once we have defined the notion of logarithm it is possible to consider defining complex powers of complex numbers. Given $\alpha \in \mathbb{C}$, and $z \neq 0$ we define the α-th power of z by

$$z^\alpha := \mathrm{e}^{\log(z^\alpha)} = \mathrm{e}^{\alpha\mathrm{Log}|z| + \alpha\mathrm{i}\arg(z)}.$$

The multi-valuedness of arg means that the same is true for z^α. If we rewrite the above as

$$z^\alpha = \mathrm{e}^{\alpha\mathrm{Log}|z| + \alpha\mathrm{i}\arg(z)} = \mathrm{e}^{\alpha\mathrm{Log}|z| + \alpha\mathrm{i}\mathrm{Arg}(z) + 2\pi\alpha k\mathrm{i}} = \mathrm{e}^{\alpha\mathrm{Log}(z)}\mathrm{e}^{2\pi\alpha k\mathrm{i}}$$

for $k \in \mathbb{Z}$ the multi-valuedness becomes more evident. The number of α powers, whether it is one, finitely many or infinitely many will depend on α.

Indeed if α is an integer for example then $\mathrm{e}^{2\pi\alpha k\mathrm{i}} = 1$, which means that in fact there is only one value of z^α. If α is rational, say $\alpha = p/q$, with p, q coprime, then z^α will have finitely many powers. It is easy to see that for $\alpha = p/q$ (with p, q coprime, and $q \in \mathbb{N}$)

$$\mathrm{e}^{2\pi\alpha k\mathrm{i}} = \mathrm{e}^{2\pi\alpha(k+q)\mathrm{i}}$$

and therefore z^α will take q different values

$$e^{\alpha \mathrm{Log}(z)} e^{2\pi\alpha k \mathrm{i}}, \qquad k = 0, 1, \ldots, q-1.$$

In the case of an irrational α it will actually take infinitely many values.

In the rational case the result obtained above is consistent with what we know about finding roots of polynomials. If we consider, for $q \in \mathbb{N}$ the equation $z^q = 1$ we know it should have q roots which correspond to

$$z = 1^{1/q}.$$

Now, using the expressions above we find

$$1^{1/q} = e^{\mathrm{Log}(1)/q} e^{2\pi \mathrm{i} k/q} = e^{2\pi \mathrm{i} k/q} \qquad k = 0, 1, \ldots, q-1.$$

7.3 Complex integration, contour integrals

For a function $f : [a, b] \to \mathbb{C}$ we define

$$\int_a^b f(t)\mathrm{d}t = \int_a^b \mathbf{Re}\, f(t)\mathrm{d}t + \mathrm{i} \int_a^b \mathbf{Im}\, f(t)\mathrm{d}t. \tag{7.14}$$

This definition means that we reduce integrating a complex-valued function to integrating two real-valued functions, and can therefore use every result we know from before, such as the Fundamental Theorem of Calculus to compute each integral.

It is easy to see that for every $f, g : [a, b] \to \mathbb{C}$ and every $\alpha, \beta \in \mathbb{C}$ we have

$$\int_a^b [\alpha f + \beta g]\mathrm{d}t = \alpha \int_a^b f(t)\mathrm{d}t + \beta \int_a^b g(t)\mathrm{d}t.$$

That $\int_a^b (f+g)\mathrm{d}t = \int_a^b f\mathrm{d}t + \int_a^b g\mathrm{d}t$ follows immediately from the definition. We show the more tedious $\int_a^b \alpha f(t)\mathrm{d}t = \alpha \int_a^b f(t)\mathrm{d}t$. We have (suppressing the limits of integration and $\mathrm{d}t$ for simplicity)

$$\alpha \int f = \alpha \left[\int \mathbf{Re}(f) + \mathrm{i} \int \mathbf{Im}(f) \right]$$

$$= \mathbf{Re}(\alpha) \int \mathbf{Re}(f) - \mathbf{Im}(\alpha) \int \mathbf{Im}(f) + \mathrm{i} \left[\mathbf{Im}(\alpha) \int \mathbf{Re}(f) + \mathbf{Re}(\alpha) \int \mathbf{Im}(f) \right]$$

$$= \int \mathbf{Re}(\alpha)\, \mathbf{Re}(f) - \mathbf{Im}(\alpha)\, \mathbf{Im}(f) + \mathrm{i} \left[\int \mathbf{Im}(\alpha)\, \mathbf{Re}(f) + \mathbf{Re}(\alpha)\, \mathbf{Im}(f) \right]$$

$$= \int \mathbf{Re}(\alpha f) + \mathrm{i} \int \mathbf{Im}(\alpha f) = \int (\alpha f).$$

Notice that in this case

$$\overline{\int_a^b f(t)\mathrm{d}t} = \int_a^b \overline{f(t)}\mathrm{d}t. \tag{7.15}$$

Indeed

$$\overline{\int_a^b f(t)\mathrm{d}t} = \int_a^b \mathbf{Re}\, f(t)\mathrm{d}t - \mathrm{i} \int_a^b \mathbf{Im}\, f(t)\mathrm{d}t = \int_a^b \mathbf{Re}\, \overline{f(t)}\mathrm{d}t + \mathrm{i} \int_a^b \mathbf{Im}\, \overline{f(t)}\mathrm{d}t = \int_a^b \overline{f(t)}\mathrm{d}t.$$

We also have the following estimate (which we will use repeteadly below)

$$\left| \int_a^b f(t)\mathrm{d}t \right| \le \int_a^b |f(t)|\mathrm{d}t. \tag{7.16}$$

To prove this result, assume that $\int_a^b f(t)\mathrm{d}t = R\mathrm{e}^{\mathrm{i}\theta}$, where $R = \left|\int_a^b f(t)\mathrm{d}t\right|$. As a result of this representation R also equals

$$R = \mathrm{e}^{-\mathrm{i}\theta}\int_a^b f(t)\mathrm{d}t = \int_a^b \mathrm{e}^{-\mathrm{i}\theta} f(t)\mathrm{d}t.$$

Now, if we write $\mathrm{e}^{-\mathrm{i}\theta} f(t) = u + \mathrm{i}v$, with u and v real valued. Then we must have

$$R = \int_a^b u\mathrm{d}t \qquad \int_a^b v\mathrm{d}t = 0.$$

Notice that $u = \mathbf{Re}[\mathrm{e}^{-\mathrm{i}\theta} f(t)] \le \left|\mathrm{e}^{-\mathrm{i}\theta} f(t)\right| \le |f(t)|$. This implies

$$R = \int_a^b u(t)\mathrm{d}t \le \int_a^b |f(t)|\mathrm{d}t.$$

But since R equals $\left|\int_a^b f(t)\mathrm{d}t\right|$ we are done.

The definition above is a natural choice for integrating functions from $\mathbb{R}$ to $\mathbb{C}$, with a far less obvious choice for integrating a function from $\mathbb{C}$ to $\mathbb{C}$. Instead, we want to study integrals of complex valued-valued functions along curves, that is, expressions of the form

$$\int_\Gamma f\mathrm{d}z,$$

where Γ is a curve in the complex plane. To define a curve in $\mathbb{C}$, consider a function $\gamma : [a,b] \to \mathbb{C}$, given by $\gamma(t) = x(t) + \mathrm{i}y(t)$. We will ask that the curve γ be C^1. The primary reason is that we want to have a well defined tangent at every point of the curve (which is also integrable). We say that the curve $\Gamma = \gamma([a,b]) \subset \mathbb{C}$ is *parametrised* by the map γ.

Definition 7.22. *Given a function $f : \Omega \subset \mathbb{C} \to \mathbb{C}$ along the path $\Gamma \subset \Omega \subset \mathbb{C}$ parametrised by $\gamma : [a,b] \to \mathbb{C}$ the integral of f over Γ is given by*

$$\int_\Gamma f\mathrm{d}z = \int_a^b f(\gamma(t))\gamma'(t)\mathrm{d}t = \int_a^b \mathbf{Re}(f(\gamma(t))\gamma'(t))\mathrm{d}t + \mathrm{i}\int_a^b \mathbf{Im}(f(\gamma(t))\gamma'(t))\mathrm{d}t.$$

Notice that we are not making any regularity assumptions on f, just that the integrals are well defined. Sometimes we will consider more than one parametrisation of a curve Γ, say γ_1 and γ_2 and will use the notation $\int_{\gamma_1} f$ and $\int_{\gamma_2} f$ in addition to $\int_\Gamma$.

On many occasions we want to consider curves that are not C^1 but perhaps just piece-wise C^1. For example a square. In this case we can think of Γ as a union of n curves Γ_j, each one C^1, and parametrised in the right direction, so that connected in the right order they describe the entire curve Γ. We can define

$$\int_\Gamma f\,\mathrm{d}z := \sum_{j=1}^n \int_{\Gamma_j} f\,\mathrm{d}z.$$

It is straightforward from the definition (details are left as an Exercise) that given a curve Γ, and two functions $f, g : \mathbb{C} \to \mathbb{C}$ and $\alpha, \beta \in \mathbb{C}$ we have

$$\int_\Gamma (\alpha f(z) + \beta g(z))\mathrm{d}z = \alpha\int_\Gamma f(z)\mathrm{d}z + \beta\int_\Gamma g(z)\mathrm{d}z.$$

If we allow for $\gamma'(t)$ not to exist at finitely many points, this can be defined as a single integral, with clearly both formulations being equivalent.

