Calculus 1 to 4 (2004–2006)

Axel Schüler

January 3, 2007

Contents

Chapter 1

Real and Complex Numbers

Basics

Notations

$\mathbb{R}$	Real numbers
$\mathbb{C}$	Complex numbers
$\mathbb{Q}$	Rational numbers
$\mathbb{N} = \{1, 2, \dots\}$	positive integers (natural numbers)
$\mathbb{Z}$	Integers

We know that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$. We write $\mathbb{R}_+$, $\mathbb{Q}_+$ and $\mathbb{Z}_+$ for the non-negative real, rational, and integer numbers $x \geq 0$, respectively. The notions $A \subset B$ and $A \subseteq B$ are equivalent. If we want to point out that B is strictly bigger than A we write $A \subsetneq B$.

We use the following symbols

$:=$	defining equation
$\curvearrowright$, $\Rightarrow$	implication, "if ..., then ..."
$\Longleftrightarrow$	"if and only if", equivalence
$\forall$	for all
$\exists$	there exists

Let $a < b$ fixed real numbers. We denote the *intervals* as follows

$[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$	closed interval
$(a, b) := \{x \in \mathbb{R} \mid a < x < b\}$	open interval
$[a, b) := \{x \in \mathbb{R} \mid a \leq x < b\}$	half-open interval
$(a, b] := \{x \in \mathbb{R} \mid a < x \leq b\}$	half-open interval
$[a, \infty) := \{x \in \mathbb{R} \mid a \leq x\}$	closed half-line
$(a, \infty) := \{x \in \mathbb{R} \mid a < x\}$	open half-line
$(-\infty, b] := \{x \in \mathbb{R} \mid x \leq b\}$	closed half-line
$(-\infty, b) := \{x \in \mathbb{R} \mid x < b\}$	open half-line

(a) Sums and Products

Let us recall the meaning of the sum sign $\sum$ and the product sign $\prod$. Suppose $m \leq n$ are integers, and a_k, $k = m, \ldots, n$ are real numbers. Then we set

$$\sum_{k=m}^{n} a_k := a_m + a_{m+1} + \cdots + a_n, \qquad \prod_{k=m}^{n} a_k := a_m a_{m+1} \cdots a_n.$$

In case $m = n$ the sum and the product consist of one summand and one factor only, respectively. In case $n < m$ it is customary to set

$$\sum_{k=m}^{n} a_k := 0, \text{(empty sum)} \qquad \prod_{k=m}^{n} a_k := 1 \quad \text{(empty product)}.$$

The following rules are obvious: If $m \leq n \leq p$ and $d \in \mathbb{Z}$ are integers then

$$\sum_{k=m}^{n} a_k + \sum_{k=n+1}^{p} a_k = \sum_{k=m}^{p} a_k, \qquad \sum_{k=m}^{n} a_k = \sum_{k=m+d}^{n+d} a_{k-d} \quad \text{(index shift)}.$$

We have for $a \in \mathbb{R}$, $\sum_{k=m}^{n} a = (n - m + 1)a$.

(b) Mathematical Induction

Mathematical induction is a powerful method to prove theorems about natural numbers.

Theorem 1.1 (Principle of Mathematical Induction) *Let $n_0 \in \mathbb{Z}$ be an integer. To prove a statement $A(n)$ for all integers $n \geq n_0$ it is sufficient to show:*

(I) *$A(n_0)$ is true.*
(II) *For any $n \geq n_0$: If $A(n)$ is true, so is $A(n+1)$* (Induction step).

It is easy to see how the principle works: First, $A(n_0)$ is true. Apply (II) to $n = n_0$ we obtain that $A(n_0 + 1)$ is true. Successive application of (II) yields $A(n_0 + 2)$, $A(n_0 + 3)$ are true and so on.

Example 1.1 (a) For all nonnegative integers n we have $\sum_{k=1}^{n}(2k - 1) = n^2$.

Proof. We use induction over n. In case $n = 0$ we have an empty sum on the left hand side (lhs) and $0^2 = 0$ on the right hand side (rhs). Hence, the statement is true for $n = 0$.
Suppose it is true for some fixed n. We shall prove it for $n + 1$. By the definition of the sum and by induction hypothesis, $\sum_{k=1}^{n}(2k - 1) = n^2$, we have

$$\sum_{k=1}^{n+1}(2k-1) = \sum_{k=1}^{n}(2k-1) + 2(n+1) - 1 \underset{\text{ind. hyp.}}{=} n^2 + 2n + 1 = (n+1)^2.$$

This proves the claim for $n + 1$. ∎

(b) For all positive integers $n \geq 8$ we have $2^n > 3n^2$.
Proof. In case $n = 8$ we have

$$2^n = 2^8 = 256 > 192 = 3 \cdot 64 = 3 \cdot 8^2 = 3n^2;$$

and the statement is true in this case.
Suppose it is true for some fixed $n \geq 8$, i. e. $2^n > 3n^2$ (induction hypothesis). We will show that the statement is true for $n+1$, i. e. $2^{n+1} > 3(n+1)^2$ (induction assertion). Note that $n \geq 8$ implies

$$\begin{aligned}
n - 1 \geq 7 > 2 \quad &\Longrightarrow (n-1)^2 > 4 > 2 \quad &&\Longrightarrow n^2 - 2n - 1 > 0 \\
\Longrightarrow 3(n^2 - 2n - 1) > 0 \quad &\Longrightarrow 3n^2 - 6n - 3 > 0 \quad &&\mid +3n^2 + 6n + 3 \\
\Longrightarrow 6n^2 > 3n^2 + 6n + 3 \quad &\Longrightarrow 2 \cdot 3n^2 > 3(n^2 + 2n + 1) \\
\Longrightarrow 2 \cdot 3n^2 > 3(n+1)^2. \quad &&&(1.1)
\end{aligned}$$

By induction assumption, $2^{n+1} = 2 \cdot 2^n > 2 \cdot 3n^2$. This together with (1.1) yields $2^{n+1} > 3(n+1)^2$. Thus, we have shown the induction assertion. Hence the statement is true for all positive integers $n \geq 8$. ∎

For a positive integer $n \in \mathbb{N}$ we set

$$n! := \prod_{k=1}^{n} k, \qquad \text{read: “}n\text{ factorial,”} \quad 0! = 1! = 1.$$

(c) Binomial Coefficients

For non-negative integers $n, k \in \mathbb{Z}_+$ we define

$$\binom{n}{k} := \prod_{i=1}^{k} \frac{n-i+1}{i} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 2 \cdot 1}.$$

The numbers $\binom{n}{k}$ (read: “n choose k”) are called *binomial coefficients* since they appear in the binomial theorem, see Proposition 1.4 below. It just follows from the definition that

$$\begin{aligned}
&\binom{n}{k} = 0 \quad \text{for } k > n, \\
&\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k} \quad \text{for } 0 \leq k \leq n.
\end{aligned}$$

Lemma 1.2 *For $0 \leq k \leq n$ we have:*

$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

Proof. For $k = n$ the formula is obvious. For $0 \le k \le n-1$ we have

$$\begin{aligned}\binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-k-1)!} \\ &= \frac{(k+1)n! + (n-k)n!}{(k+1)!(n-k)!} = \frac{(n+1)!}{(k+1)!(n-k)!} = \binom{n+1}{k+1}.\end{aligned}$$

■

We say that X is an n-set if X has exactly n elements. We write $\operatorname{Card} X = n$ (from "cardinality") to denote the number of elements in X.

Lemma 1.3 *The number of k-subsets of an n-set is $\binom{n}{k}$.*

The Lemma in particular shows that $\binom{n}{k}$ is always an integer (which is not obvious by its definition).

Proof. We denote the number of k-subsets of an n set X_n by C_k^n. It is clear that $C_0^n = C_n^n = 1$ since $\varnothing$ is the only 0-subset of X_n and X_n itself is the only n-subset of X_n. We use induction over n. The case $n = 1$ is obvious since $C_0^1 = C_1^1 = \binom{1}{0} = \binom{1}{1} = 1$. Suppose that the claim is true for some fixed n. We will show the statement for the $(n+1)$-set $X = \{1, \dots, n+1\}$ and all k with $1 \le k \le n$. The family of $(k+1)$-subsets of X splits into two disjoint classes. In the first class $\mathcal{A}_1$ every subset contains $n+1$; in the second class $\mathcal{A}_2$, not. To form a subset in $\mathcal{A}_1$ one has to choose another k elements out of $\{1, \dots, n\}$. By induction assumption the number is $\operatorname{Card} \mathcal{A}_1 = C_k^n = \binom{n}{k}$. To form a subset in $\mathcal{A}_2$ one has to choose $k+1$ elements out of $\{1, \dots, n\}$. By induction assumption this number is $\operatorname{Card} \mathcal{A}_2 = C_{k+1}^n = \binom{n}{k+1}$. By Lemma 1.2 we obtain

$$C_{k+1}^{n+1} = \operatorname{Card} \mathcal{A}_1 + \operatorname{Card} \mathcal{A}_2 = \binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

which proves the induction assertion. ■

Proposition 1.4 (Binomial Theorem) *Let $x, y \in R$ and $n \in \mathbb{N}$. Then we have*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k.$$

Proof. We give a direct proof. Using the distributive law we find that each of the 2^n summands of product $(x+y)^n$ has the form $x^{n-k}\, y^k$ for some $k = 0, \dots, n$. We number the n factors as $(x+y)^n = f_1 \cdot f_2 \cdots f_n$, $f_1 = f_2 = \cdots = f_n = x + y$. Let us count how often the summand $x^{n-k}\, y^k$ appears. We have to choose k factors y out of the n factors $f_1, \dots, f_n$. The remaining $n - k$ factors must be x. This gives a 1-1-correspondence between the k-subsets of $\{f_1, \dots, f_n\}$ and the different summands of the form $x^{n-k}\, y^k$. Hence, by Lemma 1.3 their number is $C_k^n = \binom{n}{k}$. This proves the proposition. ■

1.1 Real Numbers

In this lecture course we ***assume the system of real numbers to be given***. Recall that the set of integers is $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ while the fractions of integers $\mathbb{Q} = \{\frac{m}{n} \mid m, n \in \mathbb{Z},\ n \neq 0\}$ form the set of rational numbers.

A satisfactory discussion of the main concepts of analysis such as convergence, continuity, differentiation and integration must be based on an accurately defined number concept.
An existence proof for the real numbers is given in [Rud76, Appendix to Chapter 1]. The author explicitly constructs the real numbers $\mathbb{R}$ starting from the rational numbers $\mathbb{Q}$.
The aim of the following two sections is to formulate the axioms which are sufficient to derive all properties and theorems of the real number system.
The rational numbers are inadequate for many purposes, both as a field and an ordered set. For instance, there is no rational x with $x^2 = 2$. This leads to the introduction of irrational numbers which are often written as infinite decimal expansions and are considered to be "approximated" by the corresponding finite decimals. Thus the sequence

$$1,\ 1.4,\ 1.41,\ 1.414,\ 1.4142,\ \dots$$

"tends to $\sqrt{2}$." But unless the irrational number $\sqrt{2}$ has been clearly defined, the question must arise: What is it that this sequence "tends to"?
This sort of question can be answered as soon as the so-called "real number system" is constructed.

Example 1.2 As shown in the exercise class, there is no rational number x with $x^2 = 2$. Set

$$A = \{x \in \mathbb{Q}_+ \,|\, x^2 < 2\} \quad \text{and} \quad B = \{x \in \mathbb{Q}_+ \mid x^2 > 2\}.$$

Then $A \cup B = \mathbb{Q}_+$ and $A \cap B = \varnothing$. One can show that in the rational number system, A has no largest element and B has no smallest element, for details see Appendix A or Rudin's book [Rud76, Example 1.1, page 2]. This example shows that the system of rational numbers has certain gaps in spite of the fact that between any two rationals there is another: If $r < s$ then $r < (r+s)/2 < s$. The real number system fills these gaps. This is the principal reason for the fundamental role which it plays in analysis.
We start with the brief discussion of the general concepts of *ordered set* and *field*.

1.1.1 Ordered Sets

Definition 1.1 (a) Let S be a set. An *order* (or *total order*) on S is a relation, denoted by $<$, with the following properties. Let $x, y, z \in S$.

(i) One and only one of the following statements is true.

$$x < y, \quad x = y, \quad y < x \quad (\textit{trichotomy})$$

(ii) $x < y$ and $y < z$ implies $x < z$ (*transitivity*).

In this case S is called an *ordered set*.
(b) Suppose $(S, <)$ is an ordered set, and $E \subseteq S$. If there exists a $\beta \in S$ such that $x \leq \beta$ for all $x \in E$, we say that E is *bounded above*, and call β an *upper bound* of E. *Lower bounds* are defined in the same way with $\geq$ in place of $\leq$.
If E is both bounded above and below, we say that E is *bounded*.

The statement $x < y$ may be read as "x is less than y" or "x precedes y". It is convenient to write $y > x$ instead of $x < y$. The notation $x \leq y$ indicates $x < y$ or $x = y$. In other words, $x \leq y$ is the negation of $x > y$. For example, $\mathbb{R}$ is an ordered set if $r < s$ is defined to mean that $s - r > 0$ is a positive real number.

Example 1.3 (a) The intervals $[a, b]$, $(a, b]$, $[a, b)$, (a, b), $(-\infty, b)$, and $(-\infty, b]$ are bounded above by b and all numbers greater than b.
(b) $E := \{\frac{1}{n} \mid n \in \mathbb{N}\} = \{1, \frac{1}{2}, \frac{1}{3}, \dots\}$ is bounded above by any $\alpha \geq 1$. It is bounded below by 0.

Definition 1.2 Suppose S is an ordered set, $E \subseteq S$, an E is bounded above. Suppose there exists an $\alpha \in S$ such that

(i) α is an upper bound of E.
(ii) If β is an upper bound of E then $\alpha \leq \beta$.

Then α is called the *supremum of* E (or *least upper bound*) of E. We write

$$\alpha = \sup E.$$

An equivalent formulation of (ii) is the following:

(ii)′ If $\beta < \alpha$ then β is not an upper bound of E.

The *infimum* (or *greatest lower bound*) of a set E which is bounded below is defined in the same manner: The statement

$$\alpha = \inf E$$

means that α is a lower bound of E and for all lower bounds β of E we have $\beta \leq \alpha$.

Example 1.4 (a) If $\alpha = \sup E$ exists, then α may or may not belong to E. For instance consider $[0, 1)$ and $[0, 1]$. Then

$$1 = \sup[0, 1) = \sup[0, 1],$$

however $1 \notin [0, 1)$ but $1 \in [0, 1]$. We will show that $\sup[0, 1] = 1$. Obviously, 1 is an upper bound of $[0, 1]$. Suppose that $\beta < 1$, then β is not an upper bound of $[0, 1]$ since $\beta \ngeq 1$. Hence $1 = \sup[0, 1]$.
We show will show that $\sup[0, 1) = 1$. Obviously, 1 is an upper bound of this interval. Suppose that $\beta < 1$. Then $\beta < \frac{\beta+1}{2} < 1$. Since $\frac{\beta+1}{2} \in [0, 1)$, β is not an upper bound. Consequently, $1 = \sup[0, 1)$.
(b) Consider the sets A and B of Example 1.2 as subsets of the ordered set $\mathbb{Q}$. Since $A \cup B = \mathbb{Q}_+$ (there is no rational number with $x^2 = 2$) the upper bounds of A are exactly the elements of B.

Indeed, if $a \in A$ and $b \in B$ then $a^2 < 2 < b^2$. Taking the square root we have $a < b$. Since B contains no smallest member, *A has no supremum in $\mathbb{Q}_+$*.
Similarly, B is bounded below by any element of A. Since A has no largest member, *B has no infimum in $\mathbb{Q}$*.

Remarks 1.1 (a) It is clear from (ii) and the trichotomy of $<$ that there is at most one such α. Indeed, suppose α' also satisfies (i) and (ii), by (ii) we have $\alpha \leq \alpha'$ and $\alpha' \leq \alpha$; hence $\alpha = \alpha'$.
(b) If $\sup E$ exists *and belongs to E*, we call it the *maximum* of E and denote it by $\max E$. Hence, $\max E = \sup E$ and $\max E \in E$. Similarly, if the infimum of E exists *and belongs to E* we call it the *minimum* and denote it by $\min E$; $\min E = \inf E$, $\min E \in E$.

bounded subset of $\mathbb{Q}$	an upper bound	$\sup$	$\max$
$[0,1]$	2	1	1
$[0,1)$	2	1	—
A	2	—	—

(c) Suppose that α is an upper bound of E and $\alpha \in E$ then $\alpha = \max E$, that is, property (ii) in Definition 1.2 is automatically satisfied. Similarly, if $\beta \in E$ is a lower bound, then $\beta = \min E$.
(d) If E is a finite set it has always a maximum and a minimum.

1.1.2 Fields

Definition 1.3 A *field* is a set F with two operations, called *addition* and *multiplication* which satisfy the following so-called "field axioms" (A), (M), and (D):

(A) Axioms for addition

(A1) If $x \in F$ and $y \in F$ then their sum $x + y$ is in F.
(A2) Addition is commutative: $x + y = y + x$ for all $x, y \in F$.
(A3) Addition is associative: $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$.
(A4) F contains an element 0 such that $0 + x = x$ for all $x \in F$.
(A5) To every $x \in F$ there exists an element $-x \in F$ such that $x + (-x) = 0$.

(M) Axioms for multiplication

(M1) If $x \in F$ and $y \in F$ then their product xy is in F.
(M2) Multiplication is commutative: $xy = yx$ for all $x, y \in F$.
(M3) Multiplication is associative: $(xy)z = x(yz)$ for all $x, y, z \in F$.
(M4) F contains an element 1 such that $1x = x$ for all $x \in F$.
(M5) If $x \in F$ and $x \neq 0$ then there exists an element $1/x \in F$ such that $x \cdot (1/x) = 1$.

(D) The distributive law

$$x(y + z) = xy + xz$$

holds for all $x, y, z \in F$.

Remarks 1.2 (a) One usually writes

$$x - y,\ \frac{x}{y},\ x + y + z,\ xyz,\ x^2,\ x^3,\ 2x, \ldots$$

in place of

$$x + (-y),\ x \cdot \frac{1}{y},\ (x + y) + z,\ (xy)z,\ x \cdot x,\ x \cdot x \cdot x,\ 2x, \ldots$$

(b) The field axioms clearly hold in $\mathbb{Q}$ if addition and multiplication have their customary meaning. Thus $\mathbb{Q}$ is a field. The integers $\mathbb{Z}$ form *not* a field since $2 \in \mathbb{Z}$ has no multiplicative inverse (axiom (M5) is not fulfilled).
(c) The smallest field is $\mathbb{F}_2 = \{0, 1\}$ consisting of the neutral element 0 for addition and the neutral element 1 for multiplication. Multiplication and addition are defined as follows

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

. It is easy to check the field axioms (A), (M), and (D) directly.
(d) (A1) to (A5) and (M1) to (M5) mean that both $(F, +)$ and $(F \setminus \{0\}, \cdot)$ are *commutative (or abelian) groups*, respectively.