Example 7.23. *Let $f : \mathbb{C} \to \mathbb{C}$ be given by $f(z) = f(x+\mathrm{i}y) = x^4 + \mathrm{i}y^4$ and the curve joining the origin in a straight line to the point $1+\mathrm{i}$, parametrized by $\gamma : [0,1] \to \mathbb{C}$, $\gamma(t) = (1+\mathrm{i})t$. Notice that $\gamma'(t) = 1+\mathrm{i}$ and so we have*

$$\int_\Gamma f = \int_0^1 (t^4 + \mathrm{i}t^4)(1+\mathrm{i})\mathrm{d}t = \int_0^1 2\mathrm{i}t^4\mathrm{d}t = \frac{2}{5}\mathrm{i}.$$

In the next Lemma we want to show that $\int_\Gamma f$ depends only on the orientation of the parametrisation of the curve. More precisely

Lemma 7.24. *Let Γ be a curve in $\mathbb{C}$, parametrised by $\gamma : [a,b] \to \mathbb{C}$, that is $\gamma([a,b]) = \Gamma$. Given $f : \Omega \subset \mathbb{C} \to \mathbb{C}$ and $\Gamma \subset \Omega$ we have:*

1. *if γ^- represents the parametrisation of γ in the opposite direction, then*

$$\int_{\gamma^-} f = -\int_\gamma f.$$

If a curve Γ has attached a sense of direction we will call it a directed curve. In this case we will denote by $-\Gamma$ the same curve swept in the opposite direction. Without the need to specify the parametrisation we can reformulate the above result by

$$\int_\Gamma f \mathrm{d}z = -\int_{-\Gamma} f \mathrm{d}z.$$

2. *If $\tilde{\gamma} : [\tilde{a}, \tilde{b}] \to \mathbb{C}$ is another parametrisation of Γ that preserves the orientation then*

$$\int_{\tilde{\gamma}} f = \int_\gamma f.$$

We refer to this fact as reparametrisation invariance. [In practise, with the regularity we are demanding on the curves, this means that there exists $\phi : [\tilde{a}, \tilde{b}] \to [a,b]$, bijective and increasing, such that $\tilde{\gamma} = \gamma(\phi)$.]

Proof. 1. Notice that if $\gamma : [a,b] \to \mathbb{C}$ parametrises the curve in one direction then γ^- is given by $\gamma^- : [a,b] \to \mathbb{C}$ with $\gamma^-(t) = \gamma(a+b-t)$. Therefore

$$\int_{\gamma^-} f = \int_a^b f(\gamma^-(t)) \left(\gamma^-\right)'(t)\mathrm{d}t = \int_a^b f(\gamma(a+b-t))(-\gamma'(a+b-t))\mathrm{d}t$$
$$= \int_b^a f(\gamma(s))(-\gamma'(s))(-1)\mathrm{d}s = -\int_a^b f(\gamma(s))\gamma'(s)\mathrm{d}s = -\int_\gamma f.$$

2. The proof is very similar to part one.

$$\int_{\tilde{\gamma}} f = \int_{\tilde{a}}^{\tilde{b}} f(\tilde{\gamma}(t))\tilde{\gamma}'(t)\mathrm{d}t = \int_{\tilde{a}}^{\tilde{b}} f(\gamma(\phi(t)))\gamma'(\phi(t))\phi'(t)\mathrm{d}t = \int_a^b f(\gamma(s))\gamma'(s)\mathrm{d}s = \int_\gamma f,$$

where we have made the change of variables $\phi(t) = s$ and therefore $\phi'(t)\mathrm{d}t = \mathrm{d}s$ □

Consider the function $f(z) = 1$ as a complex-valued function and a curve $\gamma : [a,b] \to \mathbb{C}$. Then

$$\int_\gamma f \mathrm{d}z = \int_a^b \gamma'(t)\mathrm{d}t.$$

Here $\gamma'(t)$ is a complex valued number and the integral will be a complex number. For example if we take γ just like in Example 7.23 we have $\gamma'(t) = 1 + \mathrm{i}$ and $\int_\gamma f \mathrm{d}z = \int_0^1 (1+\mathrm{i})\mathrm{d}t = 1 + \mathrm{i}$. This is because we are considering $\mathrm{d}z$ as complex valued, given by $\gamma'(t)\mathrm{d}t$.

We could consider defining the integral

$$\int_\gamma |\mathrm{d}z| := \int_a^b |\gamma'(t)|\mathrm{d}t = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2}\mathrm{d}t = l(\gamma),$$

where $\gamma : [a,b] \to \mathbb{C}$ is given by $\gamma(t) = x(t) + \mathrm{i}y(t)$, and $l(\gamma)$ stands for the length of the curve γ.

Similarly for $f : \mathbb{C} \to \mathbb{C}$ we can define

$$\int_\gamma |f||\mathrm{d}z| := \int_a^b |f(\gamma(t))||\gamma'(t)|\mathrm{d}t.$$

Notice that $\int_\gamma |f||\mathrm{d}z| \geq 0$ and that we have

$$\left|\int_\gamma f\mathrm{d}z\right| \leq \int_\gamma |f||\mathrm{d}z|.$$

To show this notice that (using (7.16))

$$\left|\int_\gamma f\mathrm{d}z\right| = \left|\int_a^b f(\gamma(t))\gamma'(t)\mathrm{d}t\right| \leq \int_a^b |f(\gamma(t))||\gamma'(t)|\mathrm{d}t = \int_\gamma |f||\mathrm{d}z|.$$

We can further estimate the right-hand side

$$\int_\gamma |f||\mathrm{d}z| \leq \max_{z\in\Gamma}|f(z)| \int_\gamma |\mathrm{d}z| = \max_{z\in\Gamma}|f(z)|l(\gamma).$$

Therefore we obtain

$$\left|\int_\gamma f\mathrm{d}z\right| \leq \max_{z\in\Gamma}|f(z)|l(\gamma).$$

Definition 7.25. *Given $f : \mathbb{C} \to \mathbb{C}$ and a curve $\gamma : [a, b] \to \mathbb{C}$ we define*

$$\int_\gamma f\mathrm{d}\bar{z} := \int_a^b f(\gamma(t))\overline{\gamma'(t)}\mathrm{d}t.$$

Observe that in general

$$\overline{\int_\gamma f(z)\mathrm{d}z}$$

is not equal to

$$\int_\gamma \overline{f(z)}\mathrm{d}z,$$

unlike when we considered functions $f : [a, b] \to \mathbb{C}$; see (7.15). Instead we have

$$\overline{\int_\gamma f(z)\mathrm{d}z} = \overline{\int_a^b f(\gamma(t))\gamma'(t)\mathrm{d}t} = \int_a^b \overline{f(\gamma(t))\gamma'(t)}\mathrm{d}t = \int_a^b \overline{f(\gamma(t))}\,\overline{\gamma'(t)}\mathrm{d}t = \int_\gamma \overline{f(z)}\mathrm{d}\bar{z}.$$

We compute a few more examples of integrals along curves.

Example 7.26. *Integrate $f(z) = \bar{z}$ (the definition does not require functions to be analytic) along the circle of centred at $1 + \mathrm{i}$ of radius 2 (oriented counterclockwise).*

First we describe the curve γ. Notice that $2\mathrm{e}^{\mathrm{i}t}$ for $t \in [0, 2\pi]$ describes a circle or raidus two centred at the origin and with the required orientation. Therefore $\gamma(t) = (1 + \mathrm{i}) + 2\mathrm{e}^{\mathrm{i}t}$ for $t \in [0, 2\pi)$. We have $\gamma'(t) = 2\mathrm{i}\mathrm{e}^{\mathrm{i}t}$. Therefore the integral becomes

$$\int_\gamma f(z)\mathrm{d}z = \int_0^{2\pi} \overline{((1 + \mathrm{i}) + 2\mathrm{e}^{\mathrm{i}t})}2\mathrm{i}\mathrm{e}^{\mathrm{i}t}\mathrm{d}t = 2(1 - \mathrm{i})\mathrm{i}\int_0^{2\pi} \mathrm{e}^{\mathrm{i}t}\mathrm{d}t + \int_0^{2\pi} 4i\mathrm{d}t = 8\pi\mathrm{i},$$

since $\int_0^{2\pi} \mathrm{e}^{\mathrm{i}t}\mathrm{d}t = 0$. Indeed

$$\int_0^{2\pi} \mathrm{e}^{\mathrm{i}t}\mathrm{d}t = \int_0^{2\pi} \cos t\mathrm{d}t + \mathrm{i}\int_0^{2\pi} \sin t\mathrm{d}t = 0.$$

In fact

$$\int_0^{2\pi} \mathrm{e}^{\mathrm{i}nt}\mathrm{d}t = 0, \qquad \textit{for all } n \neq 0.$$

Example 7.27. *Integrate $f(z) = z$ along the circle of centred at $1 + \mathrm{i}$ of radius 2 (oriented counterclockwise). As before $\gamma(t) = (1 + \mathrm{i}) + 2\mathrm{e}^{\mathrm{i}t}$ for $t \in [0, 2\pi]$. We have $\gamma'(t) = 2\mathrm{i}\mathrm{e}^{\mathrm{i}t}$. Therefore the integral becomes*

$$\int_\gamma f(z)\mathrm{d}z = \int_0^{2\pi} \Big((1+\mathrm{i}) + 2\mathrm{e}^{\mathrm{i}t}\Big)2\mathrm{i}\mathrm{e}^{\mathrm{i}t}\mathrm{d}t = 2(1+\mathrm{i})\mathrm{i}\int_0^{2\pi} \mathrm{e}^{\mathrm{i}t}\mathrm{d}t + \int_0^{2\pi} 4i\mathrm{e}^{2\mathrm{i}t}\mathrm{d}t = 0,$$

using that $\int_0^{2\pi} \mathrm{e}^{\mathrm{i}nt}\mathrm{d}t = 0, \qquad$ for all $n \neq 0$.