Proposition 1.5 *The axioms of addition imply the following statements.*
(a) *If* $x + y = x + z$ *then* $y = z$ *(Cancellation law).*
(b) *If* $x + y = x$ *then* $y = 0$ *(The element* 0 *is unique).*
(c) *If* $x + y = 0$ *the* $y = -x$ *(The inverse* $-x$ *is unique).*
(d) $-(-x) = x$.

Proof. If $x + y = x + z$, the axioms (A) give

$$y \underset{\text{A 4}}{=} 0 + y \underset{\text{A 5}}{=} (-x + x) + y \underset{\text{A 3}}{=} -x + (x + y) \underset{\text{assump.}}{=} -x + (x + z)$$
$$\underset{\text{A 3}}{=} (-x + x) + z \underset{\text{A 5}}{=} 0 + z \underset{\text{A 4}}{=} z.$$

This proves (a). Take $z = 0$ in (a) to obtain (b). Take $z = -x$ in (a) to obtain (c). Since $-x + x = 0$, (c) with $-x$ in place of x and x in place of y, gives (d). ∎

Proposition 1.6 *The axioms for multiplication imply the following statements.*
(a) *If* $x \neq 0$ *and* $xy = xz$ *then* $y = z$ *(Cancellation law).*
(b) *If* $x \neq 0$ *and* $xy = x$ *then* $y = 1$ *(The element* 1 *is unique).*
(c) *If* $x \neq 0$ *and* $xy = 1$ *then* $y = 1/x$ *(The inverse* $1/x$ *is unique).*
(d) *If* $x \neq 0$ *then* $1/(1/x) = x$.

The proof is so similar to that of Proposition 1.5 that we omit it.

Proposition 1.7 *The field axioms imply the following statements, for any* $x, y, z \in F$
(a) $0x = 0$.
(b) *If* $xy = 0$ *then* $x = 0$ *or* $y = 0$.
(c) $(-x)y = -(xy) = x(-y)$.
(d) $(-x)(-y) = xy$.

Proof. $0x + 0x = (0 + 0)x = 0x$. Hence 1.5 (b) implies that $0x = 0$, and (a) holds.
Suppose to the contrary that both $x \neq 0$ and $y \neq 0$ then (a) gives

$$1 = \frac{1}{y} \cdot \frac{1}{x} xy = \frac{1}{y} \cdot \frac{1}{x} 0 = 0,$$

a contradiction. Thus (b) holds.
The first equality in (c) comes from

$$(-x)y + xy = (-x + x)y = 0y = 0,$$

combined with 1.5 (b); the other half of (c) is proved in the same way. Finally,

$$(-x)(-y) = -[x(-y)] = -[-xy] = xy$$

by (c) and 1.5 (d). ∎

1.1.3 Ordered Fields

In analysis dealing with equations is as important as dealing with inequalities. Calculations with inequalities are based on the ordering axioms. It turns out that all can be reduced to the notion of positivity.
In F there are distinguished positive elements ($x > 0$) such that the following axioms are valid.

Definition 1.4 An *ordered field* is a field F which is also an ordered set, such that for all $x, y, z \in F$

(O) Axioms for ordered fields

(O1) $x > 0$ and $y > 0$ implies $x + y > 0$,
(O2) $x > 0$ and $y > 0$ implies $xy > 0$.

If $x > 0$ we call x *positive*; if $x < 0$, x is *negative*.

For example $\mathbb{Q}$ and $\mathbb{R}$ are ordered fields, if $x > y$ is defined to mean that $x - y$ is positive.

Proposition 1.8 *The following statements are true in every ordered field* F.
(a) *If* $x < y$ *and* $a \in F$ *then* $a + x < a + y$.
(b) *If* $x < y$ *and* $x' < y'$ *then* $x + x' < y + y'$.

Proof. (a) By assumption $(a+y)-(a+x)=y-x>0$. Hence $a+x<a+y$.
(b) By assumption and by (a) we have $x+x'<y+x'$ and $y+x'<y+y'$. Using transitivity, see Definition 1.1 (ii), we have $x+x'<y+y'$. ∎

Proposition 1.9 *The following statements are true in every ordered field.*
(a) *If* $x>0$ *then* $-x<0$, *and if* $x<0$ *then* $-x>0$.
(b) *If* $x>0$ *and* $y<z$ *then* $xy<xz$.
(c) *If* $x<0$ *and* $y<z$ *then* $xy>xz$.
(d) *If* $x\neq 0$ *then* $x^2>0$. *In particular,* $1>0$.
(e) *If* $0<x<y$ *then* $0<1/y<1/x$.

Proof. (a) If $x>0$ then $0=-x+x>-x+0=-x$, so that $-x<0$. If $x<0$ then $0=-x+x<-x+0=-x$ so that $-x>0$. This proves (a).
(b) Since $z>y$, we have $z-y>0$, hence $x(z-y)>0$ by axiom (O2), and therefore

$$xz=x(z-y)+xy \underset{Prp.\,1.8}{>} 0+xy=xy.$$

(c) By (a), (b) and Proposition 1.7 (c)

$$-[x(z-y)]=(-x)(z-y)>0,$$

so that $x(z-y)<0$, hence $xz<xy$.
(d) If $x>0$ axiom 1.4 (ii) gives $x^2>0$. If $x<0$ then $-x>0$, hence $(-x)^2>0$ But $x^2=(-x)^2$ by Proposition 1.7 (d). Since $1^2=1$, $1>0$.
(e) If $y>0$ and $v\leq 0$ then $yv\leq 0$. But $y\cdot(1/y)=1>0$. Hence $1/y>0$, likewise $1/x>0$. If we multiply $x<y$ by the the positive quantity $(1/x)(1/y)$, we obtain $1/y<1/x$. ∎

Remarks 1.3 (a) The finite field $\mathbb{F}_2=\{0,1\}$, see Remarks 1.2, is not an ordered field since $1+1=0$ which contradicts $1>0$.
(b) The field of complex numbers $\mathbb{C}$ (see below) is not an ordered field since $\mathrm{i}^2=-1$ contradicts Proposition 1.9 (a), (d).

1.1.4 Embedding of natural numbers into the real numbers

Let F be an ordered field. We want to recover the integers inside F. In order to distinguish 0 and 1 in F from the integers 0 and 1 we temporarily write 0_F and 1_F. For a positive integer $n\in\mathbb{N}$, $n\geq 2$ we define

$$n_F := 1_F+1_F+\cdots+1_F \quad (n \text{ times}).$$

Lemma 1.10 *We have* $n_F>0_F$ *for all* $n\in\mathbb{N}$.

Proof. We use induction over n. By Proposition 1.9 (d) the statement is true for $n = 1$. Suppose it is true for a fixed n, i. e. $n_F > 0_F$. Moreover $1_F > 0_F$. Using axiom (O2) we obtain $(n+1)1_F = n_F + 1_F > 0$. ∎

From Lemma 1.10 it follows that $m \neq n$ implies $n_F \neq m_F$. Indeed, let n be greater than m, say $n = m + k$ for some $k \in \mathbb{N}$, then $n_F = m_F + k_F$. Since $k_F > 0$ it follows from 1.8 (a) that $n_F > m_F$. In particular, $n_F \neq m_F$. Hence, the mapping

$$\mathbb{N} \to F, \qquad n \mapsto n_F$$

is a one-to-one correspondence (injective). In this way the positive integers are embedded into the real numbers. Addition and multiplication of natural numbers and of its embeddings are the same:

$$n_F + m_F = (n+m)_F, \qquad n_F\, m_F = (nm)_F.$$

From now on we identify a natural number with the associated real number. We write n for n_F.

Definition 1.5 (The Archimedean Axiom) An ordered field F is called *Archimedean* if for all $x, y \in F$ with $x > 0$ and $y > 0$ there exists $n \in \mathbb{N}$ such that $nx > y$.

An equivalent formulation is: The subset $\mathbb{N} \subset F$ of positive integers is not bounded above. Choose $x = 1$ in the above definition, then for any $y \in F$ there in an $n \in \mathbb{N}$ such that $n > y$; hence $\mathbb{N}$ is not bounded above.
Suppose $\mathbb{N}$ is not bounded and $x > 0, y > 0$ are given. Then y/x is not an upper bound for $\mathbb{N}$, that is there is some $n \in \mathbb{N}$ with $n > y/x$ or $nx > y$.

1.1.5 The completeness of $\mathbb{R}$

Using the axioms so far we are not yet able to prove the existence of irrational numbers. We need the completeness axiom.

Definition 1.6 (Order Completeness) An ordered set S is said to be *order complete* if for every non-empty bounded subset $E \subset S$ has a supremum $\sup E$ in S.

(C) Completeness Axiom

The real numbers are order complete, i. e. every bounded subset $E \subset \mathbb{R}$ has a supremum.

The set $\mathbb{Q}$ of rational numbers is not order complete since, for example, the bounded set $A = \{x \in \mathbb{Q}_+ \mid x^2 < 2\}$ has no supremum in $\mathbb{Q}$. Later we will define $\sqrt{2} := \sup A$. The existence of $\sqrt{2}$ in $\mathbb{R}$ is furnished by the completeness axiom (C).
Axiom (C) implies that every bounded subset $E \subset \mathbb{R}$ has an infimum. This is an easy consequence of Homework 1.4 (a).
We will see that an order complete field is always Archimedean.

Proposition 1.11 (a) $\mathbb{R}$ *is Archimedean.*
(b) *If* $x, y \in \mathbb{R}$, *and* $x < y$ *then there is a* $p \in \mathbb{Q}$ *with* $x < p < y$.

Part (b) may be stated by saying that $\mathbb{Q}$ *is dense in* $\mathbb{R}$.
Proof. (a) Let $x, y > 0$ be real numbers which do not fulfill the Archimedean property. That is, if $A := \{nx \mid n \in \mathbb{N}\}$, then y would be an upper bound of A. Then (C) furnishes that A has a supremum $\alpha = \sup A$. Since $x > 0$, $\alpha - x < \alpha$ and $\alpha - x$ is not an upper bound of A. Hence $\alpha - x < mx$ for some $m \in \mathbb{N}$. But then $\alpha < (m+1)x$, which is impossible, since α is an upper bound of A.
(b) See [Rud76, Theorem 29]. ∎

Remarks 1.4 (a) If $x, y \in \mathbb{Q}$ with $x < y$, then there exists $z \in \mathbb{R} \setminus \mathbb{Q}$ with $x < z < y$; chose $z = x + (y - x)/\sqrt{2}$.
Ex class: (b) We shall show that $\inf\left\{\frac{1}{n} \mid n \in \mathbb{N}\right\} = 0$. Since $n > 0$ for all $n \in \mathbb{N}$, $\frac{1}{n} > 0$ by Proposition 1.9 (e) and 0 is a lower bound. Suppose $\alpha > 0$. Since $\mathbb{R}$ is Archimedean, we find $m \in \mathbb{N}$ such that $1 < m\alpha$ or, equivalently $1/m < \alpha$. Hence, α is not a lower bound for E which proves the claim.
(c) Axiom (C) is equivalent to the Archimedean property together with the *topological* completeness ("Every Cauchy sequence in $\mathbb{R}$ is convergent," see Proposition 2.18).
(d) Axiom (C) is equivalent to the *axiom of nested intervals*, see Proposition 2.11 below:

> Let $I_n := [a_n, b_n]$ a sequence of closed nested intervals, that is $(I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots)$ such that for all $\varepsilon > 0$ there exists n_0 such that $0 \le b_n - a_n < \varepsilon$ for all $n \ge n_0$.
> Then there exists a unique real number $a \in \mathbb{R}$ which is a member of all intervals, i. e. $\{a\} = \bigcap_{n\in\mathbb{N}} I_n$.

1.1.6 The Absolute Value

For $x \in \mathbb{R}$ one defines

$$|\,x\,| := \begin{cases} x, & \text{if} \quad x \ge 0, \\ -x, & \text{if} \quad x < 0. \end{cases}$$

Lemma 1.12 *For $a, x, y \in \mathbb{R}$ we have*
(a) $|\,x\,| \ge 0$ *and* $|\,x\,| = 0$ *if and only if* $x = 0$. *Further* $|\,-x\,| = |\,x\,|$.
(b) $\pm x \le |\,x\,|$, $|\,x\,| = \max\{x, -x\}$, *and* $|\,x\,| \le a \iff (x \le a \quad \text{and} \quad -x \le a)$.
(c) $|\,xy\,| = |\,x\,|\,|\,y\,|$ *and* $\left|\frac{x}{y}\right| = \frac{|x|}{|y|}$ *if* $y \neq 0$.
(d) $|\,x + y\,| \le |\,x\,| + |\,y\,|$ *(triangle inequality).*
(e) $||\,x\,| - |\,y\,|| \le |\,x + y\,|$.

Proof. (a) By Proposition 1.9 (a), $x < 0$ implies $|\,x\,| = -x > 0$. Also, $x > 0$ implies $|\,x\,| > 0$. Putting both together we obtain, $x \neq 0$ implies $|\,x\,| > 0$ and thus $|\,x\,| = 0$ implies $x = 0$. Moreover $|\,0\,| = 0$. This shows the first part.
The statement $|\,x\,| = |\,-x\,|$ follows from (b) and $-(-x) = x$.
(b) Suppose first that $x \ge 0$. Then $x \ge 0 \ge -x$ and we have $\max\{x, -x\} = x = |\,x\,|$. If $x < 0$ then $-x > 0 > x$ and

$\max\{-x, x\} = -x = |\,x\,|$. This proves $\max\{x, -x\} = |\,x\,|$. Since the maximum is an upper bound, $|\,x\,| \geq x$ and $|\,x\,| \geq -x$. Suppose now a is an upper bound of $\{x, -x\}$. Then $|\,x\,| = \max\{x, -x\} \leq a$. On the other hand, $\max\{x, -x\} \leq a$ implies that a is an upper bound of $\{x, -x\}$ since $\max$ is.
One proves the first part of (c) by verifying the four cases (i) $x, y \geq 0$, (ii) $x \geq 0$, $y < 0$, (iii) $x < 0$, $y \geq 0$, and (iv) $x, y < 0$ separately. (i) is clear. In case (ii) we have by Proposition 1.9 (a) and (b) that $xy \leq 0$, and Proposition 1.7 (c)

$$|\,x\,|\,|\,y\,| = x(-y) = -(xy) = |\,xy\,|\,.$$

The cases (iii) and (iv) are similar. To the second part.
Since $x = \frac{x}{y} \cdot y$ we have by the first part of (c), $|\,x\,| = \left|\,\frac{x}{y}\,\right|\,|\,y\,|$. The claim follows by multiplication with $\frac{1}{|\,y\,|}$.
(d) By (b) we have $\pm x \leq |\,x\,|$ and $\pm y \leq |\,y\,|$. It follows from Proposition 1.8 (b) that

$$\pm(x + y) \leq |\,x\,| + |\,y\,|\,.$$

By the second part of (b) with $a = |\,x\,| + |\,y\,|$, we obtain $|\,x + y\,| \leq |\,x\,| + |\,y\,|$.
(e) Inserting $u := x + y$ and $v := -y$ into $|\,u + v\,| \leq |\,u\,| + |\,v\,|$ one obtains

$$|\,x\,| \leq |\,x + y\,| + |\,-y\,| = |\,x + y\,| + |\,y\,|\,.$$

Adding $-|\,y\,|$ on both sides one obtains $|\,x\,| - |\,y\,| \leq |\,x + y\,|$. Changing the role of x and y in the last inequality yields $-(|\,x\,| - |\,y\,|) \leq |\,x + y\,|$. The claim follows again by (b) with $a = |\,x + y\,|$. ∎

1.1.7 Supremum and Infimum revisited

The following equivalent definition for the supremum of sets of real numbers is often used in the sequel. Note that

$$\begin{aligned} & x \leq \beta \qquad \forall\, x \in M \\ \Longrightarrow\ & \sup M \leq \beta. \end{aligned}$$

Similarly, $\alpha \leq x$ for all $x \in M$ implies $\alpha \leq \inf M$.

Remarks 1.5 (a) Suppose that $E \subset \mathbb{R}$. Then α is the supremum of E if and only if

(1) α is an upper bound for E,
(2) For all $\varepsilon > 0$ there exists $x \in E$ with $\alpha - \varepsilon < x$.

Using the Archimedean axiom (2) can be replaced by

(2') For all $n \in \mathbb{N}$ there exists $x \in E$ such that $\alpha - \frac{1}{n} < x$.

(b) Let $M \subset \mathbb{R}$ and $N \subset \mathbb{R}$ nonempty subsets which are bounded above.
Then $M + N := \{m + n \mid m \in M, n \in N\}$ is bounded above and

$$\sup(M + N) = \sup M + \sup N.$$

(c) Let $M \subset \mathbb{R}_+$ and $N \subset \mathbb{R}_+$ nonempty subsets which are bounded above.
Then $MN := \{mn \mid m \in M, n \in N\}$ is bounded above and

$$\sup(MN) = \sup M \sup N.$$

1.1.8 Powers of real numbers

We shall prove the existence of nth roots of positive reals. We already know x^n, $n \in \mathbb{Z}$. It is recursively defined by $x^n := x^{n-1} \cdot x$, $x^1 := x$, $n \in \mathbb{N}$ and $x^n := \frac{1}{x^{-n}}$ for $n < 0$.

Proposition 1.13 (Bernoulli's inequality) *Let $x \geq -1$ and $n \in \mathbb{N}$. Then we have*

$$(1 + x)^n \geq 1 + nx.$$

Equality holds if and only if $x = 0$ or $n = 1$.

Proof. We use induction over n. In the cases $n = 1$ and $x = 0$ we have equality. The strict inequality (with an $>$ sign in place of the $\geq$ sign) holds for $n_0 = 2$ and $x \neq 0$ since $(1 + x)^2 = 1 + 2x + x^2 > 1 + 2x$. Suppose the strict inequality is true for some fixed $n \geq 2$ and $x \neq 0$. Since $1 + x \geq 0$ by Proposition 1.9 (b) multiplication of the induction assumption by this factor yields

$$(1 + x)^{n+1} \geq (1 + nx)(1 + x) = 1 + (n + 1)x + nx^2 > 1 + (n + 1)x.$$

This proves the strict assertion for $n + 1$. We have equality only if $n = 1$ or $x = 0$. ∎

Lemma 1.14 (a) *For $x, y \in \mathbb{R}$ with $x, y > 0$ and $n \in \mathbb{N}$ we have*

$$x < y \iff x^n < y^n.$$

(b) *For $x, y \in \mathbb{R}_+$ and $n \in \mathbb{N}$ we have*

$$nx^{n-1}(y - x) \leq y^n - x^n \leq ny^{n-1}(y - x). \tag{1.2}$$

We have equality if and only if $n = 1$ or $x = y$.