Theorem 7.28. *Assume that $F : \Omega \subset \mathbb{C} \to \mathbb{C}$ is analytic (Ω open) and set $f(z) = \frac{\mathrm{d}F}{\mathrm{d}z}$, with f continuous. Let $\gamma : [a, b] \to \Omega$ be a C^1 curve. Then*

$$\int_\gamma f\mathrm{d}z = F(\gamma(b)) - F(\gamma(a)).$$

Proof. We have

$$\int_\gamma f\mathrm{d}z = \int_a^b f(\gamma(t))\gamma'(t)\mathrm{d}t = \int_a^b \frac{\mathrm{d}F}{\mathrm{d}z}(\gamma(t))\gamma'(t)\mathrm{d}t = \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}F(\gamma(t))\mathrm{d}t = F(\gamma(b)) - F(\gamma(a)).$$

□

We remark that there are no assumptions made about Ω other than it is open. That is, all we need for the result to be true, is that f is analytic in an open neighborhood of the curve. The notion of simply connected (for a domain) will be defined later, but we emphasize that there is no such requirement on Ω above result to be true.

7.3.1 Links with Green's and Gauss' Theorems

We want to connect the notion of contour integral with the notions introduced in last year's modules. The line integral we have just defined has many similarities with the notion of tangential line integral for a vector field $\mathbf{F}$. There the definition read

$$\int_C \mathbf{F} \cdot d\mathbf{r} := \int_\alpha^\beta \mathbf{F}(\mathbf{r}(t)) \cdot \frac{d\mathbf{r}}{dt}dt,$$

where C is a curve parametrised by $\mathbf{r} : [\alpha, \beta] \to \mathbb{R}^n$ with $\mathbf{r}(\alpha) = p$ and $\mathbf{r}(\beta) = q$. For closed curves that integral is usually referred as circulation.

We also recall the flux integral, which is given by

$$\int_C \mathbf{F} \cdot \mathbf{n}\mathrm{d}t.$$

Here $\mathbf{n}$ represents the normal, with the following convention. If the curve C is parametrised by $\mathbf{r}(t) = (x(t), y(t))$, and $\mathbf{r}'(t) = (x'(t), y'(t))$ has the same direction of the tangent, we choose

$$\mathbf{n}(t) := \mathbf{r}'(t)^\perp = (y'(t), -x'(t)).$$

When considering the curves determining the boundary of a regular domain we will consider them as positively oriented. That is, choose the orientation so that the corresponding $\mathbf{n}$ as defined above corresponds (i.e. has the same direction as) to the outward normal.

The following results (considered here only for two dimensions) correspond to Green's and Gauss' Theorems. For a positively oriented regular domain Ω we have

$$\iint_\Omega \operatorname{curl}\mathbf{F}\mathrm{d}x\mathrm{d}y = \oint_{\partial\Omega} \mathbf{F} \cdot \mathrm{d}\mathbf{r}$$

and

$$\iint_{\Omega} \mathrm{div}\mathbf{F}\mathrm{d}x\mathrm{d}y = \oint_{\partial\Omega} \mathbf{F}\cdot\mathbf{n}\mathrm{d}t.$$

Now let's consider our contour integral $\int_\gamma f(z)\mathrm{d}z$ for a function $f = u + \mathrm{i}v$ and a curve $\gamma(t) = \gamma_1(t) + \mathrm{i}\gamma_2(t)$. We have

$$\int_\gamma f(z)\mathrm{d}z = \int_a^b [u(\gamma(t)) + \mathrm{i}v(\gamma(t))][\gamma_1'(t) + \mathrm{i}\gamma_2'(t)]\mathrm{d}t$$

$$= \int_a^b u(\gamma(t))\gamma_1'(t) - v(\gamma(t))\gamma_2'(t)\mathrm{d}t + \mathrm{i}\int_a^b u(\gamma(t))\gamma_2'(t) + v(\gamma(t))\gamma_1'(t)\mathrm{d}t$$

$$= \int_a^b (u,-v)\cdot(\gamma_1',\gamma_2')\mathrm{d}t + \mathrm{i}\int_a^b (u,-v)\cdot(\gamma_2',-\gamma_1')\mathrm{d}t$$

$$= \int_\gamma (u,-v)\cdot\mathrm{d}\mathbf{r} + \mathrm{i}\int_\gamma (u,-v)\cdot\mathbf{n}\mathrm{d}t,$$

and so if we define the vector field $\underline{f} = (u,-v)$, we have just shown that

$$\int_\gamma f\mathrm{d}z = \text{circulation}(\underline{f}) + \mathrm{i}\,\text{flux}(\underline{f}).$$

Using the above expression, together with Green's and Gauss' Theorem we can prove the following result.

Theorem 7.29 (Cauchy's Theorem)**.** *Let $f : \Omega \to \mathbb{C}$ be an analytic function, with Ω a simply connected domain. Let γ be a C^1 closed curve in Ω. Then*

$$\int_\gamma f\mathrm{d}z = 0.$$

Before we prove the result we define simply connected. Loosely speaking means that the domain contains no holes. A set of more formal definitions is as follows.

Definition 7.30. *A set $\Omega \subset \mathbb{C}$ is connected if it cannot be expressed as the union of non-empty open sets Ω_1 and Ω_2 such that $\Omega_1 \cap \Omega_2 = \emptyset$. An open, connected set $\Omega \subset \mathbb{C}$ is called simply connected if every closed curve in Ω can be continuously deformed to a point.*

Proof. The proof presented here assumes that the curve a simple, regular curve and that f' is continuous. If the domain is simply connected, the region inside the curve does not have any holes, and f is analytic in it. We know

$$\int_\gamma f\mathrm{d}z = \text{circulation}(\underline{f}) + \mathrm{i}\,\text{flux}(\underline{f})$$

$$= \iint_A \mathrm{curl}\,\underline{f}\mathrm{d}x\mathrm{d}y + \mathrm{i}\iint_A \mathrm{div}\underline{f}\mathrm{d}x\mathrm{d}y\,,$$

where A is the region encircled by γ. We claim that both terms are actually 0, because $\mathrm{curl}\,\underline{f} = \mathrm{div}\underline{f} = 0$. Since $\underline{f} = (u,-v)$ we have

$$\mathrm{div}\underline{f} = u_x - v_y \qquad \mathrm{curl}\,\underline{f} = -v_x - u_y$$

but since $f = u + \mathrm{i}v$ is analytic it satisfies the Cauchy–Riemann equations,

$$u_x = v_y \qquad v_x = -u_y$$

which imply the result. ☐

Notice that Cauchy's Theorem applies to Example 7.27, where the function is analytic, but obviously not to Example 7.26, where the function is not analytic.

Cauchy's Theorem works for more general *curves*. Consider the shaded region Ω in Figure 7.2. If we think of its boundary as a *one curve* Γ, even though it is formed by two separate curves we have

$$\int_\Gamma f \mathrm{d}z = 0,$$

provided that Γ is oriented positively. That means that the exterior curve, that we denote by γ_1 needs to be oriented counter-clockwise, while the interior curve, denoted by γ_2 has to be oriented clockwise.

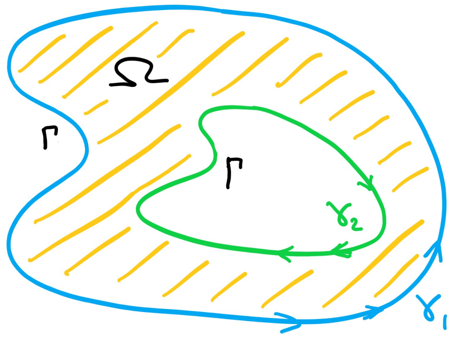

Figure 7.2: Region bound by two positively oriented curves

An equivalent formulation of this fact, which will be extremely useful is known as the deformation of contour Theorem.

Theorem 7.31. *Let $\Omega \subset \mathbb{C}$ be a region bounded by two closed simple curves γ_1 (the exterior curve) and γ_2 (the interior). Assume they are oriented positively, and let f be an analytic function in $\Omega \cup \gamma_1 \cup \gamma_2$. Then*

$$\int_{\gamma_1} f \mathrm{d}z + \int_{\gamma_2} f \mathrm{d}z = 0.$$

If we denote by γ_2^- the anti-clockwise parametrization of γ_2, then the result can be rephrased as

$$\int_{\gamma_1} f \mathrm{d}z = \int_{\gamma_2^-} f \mathrm{d}z,$$

that is the integral is the same along both curves when both are parametrised counter-clockwise.

Proof. The proof is based on creating two new contours of integration, the boundaries of two simply connected regions where f is analytic so that we can apply Cauchy's Theorem 7.29.

To achieve this we add two new curves to the previous picture, now in yellow in Figure 7.3. They join the points A (in γ_1) with D (in γ_2) and the points B (in γ_1) with C (in γ_2). The two curves we want to consider are denoted by ρ and η. Each one of them is piecewise C^1 and formed by four sections. Each one of these curves is oriented positively with respect to the region they enclose, that is, they are both oriented counter-clockwise.