Proof. (a) Observe that

$$y^n - x^n = (y - x) \sum_{k=1}^{n} y^{n-k} x^{k-1} = c(y - x)$$

with $c := \sum_{k=1}^{n} y^{n-k}\, x^{k-1} > 0$ since $x, y > 0$. The claim follows.
(b) We have

$$\begin{aligned} y^n - x^n - nx^{n-1}(y-x) &= (y-x)\sum_{k=1}^{n}\left(y^{n-k}x^{k-1} - x^{n-1}\right) \\ &= (y-x)\sum_{k=1}^{n} x^{k-1}\left(y^{n-k} - x^{n-k}\right) \geq 0 \end{aligned}$$

since by (a) $y - x$ and $y^{n-k} - x^{n-k}$ have the same sign. The proof of the second inequality is quite analogous. ∎

Proposition 1.15 *For every real $x > 0$ and every positive integer $n \in \mathbb{N}$ there is one and only one $y > 0$ such that $y^n = x$.*

This number y is written $\sqrt[n]{x}$ or $x^{\frac{1}{n}}$, and it is called "the nth root of x".
Proof. The uniqueness is clear since by Lemma 1.14 (a) $0 < y_1 < y_2$ implies $0 < y_1^n < y_2^n$.
Set

$$E := \{t \in \mathbb{R}_+ \mid t^n < x\}.$$

Observe that $E \neq \varnothing$ since $0 \in E$. We show that E is bounded above. By Bernoulli's inequality and since $0 < x < nx$ we have

$$\begin{aligned} t \in E &\Leftrightarrow t^n < x < 1 + nx < (1+x)^n \\ &\underset{\text{Lemma 1.14}}{\Longrightarrow} t < 1 + x \end{aligned}$$

Hence, $1 + x$ is an upper bound for E. By the order completeness of $\mathbb{R}$ there exists $y \in \mathbb{R}$ such that $y = \sup E$. We have to show that $y^n = x$. For, we will show that each of the inequalities $y^n > x$ and $y^n < x$ leads to a contradiction.
Assume $y^n < x$ and consider $(y+h)^n$ with "small" h ($0 < h < 1$). Lemma 1.14 (b) implies

$$0 \leq (y+h)^n - y^n \leq n\,(y+h)^{n-1}(y+h-y) < h\,n(y+1)^{n-1}.$$

Choosing h small enough that $h\,n(y+1)^{n-1} < x - y^n$ we may continue

$$(y+h)^n - y^n \leq x - y^n.$$

Consequently, $(y+h)^n < x$ and therefore $y + h \in E$. Since $y + h > y$, this contradicts the fact that y is an upper bound of E.
Assume $y^n > x$ and consider $(y-h)^n$ with "small" h ($0 < h < 1$). Again by Lemma 1.14 (b) we have

$$0 \leq y^n - (y-h)^n \leq n\,y^{n-1}(y-y+h) < h\,ny^{n-1}.$$

Choosing h small enough that $h\,ny^{n-1} < y^n - x$ we may continue

$$y^n - (y-h)^n \leq y^n - x.$$

Consequently, $x < (y-h)^n$ and therefore $t^n < x < (y-h)^n$ for all t in E. Hence $y-h$ is an upper bound for E smaller than y. This contradicts the fact that y is the *least* upper bound. Hence $y^n = x$, and the proof is complete. ∎

Remarks 1.6 (a) If a and b are positive real numbers and $n \in \mathbb{N}$ then $(ab)^{1/n} = a^{1/n}\, b^{1/n}$.
Proof. Put $\alpha = a^{1/n}$ and $\beta = b^{1/n}$. Then $ab = \alpha^n \beta^n = (\alpha\beta)^n$, since multiplication is commutative. The uniqueness assertion of Proposition 1.15 shows therefore that $(ab)^{1/n} = \alpha\beta = a^{1/n}\, b^{1/n}$. ∎

(b) Fix $b > 0$. If $m, n, p, q \in \mathbb{Z}$ and $n > 0$, $q > 0$, and $r = m/n = p/q$. Then we have

$$(b^m)^{1/n} = (b^p)^{1/q}. \tag{1.3}$$

Hence it makes sense to define $b^r = (b^m)^{1/n}$.
(c) Fix $b > 1$. If $x \in \mathbb{R}$ define

$$b^x = \sup\{b^p \mid p \in \mathbb{Q},\, p < x\}. \tag{1.4}$$

For $0 < b < 1$ set

$$b^x = \frac{1}{\left(\frac{1}{b}\right)^x}.$$

Without proof we give the familiar laws for powers and exponentials. Later we will redefine the power b^x with real exponent. Then we are able to give easier proofs.
(d) If $a, b > 0$ and $x, y \in \mathbb{R}$, then
(i) $b^{x+y} = b^x b^y$, $\quad b^{x-y} = \frac{b^x}{b^y}$, $\quad$ (ii) $b^{xy} = (b^x)^y$, $\quad$ (iii) $(ab)^x = a^x b^x$.

1.1.9 Logarithms

Fix $b > 1$, $y > 0$. Similarly as in the preceding subsection, one can prove the existence of a unique real x such that $b^x = y$. This number x is called the *logarithm of y to the base b*, and we write $x = \log_b y$. Knowing existence and uniqueness of the logarithm, it is not difficult to prove the following properties.

Lemma 1.16 *For any $a > 0$, $a \neq 1$ we have*
(a) $\log_a(bc) = \log_a b + \log_a c$ *if* $b, c > 0$*;*
(b) $\log_a(b^c) = c \log_a b$, *if* $b > 0$*;*
(c) $\log_a b = \dfrac{\log_d b}{\log_d a}$ *if* $b, d > 0$ *and* $d \neq 1$.

Later we will give an alternative definition of the logarithm function.

Review of Trigonometric Functions

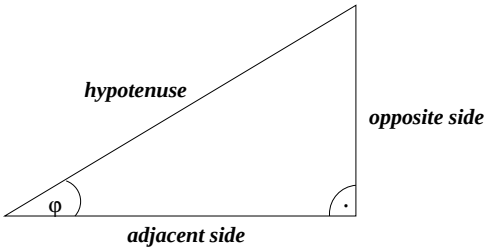
(a) Degrees and Radians

The following table gives some important angles in degrees and radians. The precise definition of π is given below. For a moment it is just an abbreviation to measure angles. Transformation of angles α_{deg} measured in degrees into angles measured in radians goes by $\alpha_{\text{rad}} = \alpha_{\text{deg}} \frac{2\pi}{360^\circ}$.

Degrees	0°	30°	45°	60°	90°	120°	135°	150°	180°	270°	360°
Radians	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π

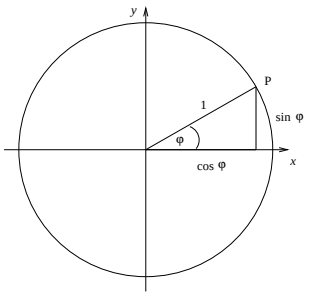
(b) Sine and Cosine

The sine, cosine, and tangent functions are defined in terms of ratios of sides of a right triangle:

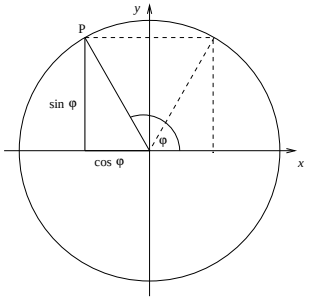


$$\cos\varphi = \frac{\text{length of the side adjacent to } \varphi}{\text{length of the hypotenuse}},$$
$$\sin\varphi = \frac{\text{length of the side opposite to } \varphi}{\text{length of the hypotenuse}},$$
$$\tan\varphi = \frac{\text{length of the side opposite to } \varphi}{\text{length of the side adjacent to } \theta}.$$

Let φ be any angle between 0° and 360°. Further let P be the point on the unit circle (with center in $(0,0)$ and radius 1) such that the ray from P to the origin $(0,0)$ and the positive x-axis make an angle φ. Then $\cos\varphi$ and $\sin\varphi$ are defined to be the x-coordinate and the y-coordinate of the point P, respectively.

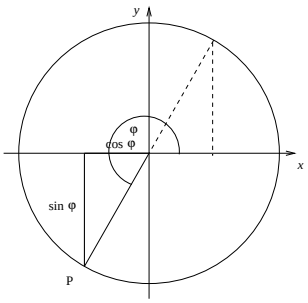


If the angle φ is between 0° and 90° this new definition coincides with the definition using the right triangle since the hypotenuse which is a radius of the unit circle has now length 1.



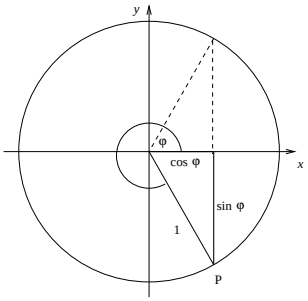
If $90^\circ < \varphi < 180^\circ$ we find

$$\cos\varphi = -\cos(180^\circ - \varphi) < 0,$$
$$\sin\varphi = \sin(180^\circ - \varphi) > 0.$$



If $180° < \varphi < 270°$ we find

$$\cos\varphi = -\cos(\varphi - 180°) < 0,$$
$$\sin\varphi = -\sin(\varphi - 180°) < 0.$$



If $270° < \varphi < 360°$ we find

$$\cos\varphi = \cos(360° - \varphi) > 0,$$
$$\sin\varphi = -\sin(360° - \varphi) < 0.$$

For angles greater than $360°$ or less than $0°$ define

$$\cos\varphi = \cos(\varphi + k{\cdot}360°), \quad \sin\varphi = \sin(\varphi + k{\cdot}360°),$$

where $k \in \mathbb{Z}$ is chosen such that $0° \le \varphi + k\,360° < 360°$. Thinking of φ to be given in radians, cosine and sine are functions defined for all real φ taking values in the closed interval $[-1, 1]$. If $\varphi \neq \frac{\pi}{2} + k\pi$, $k \in \mathbb{Z}$ then $\cos\varphi \neq 0$ and we define

$$\tan\varphi := \frac{\sin\varphi}{\cos\varphi}.$$

If $\varphi \neq k\pi$, $k \in \mathbb{Z}$ then $\sin\varphi \neq 0$ and we define

$$\cot\varphi := \frac{\cos\varphi}{\sin\varphi}.$$

In this way we have defined cosine, sine, tangent, and cotangent for arbitrary angles.

(c) Special Values

x in degrees	$0°$	$30°$	$45°$	$60°$	$90°$	$120°$	$135°$	$150°$	$180°$	$270°$	$360°$
x in radians	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{3\pi}{4}$	$\frac{5\pi}{6}$	π	$\frac{3\pi}{2}$	2π
$\sin x$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	0	-1	0
$\cos x$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{3}}{2}$	-1	0	1
$\tan x$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	/	$-\sqrt{3}$	-1	$-\frac{\sqrt{3}}{3}$	0	/	0

Recall the addition formulas for cosine and sine and the trigonometric pythagoras.

$$\begin{aligned} \cos(x+y) &= \cos x\cos y - \sin x\sin y, \\ \sin(x+y) &= \sin x\cos y + \cos x\sin y. \end{aligned} \tag{1.5}$$

$$\sin^2 x + \cos^2 x = 1. \tag{1.6}$$

1.2 Complex numbers

Some algebraic equations do not have solutions in the real number system. For instance the quadratic equation $x^2 - 4x + 8 = 0$ gives 'formally'

$$x_1 = 2 + \sqrt{-4} \quad \text{and} \quad x_2 = 2 - \sqrt{-4}.$$

We will see that one can work with this notation.

Definition 1.7 A *complex number* is an ordered pair (a, b) of real numbers. "Ordered" means that $(a, b) \neq (b, a)$ if $a \neq b$. Two complex numbers $x = (a, b)$ and $y = (c, d)$ are said to be equal if and only if $a = c$ and $b = d$. We define

$$\begin{aligned} x + y &:= (a + c, b + d), \\ xy &:= (ac - bd, ad + bc). \end{aligned}$$

Theorem 1.17 *These definitions turn the set of all complex numbers into a field, with* $(0, 0)$ *and* $(1, 0)$ *in the role of* 0 *and* 1.

Proof. We simply verify the field axioms as listed in Definition 1.3. Of course, we use the field structure of $\mathbb{R}$.
Let $x = (a, b)$, $y = (c, d)$, and $z = (e, f)$. (A1) is clear.
(A2) $x + y = (a + c, b + d) = (c + a, d + b) = y + x$.
(A3) $(x+y)+z = (a+c, b+d)+(e, f) = (a+c+e, b+d+f) = (a, b)+(c+e, d+f) = x+(y+z)$.
(A4) $x + 0 = (a, b) + (0, 0) = (a, b) = x$.
(A5) Put $-x := (-a, -b)$. Then $x + (-x) = (a, b) + (-a, -b) = (0, 0) = 0$.
(M1) is clear.
(M2) $xy = (ac - bd, ad + bc) = (ca - db, da + cb) = yx$.
(M3) $(xy)z = (ac - bd, ad + bc)(e, f) = (ace - bde - adf - bcf, acf - bdf + ade + bce) = (a, b)(ce - df, cf + de) = x(yz)$.
(M4) $x \cdot 1 = (a, b)(1, 0) = (a, b) = x$.
(M5) If $x \neq 0$ then $(a, b) \neq (0, 0)$, which means that at least one of the real numbers a, b is different from 0. Hence $a^2 + b^2 > 0$ and we can define

$$\frac{1}{x} := \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right).$$

Then

$$x \cdot \frac{1}{x} = (a, b) \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right) = (1, 0) = 1.$$

(D)

$$\begin{aligned} x(y + z) &= (a, b)(c + e, d + f) = (ac + ae - bd - bf, ad + af + bc + be) \\ &= (ac - bd, ad + bc) + (ae - bf, af + be) \\ &= xy + yz. \end{aligned}$$

■

Remark 1.7 For any two real numbers a and b we have $(a,0)+(b,0)=(a+b,0)$ and $(a,0)(b,0)=(ab,0)$. This shows that the complex numbers $(a,0)$ have the same arithmetic properties as the corresponding real numbers a. We can therefore identify $(a,0)$ with a. This gives us the real field as a subfield of the complex field.
Note that we have defined the complex numbers without any reference to the mysterious square root of -1. We now show that the notation (a,b) is equivalent to the more customary $a+b\mathrm{i}$.

Definition 1.8 $\mathrm{i} := (0,1)$.

Lemma 1.18 (a) $\mathrm{i}^2=-1$. (b) *If* $a,b\in\mathbb{R}$ *then* $(a,b)=a+b\mathrm{i}$.

Proof. (a) $\mathrm{i}^2=(0,1)(0,1)=(-1,0)=-1$.
(b) $a+b\mathrm{i}=(a,0)+(b,0)(0,1)=(a,0)+(0,b)=(a,b)$. ∎

Definition 1.9 If a,b are real and $z=a+b\mathrm{i}$, then the complex number $\overline{z}:=a-b\mathrm{i}$ is called the *conjugate* of z. The numbers a and b are the *real part* and the *imaginary part* of z, respectively. We shall write $a=\operatorname{Re} z$ and $b=\operatorname{Im} z$.

Proposition 1.19 *If* z *and* w *are complex, then*
(a) $\overline{z+w}=\overline{z}+\overline{w}$,
(b) $\overline{zw}=\overline{z}\cdot\overline{w}$,
(c) $z+\overline{z}=2\operatorname{Re} z, \qquad z-\overline{z}=2\mathrm{i}\operatorname{Im} z$,
(d) $z\overline{z}$ *is positive real except when* $z=0$.

Proof. (a), (b), and (c) are quite trivial. To prove (d) write $z=a+b\mathrm{i}$ and note that $z\overline{z}=a^2+b^2$. ∎

Definition 1.10 If z is complex number, its *absolute value* $|z|$ is the (nonnegative) root of $z\overline{z}$; that is $|z|:=\sqrt{z\overline{z}}$.

The existence (and uniqueness) of $|x|$ follows from Proposition 1.19 (d). Note that when x is real, then $x=\overline{x}$, hence $|x|=\sqrt{x^2}$. Thus $|x|=x$ if $x>0$ and $|x|=-x$ if $x<0$. We have recovered the definition of the absolute value for real numbers, see Subsection 1.1.6.

Proposition 1.20 *Let* z *and* w *be complex numbers . Then*
(a) $|z|>0$ *unless* $z=0$,
(b) $|\overline{z}|=|z|$,
(c) $|zw|=|z|\,|w|$,
(d) $|\operatorname{Re} z|\le|z|$,
(e) $|z+w|\le|z|+|w|$.

Proof. (a) and (b) are trivial. Put $z=a+b\mathrm{i}$ and $w=c+d\mathrm{i}$, with a,b,c,d real. Then

$$|zw|^2=(ac-bd)^2+(ad+bc)^2=(a^2+b^2)(c^2+d^2)=|z|^2\,|w|^2$$

or $|\,zw\,|^2 = (|\,z\,|\,|\,w\,|)^2$. Now (c) follows from the uniqueness assertion for roots.
To prove (d), note that $a^2 \le a^2 + b^2$, hence

$$|\,a\,| = \sqrt{a^2} \le \sqrt{a^2+b^2} = |\,z\,|\,.$$

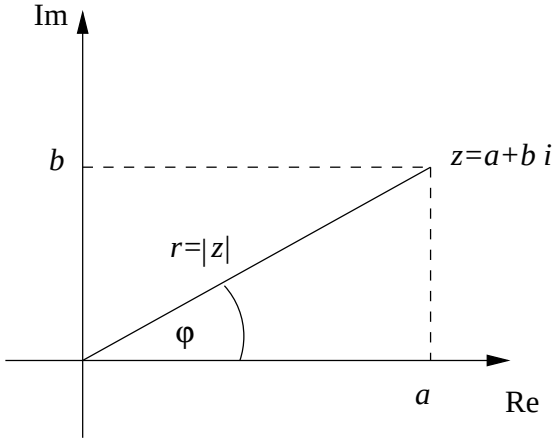
To prove (e), note that $\overline{z}\,w$ is the conjugate of $z\,\overline{w}$, so that $z\,\overline{w} + \overline{z}w = 2\,\mathrm{Re}\,(z\,\overline{w})$. Hence

$$\begin{aligned} |\,z+w\,|^2 &= (z+w)(\overline{z}+\overline{w}) = z\,\overline{z} + z\,\overline{w} + \overline{z}\,w + w\,\overline{w} \\ &= |\,z\,|^2 + 2\,\mathrm{Re}\,(z\,\overline{w}) + |\,w\,|^2 \\ &\le |\,z\,|^2 + 2\,|\,z\,|\,|\,w\,| + |\,w\,|^2 = (|\,z\,| + |\,w\,|)^2. \end{aligned}$$

Now (e) follows by taking square roots. ∎

1.2.1 The Complex Plane and the Polar form

There is a bijective correspondence between complex numbers and the points of a plane.