By Cauchy's Theorem

$$\int_\rho f \mathrm{d}z = \int_{\rho_1} f \mathrm{d}z + \int_{\rho_2} f \mathrm{d}z + \int_{\rho_3} f \mathrm{d}z + \int_{\rho_4} f \mathrm{d}z = 0, \tag{7.17}$$

$$\int_\eta f \mathrm{d}z = \int_{\eta_1} f \mathrm{d}z + \int_{\eta_2} f \mathrm{d}z + \int_{\eta_3} f \mathrm{d}z + \int_{\eta_4} f \mathrm{d}z = 0. \tag{7.18}$$

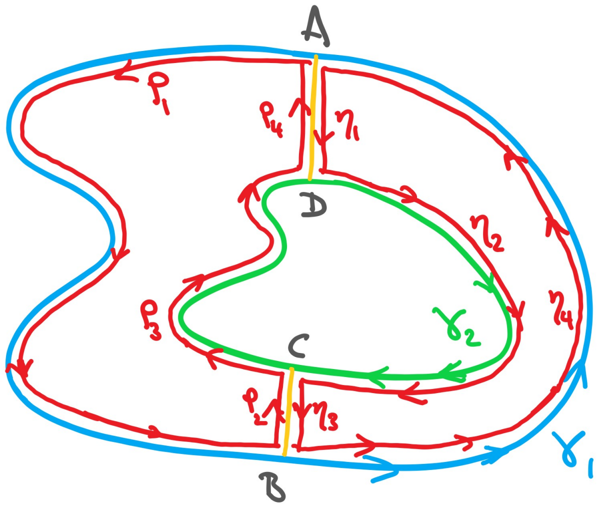

Figure 7.3: two positively oriented curves

We observe that η_1 and ρ_4 correspond to the same curve but with parametrisations in opposite directions. Similarly for η_3 and ρ_2. Therefore

$$\int_{\eta_1} f \mathrm{d}z + \int_{\rho_4} f \mathrm{d}z = 0 \qquad \int_{\eta_3} f \mathrm{d}z + \int_{\rho_2} f \mathrm{d}z = 0.$$

Adding (7.17) and (7.18) and using the above identities we find

$$\int_{\rho_1} f \mathrm{d}z + \int_{\rho_3} f \mathrm{d}z + \int_{\eta_2} f \mathrm{d}z + \int_{\eta_4} f \mathrm{d}z = 0$$

Also notice that ρ_1 and η_4 together build γ_1, while ρ_3 and η_2 build γ_2. Therefore, the above equality can be rewritten as

$$\int_{\gamma_1} f \mathrm{d}z + \int_{\gamma_2} f \mathrm{d}z = 0.$$

Since

$$\int_{\gamma_2} f \mathrm{d}z = -\int_{\gamma_2^-} f \mathrm{d}z$$

we obtain

$$\int_{\gamma_1} f \mathrm{d}z = \int_{\gamma_2^-} f \mathrm{d}z$$

as required. □

We now compute one of the fundamental contour integrals. We will show that

$$\int_{\partial B_r(a)} (z-a)^n \mathrm{d}z = \begin{cases} 2\pi \mathrm{i} & n = -1, \\ 0 & n \neq 1, \end{cases} \tag{7.19}$$

where $\partial B_r(a)$ denotes the boundary of the ball of radius r, parametrised counter-clockwise (i.e. positively oriented with respect to $B_r(a)$).

Observe that the result is uniform with respect to r. That is a natural consequence of Theorem 7.31, given than the functions we are integrating only fail to be analytic at one point (at most, depending on n). In fact we could have chosen any curve that wraps around a once and obtain the same result.

Now, to compute the integral above, notice that we can parametrise the curve as $\gamma(t) = a + r\mathrm{e}^{\mathrm{i}t}$, for $t \in [0, 2\pi)$. Therefore we have (since $\gamma'(t) = \mathrm{i}r\mathrm{e}^{\mathrm{i}t}$)

$$\int_{\partial B_r(a)} (z-a)^n \mathrm{d}z = \int_0^{2\pi} (r\mathrm{e}^{\mathrm{i}t})^n \mathrm{i}r\mathrm{e}^{\mathrm{i}t}\mathrm{d}t = \mathrm{i}r^{n+1}\int_0^{2\pi} \mathrm{e}^{\mathrm{i}(n+1)t}\mathrm{d}t.$$

Notice that in the case $n = -1$ that expression equals $2\pi\mathrm{i}$. When $n \neq -1$ notice that we obtain 0, since for all $k \neq 0$ we have

$$\int_0^{2\pi} \mathrm{e}^{\mathrm{i}kt}\mathrm{d}t = \frac{1}{k}\mathrm{e}^{\mathrm{i}kt}\Big|_0^{2\pi} = \frac{1}{k} - \frac{1}{k} = 0.$$

We restate, in the notation that will be most convenient for the next few results, the fundamental integral above in the case $n = -1$, noting that the result does not depend on r. We have

$$\int_{\partial B_r(z)} \frac{1}{w-z}\mathrm{d}w = 2\pi\mathrm{i}.$$

Definition 7.32. *Given a simple closed C^1 curve γ we denote by $I(\gamma)$ the interior region to γ. We denote by $O(\gamma)$ the exterior region to γ.*

Notice that by the deformation of contours Theorem we have

$$\int_\gamma \frac{1}{w-z}\mathrm{d}w = \int_{\partial B_r(z)} \frac{1}{w-z}\mathrm{d}w = 2\pi\mathrm{i} \tag{7.20}$$

for every $z \in I(\gamma)$ and every r sufficiently small so that $B_r(z) \subset I(\gamma)$.

Theorem 7.33. *Let $\gamma : [a,b] \to \mathbb{C}$ be a positively oriented simple closed C^1 curve. Assume that f is analytic in γ and on the interior of γ, $I(\gamma)$. Then*

$$f(z) = \frac{1}{2\pi\mathrm{i}}\int_\gamma \frac{f(w)}{w-z}\mathrm{d}w \qquad \textit{for all } z \in I(\gamma). \tag{7.21}$$

Proof. Fix $z \in I(\gamma)$, and choose r small enough so that $B_r(z) \subset I(\gamma)$. By the deformation of contours theorem we have

$$\frac{1}{2\pi\mathrm{i}}\int_\gamma \frac{f(w)}{w-z}\mathrm{d}w = \frac{1}{2\pi\mathrm{i}}\int_{\partial B_r(z)} \frac{f(w)}{w-z}\mathrm{d}w,$$

reducing the problem to considering γ as a $\partial B_r(z)$. Observe that the integral is the same for every r sufficiently small, and later on we will exploit this fact by talking limits as r tends to zero. For now, we have

$$\frac{1}{2\pi\mathrm{i}}\int_{\partial B_r(z)} \frac{f(w)}{w-z}\mathrm{d}w = \frac{1}{2\pi\mathrm{i}}\int_{\partial B_r(z)} \frac{f(z)}{w-z}\mathrm{d}w + \frac{1}{2\pi\mathrm{i}}\int_{\partial B_r(z)} \frac{f(w)-f(z)}{w-z}\mathrm{d}w =: I + II.$$

Notice that the first integral I equals $f(z)$. Indeed, using (7.20)

$$\frac{1}{2\pi\mathrm{i}}\int_{\partial B_r(z)} \frac{f(z)}{w-z}\mathrm{d}w = f(z)\frac{1}{2\pi\mathrm{i}}\int_{\partial B_r(z)} \frac{1}{w-z}\mathrm{d}w = f(z).$$

All that remains to is to show that $II = 0$. Notice that since f is analytic in $I(\gamma)$, given any $\varepsilon > 0$ we can find r sufficiently small so that

$$|f(w) - f(z)| \leq \varepsilon \qquad \text{for all } w \in \partial B_r(z).$$

We parametrise $\partial B_r(z)$ counterclockwise by $\gamma(t) = z + r\mathrm{e}^{\mathrm{i}t}$ for $t \in [0, 2\pi]$. We have $\gamma'(t) = \mathrm{i}r\mathrm{e}^{\mathrm{i}t}$ and therefore

$$\begin{aligned} |II| = \left|\frac{1}{2\pi\mathrm{i}}\int_{\partial B_r(z)} \frac{f(w)-f(z)}{w-z}\mathrm{d}w\right| &\leq \left|\frac{1}{2\pi\mathrm{i}}\int_0^{2\pi} \frac{f(z+r\mathrm{e}^{\mathrm{i}t})-f(z)}{r\mathrm{e}^{\mathrm{i}t}}\mathrm{i}r\mathrm{e}^{\mathrm{i}t}\mathrm{d}t\right| \\ &\leq \frac{1}{2\pi}\int_0^{2\pi} |f(z+r\mathrm{e}^{\mathrm{i}t}) - f(z)|\mathrm{d}t \leq \varepsilon. \end{aligned}$$

Since ε is arbitrary we obtain the desired result. □

Remark 7.34. *The formula*

$$f(z) = \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{w-z} \mathrm{d}w \qquad \text{for all } z \in I(\gamma)$$

has remarkable consequences for analytic functions. First notice that it claims that we can recover the value of f at any point by integration along a curve around that point (provided the curve is sufficiently regular, positively oriented, and contained in $I(\gamma)$). This is a very significant difference with respect to smooth functions in $\mathbb{R}^2$ for example.

Notice that since the curve γ is a compact set, for any point $z \in I(\gamma)$ the expression $w - z$ found in the denominator in Cauchy's formula is bounded away from zero, suggesting that we can differentiate the formula with respect to z to obtain

$$f'(z) = \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{(w-z)^2} \mathrm{d}w.$$

Of course we need to justify moving the derivative inside the integral sign. We assumed that f was analytic, which means that $f'(z)$ exists. The expression above would produce a formula for it, a way to compute it. The key observation is that without assuming that f has more derivatives it seems that the right hand side can be differentiated arbitrarily many times, which would suggest that f has infinitely many derivatives. This is indeed the case as we will show in the next Theorem.

Theorem 7.35. *Let $\gamma : [a, b] \to \mathbb{C}$ be a positively oriented simple closed C^1 curve. Assume that f is analytic in γ and on the interior of γ, $I(\gamma)$. Then $f^{(n)}(z)$ exists for all $n \in \mathbb{N}$ and the derivative is given by*

$$f^{(n)}(z) = \frac{n!}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{(w-z)^{(n+1)}} \mathrm{d}w \qquad \text{for all } z \in I(\gamma). \tag{7.22}$$

Proof. Notice that Theorem 7.33 would correspond to the case $n = 0$ in the current Theorem. In order to prove the result for $n = 1$ we consider the incremental quotient, and use (7.21) to obtain

$$\frac{f(z+h) - f(z)}{h} = \frac{1}{h} \left[\frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{w-z-h} \mathrm{d}w - \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{w-z} \mathrm{d}w \right].$$

By the deformation of contours Theorem we can choose γ as $\partial B_{2r}(z)$, with $B_{2r}(z) \subset I(\gamma)$. We have, operating on the right-hand side

$$\frac{f(z+h) - f(z)}{h} = \frac{1}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} \frac{f(w)}{(w-z-h)(w-z)} \mathrm{d}w$$

$$= \frac{1}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} \frac{f(w)}{(w-z)^2} \mathrm{d}w + \frac{1}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} f(w) \left[\frac{1}{(w-z-h)(w-z)} - \frac{1}{(w-z)^2} \right] \mathrm{d}w$$

$$= \frac{1}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} \frac{f(w)}{(w-z)^2} \mathrm{d}w + \frac{1}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} \left[\frac{h f(w)}{(w-z-h)(w-z)^2} \right] \mathrm{d}w.$$

To conclude the proof all that we need to do is show that the limit of the last integral as h tends to zero is zero, that is (ignoring factors of $2\pi\mathrm{i}$)

$$\lim_{h \to 0} \int_{\partial B_{2r}(z)} \left[\frac{h f(w)}{(w-z-h)(w-z)^2} \right] \mathrm{d}w = 0,$$

and recall that we are able to choose r arbitrarily small without affecting the value of the integrals above.

First we choose $|h| < r$ so that for all $w \in \partial B_{2r}(z)$ we have

$$|w - z - h| \geq |w - z| - |h| > r.$$

Here we have used the reverse triangle inequality in the first case, and the fact that $|w-z| = 2r$ for points $w \in \partial B_{2r}(z)$. Choosing $\gamma(t) = z + 2re^{it}$ for $t \in [0, 2\pi)$, we have $\gamma'(t) = 2rie^{it}$, and therefore $|\gamma'(t)| \le 2r$. Since f is analytic, in particular it is continuous and therefore there exists $M > 0$ such that $|f(w)| \le M$ for all $w \in \partial B_{2r}(z)$. Using these facts we have

$$\left| \int_{\partial B_{2r}(z)} \left[\frac{hf(w)}{(w-z-h)(w-z)^2} \right] \mathrm{d}w \right| \le \int_0^{2\pi} \frac{hM}{r(2r)^2} 2r\mathrm{d}t = \frac{\pi M}{r^2} h,$$

which goes to zero as h goes to zero, proving the result for $n = 1$. The general case is proven by induction. If we assume the result for $n = 1, 2, \cdots, k-1$ we want to prove it for $n = k$. That is, in particular we assume

$$f^{(k-1)}(z) = \frac{(k-1)!}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{(w-z)^{(k)}} \mathrm{d}w \qquad \text{for all } z \in I(\gamma).$$

We write the corresponding incremental quotient, just as before

$$\frac{f^{(k-1)}(z+h) - f^{(k-1)}(z)}{h} = \frac{1}{h} \left[\frac{(k-1)!}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{(w-z-h)^k} \mathrm{d}w - \frac{(k-1)!}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{(w-z)^k} \mathrm{d}w \right].$$

By the deformation of contours Theorem we can choose γ as $\partial B_{2r}(z)$, with $B_{2r}(z) \subset I(\gamma)$. We have, operating on the right-hand side

$$\begin{aligned}
\frac{f^{(k-1)}(z+h) - f^{(k-1)}(z)}{h} &= \frac{(k-1)!}{2\pi \mathrm{i} h} \int_{\partial B_{2r}(z)} \frac{f(w)[(w-z)^k - (w-z-h)^k]}{(w-z-h)^k (w-z)^k} \mathrm{d}w \\
&= \frac{k!}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} \frac{f(w)}{(w-z)^{(k+1)}} \mathrm{d}w \\
+ \frac{(k-1)!}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} & f(w) \left[\frac{[(w-z)^k - (w-z-h)^k]}{h(w-z-h)^k (w-z)^k} - \frac{k}{(w-z)^{(k+1)}} \right] \mathrm{d}w
\end{aligned}$$

$$= \frac{k!}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} \frac{f(w)}{(w-z)^{k+1}} \mathrm{d}w + \frac{(k-1)!}{2\pi \mathrm{i}} \int_{\partial B_{2r}(z)} f(w) \left[\frac{(w-z)^{k+1} - (w-z-h)^k (w-z) - kh(w-z-h)^k}{h(w-z-h)^k (w-z)^{k+1}} \right] \mathrm{d}w. \tag{7.23}$$

As before, all that remains is to show that the last integral tends to zero as h tends to zero. We choose h and the parametrisation as above. The result will follow if we show that

$$\left| \frac{(w-z)^{k+1} - (w-z-h)^k (w-z) - kh(w-z-h)^k}{h} \right| \le C|h|,$$

where the constant might depend on r. This is the case because, as before $|f| \le M$ and $|w-z-h| \ge |w-z| - |h| > r$ implies

$$\left| \frac{1}{(w-z-h)^k (w-z)^k} \right| \le \frac{1}{(2r)^k r^k}.$$

In order to prove (7.23), notice that the binomial formula implies

$$(w-z-h)^k = \sum_{j=0}^{k} \binom{k}{j} (w-z)^{k-j} (-h)^j$$

and therefore

$$\begin{aligned}
&(w-z)^{k+1} - (w-z-h)^k (w-z) - kh(w-z-h)^k \\
&= -\sum_{j=2}^{k} \binom{k}{j} (w-z)^{k+1-j} (-h)^j - kh \sum_{j=1}^{k} \binom{k}{j} (w-z)^{k-j} (-h)^j
\end{aligned}$$

which is of order h^2, proving the result. □

7.3.2 Consequences of Cauchy's Theorem (NOT COVERED AND NOT EXAMINABLE)

Theorem 7.36 (Taylor Series Expansion)**.** *Let f be an analytic function on $B_R(a)$ for $a \in \mathbb{C}$, $R > 0$. There there exist unique constants c_n, $n \in \mathbb{N}$ such that*

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n \qquad \text{for all } z \in B_R(a).$$

Moreover, the coefficients c_n are given by

$$c_n = \frac{1}{2\pi \mathrm{i}} \int_\gamma \frac{f(w)}{(w-a)^{n+1}} \mathrm{d}w = \frac{f^{(n)}(a)}{n!},$$

where γ is any positively oriented simple closed curve (piece-wise C^1) that is contained in $B_R(a)$ with $a \in I(\gamma)$.

Proof. Given some $z \in B_R(a)$ we will take γ to be $\partial B_r(a)$ (positively oriented), for r small enough so that $|z-a| < r < R$. We can use the Theorem of deformation of contours to prove the integrals over all curves γ as above are the same. Cauchy's formula (7.21) gives

$$f(z) = \frac{1}{2\pi \mathrm{i}} \int_{\partial B_r(a)} \frac{f(w)}{w-z} \mathrm{d}w. \tag{7.24}$$

Notice that since $|w-a| = r$ and we have chosen r so that $|z-a| < r$ we have $|z-a| < |w-a|$ for all $w \in \partial B_r(a)$. As a result

$$\frac{|z-a|}{|w-a|} < 1$$

and we can use the geometric series expansion to obtain

$$\frac{1}{w-z} = \frac{1}{w-a} \frac{1}{\left(1 - \frac{z-a}{w-a}\right)} = \frac{1}{w-a} \sum_{n=0}^{\infty} \left(\frac{z-a}{w-a}\right)^n.$$

Inserting this expression in (7.24) we obtain

$$f(z) = \frac{1}{2\pi \mathrm{i}} \int_{\partial B_r(a)} f(w) \frac{1}{w-a} \sum_{n=0}^{\infty} \left(\frac{z-a}{w-a}\right)^n \mathrm{d}w.$$

For $w \in \partial B_r(a)$ the series converges absolutely (Weierstrass M-test), and therefore we can exchange the order of the summation and integration to obtain

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{2\pi \mathrm{i}} \int_{\partial B_r(a)} \frac{f(w)}{(w-a)^{n+1}} \mathrm{d}w (z-a)^n = \sum_{n=0}^{\infty} c_n (z-a)^n,$$

obtaining the desired result. It remains to show that the coefficients are unique. Now, assume that $f(z) = \sum_{k=0}^{\infty} b_k (z-a)^k$ for some $b_k \in \mathbb{C}$. We have

$$\begin{aligned}\int_{\partial B_r(a)} \frac{f(w)}{(w-a)^{n+1}} \mathrm{d}w &= \int_{\partial B_r(a)} \sum_{k=0}^{\infty} b_k (w-a)^k \frac{1}{(w-a)^{n+1}} \mathrm{d}w \\ &= \sum_{k=0}^{\infty} b_k \int_{\partial B_r(a)} (w-a)^{k-n-1} \mathrm{d}w = 2\pi \mathrm{i} b_n,\end{aligned}$$

where we have used the fundamental integrals, together with the fact that we can commute the summation and integration. This proves that $b_n = c_n$, concluding the proof. ☐

Example 7.37. *We consider an example of a Taylor series. We consider the function $(1+z)^a$ for $a \in \mathbb{C}$ and $|z| < 1$.*

When we consider logarithms we noticed that z^n is well defined for $n \in \mathbb{N}$, but not for any a, without making any specific choice of the argument function. In this case

$$(1+z)^n = \sum_{k=0}^{n} \binom{n}{k} z^k,$$

which is a polynomial of order n, and equals the Taylor series expansion centred at the origin. This series converges for every $z \in \mathbb{C}$, not just $|z| < 1$. However, we defined

$$f(z) = (1+z)^a := \mathrm{e}^{a\mathrm{Log}(1+z)}$$

having made a choice of the argument function defining the logarithm, which meant creating a branch cut where the function was not defined. Choosing the argument in $(-\pi, \pi)$, and since our function is translated (not z^a) we obtained a function that is not defined for $z \in (-\infty, -1]$.