By the Pythagorean theorem it is clear that $|\,z\,| = \sqrt{a^2+b^2}$ is exactly the distance of z from the origin 0. The angle φ between the positive real axis and the half-line $0z$ is called the *argument* of z and is denoted by $\varphi = \arg z$. If $z \neq 0$, the argument φ is uniquely determined up to integer multiples of 2π

Elementary trigonometry gives

$$\sin\varphi = \frac{b}{|\,z\,|}, \quad \cos\varphi = \frac{a}{|\,z\,|}.$$

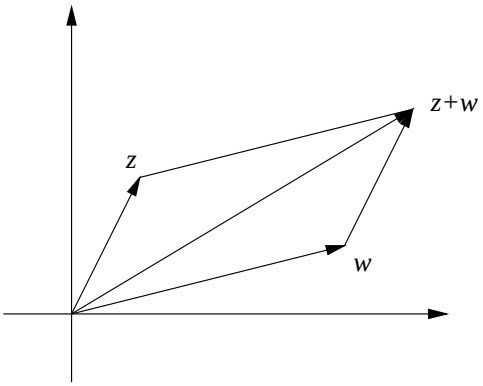
This gives with $r = |\,z\,|$, $a = r\cos\varphi$ and $b = r\sin\varphi$. Inserting these into the rectangular form of z yields

$$z = r(\cos\varphi + \mathrm{i}\sin\varphi), \tag{1.7}$$

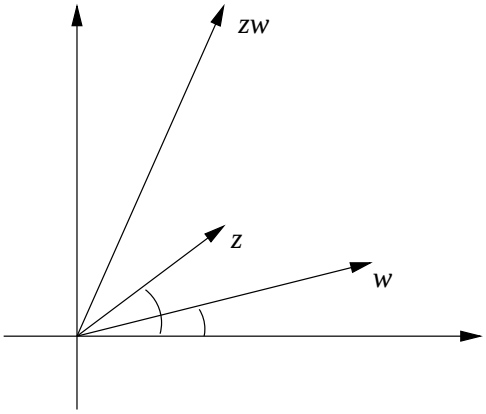
which is called the *polar form* of the complex number z.

Example 1.5 a) $z = 1 + \mathrm{i}$. Then $|\,z\,| = \sqrt{2}$ and $\sin\varphi = 1/\sqrt{2} = \cos\varphi$. This implies $\varphi = \pi/4$. Hence, the polar form of z is $1 + i = \sqrt{2}(\cos\pi/4 + \mathrm{i}\sin\pi/4)$.
b) $z = -\mathrm{i}$. We have $|\,-\mathrm{i}\,| = 1$ and $\sin\varphi = -1$, $\cos\varphi = 0$. Hence $\varphi = 3\pi/2$ and $-\mathrm{i} = 1(\cos 3\pi/2 + \mathrm{i}\sin 3\pi/2)$.
c) Computing the rectangular form of z from the polar form is easier.

$$z = 32(\cos 7\pi/6 + \mathrm{i}\sin 7\pi/6) = 32(-\sqrt{3}/2 - \mathrm{i}/2) = -16\sqrt{3} - 16\mathrm{i}.$$



The addition of complex numbers corresponds to the addition of vectors in the plane. The geometric meaning of the inequality $|\,z+w\,| \leq |\,z\,|+|\,w\,|$ is: the sum of two edges of a triangle is bigger than the third edge.



Multiplying complex numbers $z = r(\cos\varphi + \mathrm{i}\sin\varphi)$ and $w = s(\cos\psi + \mathrm{i}\sin\psi)$ in the polar form gives

$$\begin{aligned} zw &= rs(\cos\varphi + \mathrm{i}\sin\varphi)(\cos\psi + \mathrm{i}\sin\psi) \\ &= rs(\cos\varphi\cos\psi - \sin\varphi\sin\psi) + \\ &\qquad \mathrm{i}(\cos\varphi\sin\psi + \sin\varphi\cos\psi) \\ zw &= rs(\cos(\varphi+\psi) + \mathrm{i}\sin(\varphi+\psi)), \end{aligned} \tag{1.8}$$

where we made use of the addition laws for $\sin$ and $\cos$ in the last equation.

Hence, the product of complex numbers is formed by taking the product of their absolute values and the sum of their arguments.

The geometric meaning of multiplication by w is a similarity transformation of $\mathbb{C}$. More precisely, we have a rotation around 0 by the angle ψ and then a dilatation with factor s and center 0.

Similarly, if $w \neq 0$ we have

$$\frac{z}{w} = \frac{r}{s}\left(\cos(\varphi-\psi) + \mathrm{i}\sin(\varphi-\psi)\right). \tag{1.9}$$

Proposition 1.21 (De Moivre's formula) *Let $z = r(\cos\varphi + \mathrm{i}\sin\varphi)$ be a complex number with absolute value r and argument φ. Then for all $n \in \mathbb{Z}$ one has*

$$z^n = r^n(\cos(n\varphi) + \mathrm{i}\sin(n\varphi)). \tag{1.10}$$

Proof. (a) First let $n > 0$. We use induction over n to prove De Moivre's formula. For $n = 1$ there is nothing to prove. Suppose (1.10) is true for some fixed n. We will show that the assertion is true for $n+1$. Using induction hypothesis and (1.8) we find

$$z^{n+1} = z^n \cdot z = r^n(\cos(n\varphi)+\mathrm{i}\sin(n\varphi))r(\cos\varphi+\mathrm{i}\sin\varphi) = r^{n+1}(\cos(n\varphi+\varphi)+\mathrm{i}\sin(n\varphi+\varphi)).$$

This proves the induction assertion.

(b) If $n < 0$, then $z^n = 1/(z^{-n})$. Since $1 = 1(\cos 0 + \mathrm{i}\sin 0)$, (1.9) and the result of (a) gives

$$z^n = \frac{1}{z^{-n}} = \frac{1}{r^{-n}}\left(\cos(0-(-n)\varphi) + \mathrm{i}\sin(0-(-n)\varphi)\right) = r^n(\cos(n\varphi) + \mathrm{i}\sin(n\varphi)).$$

This completes the proof. ∎

Example 1.6 Compute the polar form of $z = \sqrt{3} - 3\mathrm{i}$ and compute z^{15}.
We have $|\,z\,| = \sqrt{3+9} = 2\sqrt{3}$, $\cos\varphi = 1/2$, and $\sin\varphi = -\sqrt{3}/2$. Therefore, $\varphi = -\pi/3$ and $z = 2\sqrt{3}(\cos(-\pi/3) + \sin(-\pi/3))$. By the De Moivre's formula we have

$$z^{15} = (2\sqrt{3})^{15}\left(\cos\left(-15\frac{\pi}{3}\right) + \mathrm{i}\sin\left(-15\frac{\pi}{3}\right)\right) = 2^{15}3^7\sqrt{3}(\cos(-5\pi) + \mathrm{i}\sin(-5\pi))$$
$$z^{15} = -2^{15}3^7\sqrt{3}.$$

1.2.2 Roots of Complex Numbers

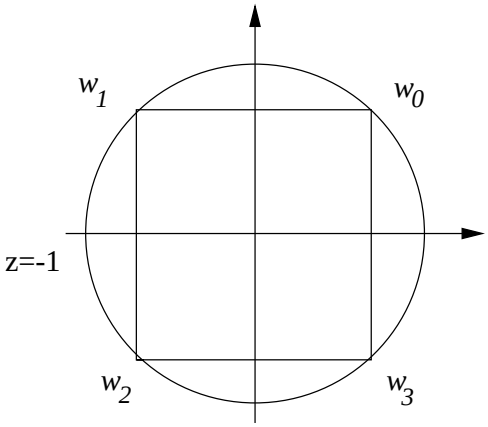
Let $z \in \mathbb{C}$ and $n \in \mathbb{N}$. A complex number w is called an *nth root of z* if $w^n = z$. In contrast to the real case, roots of complex numbers *are not unique*. We will see that there are exactly n different nth roots of z for every $z \neq 0$.
Let $z = r(\cos\varphi + \mathrm{i}\sin\varphi)$ and $w = s(\cos\psi + \mathrm{i}\sin\psi)$ an nth root of z. De Moivre's formula gives $w^n = s^n(\cos n\psi + \mathrm{i}\sin n\psi)$. If we compare w^n and z we get $s^n = r$ or $s = \sqrt[n]{r} \geq 0$. Moreover $n\psi = \varphi + 2k\pi$, $k \in \mathbb{Z}$ or $\psi = \dfrac{\varphi}{n} + \dfrac{2k\pi}{n}$, $k \in \mathbb{Z}$. For $k = 0, 1, \ldots, n-1$ we obtain different values $\psi_0, \psi_1, \ldots, \psi_{n-1}$ modulo 2π. We summarize.

Lemma 1.22 *Let $n \in \mathbb{N}$ and $z = r(\cos\varphi + \mathrm{i}\sin\varphi) \neq 0$ a complex number. Then the complex numbers*

$$w_k = \sqrt[n]{r}\left(\cos\frac{\varphi + 2k\pi}{n} + \mathrm{i}\sin\frac{\varphi + 2k\pi}{n}\right), \quad k = 0, 1, \ldots, n-1$$

are the n different nth roots of z.

Example 1.7 Compute the 4th roots of $z = -1$.



$|\,z\,| = 1 \Longrightarrow |\,w\,| = \sqrt[4]{1} = 1$, $\arg z = \varphi = 180°$. Hence

$$\psi_0 = \frac{\varphi}{4},$$
$$\psi_1 = \frac{\varphi}{4} + \frac{1 \cdot 360°}{4} = 135°,$$
$$\psi_2 = \frac{\varphi}{4} + \frac{2 \cdot 360°}{4} = 225°,$$
$$\psi_3 = \frac{\varphi}{4} + \frac{3 \cdot 360°}{4} = 315°.$$

We obtain

$$w_0 = \cos 45° + \mathrm{i}\sin 45° \quad = \frac{1}{2}\sqrt{2} + \mathrm{i}\frac{1}{2}\sqrt{2}$$
$$w_1 = \cos 135° + \mathrm{i}\sin 135° = -\frac{1}{2}\sqrt{2} + \mathrm{i}\frac{1}{2}\sqrt{2},$$
$$w_2 = \cos 225° + \mathrm{i}\sin 225° = -\frac{1}{2}\sqrt{2} - \mathrm{i}\frac{1}{2}\sqrt{2},$$
$$w_3 = \cos 315° + \mathrm{i}\sin 315° = \frac{1}{2}\sqrt{2} - \mathrm{i}\frac{1}{2}\sqrt{2}.$$

Geometric interpretation of the nth roots. The nth roots of $z \neq 0$ form a regular n-gon in the complex plane with center 0. The vertices lie on a circle with center 0 and radius $\sqrt[n]{|\,z\,|}$.

1.3 Inequalities

1.3.1 Monotony of the Power and Exponential Functions

Lemma 1.23 (a) *For $a, b > 0$ and $r \in \mathbb{Q}$ we have*

$$a < b \iff a^r < b^r \quad \text{if } r > 0,$$
$$a < b \iff a^r > b^r \quad \text{if } r < 0.$$

(b) *For $a > 0$ and $r, s \in \mathbb{Q}$ we have*

$$r < s \iff a^r < a^s \quad \text{if } a > 1,$$
$$r < s \iff a^r > a^s \quad \text{if } a < 1.$$

Proof. Suppose that $r > 0$, $r = m/n$ with integers $m, n \in \mathbb{Z}$, $n > 0$. Using Lemma 1.14 (a) twice we get

$$a < b \iff a^m < b^m \iff (a^m)^{\frac{1}{n}} < (b^m)^{\frac{1}{n}},$$

which proves the first claim. The second part $r < 0$ can be obtained by setting $-r$ in place of r in the first part and using Proposition 1.9 (e).
(b) Suppose that $s > r$. Put $x = s - r$, then $x \in \mathbb{Q}$ and $x > 0$. By (a), $1 < a$ implies $1 = 1^x < a^x$. Hence $1 < a^{s-r} = a^s/a^r$ (here we used Remark 1.6 (d)), and therefore $a^r < a^s$. Changing the roles of r and s shows that $s < r$ implies $a^s < a^r$ such that the converse direction is also true.
The proof for $a < 1$ is similar. ∎

1.3.2 The Arithmetic-Geometric mean inequality

Proposition 1.24 *Let $n \in \mathbb{N}$ and $x_1, \ldots, x_n$ be in $\mathbb{R}_+$. Then*

$$\frac{x_1 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n}. \tag{1.11}$$

We have equality if and only if $x_1 = x_2 = \cdots = x_n$.

Proof. We use forward-backward induction over n. First we show (1.11) is true for all n which are powers of 2. Then we prove that if (1.11) is true for some $n+1$, then it is true for n. Hence, it is true for all positive integers.
The inequality is true for $n = 1$. Let $a, b \geq 0$ then $(\sqrt{a} - \sqrt{b})^2 \geq 0$ implies $a + b \geq 2\sqrt{ab}$ and the inequality is true in case $n = 2$. Equality holds if and only if $a = b$. Suppose it is true for some fixed $k \in \mathbb{N}$; we will show that it is true for $2k$. Let $x_1, \ldots, x_k, y_1, \ldots, y_k \in \mathbb{R}_+$. Using induction assumption and the inequality in case $n = 2$, we have

$$\frac{1}{2k}\left(\sum_{i=1}^{k} x_i + \sum_{i=1}^{k} y_i\right) \geq \frac{1}{2}\left(\frac{1}{k}\sum_{i=1}^{k} x_i + \frac{1}{k}\sum_{i=1}^{k} y_i\right) \geq \frac{1}{2}\left(\left(\prod_{i=1}^{k} x_i\right)^{1/k} + \left(\prod_{i=1}^{k} y_i\right)^{1/k}\right)$$
$$\geq \left(\prod_{i=1}^{k} x_i \prod_{i=1}^{k} y_i\right)^{\frac{1}{2k}}.$$

This completes the 'forward' part. Assume now (1.11) is true for $n+1$. We will show it for n. Let $x_1, \ldots, x_n \in \mathbb{R}_+$ and set $A := (\sum_{i=1}^n x_i)/n$. By induction assumption we have

$$\frac{1}{n+1}\left(x_1 + \cdots + x_n + A\right) \geq \left(\prod_{i=1}^n x_i A\right)^{\frac{1}{n+1}} \iff \frac{1}{n+1}(nA + A) \geq \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n+1}} A^{\frac{1}{n+1}}$$

$$A \geq \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n+1}} A^{\frac{1}{n+1}} \iff A^{\frac{n}{n+1}} \geq \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n+1}} \iff A \geq \left(\prod_{i=1}^n x_i\right)^{1/n}.$$

It is trivial that in case $x_1 = x_2 = \cdots = x_n$ we have equality. Suppose that equality holds in a case where at least two of the x_i are different, say $x_1 < x_2$. Consider the inequality with the new set of values $x_1' := x_2' := (x_1 + x_2)/2$, and $x_i' = x_i$ for $i \geq 3$. Then

$$\left(\prod_{k=1}^n x_k\right)^{1/n} = \frac{1}{n}\sum_{k=1}^n x_k = \frac{1}{n}\sum_{k=1}^n x_k' \geq \left(\prod_{k=1}^n x_k'\right)^{1/n}.$$

$$x_1 x_2 \geq x_1' x_2' = \left(\frac{x_1 + x_2}{2}\right)^2 \iff 4x_1x_2 \geq x_1^2 + 2x_1x_2 + x_2^2 \iff 0 \geq (x_1 - x_2)^2.$$

This contradicts the choice $x_1 < x_2$. Hence, $x_1 = x_2 = \cdots = x_n$ is the only case where equality holds. This completes the proof. ∎

1.3.3 The Cauchy–Schwarz Inequality

Proposition 1.25 (Cauchy–Schwarz inequality) *Suppose that $x_1, \ldots, x_n, y_1, \ldots, y_n$ are real numbers. Then we have*

$$\left(\sum_{k=1}^n x_k y_k\right)^2 \leq \sum_{k=1}^n x_k^2 \cdot \sum_{k=1}^n y_k^2. \tag{1.12}$$

Equality holds if and only if there exists $t \in \mathbb{R}$ such that $y_k = t\,x_k$ for $k = 1, \ldots, n$ that is, the vector $y = (y_1, \ldots, y_n)$ is a scalar multiple of the vector $x = (x_1, \ldots, x_n)$.

Proof. Consider the quadratic function $f(t) = at^2 - 2bt + c$ where

$$a = \sum_{k=1}^n x_k^2, \quad b = \sum_{k=1}^n x_k y_k, \quad c = \sum_{k=1}^n y_k^2.$$

Then

$$\begin{aligned} f(t) &= \sum_{k=1}^n x_k^2 t^2 - \sum_{k=1}^n 2x_k y_k t + \sum_{k=1}^n y_k^2 \\ &= \sum_{k=1}^n \left(x_k^2 t^2 - 2x_k y_k t + y_k^2\right) = \sum_{k=1}^n (x_k t - y_k)^2 \geq 0. \end{aligned}$$

Equality holds if and only if there is a $t \in \mathbb{R}$ with $y_k = tx_k$ for all k. Suppose now, there is no such $t \in \mathbb{R}$. That is

$$f(t) > 0, \quad \text{for all } t \in \mathbb{R}.$$

In other words, the polynomial $f(t) = at^2 - 2bt + c$ has no real zeros, $t_{1,2} = \frac{1}{a}\left(b \pm \sqrt{b^2 - ac}\right)$. That is, the discriminant $D = b^2 - ac$ is negative (only complex roots); hence $b^2 < ac$:

$$\left(\sum_{k=1}^{n} x_k y_k\right)^2 < \sum_{k=1}^{n} x_k^2 \cdot \sum_{k=1}^{n} y_k^2.$$

this proves the claim. ∎

Corollary 1.26 (The Complex Cauchy–Schwarz inequality) *If $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are complex numbers, then*

$$\left|\sum_{k=1}^{n} x_k \overline{y_k}\right|^2 \leq \sum_{k=1}^{n} |\, x_k \,|^2 \sum_{k=1}^{n} |\, y_k \,|^2 . \tag{1.13}$$

Equality holds if and only if there exists a $\lambda \in \mathbb{C}$ such that $y = \lambda x$, where $y = (y_1, \ldots, y_n) \in \mathbb{C}^n$, $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$.