We want to show that in fact a binomial expansion is possible for all $a \in \mathbb{C}$. We know by Taylor's Theorem 7.36 that we have a Taylor expansion. To compute we need to work out the derivatives of $(1+z)^a$. Using the definition we have (for ($a \notin \mathbb{N}$))

$$\left(\mathrm{e}^{a\mathrm{Log}(1+z)}\right)' = \mathrm{e}^{a\mathrm{Log}(1+z)} a(\mathrm{Log}(1+z))' = (1+z)^a \frac{a}{1+z} = a(1+z)^{a-1}.$$

Notice that since by induction we have

$$\frac{\mathrm{d}^k}{\mathrm{d}z^k}\left(\mathrm{e}^{a\mathrm{Log}(1+z)}\right) = a(a-1)\cdots(a-k+1)(1+z)^{a-k}.$$

Therefore we obtain the Taylor series (centred at 0)

$$\sum_{k=0}^{\infty} \frac{a(a-1)\cdots(a-k+1)}{k!} z^k.$$

Notice that the radius of convergence of this series is 1, as we know there are issues for $z \in (-\infty, -1]$. The binomial coefficient, for integer values n and k is

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

and so extending the defition to $n \in \mathbb{C}$ we obtain

$$(1+z)^a = \sum_{k=0}^{\infty} \frac{a(a-1)\cdots(a-k+1)}{k!} z^k = \sum_{k=0}^{\infty} \binom{a}{k} z^k.$$

We can obtain similar expansions centred at different points

$$(1+z)^a = \sum_{k=0}^{\infty} \frac{a(a-1)\cdots(a-k+1)}{k!}(1+z_0)^{a-k}(z-z_0)^k,$$

which would naturally a radius of convergence R equal to the distance from the point z_0 to the half line $\{x \leq -1\}$, where we have made a brach cut for the Log function. You may ignore the issue of the radius of convergence for this series for the exam.

The following result is also a direct consequence of Cauchy's formula.

Theorem 7.38 (Liouville's Theorem)**.** *Let $f : \mathbb{C} \to \mathbb{C}$ be an analytic, bounded function. Then f is constant.*

Proof. Assume that $|f(z)| \leq M$ for all $z \in \mathbb{C}$. Let $a \neq b$ be two points in $\mathbb{C}$. Choose R large enough so that $2\max\{|a|,|b|\} < R$. That means that if we consider $w \in \partial B_R(0)$, that is $|w| = R$ then

$$|w-a| > \frac{R}{2} \qquad\qquad |w-b| > \frac{R}{2}.$$

Since f is analytic in $\mathbb{C}$ we can use Cauchy's formula to compute $f(a)$ and $f(b)$ using $\partial B_R(0)$ as the curve γ (of course positively oriented!). We have

$$f(a) - f(b) = \frac{1}{2\pi \mathrm{i}} \int_{\partial B_R(0)} \frac{f(w)}{w-a} \mathrm{d}w - \frac{1}{2\pi \mathrm{i}} \int_{\partial B_R(0)} \frac{f(w)}{w-b} \mathrm{d}w$$

$$= \frac{1}{2\pi \mathrm{i}} \int_{\partial B_R(0)} f(w) \left(\frac{1}{w-a} - \frac{1}{w-b} \right) \mathrm{d}w = \frac{a-b}{2\pi \mathrm{i}} \int_{\partial B_R(0)} \frac{f(w)}{(w-a)(w-b)} \mathrm{d}w.$$

Therefore

$$|f(a) - f(b)| \leq \frac{|a-b|}{2\pi} \frac{M}{R^2/4} \int_{\partial B_r(0)} 1 \mathrm{d}w = \frac{|a-b| 4M}{R},$$

as $\int_{\partial B_R(0} 1 \mathrm{d}w$ is just the length of the curve, which equals $2\pi R$. Notice that since R is arbitrary (provided that it is big enough, as indicated above) we can send R to infinity, showing that $|f(a) - f(b)| = 0$ for any a and b in $\mathbb{C}$, therefore proving that the function is constant. □

A fundamental consequence of Liouville's Theorem is Fundamental Theorem of Algebra.

Theorem 7.39 (Fundamental Theorem of Algebra)**.** *Every non-constant polynomial p on $\mathbb{C}$ has a root, that is, there exists $a \in \mathbb{C}$ such that $p(a) = 0$.*

Proof. We will prove the result by contradiction. Assume that $|p(z)| \neq 0$ for every $z \in \mathbb{C}$. Define $f : \mathbb{C} \to \mathbb{C}$ by $f(z) = \frac{1}{p(z)}$. Now, since p does not vanish, the function f is analytic in all of $\mathbb{C}$, since it is the composition of two holomorphic functions ($1/z$ is holomophic outside the origin).

Notice that if we assume $p(z) = \sum_{k=0}^{n} c_k z^k$, with $c_n \neq 0$ ($n > 0$), then at infinity the polynomial behaves like $c_n z^n$, as that is the highest power. That means $|p(z)|$ goes to infinity as z goes to infinity, and satisfies $|p(z)| > 1$ for all $|z| > R$ for some $R > 0$. As a result the function $f(z) = \frac{1}{p(z)}$ is bounded in $\mathbb{C}$. It is less than 1 for all $|z| > R$ based on our analysis of p, and it is bounded on the compact set $|z| \leq R$ since it is continuous.

Liouville's Theorem implies that f is in fact constant, which would force p to be constant, which is a contradiction. □

Theorem 7.40. *Let $f_n : \Omega \to \mathbb{C}$ be a sequence of analytic functions on an open set Ω. If f_n converges uniformly to f, then f is analytic.*

Recall that for a function to be analytic at one point we require that the function be differentiable in a neighbourhood of the point, and therefore the assumption on Ω being open is natural. Being analytic is a local property, and requiring that the uniform convergence holds only on compact sets would suffice.

Proof. Let $z \in \Omega$. Choose $r > 0$ sufficiently small so that $B_r(z) \subset \Omega$. Since f_n is analytic in Ω we can apply Cauchy's formula to obtain

$$f_n(z) = \frac{1}{2\pi \mathrm{i}} \int_{\partial B_r(z)} \frac{f_n(w)}{w-z} \mathrm{d}w.$$

Taking limits as n goes to infinity, and assuming that we can move the limit inside the integral we would obtain

$$f(z) = \frac{1}{2\pi \mathrm{i}} \int_{\partial B_r(z)} \frac{f(w)}{w-z} \mathrm{d}w.$$

We have seen before that this implies that f is differentiable (in fact infinitely differentiable) and obtained an expression for its derivative (see Theorem 7.35). So the only thing left is to justify moving the limit inside the integral. Notice that this is really a one dimensional integral and we can apply the results learnt earlier in the year. Taking $\gamma(t) = z + r\mathrm{e}^{\mathrm{i}t}$ for $t \in [0, 2\pi)$, we have $\gamma'(t) = \mathrm{i}r\mathrm{e}^{\mathrm{i}t}$ and so

$$\int_{\partial B_r(z)} \frac{f_n(w)}{w-z}\mathrm{d}w = \int_0^{2\pi} \frac{f_n(z+r\mathrm{e}^{\mathrm{i}t})}{r\mathrm{e}^{\mathrm{i}t}}\mathrm{i}r\mathrm{e}^{\mathrm{i}t}\mathrm{d}t = \mathrm{i}\int_0^{2\pi} f_n(z+r\mathrm{e}^{\mathrm{i}t})\mathrm{d}t. \tag{7.25}$$

For fixed z, as a function of t we have that $f_n(z + r\mathrm{e}^{\mathrm{i}t})$ converges uniformly to $f(z + r\mathrm{e}^{\mathrm{i}t})$ and applying Theorem 2.16 we can move the limit inside the integral, obtaining

$$\lim_{n\to\infty}\int_{\partial B_r(z)} \frac{f_n(w)}{w-z}\mathrm{d}w = \lim_{n\to\infty}\mathrm{i}\int_0^{2\pi} f_n(z+r\mathrm{e}^{\mathrm{i}t})\mathrm{d}t = \mathrm{i}\int_0^{2\pi} f(z+r\mathrm{e}^{\mathrm{i}t})\mathrm{d}t.$$

Notice that we have (reading the expression (7.25) backwards, now for f instead of for f_n)

$$\mathrm{i}\int_0^{2\pi} f(z+r\mathrm{e}^{\mathrm{i}t})\mathrm{d}t = \int_{\partial B_r(z)} \frac{f(w)}{w-z}\mathrm{d}w,$$

obtaining the result. □

7.3.3 Applications of Cauchy's formula to evaluate integrals in $\mathbb{R}$ (NOT COVERED AND NOT EXAMINABLE)

We present various examples that illustrate a more general theory (of residues) for computing integrals of functions over $\mathbb{R}$.

Consider for example

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\mathrm{d}x.$$

The idea is to consider the contours γ_1 and γ_2 in Figure 7.4.

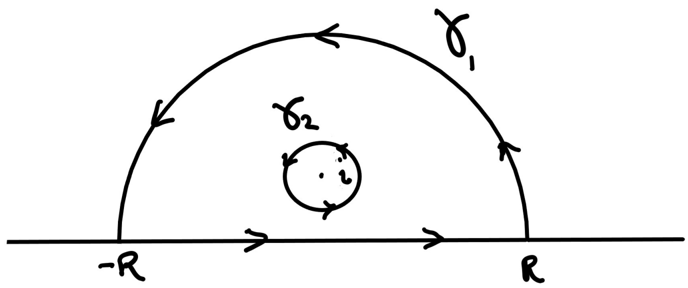

Figure 7.4: Contours

γ_1 is formed by the segment joining $-R$ and R, together with the half circle or radius R. The contour γ_2 is a circle centred at i and of radius $r < 1$. To understand the choice of curves, notice that we can rewrite the integral as

$$\int_{-\infty}^{\infty} \frac{1}{(x-\mathrm{i})(x+\mathrm{i})}\mathrm{d}x.$$

Notice that in the region enclosed by γ_2 the function (the integrand extended to a function on $\mathbb{C}$)

$$f(z) := \frac{1}{(z-\mathrm{i})(z+\mathrm{i})}$$

is analytic except for at $z = \mathrm{i}$. By the deformation of contours Theorem we know that

$$\int_{\gamma_1} f(z)\mathrm{d}z = \int_{\gamma_2} f(z)\mathrm{d}z,$$

since the two curves have the same orientation. Now

$$\int_{\gamma_1} f(z)\mathrm{d}z = \int_{-R}^{R} f(z)\mathrm{d}z + \int_{\text{arc}} f(z)\mathrm{d}z.$$

We parametrise the arc by $R\mathrm{e}^{\mathrm{i}t}$ for $t \in [0,\pi)$. We have

$$\int_{\text{arc}} f(z)\mathrm{d}z = \int_{\text{arc}} \frac{1}{1+z^2}\mathrm{d}z = \int_0^{\pi} \frac{1}{1+R^2\mathrm{e}^{2\mathrm{i}t}} R\mathrm{i}\mathrm{e}^{\mathrm{i}t}\mathrm{d}t.$$

Therefore

$$\left|\int_{\text{arc}} f(z)\mathrm{d}z\right| \leq \int_0^{\pi} \frac{R}{R^2-1}\mathrm{d}t = \pi\frac{R}{R^2-1}.$$

As R tends to infinity the $\int_{\text{arc}} f(z)\mathrm{d}z$ equals zero. Therefore

$$\int_{-\infty}^{\infty} f(z)\mathrm{d}z = \int_{\gamma_1} f(z)\mathrm{d}z = \int_{\gamma_2} f(z)\mathrm{d}z.$$

Now

$$\int_{\gamma_2} f(z)\mathrm{d}z = \int_{\partial B_r(\mathrm{i})} \frac{1}{z+i}\frac{1}{z-i}\mathrm{d}z.$$

Recall that by Cauchy's formula if $g(z)$ is analytic in the interior of a positively oriented curve then

$$\int_{\gamma} g(z)\frac{1}{z-a}\mathrm{d}z = 2\pi\mathrm{i}g(a).$$

Therefore, taking $g(z) = \frac{1}{z+i}$ we obtain

$$\int_{\partial B_r(\mathrm{i})} \frac{1}{z+i}\frac{1}{z-i}\mathrm{d}z = 2\pi\mathrm{i}\frac{1}{2\mathrm{i}} = \pi,$$

which yields

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\mathrm{d}x = \pi.$$

As a second example, let's compute

$$\int_{\infty}^{\infty} \frac{1}{1+x^4}\mathrm{d}x.$$

Notice that the function

$$\frac{1}{1+z^4}$$

has four singularities at the points

$$\mathrm{e}^{\pi\mathrm{i}/4} \qquad \mathrm{e}^{3\pi\mathrm{i}/4} \qquad \mathrm{e}^{-\pi\mathrm{i}/4} \qquad \mathrm{e}^{-3\pi\mathrm{i}/4},$$

and so if we choose a contour similar to the one above (an expanding semi-circle) there will be two singularities in the interior. We obtain the picture 7.5.

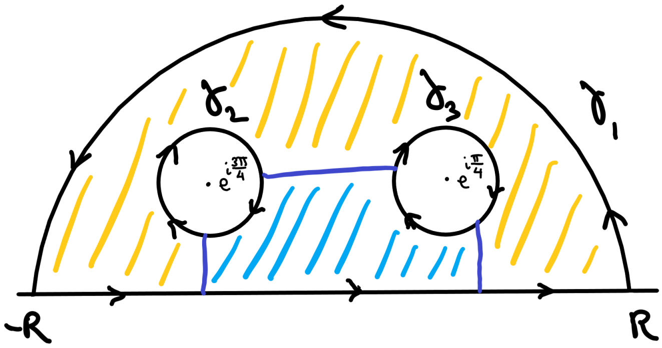

Figure 7.5: Contours

Now γ_1 is built out of the line joining $-R$ and R, together with the semi-circle or radius R centred at 0. γ_2 and γ_3 correspond to circles centred at $e^{3\pi i/4}$ and $e^{\pi i/4}$, oriented clock-wise (positively with respect to both the blue-shaded and yellow-shaded regions). Notice that with those orientations we have

$$\int_{\gamma_1} \frac{1}{1+z^4} dz + \int_{\gamma_2} \frac{1}{1+z^4} dz + \int_{\gamma_3} \frac{1}{1+z^4} dz = 0.$$

We start by considering the integral over γ_1. Wave

$$\int_{\gamma_1} \frac{1}{1+z^4} dz = \int_{-R}^{R} \frac{1}{1+z^4} dz + \int_{\text{arc}} \frac{1}{1+z^4} dz.$$

We will show that the integral over the arc goes to zero as R goes to infinity. Indeed

$$\left| \int_{\text{arc}} \frac{1}{1+z^4} dz \right| \leq \int_{\text{arc}} \frac{1}{R^4-1} |dz| = \frac{\pi R}{R^4-1},$$

which goes to zero as R goes to infinity. Above we have used that $\int_{\text{arc}} |dz|$ =length(arc)= πR.

We now consider the integral over γ_2. We have (denoting by γ_2^- the anti-clockwise parametrisation of the circle)

$$\int_{\gamma_2} \frac{1}{1+z^4} dz = -\int_{\gamma_2^-} \frac{1}{(z-e^{i\pi/4})(z-e^{3i\pi/4})(z-e^{-i\pi/4})(z-e^{-3i\pi/4})} dz$$
$$= -\int_{\gamma_2^-} \frac{g(z)}{(z-e^{3i\pi/4})} dz = -2\pi i g(e^{3i\pi/4}),$$

where the above defines g as

$$g(z) = \frac{1}{(z-e^{i\pi/4})(z-e^{-i\pi/4})(z-e^{-3i\pi/4})},$$

and we have used Cauchy's formula as g is analytic inside the curve γ_2^-. Now

$$g(e^{3i\pi/4}) = \frac{1}{(e^{3i\pi/4}-e^{i\pi/4})(e^{3i\pi/4}-e^{-i\pi/4})(e^{3i\pi/4}-e^{-3i\pi/4})} = \frac{1}{(-\sqrt{2})(-\sqrt{2}+\sqrt{2}i)(\sqrt{2}i)}.$$

Now we consider the integral over γ_3

$$\int_{\gamma_3} \frac{1}{1+z^4} dz = -\int_{\gamma_3^-} \frac{1}{(z-e^{i\pi/4})(z-e^{3i\pi/4})(z-e^{-i\pi/4})(z-e^{-3i\pi/4})} dz$$

$$= -\int_{\gamma_3^-} \frac{h(z)}{(z - e^{i\pi/4})} dz = -2\pi i h(e^{i\pi/4}),$$

where the above defines h as

$$h(z) = \frac{1}{(z - e^{3i\pi/4})(z - e^{-i\pi/4})(z - e^{-3i\pi/4})},$$

and we have used Cauchy's formula as h is analytic inside the curve γ_3^-. Now

$$h(e^{i\pi/4}) = \frac{1}{(e^{i\pi/4} - e^{3i\pi/4})(e^{i\pi/4} - e^{-i\pi/4})(e^{i\pi/4} - e^{-3i\pi/4})} = \frac{1}{\sqrt{2}(\sqrt{2}i)(\sqrt{2} + \sqrt{2}i)}.$$

Since we have

$$\int_{-\infty}^{\infty} \frac{1}{1+z^4} dz = -\int_{\gamma_2} \frac{1}{1+z^4} d \ - \int_{\gamma_3} \frac{1}{1+z^4} dz$$

we obtain

$$\int_{-\infty}^{\infty} \frac{1}{1+z^4} dz = 2\pi i \frac{1}{(-\sqrt{2})(-\sqrt{2} + \sqrt{2}i)(\sqrt{2}i)} + 2\pi i \frac{1}{\sqrt{2}(\sqrt{2}i)(\sqrt{2} + \sqrt{2}i)} = \frac{\pi\sqrt{2}}{2}.$$

In addition to being able to integrate quotients involving polynomials, we can integration some trigonometric functions. For example

$$\int_{-\infty}^{\infty} \frac{\cos(3x)}{4 + x^2} dx.$$

We can rewrite this integral as

$$\mathbf{Re} \int_{-\infty}^{\infty} \frac{e^{3iz}}{(z - 2i)(z + 2i)} dz,$$

and we can actually drop the $\mathbf{Re}$ part as the imaginary part will be an odd integrand and it will vanish. We consider the contours (notice they are both oriented counter-clockwise)

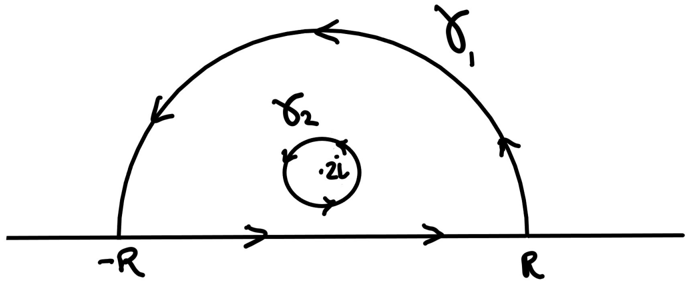

Figure 7.6: Contours

As before,

$$\int_{\gamma_1} \frac{e^{3iz}}{(z - 2i)(z + 2i)} dz = \int_{\gamma_2} \frac{e^{3iz}}{(z - 2i)(z + 2i)} dz dz.$$

Also

$$\int_{\gamma_1} \frac{e^{3iz}}{(z - 2i)(z + 2i)} dz = \int_{-R}^{R} \frac{e^{3iz}}{(z - 2i)(z + 2i)} dz + \int_{\text{arc}} \frac{e^{3iz}}{(z - 2i)(z + 2i)} |dz|.$$