Proof. Using the generalized triangle inequality $|\, z_1 + \cdots + z_n \,| \leq |\, z_1 \,| + \cdots + |\, z_n \,|$ and the real Cauchy–Schwarz inequality we obtain

$$\left|\sum_{k=1}^{n} x_k \,\overline{y_k}\right|^2 \leq \left(\sum_{k=1}^{n} |\, x_k \, y_k \,|\right)^2 = \left(\sum_{k=1}^{n} |\, x_k \,|\; |\, y_k \,|\right)^2 \leq \sum_{k=1}^{n} |\, x_k \,|^2 \cdot \sum_{k=1}^{n} |\, y_k \,|^2 .$$

This proves the inequality.
The right "less equal" is an equality if there is a $t \in \mathbb{R}$ such that $|\, y \,| = t\,|\, x \,|$. In the first "less equal" sign we have equality if and only if all $z_k = x_k \overline{y}_k$ have the same argument; that is $\arg y_k = \arg x_k + \varphi$. Putting both together yields $y = \lambda x$ with $\lambda = t(\cos\varphi + \mathrm{i}\sin\varphi)$. ∎

1.4 Appendix A

In this appendix we collect some assitional facts which were not covered by the lecture.
We now show that the equation

$$x^2 = 2 \tag{1.14}$$

is not satisfied by any rational number x.
Suppose to the contrary that there were such an x, we could write $x = m/n$ with integers m and n, $n \neq 0$ that are not both even. Then (1.14) implies

$$m^2 = 2n^2. \tag{1.15}$$

This shows that m^2 is even and hence m is even. Therefore m^2 is divisible by 4. It follows that the right hand side of (1.15) is divisible by 4, so that n^2 is even, which implies that n is even. But this contradicts our choice of m and n. Hence (1.14) is impossible for rational x.
We shall show that A contains no largest element and B contains no smallest. That is for every $p \in A$ we can find a rational $q \in A$ with $p < q$ and for every $p \in B$ we can find a rational $q \in B$ such that $q < p$.
Suppose that p is in A. We associate with $p > 0$ the rational number

$$q = p + \frac{2 - p^2}{p + 2} = \frac{2p + 2}{p + 2}. \tag{1.16}$$

Then

$$q^2 - 2 = \frac{4p^2 + 8p + 4 - 2p^2 - 8p - 8}{(p + 2)^2} = \frac{2(p^2 - 2)}{(p + 2)^2}. \tag{1.17}$$

If p is in A then $2 - p^2 > 0$, (1.16) shows that $q > p$, and (1.17) shows that $q^2 < 2$. If p is in B then $2 < p^2$, (1.16) shows that $q < p$, and (1.17) shows that $q^2 > 2$.

A Non-Archimedean Ordered Field

The fields $\mathbb{Q}$ and $\mathbb{R}$ are Archimedean, see below. But there exist ordered fields without this property. Let $F := \mathbb{R}(t)$ the field of rational functions $f(t) = p(t)/q(t)$ where p and q are polynomials with real coefficients. Since p and q have only finitely many zeros, for large t, $f(t)$ is either positive or negative. In the first case we set $f > 0$. In this way $\mathbb{R}(t)$ becomes an ordered field. But $t > n$ for all $n \in \mathbb{N}$ since the polynomial $f(t) = t - n$ becomes positive for large t (and fixed n).
Our aim is to define b^x for arbitrary *real* x.

Lemma 1.27 *Let b, p be real numbers with $b > 1$ and $p > 0$. Set*

$$M = \{b^r \mid r \in \mathbb{Q}, r < p\}, \quad M' = \{b^s \mid s \in \mathbb{Q}, p < s\}.$$

Then

$$\sup M = \inf M'.$$

Proof. (a) M is bounded above by arbitrary b^s, $s \in Q$, with $s > p$, and M' is bounded below by any b^r, $r \in \mathbb{Q}$, with $r < p$. Hence $\sup M$ and $\inf M'$ both exist.
(b) Since $r < p < s$ implies $a^r < b^s$ by Lemma 1.23, $\sup M \leq b^s$ for all $b^s \in M'$. Taking the infimum over all such b^s, $\sup M \leq \inf M'$.
(c) Let $s = \sup M$ and $\varepsilon > 0$ be given. We want to show that $\inf M' < s + \varepsilon$. Choose $n \in \mathbb{N}$ such that

$$1/n < \varepsilon/(s(b - 1)). \tag{1.18}$$

By Proposition 1.11 there exist $r, s \in \mathbb{Q}$ with

$$r < p < s \quad \text{and} \quad s - r < \frac{1}{n}. \tag{1.19}$$

Using $s - r < 1/n$, Bernoulli's inequality (part 2), and (1.18), we compute

$$b^s - b^r = b^r(b^{s-r} - 1) \le s(b^{\frac{1}{n}} - 1) \le s\frac{1}{n}(b-1) < \varepsilon.$$

Hence

$$\inf M' \le b^s < b^r + \varepsilon \le \sup M + \varepsilon.$$

Since ε was arbitrary, $\inf M' \le \sup M$, and finally, with the result of (b), $\inf M' = \sup M$. ∎

Corollary 1.28 *Suppose $p \in \mathbb{Q}$ and $b > 1$ is real. Then*

$$b^p = \sup\{b^r \mid r \in \mathbb{Q}, r < p\}.$$

Proof. For all rational numbers $r, p, s \in \mathbb{Q}$, $r < p < s$ implies $a^r < a^p < a^s$. Hence $\sup M \le a^p \le \inf M'$. By the lemma, these three numbers coincide. ∎

Inequalities

Now we extend Bernoulli's inequality to rational exponents.

Proposition 1.29 (Bernoulli's inequality) *Let $a \ge -1$ real and $r \in \mathbb{Q}$. Then*
(a) $(1+a)^r \ge 1 + ra$ *if* $r \ge 1$,
(b) $(1+a)^r \le 1 + ra$ *if* $0 \le r \le 1$.
Equality holds if and only if $a = 0$ or $r = 1$.

Proof. (b) Let $r = m/n$ with $m \le n$, $m, n \in \mathbb{N}$. Apply (1.11) to $x_i := 1 + a$, $i = 1, \dots, m$ and $x_i := 1$ for $i = m+1, \dots, n$. We obtain

$$\frac{1}{n}\left(m(1+a) + (n-m)1\right) \ge \left((1+a)^m \cdot 1^{n-m}\right)^{\frac{1}{n}}$$
$$\frac{m}{n}a + 1 \ge (1+a)^{\frac{m}{n}},$$

which proves (b). Equality holds if $n = 1$ or if $x_1 = \cdots = x_n$ i. e. $a = 0$.
(a) Now let $s \ge 1$, $z \ge -1$. Setting $r = 1/s$ and $a := (1+z)^{1/r} - 1$ we obtain $r \le 1$ and $a \ge -1$. Inserting this into (b) yields

$$(1+a)^r \le \left((1+z)^{\frac{1}{r}}\right)^r \le 1 + r\left((1+z)^s - 1\right)$$
$$z \le r\left((1+z)^s - 1\right)$$
$$1 + sz \le (1+z)^s.$$

This completes the proof of (a). ∎

Corollary 1.30 (Bernoulli's inequality) *Let $a \geq -1$ real and $x \in \mathbb{R}$. Then*
(a) $(1+a)^x \geq 1 + xa$ *if* $x \geq 1$,
(b) $(1+a)^x \leq 1 + xa$ *if* $x \leq 1$. *Equality holds if and only if* $a = 0$ *or* $x = 1$.

Proof. (a) First let $a > 0$. By Proposition 1.29 (a) $(1+a)^r \geq 1 + ra$ if $r \in \mathbb{Q}$. Hence,

$$(1+a)^x = \sup\{(1+a)^r \mid r \in \mathbb{Q},\, r < x\} \geq \sup\{1 + ra \mid r \in \mathbb{Q},\, r < x\} = 1 + xa.$$

Now let $-1 \leq a < 0$. Then $r < x$ implies $ra > xa$, and Proposition 1.29 (a) implies

$$(1+a)^r \geq 1 + ra > 1 + xa. \tag{1.20}$$

By definition of the power with a real exponent, see (1.4)

$$(1+a)^x = \frac{1}{\sup\{(1/(a+1))^r \mid r \in \mathbb{Q},\, r < x\}} \underset{\text{HW 2.1}}{=} \inf\{(1+a)^r \mid r \in \mathbb{Q},\, r < x\}.$$

Taking in (1.20) the infimum over all $r \in \mathbb{Q}$ with $r < x$ we obtain

$$(1+a)^x = \inf\{(1+a)^r \mid r \in \mathbb{Q},\, r < x\} \geq 1 + xa.$$

(b) The proof is analogous, so we omit it. ∎

Proposition 1.31 (Young's inequality) *If $a, b \in \mathbb{R}_+$ and $p > 1$, then*

$$ab \leq \frac{1}{p}\, a^p + \frac{1}{q}\, b^q, \tag{1.21}$$

where $1/p + 1/q = 1$. Equality holds if and only if $a^p = b^q$.

Proof. First note that $1/q = 1 - 1/p$. We reformulate Bernoulli's inequality for $y \in \mathbb{R}_+$ and $p > 1$

$$y^p - 1 \geq p(y-1) \iff \frac{1}{p}(y^p - 1) + 1 \geq y \iff \frac{1}{p} y^p + \frac{1}{q} \geq y.$$

If $b = 0$ the statement is always true. If $b \neq 0$ insert $y := ab/b^q$ into the above inequality:

$$\begin{aligned} \frac{1}{p}\left(\frac{ab}{b^q}\right)^p + \frac{1}{q} &\geq \frac{ab}{b^q} \\ \frac{1}{p}\frac{a^p b^p}{b^{pq}} + \frac{1}{q} &\geq \frac{ab}{b^q} \qquad \mid \cdot\, b^q \\ \frac{1}{p} a^p + \frac{1}{q} b^q &\geq ab, \end{aligned}$$

since $b^{p+q} = b^{pq}$. We have equality if $y = 1$ or $p = 1$. The later is impossible by assumption. $y = 1$ is equivalent to $b^q = ab$ or $b^{q-1} = a$ or $b^{(q-1)p} = a^p$ ($b \neq 0$). If $b = 0$ equality holds if and only if $a = 0$. ∎

Proposition 1.32 (Hölder's inequality) *Let $p > 1$, $1/p + 1/q = 1$, and $x_1, \ldots, x_n \in \mathbb{R}_+$ and $y_1, \ldots, y_n \in \mathbb{R}_+$ non-negative real numbers. Then*

$$\sum_{k=1}^{n} x_k y_k \leq \left(\sum_{k=1}^{n} x_k^p \right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} y_k^q \right)^{\frac{1}{q}}. \tag{1.22}$$

We have equality if and only if there exists $c \in \mathbb{R}$ such that for all $k = 1, \ldots, n$, $x_k^p / y_k^q = c$ (they are proportional).

Proof. Set $A := (\sum_{k=1}^n x_k^p)^{\frac{1}{p}}$ and $B := (\sum_{k=1}^n y_k^q)^{\frac{1}{q}}$. The cases $A = 0$ and $B = 0$ are trivial. So we assume $A, B > 0$. By Young's inequality we have

$$\begin{aligned}
\frac{x_k}{A} \cdot \frac{y_k}{B} &\leq \frac{1}{p} \frac{x_k^p}{A^p} + \frac{1}{q} \frac{y_k^q}{B^q} \\
\Longrightarrow \frac{1}{AB} \sum_{k=1}^{n} x_k y_k &\leq \frac{1}{pA^p} \sum_{k=1}^{n} x_k^p + \frac{1}{qB^q} \sum_{k=1}^{n} y_k^q \\
&= \frac{1}{p \sum x_k^p} \sum x_k^p + \frac{1}{q \sum y_k^q} \sum y_k^q \\
&= \frac{1}{p} + \frac{1}{q} = 1 \\
\Longrightarrow \sum_{k=1}^{n} x_k y_k &\leq \left(\sum_{k=1}^{n} x_k^p \right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} y_k^q \right)^{\frac{1}{q}}.
\end{aligned}$$

Equality holds if and only if $x_k^p / A^p = y_k^q / B^q$ for all $k = 1, \ldots, n$. Therefore, $x_k^p / y_k^q = \text{const.}$ ∎

Corollary 1.33 (Complex Hölder's inequality) *Let $p > 1$, $1/p + 1/q = 1$ and $x_k, y_k \in \mathbb{C}$, $k = 1, \ldots, n$. Then*

$$\sum_{k=1}^{n} |\, x_k y_k \,| \leq \left(\sum_{k=1}^{n} |\, x_k \,|^p \right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} |\, y_k \,|^q \right)^{\frac{1}{q}}.$$

Equality holds if and only if $|\, x_k \,|^p \,/\, |\, y_k \,|^q = \text{const.}$ for $k = 1, \ldots, n$.

Proof. Set $x_k := |\, x_k \,|$ and $y_k := |\, y_k \,|$ in (1.22). This will prove the statement. ∎

Proposition 1.34 (Minkowski's inequality) *If $x_1, \ldots, x_n \in \mathbb{R}_+$ and $y_1, \ldots, y_n \in \mathbb{R}_+$ and $p \geq 1$ then*

$$\left(\sum_{k=1}^{n} (x_k + y_k)^p \right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^{n} x_k^p \right)^{\frac{1}{p}} + \left(\sum_{k=1}^{n} y_k^p \right)^{\frac{1}{p}}. \tag{1.23}$$

Equality holds if $p = 1$ or if $p > 1$ and $x_k / y_k = \text{const.}$

Proof. The case $p = 1$ is obvious. Let $p > 1$. As before let $q > 0$ be the unique positive number with $1/p + 1/q = 1$. We compute

$$\begin{aligned}
&\sum_{k=1}^{n}(x_k + y_k)^p = \sum_{k=1}^{n}(x_k + y_k)(x_k + y_k)^{p-1} = \sum_{k=1}^{n} x_k(x_k + y_k)^{p-1} + \sum_{k=1}^{n} y_k(x_k + y_k)^{p-1} \\
&\underset{(1.22)}{\leq} \left(\sum x_k^p\right)^{\frac{1}{p}} \left(\sum_k (x_k + y_k)^{(p-1)q}\right)^{\frac{1}{q}} + \left(\sum y_k^p\right)^{\frac{1}{p}} \left(\sum_k (x_k + y_k)^{(p-1)q}\right)^{\frac{1}{q}} \\
&\leq \left(\left(\sum x_k^p\right)^{1/p} + \left(\sum y_k^q\right)^{1/q}\right) \left(\sum (x_k + y_k)^p\right)^{1/q}.
\end{aligned}$$

We can assume that $\sum(x_k + y_k)^p > 0$. Using $1 - \dfrac{1}{q} = \dfrac{1}{p}$ by taking the quotient of the last inequality by $\left(\sum(x_k + y_k)^p\right)^{1/q}$ we obtain the claim.
Equality holds if $x_k^p/(x_k + y_k)^{(p-1)q} = \text{const.}$ and $y_k^p/(x_k + y_k)^{(p-1)q)} = \text{const.}$; that is $x_k/y_k = \text{const.}$ ∎

Corollary 1.35 (Complex Minkowski's inequality) *If $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{C}$ and $p \geq 1$ then*

$$\left(\sum_{k=1}^{n} |\, x_k + y_k \,|^p\right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^{n} |\, x_k \,|^p\right)^{\frac{1}{p}} + \left(\sum_{k=1}^{n} |\, y_k \,|^p\right)^{\frac{1}{p}}. \tag{1.24}$$

Equality holds if $p = 1$ or if $p > 1$ and $x_k/y_k = \lambda > 0$.

Proof. Using the triangle inequality gives $|\, x_k + y_k \,| \leq |\, x_k \,| + |\, y_k \,|$; hence $\sum_{k=1}^{n} |\, x_k + y_k \,|^p \leq \sum_{k=1}^{n} (|\, x_k \,| + |\, y_k \,|)^p$. The real version of Minkowski's inequality now proves the assertion. ∎

If $x = (x_1, \ldots, x_n)$ is a vector in $\mathbb{R}^n$ or $\mathbb{C}^n$, the (non-negative) number

$$\|x\|_p := \left(\sum_{k=1}^{n} |\, x_k \,|^p\right)^{\frac{1}{p}}$$

is called the *p-norm* of the vector x. Minkowski's inequalitie then reads as

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p$$

which is the triangle inequality for the p-norm.

Chapter 2

Sequences and Series

This chapter will deal with one of the main notions of calculus, the **limit of a sequence**. Although we are concerned with real sequences, almost all notions make sense in arbitrary metric spaces like $\mathbb{R}^n$ or $\mathbb{C}^n$.
Given $a \in \mathbb{R}$ and $\varepsilon > 0$ we define the *ε-neighborhood of a* as

$$U_\varepsilon(a) := (a-\varepsilon, a+\varepsilon) = \{x \in \mathbb{R} \mid a-\varepsilon < x < a+\varepsilon\} = \{x \in \mathbb{R} \mid |\, x-a \,| < \varepsilon\}.$$

2.1 Convergent Sequences

A *sequence* is a mapping $x\colon \mathbb{N} \to \mathbb{R}$. To every $n \in \mathbb{N}$ we associate a real number x_n. We write this as $(x_n)_{n\in\mathbb{N}}$ or $(x_1, x_2, \dots)$. For different sequences we use different letters as (a_n), (b_n), (y_n).

Example 2.1 (a) $x_n = \frac{1}{n}$, $\left(\frac{1}{n}\right)$, $(x_n) = \left(1, \frac{1}{2}, \frac{1}{3}, \dots\right)$;
(b) $x_n = (-1)^n + 1$, $(x_n) = (0, 2, 0, 2, \dots)$;
(c) $x_n = a$ ($a \in \mathbb{R}$ fixed), $(x_n) = (a, a, \dots)$ (constant sequence),
(d) $x_n = 2n-1$, $(x_n) = (1, 3, 5, 7, \dots)$ the sequence of odd positive integers.
(e) $x_n = a^n$ ($a \in \mathbb{R}$ fixed), $(x_n) = (a, a^2, a^3, \dots)$ (geometric sequence);

Definition 2.1 A sequence (x_n) is said to be *convergent to* $x \in \mathbb{R}$ if

> For every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies
>
> $$|\, x_n - x \,| < \varepsilon.$$

x is called the *limit* of (x_n) and we write

$$x = \lim_{n\to\infty} x_n \quad \text{or simply} \quad x = \lim x_n \quad \text{or} \quad x_n \to x.$$

If there is no such x with the above property, the sequence (x_n) is said to be *divergent*.
In other words: (x_n) converges to x if any neighborhood $U_\varepsilon(x)$, $\varepsilon > 0$, contains "almost all" elements of the sequence (x_n). "Almost all" means "all but finitely many." Sometimes we say "for sufficiently large n" which means the same.

This is an equivalent formulation since $x_n \in U_\varepsilon(x)$ means $x-\varepsilon < x_n < x+\varepsilon$, hence $|\, x - x_n \,| < \varepsilon$. The n_0 in question need not to be the smallest possible.
We write

$$\lim_{n\to\infty} x_n = +\infty \tag{2.1}$$

if for all $E > 0$ there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies $x_n \geq E$. Similarly, we write

$$\lim_{n\to\infty} x_n = -\infty \tag{2.2}$$

if for all $E > 0$ there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies $x_n \leq -E$. In these cases we say that $+\infty$ and $-\infty$ are *improper limits* of (x_n). Note that in both cases (x_n) is divergent.

Example 2.2 This is Example 2.1 continued.
(a) $\lim_{n\to\infty} \frac{1}{n} = 0$. Indeed, let $\varepsilon > 0$ be fixed. We are looking for some n_0 with $\left|\, \frac{1}{n} - 0 \,\right| < \varepsilon$ for all $n \geq n_0$. This is equivalent to $1/\varepsilon < n$. Choose $n_0 > 1/\varepsilon$ (which is possible by the Archimedean property). Then for all $n \geq n_0$ we have

$$n \geq n_0 > \frac{1}{\varepsilon} \Longrightarrow \frac{1}{n} < \varepsilon \Longrightarrow |\, x_n - 0 \,| < \varepsilon.$$

Therefore, (x_n) tends to 0 as $n \to \infty$.
(b) $x_n = (-1)^n + 1$ is divergent. Suppose to the contrary that x is the limit. To $\varepsilon = 1$ there is n_0 such that for $n \geq n_0$ we have $|\, x_n - x \,| < 1$. For even $n \geq n_0$ this implies $|\, 2 - x \,| < 1$ for odd $n \geq n_0$, $|\, 0 - x \,| = |\, x \,| < 1$. The triangle inequality gives

$$2 = |\, (2 - x) + x \,| \leq |\, 2 - x \,| + |\, x \,| < 1 + 1 = 2.$$

This is a contradiction. Hence, (x_n) is divergent.
(c) $x_n = a$. $\lim x_n = a$ since $|\, x_n - a \,| = |\, a - a \,| = 0 < \varepsilon$ for all $\varepsilon > 0$ and all $n \in \mathbb{N}$.
(d) $\lim(2n - 1) = +\infty$. Indeed, suppose that $E > 0$ is given. Choose $n_0 > \frac{E}{2} + 1$. Then

$$n \geq n_0 \Longrightarrow n > \frac{E}{2} + 1 \Longrightarrow 2n - 2 > E \Longrightarrow x_n = 2n - 1 > 2n - 2 > E.$$

This proves the claim. Similarly, one can show that $\lim -n^3 = -\infty$. But both $((-n)^n)$ and $(1, 2, 1, 3, 1, 4, 1, 5, \dots)$ have no improper limit. Indeed, the first one becomes arbitrarily large for even n and arbitrarily small for odd n. The second one becomes large for eveb n but is constant for odd n.
(e) $x_n = a^n$, $(a \geq 0)$.

$$\lim_{n\to\infty} a^n = \begin{cases} 1, & \text{if} \quad a = 1, \\ 0, & \text{if} \quad 0 \leq a < 1. \end{cases}$$

(a^n) is divergent for $a > 1$. Moreover, $\lim a^n = +\infty$. To prove this let $E > 0$ be given. By the Archimedean property of $\mathbb{R}$ and since $a - 1 > 0$ we find $m \in \mathbb{N}$ such that $m(a - 1) > E$. Bernoulli's inequality gives

$$a^m \geq m(a - 1) + 1 > m(a - 1) > E.$$

By Lemma 1.23 (b), $n \geq m$ implies

$$a^n \geq a^m > E.$$

This proves the claim.
Clearly (a^n) is convergent in cases $a = 0$ and $a = 1$ since it is constant then. Let $0 < a < 1$ and put $b = \frac{1}{a} - 1$; then $b > 0$. Bernoulli's inequality gives

$$\begin{aligned} \frac{1}{a^n} = \left(\frac{1}{a}\right)^n &= (b+1)^n \geq 1 + nb > nb \\ \Longrightarrow 0 < a^n &< \frac{1}{nb}. \end{aligned} \tag{2.3}$$

Let $\varepsilon > 0$. Choose $n_0 > \dfrac{1}{\varepsilon b}$. Then $\varepsilon > \dfrac{1}{n_0 b}$ and $n \geq n_0$ implies

$$|\, a^n - 0 \,| = |\, a^n \,| = a^n \underset{(2.3)}{<} \frac{1}{nb} \leq \frac{1}{n_0 b} < \varepsilon.$$

Hence, $a^n \to 0$.

Proposition 2.1 *The limit of a convergent sequence is uniquely determined.*

Proof. Suppose that $x = \lim x_n$ and $y = \lim x_n$ and $x \neq y$. Put $\varepsilon := |\, x - y \,|/2 > 0$. Then

$$\begin{aligned} &\exists n_1 \in \mathbb{N}\ \forall n \geq n_1 : |\, x - x_n \,| < \varepsilon, \\ &\exists n_2 \in \mathbb{N}\ \forall n \geq n_2 : |\, y - x_n \,| < \varepsilon. \end{aligned}$$

Choose $m \geq \max\{n_1, n_2\}$. Then $|\, x - x_m \,| < \varepsilon$ and $|\, y - x_m \,| < \varepsilon$. Hence,

$$|\, x - y \,| \leq |\, x - x_m \,| + |\, y - x_m \,| < 2\varepsilon = |\, x - y \,|.$$

This contradiction establishes the statement. ∎

Proposition 2.1 holds in arbitrary metric spaces.

Definition 2.2 A sequence (x_n) is said to be *bounded* if the set of its elements is a bounded set; i. e. there is a $C \geq 0$ such that

$$|\, x_n \,| \leq C \quad \text{for all } n \in \mathbb{N}.$$

Similarly, (x_n) is said to be *bounded above* or *bounded below* if there exists $C \in \mathbb{R}$ such that $x_n \leq C$ or $x_n \geq C$, respectively, for all $n \in \mathbb{N}$

Proposition 2.2 *If (x_n) is convergent, then (x_n) is bounded.*

Proof. Let $x = \lim x_n$. To $\varepsilon = 1$ there exists $n_0 \in \mathbb{N}$ such that $|\, x - x_n \,| < 1$ for all $n \geq n_0$. Then $|\, x_n \,| = |\, x_n - x + x \,| \leq |\, x_n - x \,| + |\, x \,| < |\, x \,| + 1$ for all $n \geq n_0$. Put

$$C := \max\{|\, x_1 \,|, \ldots, |\, x_{n_0-1} \,|, |\, x \,| + 1\}.$$

Then $|\, x_n \,| \leq C$ for all $n \in \mathbb{N}$. ∎

The reversal statement is not true; there are bounded sequences which are not convergent, see Example 2.1 (b).

Ex Class: If (x_n) has an improper limit, then (x_n) is divergent.

Proof. Suppose to the contrary that (x_n) is convergent; then it is bounded, say $|\, x_n \,| \leq C$ for all n. This contradicts $x_n > E$ as well as $x_n < -E$ for $E = C$ and sufficiently large n. Hence, (x_n) has no improper limits, a contradiction. ∎

2.1.1 Algebraic operations with sequences

The sum, difference, product, quotient and absolute value of sequences (x_n) and (y_n) are defined as follows

$$(x_n) \pm (y_n) := (x_n \pm y_n), \qquad (x_n) \cdot (y_n) := (x_n\, y_n),$$
$$\frac{(x_n)}{(y_n)} := \left(\frac{x_n}{y_n}\right), (y_n \neq 0) \qquad |\,(x_n)\,| := (|\, x_n \,|).$$

Proposition 2.3 *If (x_n) and (y_n) are convergent sequences and $c \in \mathbb{R}$, then their sum, difference, product, quotient (provided $y_n \neq 0$ and $\lim y_n \neq 0$), and their absolute values are also convergent:*

(a) $\lim(x_n \pm y_n) = \lim x_n \pm \lim y_n$;

(b) $\lim(cx_n) = c \lim x_n$, $\lim(x_n + c) = \lim x_n + c$.

(c) $\lim(x_n\, y_n) = \lim x_n \cdot \lim y_n$;

(d) $\lim \frac{x_n}{y_n} = \frac{\lim x_n}{\lim y_n}$ *if* $y_n \neq 0$ *for all* n *and* $\lim y_n \neq 0$;

(e) $\lim |\, x_n \,| = |\, \lim x_n \,|$.

Proof. Let $x_n \to x$ and $y_n \to y$.

(a) Given $\varepsilon > 0$ then there exist integers n_1 and n_2 such that

$$n \geq n_1 \text{ implies } |\, x_n - x \,| < \varepsilon/2 \text{ and } n \geq n_2 \text{ implies } |\, y_n - y \,| < \varepsilon/2.$$

If $n_0 := \max\{n_1, n_2\}$, then $n \geq n_0$ implies

$$|\,(x_n + y_n) - (x + y)\,| \leq |\, x_n - x \,| + |\, y_n - y \,| \leq \varepsilon.$$

The proof for the difference is quite similar.

(b) follows from $|\, cx_n - cx \,| = |\, c \,|\, |\, x_n - x \,|$ and $|\,(x_n + c) - (x + c)\,| = |\, x_n - x \,|$.

(c) We use the identity

$$x_n y_n - xy = (x_n - x)(y_n - y) + x(y_n - y) + y(x_n - x). \tag{2.4}$$

Given $\varepsilon > 0$ there are integers n_1 and n_2 such that

$n \geq n_1$ implies $|\, x_n - x \,| < \sqrt{\varepsilon}$ and $n \geq n_2$ implies $|\, y_n - y \,| < \sqrt{\varepsilon}$.

If we take $n_0 = \max\{n_1, n_2\}$, $n \geq n_0$ implies

$$|\, (x_n - x)(y_n - y) \,| < \varepsilon,$$

so that

$$\lim_{n\to\infty} (x_n - x)(y_n - y) = 0.$$

Now we apply (a) and (b) to (2.4) and conclude that

$$\lim_{n\to\infty} (x_n y_n - xy) = 0.$$

(d) Choosing n_1 such that $|\, y_n - y \,| < |\, y \,|/2$ if $n \geq n_1$, we see that

$$|\, y \,| \leq |\, y - y_n \,| + |\, y_n \,| < |\, y \,|/2 + |\, y_n \,| \implies |\, y_n \,| > |\, y \,|/2.$$

Given $\varepsilon > 0$, there is an integer $n_2 > n_1$ such that $n \geq n_2$ implies

$$|\, y_n - y \,| < |\, y \,|^2 \, \varepsilon/2.$$

Hence, for $n \geq n_2$,

$$\left| \frac{1}{y_n} - \frac{1}{y} \right| = \left| \frac{y_n - y}{y_n y} \right| < \frac{2}{|\, y \,|^2} \, |\, y_n - y \,| < \varepsilon$$

and we get $\lim(\frac{1}{y_n}) = \dfrac{1}{\lim y_n}$. The general case can be reduced to the above case using (c) and $(x_n/y_n) = (x_n \cdot 1/y_n)$.

(e) By Lemma 1.12 (e) we have $|\, |\, x_n \,| - |\, x \,|\, | \leq |\, x_n - x \,|$. Given $\varepsilon > 0$, there is n_0 such that $n \geq n_0$ implies $|\, x_n - x \,| < \varepsilon$. By the above inequality, also $|\, |\, x_n \,| - |\, x \,|\, | \leq \varepsilon$ and we are done. ∎

Example 2.3 (a) $z_n := \frac{n+1}{n}$. Set $x_n = 1$ and $y_n = 1/n$. Then $z_n = x_n + y_n$ and we already know that $\lim x_n = 1$ and $\lim y_n = 0$. Hence, $\lim \frac{n+1}{n} = \lim 1 + \lim \frac{1}{n} = 1 + 0 = 1$.

(b) $a_n = \frac{3n^2+13n}{n^2-2}$. We can write this as

$$a_n = \frac{3 + \frac{13}{n}}{1 - \frac{2}{n^2}}.$$

Since $\lim 1/n = 0$, by Proposition 2.3, we obtain $\lim 1/n^2 = 0$ and $\lim 13/n = 0$. Hence $\lim 2/n^2 = 0$ and $\lim (3 + 13/n) = 3$. Finally,

$$\lim_{n\to\infty} \frac{3n^2 + 13n}{n^2 - 2} = \frac{\lim_{n\to\infty} \left(3 + \frac{13}{n}\right)}{\lim_{n\to\infty} \left(1 - \frac{2}{n}\right)} = \frac{3}{1} = 3.$$

(c) We introduce the notion of a polynomial and and a rational function.

Given $a_0, a_1, \ldots, a_n \in \mathbb{R}$, $a_n \neq 0$. The function $p\colon \mathbb{R} \to \mathbb{R}$ given by $p(t) := a_n t^n + a_{n-1} t^{n-1} + \cdots + a_1 t + a_0$ is called a *polynomial*. The positive integer n is the *degree* of the polynomial

$p(t)$, and $a_1, \dots a_n$ are called the *coefficients* of $p(t)$. The set of all real polynomials forms a real vector space denoted by $\mathbb{R}[x]$.
Given two polynomials p and q; put $D := \{t \in \mathbb{R} \mid q(t) \neq 0\}$. Then $r = \frac{p}{q}$ is a called a *rational function* where $r\colon D \to \mathbb{R}$ is defined by

$$r(t) := \frac{p(t)}{q(t)}.$$

Polynomials are special rational functions with $q(t) \equiv 1$. the set of rational functions with real coefficients form both a real vector space and a field. It is denoted by $\mathbb{R}(x)$.

Lemma 2.4 (a) *Let $a_n \to 0$ be a sequence tending to zero with $a_n \neq 0$ for every n. Then*

$$\lim_{n\to\infty} \frac{1}{a_n} = \begin{cases} +\infty, & \text{if } a_n > 0 \text{ for almost all } n; \\ -\infty, & \text{if } a_n < 0 \text{ for almost all } n. \end{cases}$$

(b) *Let $y_n \to a$ be a sequence converging to a and $a > 0$. Then $y_n > 0$ for almost all $n \in \mathbb{N}$.*

Proof. (a) We will prove the case with $-\infty$. Let $\varepsilon > 0$. By assumption there is a positive integer n_0 such that $n \geq n_0$ implies $-\varepsilon < a_n < 0$. Tis implies $0 < -a_n < \varepsilon$ and further $\frac{1}{a_n} < -\frac{1}{\varepsilon} < 0$. Suppose $E > 0$ is given; choose $\varepsilon = 1/E$ and n_0 as above. Then by the previous argument, $n \geq n_0$ implies

$$\frac{1}{a_n} < -\frac{1}{\varepsilon} = -E.$$

This shows $\lim_{n\to\infty} \frac{1}{a_n} = -\infty$.
(b) To $\varepsilon = a$ there exists n_0 such that $n \geq n_0$ implies $|\, y_n - a \,| < a$. That is $-a < y_n - a < a$ or $0 < y_n < 2a$ which proves the claim. ∎

Lemma 2.5 *Suppose that $p(t) = \sum_{k=0}^{r} a_k t^k$ and $q(t) = \sum_{k=0}^{s} b_k t^k$ are real polynomials with $a_r \neq 0$ and $b_s \neq 0$. Then*

$$\lim_{n\to\infty} \frac{p(n)}{q(n)} = \begin{cases} 0, & r < s, \\ \frac{a_r}{b_r}, & r = s, \\ +\infty, & r > s \quad \text{and} \quad \frac{a_r}{b_r} > 0, \\ -\infty, & r > s \quad \text{and} \quad \frac{a_r}{b_r} < 0. \end{cases}$$

Proof. Note first that

$$\frac{p(n)}{q(n)} = \frac{n^r \left(a_r + a_{r-1}\frac{1}{n} + \dots + a_0 \frac{1}{n^r}\right)}{n^s \left(b_s + b_{s-1}\frac{1}{n} + \dots + b_0 \frac{1}{n^s}\right)} = \frac{1}{n^{s-r}} \cdot \frac{a_r + a_{r-1}\frac{1}{n} + \dots + a_0 \frac{1}{n^r}}{b_s + b_{s-1}\frac{1}{n} + \dots + b_0 \frac{1}{n^s}} =: \frac{1}{n^{s-r}} \cdot c_n$$

Suppose that $r = s$. By Proposition 2.3, $\frac{1}{n^k} \longrightarrow 0$ for all $k \in \mathbb{N}$. By the same proposition, the limits of each summand in the numerator and denominator is 0 except for the first in each sum. Hence,

$$\lim_{n\to\infty} \frac{p(n)}{q(n)} = \frac{\lim_{n\to\infty} \left(a_r + a_{r-1}\frac{1}{n} + \dots + a_0 \frac{1}{n^r}\right)}{\lim_{n\to\infty} \left(b_s + b_{s-1}\frac{1}{n} + \dots + b_0 \frac{1}{n^s}\right)} = \frac{a_r}{b_r}.$$

Supose now that $r < s$. As in the previous case, the sequence (c_n) tends to a_r/b_s but the first factor $\frac{1}{n^{s-r}}$ tends to 0. Hence, the product sequence tends to 0.
Suppose that $r > s$ and $\frac{a_s}{b_r} > 0$. The sequence (c_n) has a positive limit. By Lemma 2.4 (b) almost all $c_n > 0$. Hence,

$$d_n := \frac{q(n)}{p(n)} = \frac{1}{n^{r-s}} \frac{1}{c_n}$$

tends to 0 by the above part and $d_n > 0$ for almost all n. By Lemma 2.4 (a), the sequence $\left(\frac{1}{d_n}\right) = \left(\frac{p(n)}{q(n)}\right)$ tends to $+\infty$ as $n \to \infty$, which proves the claim in the first case. The case $a_r/b_s < 0$ can be obtained by multiplying with -1 and noting that $\lim_{n\to\infty} x_n = +\infty$ implies $\lim_{n\to\infty}(-x_n) = -\infty$. ∎

In the German literature the next proposition is known as the 'Theorem of the two policemen'.

Proposition 2.6 (Sandwich Theorem) *Let a_n, b_n and x_n be real sequences with $a_n \le x_n \le b_n$ for all but finitely many $n \in \mathbb{N}$. Further let $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = x$. Then x_n is also convergent to x.*

Proof. Let $\varepsilon > 0$. There exist n_1, n_2, and $n_3 \in \mathbb{N}$ such $n \ge n_1$ implies $a_n \in U_\varepsilon(x)$, $n \ge n_2$ implies $b_n \in U_\varepsilon(x)$, and $n \ge n_3$ implies $a_n \le x_n \le b_n$. Choosing $n_0 = \max\{n_1, n_2, n_3\}$, $n \ge n_0$ implies $x_n \in U_\varepsilon(x)$. Hence, $x_n \to x$. ∎

Remark. (a) If two sequences (a_n) and (b_n) differ only in finitely many elements, then both sequences converge to the same limit or both diverge.
(b) Define the "shifted sequence" $b_n := a_{n+k}$, $n \in \mathbb{N}$, where k is a fixed positv integer. Then both sequences converge to the same limit or both diverge.

2.1.2 Some special sequences

Proposition 2.7 (a) *If $p > 0$, then* $\lim_{n\to\infty} \frac{1}{n^p} = 0$.
(b) *If $p > 0$, then* $\lim_{n\to\infty} \sqrt[n]{p} = 1$.
(c) $\lim_{n\to\infty} \sqrt[n]{n} = 1$.
(d) *If $a > 1$ and $\alpha \in \mathbb{R}$, then* $\lim_{n\to\infty} \frac{n^\alpha}{a^n} = 0$.

Proof. (a) Let $\varepsilon > 0$. Take $n_0 > (1/\varepsilon)^{1/p}$ (Note that the Archimedean Property of the real numbers is used here). Then $n \ge n_0$ implies $1/n^p < \varepsilon$.
(b) If $p > 1$, put $x_n = \sqrt[n]{p} - 1$. Then, $x_n > 0$ and by Bernoulli's inequality (that is by homework 4.1) we have $p^{\frac{1}{n}} = (1 + p - 1)^{\frac{1}{n}} < 1 + \frac{p-1}{n}$ that is

$$0 < x_n \le \frac{1}{n}(p - 1)$$

By Proposition 2.6, $x_n \to 0$. If $p = 1$, (b) is trivial, and if $0 < p < 1$ the result is obtained by taking reciprocals.