We consider first the integral over the arc (half circle of radius R). We have, for $R >> 4$

$$\left|\int_{\text{arc}} \frac{\mathrm{e}^{3\mathrm{i}z}}{(z-2\mathrm{i})(z+2\mathrm{i})}\mathrm{d}z\right| \leq \int_{\text{arc}} \frac{|\mathrm{e}^{3\mathrm{i}z}|}{|z^2+4|}|\mathrm{d}z| \leq \int_{\text{arc}} \frac{\mathrm{e}^{-3\,\mathbf{Im}\,z}}{R^2-4}|\mathrm{d}z| \leq \frac{\pi R}{R^2-4} \xrightarrow[R\to\infty]{} 0,$$

where we have used that along the arc, $\mathbf{Im}\,z \geq 0$ and so $\mathrm{e}^{-3\,\mathbf{Im}\,z} \leq 1$. Now for γ_2 (remember it is oriented anti-clockwise)

$$\int_{\gamma_2} \frac{\mathrm{e}^{3\mathrm{i}z}}{(z-2\mathrm{i})(z+2\mathrm{i})}\mathrm{d}z = \int_{\gamma_2} \frac{g(z)}{z-2\mathrm{i}}\mathrm{d}z = 2\pi\mathrm{i}g(2\mathrm{i}),$$

where

$$g(z) = \frac{\mathrm{e}^{3\mathrm{i}z}}{z+2\mathrm{i}}\mathrm{d}z$$

and we have used Cauchy's formula since g is analytic inside γ_2. We have

$$g(2\mathrm{i}) = \frac{\mathrm{e}^{-6}}{4\mathrm{i}}.$$

Therefore

$$\int_{-\infty}^{\infty} \frac{\mathrm{e}^{3\mathrm{i}z}}{(z-2\mathrm{i})(z+2\mathrm{i})}\mathrm{d}z = \int_{\gamma_1} \frac{\mathrm{e}^{3\mathrm{i}z}}{(z-2\mathrm{i})(z+2\mathrm{i})} = \int_{\gamma_2} \frac{\mathrm{e}^{3\mathrm{i}z}}{(z-2\mathrm{i})(z+2\mathrm{i})} = 2\pi\mathrm{i}g(2\mathrm{i}) = \frac{\pi}{2}\mathrm{e}^{-6}.$$

We use a similar approach for

$$\int_{-\infty}^{\infty} \frac{x\sin x}{1+x^2}\mathrm{d}x = \frac{1}{\mathrm{i}}\int_{-\infty}^{\infty} \frac{z\mathrm{e}^{\mathrm{i}z}}{1+z^2}\mathrm{d}z.$$

Notice the real part of the integral vanishes as the integrand is odd (hence dividing the i). We consider the following contours of integration We consider the contours (notice they are both oriented counter-clockwise) By Cauchy's Theorem since the integrand is analytic in the region between the curves we

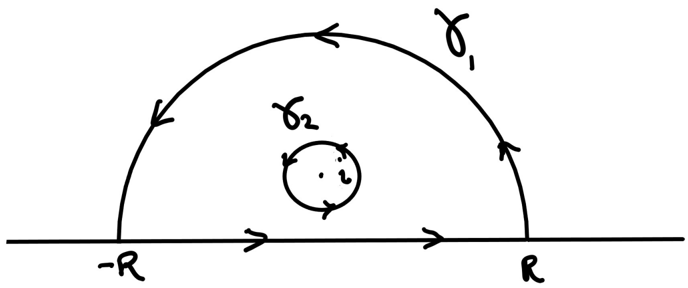

Figure 7.7: Contours

have

$$\int_{\gamma_1} \frac{z\mathrm{e}^{\mathrm{i}z}}{1+z^2}\mathrm{d}z = \int_{\gamma_2} \frac{z\mathrm{e}^{\mathrm{i}z}}{1+z^2}\mathrm{d}z.$$

Now for γ_2

$$\int_{\gamma_2} \frac{z\mathrm{e}^{\mathrm{i}z}}{(z-\mathrm{i})(z+\mathrm{i})}\mathrm{d}z = \int_{\gamma_2} \frac{h(z)}{(z-\mathrm{i})}\mathrm{d}z = 2\pi\mathrm{i}h(\mathrm{i}),$$

for

$$h(z) = \frac{z\mathrm{e}^{\mathrm{i}z}}{z+\mathrm{i}},$$

and so

$$\int_{\gamma_2} \frac{z\mathrm{e}^{\mathrm{i}z}}{(z-\mathrm{i})(z+\mathrm{i})}\mathrm{d}z = 2\pi\mathrm{i}\frac{\mathrm{i}\mathrm{e}^{-1}}{2\mathrm{i}} = \frac{\pi}{\mathrm{e}}\mathrm{i}.$$

We now consider the integral over γ_1. We look at the integral along the arc.

$$\left|\int_{\mathsf{arc}} \frac{z\mathrm{e}^{\mathrm{i}z}}{1+z^2}\mathrm{d}z\right| = \left|\int_0^{\pi} \frac{R\mathrm{e}^{\mathrm{i}t}\mathrm{e}^{-R\sin t + R\mathrm{i}\cos t}}{1+R^2\mathrm{e}^{2\mathrm{i}t}} R\mathrm{i}\mathrm{e}^{\mathrm{i}t}\mathrm{d}t\right| \le \int_0^{\pi} \frac{R^2}{R^2-1}\mathrm{e}^{-R\sin t}\mathrm{d}t$$

$$\le 2\int_0^{\pi} \mathrm{e}^{-R\sin t}\mathrm{d}t = 4\int_0^{\pi/2} \mathrm{e}^{-R\sin t}\mathrm{d}t \le 4\int_0^{\pi/2} \mathrm{e}^{-R2t/\pi}\mathrm{d}t = -4\frac{\pi}{2R}\mathrm{e}^{-R2t/\pi}\Big|_0^{\pi/2}$$

$$= \frac{-2\pi}{R}\left[\mathrm{e}^{-R} - 1\right] \xrightarrow[R\to\infty]{} 0.$$

Therefore

$$\int_{-\infty}^{\infty} \frac{x\sin x}{1+x^2}\mathrm{d}x = \frac{1}{\mathrm{i}}\int_{\gamma_1} \frac{z\mathrm{e}^{\mathrm{i}z}}{(z-\mathrm{i})(z+\mathrm{i})}\mathrm{d}z = \frac{1}{\mathrm{i}}\int_{\gamma_2} \frac{z\mathrm{e}^{\mathrm{i}z}}{(z-\mathrm{i})(z+\mathrm{i})}\mathrm{d}z = \frac{\pi}{\mathrm{e}}.$$

As the final example we consider an integrand that has a singularity along the natural path of integration

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\mathrm{d}x.$$

If we complexify the integrand, we are left with

$$\frac{1}{\mathrm{i}}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z,$$

since the real part of the integrand is odd. Since the denominator vanishes for a point in the real axis we need to modify the contours we consider.

The contour is formed of 4 curves, and since the integrand is analytic we know that

$$\int_{\gamma_1} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z + \int_{\gamma_2} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z + \int_{\gamma_3} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z + \int_{\gamma_4} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z = 0.$$

Now for γ_1

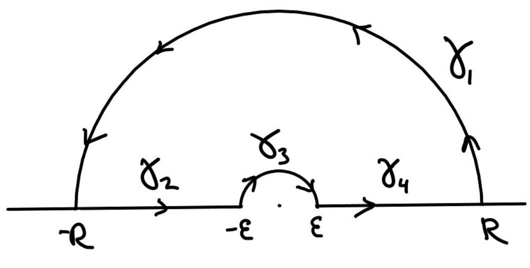

Figure 7.8: Contours

$$\int_{\gamma_1} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z = \int_0^{\pi} \frac{\mathrm{e}^{\mathrm{i}R\cos t - R\sin t}}{R\mathrm{e}^{\mathrm{i}t}}\mathrm{i}R\mathrm{e}^{\mathrm{i}t}\mathrm{d}t$$

and so

$$\left|\int_{\gamma_1} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z\right| \le \int_0^{\pi} \mathrm{e}^{R\sin t}\mathrm{d}t \xrightarrow[R\to\infty]{} 0$$

as we have seen before. As for γ_3, since it is oriented clock-wise

$$\int_{\gamma_3} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z = -\int_{\gamma_3^-} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z = -\int_0^\pi \frac{\mathrm{e}^{\mathrm{i}\varepsilon e^{\mathrm{i}t}}}{\varepsilon \mathrm{e}^{\mathrm{i}t}}\mathrm{i}\varepsilon\mathrm{e}^{\mathrm{i}t}\mathrm{d}t = -\mathrm{i}\int_0^\pi \mathrm{e}^{\mathrm{i}\varepsilon\cos t}\mathrm{e}^{-\varepsilon\sin t}\mathrm{d}t \xrightarrow[\varepsilon\to 0]{} -\pi\mathrm{i}.$$

Therefore

$$\int_{-\infty}^{\infty} \frac{\sin x}{x}\mathrm{d}x = \frac{1}{\mathrm{i}}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z$$

$$= \lim_{\varepsilon\to 0}\lim_{R\to\infty} \frac{1}{\mathrm{i}}\left[\int_{\gamma_2} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z + \int_{\gamma_4} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z\right] = \lim_{\varepsilon\to 0}\lim_{R\to\infty}\left[-\frac{1}{\mathrm{i}}\int_{\gamma_1} \frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z - \int_{\gamma_3} \frac{1}{\mathrm{i}}\frac{\mathrm{e}^{\mathrm{i}z}}{z}\mathrm{d}z\right] = \pi.$$

Bibliography

[1] Rudin, W. Principles of mathematical analysis. Third edition. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., 1976.

[2] Lebl, J. Basic Analysis: Introduction to Real Analysis. Available to download at https://www.jirka.org/ra/

[3] Ross, K. Elementary Analysis: the theory of calculus. Springer 2013

[4] Abbott, S. Understanding analysis. Springer, 2001.

[5] Gelbaum, B.R., & Olmsted, J.M.H. Counterexamples in Analysis, Holden Day Inc, 1964.