(c) Put $x_n = \sqrt[n]{n} - 1$. Then $x_n \geq 0$, and, by the binomial theorem,

$$n = (1 + x_n)^n \geq \frac{n(n-1)}{2}\, x_n^2.$$

Hence

$$0 \leq x_n \leq \sqrt{\frac{2}{n-1}}, \quad (n \geq 2).$$

By (a), $\frac{1}{\sqrt{n-1}} \to 0$. Applying the sandwich theorem again, $x_n \to 0$ and so $\sqrt[n]{n} \to 1$.
(d) Put $p = a - 1$, then $p > 0$. Let k be an integer such that $k > \alpha$, $k > 0$. For $n > 2k$,

$$(1+p)^n > \binom{n}{k} p^k = \frac{n(n-1)\cdots(n-k+1)}{k!}\, p^k > \frac{n^k p^k}{2^k k!}.$$

Hence,

$$0 < \frac{n^\alpha}{a^n} = \frac{n^\alpha}{(1+p)^n} < \frac{2^k k!}{p^k}\, n^{\alpha - k} \quad (n > 2k).$$

Since $\alpha - k < 0$, $n^{\alpha-k} \to 0$ by (a). ∎

Q 1. Let (x_n) be a convergent sequence, $x_n \to x$. Then the sequence of arithmetic means $s_n := \frac{1}{n} \sum_{k=1}^{n} x_k$ also converges to x.
Q 2. Let $(x_n) > 0$ a convergent sequence of positive numbers with and $\lim x_n = x > 0$. Then $\sqrt[n]{x_1 x_2 \cdots x_n} \to x$. *Hint:* Consider $y_n = \log x_n$.

2.1.3 Monotonic Sequences

Definition 2.3 A *real* sequence (x_n) is said to be

(a) *monotonically increasing* if $x_n \leq x_{n+1}$ for all n;
(b) *monotonically decreasing* if $x_n \geq x_{n+1}$ for all n.

The class of *monotonic sequences* consists of the increasing and decreasing sequences.

A sequence is said to be *strictly monotonically increasing or decreasing* if $x_n < x_{n+1}$ or $x_n > x_{n+1}$ for all n, respectively. We write $x_n \nearrow$ and $x_n \searrow$.

Proposition 2.8 *A monotonic and bounded sequence is convergent. More precisely, if* (x_n) *is increasing and bounded above, then* $\lim x_n = \sup\{x_n\}$. *If* (x_n) *is decreasing and bounded below, then* $\lim x_n = \inf\{x_n\}$.

Proof. Suppose $x_n \leq x_{n+1}$ for all n (the proof is analogous in the other case). Let $E := \{x_n \mid n \in \mathbb{N}\}$ and $x = \sup E$. Then $x_n \leq x$, $n \in \mathbb{N}$. For every $\varepsilon > 0$ there is an integer $n_0 \in \mathbb{N}$ such that

$$x - \varepsilon < x_{n_0} < x,$$

for otherwise $x - \varepsilon$ would be an upper bound of E. Since x_n increases, $n \geq n_0$ implies

$$x - \varepsilon < x_n < x,$$

which shows that (x_n) converges to x. ∎

Example 2.4 Let $x_n = \dfrac{c^n}{n!}$ with some fixed $c > 0$. We will show that $x_n \to 0$ as $n \to \infty$. Writing (x_n) recursively,

$$x_{n+1} = \frac{c}{n+1}\, x_n, \tag{2.5}$$

we observe that (x_n) is strictly decreasing for $n \geq c$. Indeed, $n \geq c$ implies $x_{n+1} = x_n \dfrac{c}{n+1} < x_n$. On the other hand, $x_n > 0$ for all n such that (x_n) is bounded below by 0. By Proposition 2.8, (x_n) converges to some $x \in \mathbb{R}$. Taking the limit $n \to \infty$ in (2.5), we have

$$x = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} \frac{c}{n+1} \cdot \lim_{n\to\infty} x_n = 0 \cdot x = 0.$$

Hence, the sequence tends to 0.

2.1.4 Subsequences

Definition 2.4 Let (x_n) be a sequence and $(n_k)_{k\in\mathbb{N}}$ a strictly increasing sequence of positive integers $n_k \in \mathbb{N}$. We call $(x_{n_k})_{k\in\mathbb{N}}$ a *subsequence* of $(x_n)_{n\in\mathbb{N}}$. If (x_{n_k}) converges, its limit is called a *subsequential limit* of (x_n).

Example 2.5 (a) $x_n = 1/n$, $n_k = 2^k$. then $(x_{n_k}) = (1/2, 1/4, 1/8, \dots)$.
(b) $(x_n) = (1, -1, 1, -1, \dots)$. $(x_{2k}) = (-1, -1, \dots)$ has the subsequential limit -1; $(x_{2k+1}) = (1, 1, 1, \dots)$ has the subsequential limit 1.

Proposition 2.9 *Subsequences of convergent sequences are convergent with the same limit.*

Proof. Let $\lim x_n = x$ and x_{n_k} be a subsequence. To $\varepsilon > 0$ there exists $m_0 \in \mathbb{N}$ such that $n \geq m_0$ implies $|\, x_n - x \,| < \varepsilon$. Since $n_m \geq m$ for all m, $m \geq m_0$ implies $|\, x_{n_m} - x \,| < \varepsilon$; hence $\lim x_{n_m} = x$. ∎

Definition 2.5 Let (x_n) be a sequence. We call $x \in \mathbb{R}$ a *limit point* of (x_n) if every neighborhood of x contains infinitely many elements of (x_n).

Proposition 2.10 *The point x is limit point of the sequence (x_n) if and only if x is a subsequential limit.*

Proof. If $\lim\limits_{k\to\infty} x_{n_k} = x$ then every neighborhood $U_\varepsilon(x)$ contains all but finitely many x_{n_k}; in particular, it contains infinitely many elements x_n. That is, x is a limit point of (x_n).
Suppose x is a limit point of (x_n). To $\varepsilon = 1$ there exists $x_{n_1} \in U_1(x)$. To $\varepsilon = 1/k$ there exists n_k with $x_{n_k} \in U_{1/k}(x)$ and $n_k > n_{k-1}$. We have constructed a subsequence (x_{n_k}) of (x_n) with

$$|\, x - x_{n_k} \,| < \frac{1}{k};$$

Hence, (x_{n_k}) converges to x. ∎

Question: Which sequences do have limit points? The answer is: Every *bounded* sequence has limit points.

Proposition 2.11 (Principle of nested intervals) *Let $I_n := [a_n, b_n]$ a sequence of closed nested intervals $I_{n+1} \subseteq I_n$ such that their lengths $b_n - a_n$ tend to 0:*

Given $\varepsilon > 0$ there exists n_0 such that $0 \leq b_n - a_n < \varepsilon$ for all $n \geq n_0$.

For any such interval sequence $\{I_n\}$ there exists a unique real number $x \in \mathbb{R}$ which is a member of all intervals, i. e. $\{x\} = \bigcap_{n\in\mathbb{N}} I_n$.

Proof. Since the intervals are nested, $(a_n) \nearrow$ is an increasing sequence bounded above by each of the b_k, and $(b_n) \searrow$ is decreasing sequence bounded below by each of the a_k. Consequently, by Proposition 2.8 we have

$$\exists x = \lim_{n\to\infty} a_n = \sup\{a_n\} \leq b_m, \quad \text{for all } m, \text{ and } \quad \exists y = \lim_{m\to\infty} b_m = \inf\{b_m\} \geq x.$$

Since $a_n \leq x \leq y \leq b_n$ for all $n \in \mathbb{N}$

$$\varnothing \neq [x, y] \subseteq \bigcap_{n\in\mathbb{N}} I_n.$$

We show the converse inclusion namely that $\bigcap_{n\in\mathbb{N}}[a_n, b_n] \subseteq [x, y]$. Let $p \in I_n$ for all n, that is, $a_n \leq p \leq b_n$ for all $n \in \mathbb{N}$. Hence $\sup_n a_n \leq p \leq \inf_n b_n$; that is $p \in [x, y]$. Thus, $[x, y] = \bigcap_{n\in\mathbb{N}} I_n$. We show uniqueness, that is $x = y$. Given $\varepsilon > 0$ we find n such that $y - x \leq b_n - a_n \leq \varepsilon$. Hence $y - x \leq 0$; therefore $x = y$. The intersection contains a unique point x. ∎

Proposition 2.12 (Bolzano–Weierstraß) *A bounded real sequence has a limit point.*

Proof. We use the principle of nested intervals. Let (x_n) be bounded, say $|\, x_n \,| \leq C$. Hence, the interval $[-C, C]$ contains infinitely many x_k. Consider the intervals $[-C, 0]$ and $[0, C]$. At least one of them contains infinitely many x_k, say $I_1 := [a_1, b_1]$. Suppose, we have already constructed $I_n = [a_n, b_n]$ of length $b_n - a_n = C/2^{n-2}$ which contains infinitely many x_k. Consider the two intervals $[a_n, (a_n + b_n)/2]$ and $[(a_n + b_n)/2, b_n]$ of length $C/2^{n-1}$. At least one of them still contains infinitely many x_k, say $I_{n+1} := [a_{n+1}, b_{n+1}]$. In this way we have constructed a nested sequence of intervals which length go to 0. By Proposition 2.11, there exists a unique $x \in \bigcap_{n\in\mathbb{N}} I_n$. We will show that x is a subsequential limit of (x_n) (and hence a limit point). For, choose $x_{n_k} \in I_k$; this is possible since I_k contains infinitely many x_m. Then, $a_k \leq x_{n_k} \leq b_k$ for all $k \in \mathbb{N}$. Proposition 2.6 gives $x = \lim a_k \leq \lim x_{n_k} \leq \lim b_k = x$; hence $\lim x_{n_k} = x$. ∎

Remark. The principle of nested intevals is *equivalent* to the order completeness of $\mathbb{R}$.

Example 2.6 (a) $x_n = (-1)^{n-1} + \frac{1}{n}$; set of limit points is $\{-1, 1\}$. First note that $-1 = \lim_{n\to\infty} x_{2n}$ and $1 = \lim_{n\to\infty} x_{2n+1}$ are subsequential limits of (x_n). We show, for example, that $\frac{1}{3}$ is not a limit point. Indeed, for $n \geq 4$, there exists a small neighborhood of $\frac{1}{3}$ which has no intersection with $U_{\frac{1}{4}}(1)$ and $U_{\frac{1}{4}}(-1)$. Hence, $\frac{1}{3}$ is not a limit point.
(b) $x_n = n - 5\left[\frac{n}{5}\right]$, where $[x]$ denotes the least integer less than or equal to x ($[\pi] = [3] = 3$, $[-2.8] = -3$, $[1/2] = 0$). $(x_n) = (1, 2, 3, 4, 0, 1, 2, 3, 4, 0, \ldots)$; the set of limit points is $\{0, 1, 2, 3, 4\}$
(c) One can enumerate the rational numbers in $(0, 1)$ in the following way.

$$\begin{array}{llllll} \frac{1}{2}, & & & x_1, & & \\ \frac{1}{3}, & \frac{2}{3}, & & x_2, & x_3 & \\ \frac{1}{4}, & \frac{2}{4}, & \frac{3}{4}, & x_4, & x_5, & x_6, \\ & \ldots & & & \ldots & \end{array}$$

The set of limit points is the whole interval $[0, 1]$ since in any neighborhood of any real number there is a rational number, see Proposition 1.11 (b) 1.11 (b). Any rational number of $(0, 1)$ appears infinitely often in this sequence, namely as $\frac{p}{q} = \frac{2p}{2q} = \frac{3p}{3q} = \cdots$.
(d) $x_n = n$ has no limit point. Since (x_n) is not bounded, Bolzano-Weierstraß fails to apply.

Definition 2.6 (a) Let (x_n) be a bounded sequence and A its set of limit points. Then $\sup A$ is called the *upper limit* of (x_n) and $\inf A$ is called the *lower limit* of (x_n). We write

$$\overline{\lim_{n\to\infty}}\, x_n, \quad \text{and} \quad = \underline{\lim_{n\to\infty}}\, x_n.$$

for the upper and lower limits of (x_n), respectively.
(b) If (x_n) is not bounded above, we write $\overline{\lim}\, x_n = +\infty$. If moreover $+\infty$ is the only limit point, $\lim x_n = +\infty$, and we can also write $\underline{\lim}\, x_n = +\infty$. If (x_n) is not bounded below, $\underline{\lim}\, x_n = -\infty$.

Proposition 2.13 *Let (x_n) be a bounded sequence and A the set of limit points of (x_n). Then $\overline{\lim}\, x_n$ and $\underline{\lim}\, x_n$ are also limit points of (x_n).*

Proof. Let $\overline{x} = \overline{\lim}\, x_n$. Let $\varepsilon > 0$. By the definition of the supremum of A there exists $x' \in A$ with

$$\overline{x} - \frac{\varepsilon}{2} < x' < \overline{x}.$$

Since x' is a limit point, $U_{\frac{\varepsilon}{2}}(x')$ contains infinitely many elements x_k. By construction, $U_{\frac{\varepsilon}{2}}(x') \subseteq U_\varepsilon(\overline{x})$. Indeed, $x \in U_{\frac{\varepsilon}{2}}(x')$ implies $|\, x - x' \,| < \frac{\varepsilon}{2}$ and therefore

$$|\, x - \overline{x} \,| = |\, x - x' + x' - \overline{x} \,| \leq |\, x - x' \,| + |\, x' - \overline{x} \,| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Hence, $\overline{x}$ is a limit point, too. The proof for $\underline{\lim}\, x_n$ is similar. ∎

Proposition 2.14 *Let $b \in \mathbb{R}$ be fixed. Suppose (x_n) is a sequence which is bounded above, then*

$$\begin{aligned} &x_n \leq b \quad \textit{for all but finitely many } n \textit{ implies} \\ &\overline{\lim_{n\to\infty}}\, x_n \leq b. \end{aligned} \tag{2.6}$$

Similarly, if (x_n) is bounded below, then

$$\begin{aligned} x_n \geq b \quad &\textit{for all but finitely many } n \textit{ implies} \\ \underline{\lim}_{n\to\infty}\, x_n \geq b. & \end{aligned} \tag{2.7}$$

Proof. We prove only the first part for $\overline{\lim}\, x_n$. Proving statement for $\underline{\lim}\, x_n$ is similar.
Let $t := \overline{\lim}\, x_n$. Suppose to the contrary that $t > b$. Set $\varepsilon = (t-b)/2$, then $U_\varepsilon(t)$ contains infinitely many x_n (t is a limit point) which are *all greater than* b; this contradicts $x_n \leq b$ for all but finitely many n. Hence $\overline{\lim}\, x_n \leq b$. ∎

Applying the first part to $b = \sup_n\{x_n\}$ and noting that $\inf A \leq \sup A$, we have

$$\inf_n\{x_n\} \leq \underline{\lim}_{n\to\infty}\, x_n \leq \overline{\lim}_{n\to\infty}\, x_n \leq \sup_n\{x_n\}.$$

The next proposition is a converse statement to Proposition 2.9.

Proposition 2.15 *Let (x_n) be a bounded sequence with a unique limit point x. Then (x_n) converges to x.*

Proof. Suppose to the contrary that (x_n) diverges; that is, there exists some $\varepsilon > 0$ such that infinitely many x_n are outside $U_\varepsilon(x)$. We view these elements as a subsequence $(y_k) := (x_{n_k})$ of (x_n). Since (x_n) is bounded, so is (y_k). By Proposition 2.12 there exists a limit point y of (y_k) which is in turn also a limit point of (x_n). Since $y \notin U_\varepsilon(x)$, $y \neq x$ is a second limit point; a contradiction! We conclude that (x_n) converges to x. ∎

Note that $t := \overline{\lim}\, x_n$ is uniquely characterized by the following two properties. For every $\varepsilon > 0$

$$\begin{aligned} t - \varepsilon < x_n, \quad &\text{for infinitely many } n, \\ x_n < t + \varepsilon, \quad &\text{for almost all } n. \end{aligned}$$

(See also homework 6.2) Let us consider the above examples.

Example 2.7 (a) $x_n = (-1)^{n-1} + 1/n$; $\underline{\lim}\, x_n = -1$, $\overline{\lim}\, x_n = 1$.
(b) $x_n = n - 5\left[\frac{n}{5}\right]$, $\underline{\lim}\, x_n = 0$, $\overline{\lim}\, x_n = 4$.
(c) (x_n) is the sequence of rational numbers of $(0,1)$; $\underline{\lim}\, x_n = 0$, $\overline{\lim}\, x_n = 1$.
(d) $x_n = n$; $\underline{\lim}\, x_n = \overline{\lim}\, x_n = +\infty$.

Proposition 2.16 *If $s_n \leq t_n$ for all but finitely many n, then*

$$\overline{\lim}_{n\to\infty}\, s_n \leq \overline{\lim}_{n\to\infty}\, t_n, \qquad \underline{\lim}_{n\to\infty}\, s_n \leq \underline{\lim}_{n\to\infty}\, t_n.$$

Proof. (a) We keep the notations s^* and t^* for the upper limits of (s_n) and (t_n), respectively. Set $\underline{s} = \underline{\lim}\, s_n$ and $\underline{t} = \underline{\lim}\, t_n$. Let $\varepsilon > 0$. By homework 6.3 (a)

$$\begin{aligned} & \underline{s} - \varepsilon \le s_n \quad \text{for all but finitely many } n \\ \xRightarrow[\text{by assumption}]{} \quad & \underline{s} - \varepsilon \le s_n \le t_n \quad \text{for all but finitely many } n \\ \xRightarrow[\text{by Prp. 2.14}]{} \quad & \underline{s} - \varepsilon \le \underline{\lim}\, t_n \\ \xRightarrow[\text{by first Remark in Subsection 1.1.7}]{} \quad & \sup\{\underline{s} - \varepsilon \mid \varepsilon > 0\} \le \underline{t} \\ & \underline{s} \le \underline{t}. \end{aligned}$$

(b) The proof for the lower limit follows from (a) and $-\sup E = \inf(-E)$. ∎

2.2 Cauchy Sequences

The aim of this section is to characterize convergent sequences without knowing their limits.

Definition 2.7 A sequence (x_n) is said to be a *Cauchy sequence* if:

> For every $\varepsilon > 0$ there exists a positive integer n_0 such that $|\, x_n - x_m \,| < \varepsilon$ for all $m, n \ge n_0$.

The definition makes sense in arbitrary metric spaces. The definition is equivalent to

$$\forall\, \varepsilon > 0 \;\; \exists\, n_0 \in \mathbb{N} \;\; \forall\, n \ge n_0 \;\; \forall\, k \in \mathbb{N} : |\, x_{n+k} - x_n \,| < \varepsilon.$$

Lemma 2.17 *Every convergent sequence is a Cauchy sequence.*

Proof. Let $x_n \to x$. To $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ such that $n \ge n_0$ implies $x_n \in U_{\varepsilon/2}(x)$. By triangle inequality, $m, n \ge n_0$ implies

$$|\, x_n - x_m \,| \le |\, x_n - x \,| + |\, x_m - x \,| \le \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Hence, (x_n) is a Cauchy sequence. ∎

Proposition 2.18 (Cauchy convergence criterion) *A real sequence is convergent if and only if it is a Cauchy sequence.*

Proof. One direction is Lemma 2.17. We prove the other direction. Let (x_n) be a Cauchy sequence. First we show that (x_n) is bounded. To $\varepsilon = 1$ there is a positive integer n_0 such that $m, n \ge n_0$ implies $|\, x_m - x_n \,| < 1$. In particular $|\, x_n - x_{n_0} \,| < 1$ for all $n \ge n_0$; hence $|\, x_n \,| < 1 + |\, x_{n_0} \,|$. Setting

$$C = \max\{|\, x_1 \,|, |\, x_2 \,|, \ldots, |\, x_{n_0-1} \,|, |\, x_{n_0} \,| + 1\},$$

$|\,x_n\,| < C$ for all n.
By Proposition 2.12 there exists a limit point x of (x_n); and by Proposition 2.10 a subsequence (x_{n_k}) converging to x. We will show that $\lim_{n\to\infty} x_n = x$. Let $\varepsilon > 0$. Since $x_{n_k} \to x$ we find $k_0 \in \mathbb{N}$ such that $k \geq k_0$ implies $|\,x_{n_k} - x\,| < \varepsilon/2$. Since (x_n) is a Cauchy sequence, there exists $n_0 \in \mathbb{N}$ such that $m, n \geq n_0$ implies $|\,x_n - x_m\,| < \varepsilon/2$. Put $n_1 := \max\{n_0, n_{k_0}\}$ and choose k_1 with $n_{k_1} \geq n_1 \geq n_{k_0}$. Then $n \geq n_1$ implies

$$|\,x - x_n\,| \leq \left|\, x - x_{n_{k_1}} \,\right| + \left|\, x_{n_{k_1}} - x_n \,\right| < 2 \cdot \varepsilon/2 = \varepsilon.$$

■

Example 2.8 (a) $x_n = \sum_{k=1}^{n} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$. We show that (x_n) is not a Cauchy sequence. For, consider

$$x_{2m} - x_m = \sum_{k=m+1}^{2m} \frac{1}{k} \geq \sum_{k=m+1}^{2m} \frac{1}{2m} = m \cdot \frac{1}{2m} = \frac{1}{2}.$$

Hence, there is no n_0 such that $p, n \geq n_0$ implies $|\,x_p - x_n\,| < \frac{1}{2}$.
(b) $x_n = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} = 1 - 1/2 + 1/3 - + \cdots + (-1)^{n+1} 1/n$. Consider

$$\begin{aligned} x_{n+k} - x_n &= (-1)^n \left[\frac{1}{n+1} - \frac{1}{n+2} + \frac{1}{n+3} - \cdots + (-1)^{k+1} \frac{1}{n+k} \right] \\ &= (-1)^n \left[\left(\frac{1}{n+1} - \frac{1}{n+2} \right) + \left(\frac{1}{n+3} - \frac{1}{n+4} \right) + \cdots \right. \\ &\qquad \left. + \begin{cases} \left(\frac{1}{n+k-1} - \frac{1}{n+k}\right), & k \text{ even} \\ \frac{1}{n+k}, & k \text{ odd} \end{cases} \right] \end{aligned}$$

Since all summands in parentheses are positive, we conclude

$$\begin{aligned} |\,x_{n+k} - x_n\,| &= \left(\frac{1}{n+1} - \frac{1}{n+2} \right) + \left(\frac{1}{n+3} - \frac{1}{n+4} \right) + \cdots + \begin{cases} \left(\frac{1}{n+k-1} - \frac{1}{n+k}\right), & k \text{ even} \\ \frac{1}{n+k}, & k \text{ odd} \end{cases} \\ &= \frac{1}{n+1} - \left[\left(\frac{1}{n+2} - \frac{1}{n+3} \right) + \cdots + \begin{cases} \frac{1}{n+k}, & k \text{ even} \\ \left(\frac{1}{n+k-1} - \frac{1}{n+k}\right), & k \text{ even} \end{cases} \right. \end{aligned}$$

$$|\,x_{n+k} - x_n\,| < \frac{1}{n+1},$$

since all summands in parentheses are positive. Hence, (x_n) is a Cauchy sequence and converges.

2.3 Series

Definition 2.8 Given a sequence (a_n), we associate with (a_n) a sequence (s_n), where

$$s_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n.$$

For (s_n) we also use the symbol

$$\sum_{k=1}^{\infty} a_k, \tag{2.8}$$

and we call it an *infinite series* or just a *series*. The numbers s_n are called the *partial sums* of the series. If (s_n) converges to s, we say that the series *converges*, and write

$$\sum_{k=1}^{\infty} a_k = s.$$

The number s is called the *sum* of the series.

Remarks 2.1 (a) The sum of a series should be clearly understood as *the limit of the sequence of partial sums*; it is not simply obtained by addition.
(b) If (s_n) diverges, the series is said to be *divergent*.
(c) The symbol $\sum_{k=1}^{\infty} a_k$ means both, the sequence of partial sums as well as the limit of this sequence (if it exists). Sometimes we use series of the form $\sum_{k=k_0}^{\infty} a_k$, $k_0 \in \mathbb{N}$. We simply write $\sum a_k$ if there is no ambiguity about the bounds of the index k.

Example 2.9 (Example 2.8 continued)
(1) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent. This is the *harmonic series*.
(2) $\displaystyle\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n}$ is convergent. It is an example of an *alternating series* (the summands are changing their signs, and the absolute value of the summands form a decreasing to 0 sequence).
(3) $\displaystyle\sum_{n=0}^{\infty} q^n$ is called the *geometric series*. It is convergent for $|\,q\,| < 1$ with $\sum_0^{\infty} q^n = \frac{1}{1-q}$. This is seen from $\displaystyle\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q}$, see proof of Lemma 1.14, first formula with $y = 1$, $x = q$. The series diverges for $|\,q\,| \geq 1$. The general formula in case $|\,q\,| < 1$ is

$$\sum_{n=n_0}^{\infty} cq^n = \frac{cq^{n_0}}{1-q}. \tag{2.9}$$

2.3.1 Properties of Convergent Series

Lemma 2.19 (1) *If $\sum_{n=1}^{\infty} a_n$ is convergent, then $\sum_{k=m}^{\infty} a_k$ is convergent for any $m \in \mathbb{N}$.*
(2) *If $\sum a_n$ is convergent, then the sequence $r_n := \sum_{k=n+1}^{\infty} a_k$ tends to 0 as $n \to \infty$.*

(3) *If (a_n) is a sequence of nonnegative real numbers, then $\sum a_n$ converges if and only if the partial sums are bounded.*

Proof. (1). Suppose that $\sum_{n=1}^{\infty} a_n = s$; we show that $\sum_{n=m}^{\infty} a_k = s - (a_1 + a_2 + \cdots + a_{m-1})$. Indeed, let (s_n) and (t_n) denote the nth partial sums of $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=m}^{\infty} a_k$, respectively. Then for $n > m$ one has $t_n = s_n - \sum_{k=1}^{m-1} a_k$. Taking the limit $n \to \infty$ proves the claim.
We prove (2). Suppose that $\sum_1^{\infty} a_n$ converges to s. By (1), $r_n = \sum_{k=n+1}^{\infty} a_k$ is also a convergent series for all n. We have

$$\begin{aligned} \sum_{k=1}^{\infty} a_k &= \sum_{k=1}^{n} a_k + \sum_{k=n+1}^{\infty} a_k \\ \Longrightarrow\ s &= s_n + r_n \\ \Longrightarrow\ r_n &= s - s_n \\ \Longrightarrow\ \lim_{n\to\infty} r_n &= s - s = 0. \end{aligned}$$

(3) Suppose $a_n \geq 0$, then $s_{n+1} = s_n + a_{n+1} \geq s_n$. Hence, (s_n) is an increasing sequence. By Proposition 2.8, (s_n) converges.
The other direction is trivial since every convergent sequence is bounded. ∎

Proposition 2.20 (Cauchy criterion) *$\sum a_n$ converges if and only if for every $\varepsilon > 0$ there is an integer $n_0 \in \mathbb{N}$ such that*

$$\left| \sum_{k=m}^{n} a_k \right| < \varepsilon \tag{2.10}$$

if $m, n \geq n_0$.

Proof. Clear from Proposition 2.18. Consider the sequence of partial sums $s_n = \sum_{k=1}^{n} a_k$ and note that for $n \geq m$ one has $| s_n - s_{m-1} | = | \sum_{k=m}^{n} a_k |$. ∎

Corollary 2.21 *If $\sum a_n$ converges, then (a_n) converges to 0.*

Proof. Take $m = n$ in (2.10); this yields $| a_n | < \varepsilon$. Hence (a_n) tends to 0. ∎

Proposition 2.22 (Comparison test) (a) *If $| a_n | \leq C b_n$ for some $C > 0$ and for almost all $n \in \mathbb{N}$, and if $\sum b_n$ converges, then $\sum a_n$ converges.*
(b) *If $a_n \geq C d_n \geq 0$ for some $C > 0$ and for almost all n, and if $\sum d_n$ diverges, then $\sum a_n$ diverges.*

Proof. (a) Suppose $n \geq n_1$ implies $|\, a_n \,| \leq C b_n$. Given $\varepsilon > 0$, there exists $n_0 \geq n_1$ such that $m, n \geq n_0$ implies

$$\sum_{k=m}^{n} b_k < \frac{\varepsilon}{C}$$

by the Cauchy criterion. Hence

$$\left| \sum_{k=m}^{n} a_k \right| \leq \sum_{k=m}^{n} |\, a_k \,| \leq \sum_{k=m}^{n} C b_k < \varepsilon,$$

and (a) follows by the Cauchy criterion.
(b) follows from (a), for if $\sum a_n$ converges, so must $\sum d_n$. ∎

2.3.2 Operations with Convergent Series

Definition 2.9 If $\sum a_n$ and $\sum b_n$ are series, we define sums and differences as follows $\sum a_n \pm \sum b_n := \sum (a_n \pm b_n)$ and $c \sum a_n := \sum c a_n$, $c \in \mathbb{R}$.
Let $c_n := \sum_{k=1}^{n} a_k b_{n-k+1}$, then $\sum c_n$ is called the *Cauchy product* of $\sum a_n$ and $\sum b_n$.

If $\sum a_n$ and $\sum b_n$ are convergent, it is easy to see that $\sum_1^\infty (a_n + b_n) = \sum_1^\infty a_n + \sum_1^\infty b_n$ and $\sum_1^n c a_n = c \sum_1^n a_n$.
Caution, the product series $\sum c_n$ need not to be convergent. Indeed, let $a_n := b_n := (-1)^n / \sqrt{n}$. One can show that $\sum a_n$ and $\sum b_n$ are convergent (see Proposition 2.29 below), however, $\sum c_n$ is not convergent, when $c_n = \sum_{k=1}^{n} a_k b_{n-k+1}$. Proof: By the arithmetic-geometric mean inequality, $|\, a_k b_{n-k+1} \,| = \frac{1}{\sqrt{k(n+1-k)}} \geq \frac{2}{n+1}$. Hence, $|\, c_n \,| \geq \sum_{k=1}^{n} \frac{2}{n+1} = \frac{2n}{n+1}$. Since c_n doesn't converge to 0 as $n \to \infty$, $\sum_{n=0}^{\infty} c_n$ diverges by Corollary 2.21

2.3.3 Series of Nonnegative Numbers

Proposition 2.23 (Compression Theorem) *Suppose $a_1 \geq a_2 \geq \cdots \geq 0$. Then the series $\sum_{n=1}^{\infty} a_n$ converges if and only if the series*

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots \tag{2.11}$$

converges.

Proof. By Lemma 2.19 (3) it suffices to consider boundedness of the partial sums. Let

$$s_n = a_1 + \cdots + a_n,$$
$$t_k = a_1 + 2a_2 + \cdots + 2^k a_{2^k}.$$

For $n < 2^k$

$$\begin{aligned} s_n &\le a_1 + (a_2 + a_3) + \cdots + (a_{2^k} + \cdots + a_{2^{k+1}-1}) \\ s_n &\le a_1 + 2a_2 + \cdots + 2^k a_{2^k} = t_k. \end{aligned} \tag{2.12}$$

On the other hand, if $n > 2^k$,

$$\begin{aligned} s_n &\ge a_1 + a_2 + (a_3 + a_4) + \cdots + (a_{2^{k-1}+1} + \cdots + a_{2^k}) \\ s_n &\ge \frac{1}{2}a_1 + a_2 + 2a_4 + \cdots + 2^{k-1}a_{2^k} \\ s_n &\ge \frac{1}{2}t_k. \end{aligned} \tag{2.13}$$

By (2.12) and (2.13), the sequences s_n and t_k are either both bounded or both unbounded. This completes the proof. ∎

Example 2.10 (a) $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if $p > 1$ and diverges if $p \le 1$.
If $p \le 0$, divergence follows from Corollary 2.21. If $p > 0$ Proposition 2.23 is applicable, and we are led to the series

$$\sum_{k=0}^{\infty} 2^k \frac{1}{2^{kp}} = \sum_{k=0}^{\infty} \left(\frac{1}{2^{p-1}}\right)^k.$$

This is a geometric series wit $q = \frac{1}{2^{p-1}}$. It converges if and only if $2^{p-1} > 1$ if and only if $p > 1$.
(b) If $p > 1$,

$$\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p} \tag{2.14}$$

converges; if $p \le 1$, the series diverges. "$\log n$" denotes the logarithm to the base e.
If $p < 0$, $\frac{1}{n(\log n)^p} > \frac{1}{n}$ and divergence follows by comparison with the harmonic series. Now let $p > 0$. By Lemma 1.23 (b), $\log n < \log(n+1)$. Hence $(n(\log n)^p)$ increases and $1/(n(\log n))^p$ decreases; we can apply Proposition 2.23 to (2.14). This leads us to the series

$$\sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k(\log 2^k)^p} = \sum_{k=1}^{\infty} \frac{1}{(k \log 2)^p} = \frac{1}{(\log 2)^p} \sum_{k=1}^{\infty} \frac{1}{k^p},$$

and the assertion follows from example (a).

This procedure can evidently be continued. For instance $\sum_{n=3}^{\infty} \frac{1}{n \log n \log\log n}$ diverges, whereas $\sum_{n=3}^{\infty} \frac{1}{n \log n (\log\log n)^2}$ converges.

2.3.4 The Number e

Leonhard Euler (Basel 1707 – St. Petersburg 1783) was one of the greatest mathematicians. He made contributions to Number Theorie, Ordinary Differential Equations, Calculus of Variations, Astronomy, Mechanics. Fermat (1635) claimed that all numbers of the form $f_n = 2^{2^n}+1$, $n \in \mathbb{N}$, are prime numbers. This is obviously true for the first 5 numbers $(3, 5, 17, 257, 65537)$. Euler showed that $641 \mid 2^{32}+1$. In fact, it is open whether any other element f_n is a prime number. Euler showed that the equation $x^3 + y^3 = z^3$ has no solution in positive integers x, y, z. This is the special case of Fermat's last theorem. It is known that the limit $\gamma = \lim\limits_{n\to\infty}\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\log n\right)$ exists and gives a finite number γ, the so called *Euler constant*. It is not known whether γ is rational or not. Further, *Euler numbers* E_r play a role in calculating the series $\sum_n (-1)^n \frac{1}{(2n+1)^r}$. Soon, we will speak about the Euler formula $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$. More about live and work of famous mathematicians is to be found in
`www-history.mcs.st-andrews.ac.uk/`
We define

$$\mathrm{e} := \sum_{n=0}^{\infty} \frac{1}{n!}, \tag{2.15}$$

where $0! = 1! = 1$ by definition. Since

$$\begin{aligned} s_n &= 1+1+\frac{1}{1\cdot 2}+\frac{1}{1\cdot 2\cdot 3}+\cdots+\frac{1}{1\cdot 2\cdots n} \\ &< 1+1+\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^{n-1}} < 3, \end{aligned}$$

the series converges (by the comparing it with the geometric series with $q = \frac{1}{2}$) and the definition makes sense. In fact, the series converges very rapidly and allows us to compute e with great accuracy. It is of interest to note that e can also defined by means of another limit process. e is called the *Euler number*.

Proposition 2.24

$$\mathrm{e} = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n. \tag{2.16}$$

Proof. Let

$$s_n = \sum_{k=0}^{n}\frac{1}{k!}, \quad t_n = \left(1+\frac{1}{n}\right)^n.$$

By the binomial theorem,

$$\begin{aligned} t_n &= 1 + n\frac{1}{n} + \frac{n(n-1)}{2!}\cdot\frac{1}{n^2} + \frac{n(n-1)(n-2)}{3!}\cdot\frac{1}{n^3} + \cdots + \frac{n(n-1)\cdots 1}{n!}\cdot\frac{1}{n^n} \\ &= 1+1+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\cdots \\ &\qquad + \frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{n-1}{n}\right) \\ &\le 1+1+\frac{1}{2!}+\frac{1}{3!}+\cdots+\frac{1}{n!} \end{aligned}$$

Hence, $t_n \le s_n$, so that by Proposition 2.16

$$\overline{\lim_{n\to\infty}}\, t_n \le \overline{\lim_{n\to\infty}}\, s_n = \lim_{n\to\infty} s_n = \mathrm{e}. \tag{2.17}$$

Next if $n \ge m$,

$$t_n \ge 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \cdots + \frac{1}{m!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{m-1}{n}\right).$$

Let $n \to \infty$, keeping m fixed again by Proposition 2.16 we get

$$\underline{\lim_{n\to\infty}}\, t_n \ge 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{m!} = s_m.$$

Letting $m \to \infty$, we finally get

$$\mathrm{e} \le \underline{\lim_{n\to\infty}}\, t_n. \tag{2.18}$$

The proposition follows from (2.17) and (2.18). ∎

The rapidity with which the series $\sum 1/n!$ converges can be estimated as follows.

$$\begin{aligned}\mathrm{e} - s_n &= \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \cdots \\ &< \frac{1}{(n+1)!}\left[1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \cdots\right] = \frac{1}{(n+1)!}\cdot\frac{1}{1-\frac{1}{n+1}} = \frac{1}{n!\,n}\end{aligned}$$

so that

$$0 < \mathrm{e} - s_n < \frac{1}{n!\,n}. \tag{2.19}$$

We use the preceding inequality to compute e. For $n = 9$ we find

$$s_9 = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} + \frac{1}{5040} + \frac{1}{40,320} + \frac{1}{362,880} = 2.718281526... \tag{2.20}$$

By (2.19)

$$\mathrm{e} - s_9 < \frac{3.1}{10^7}$$

such that the first six digits of e in (2.20) are correct.

Example 2.11

(a) $$\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = \lim_{n\to\infty}\left(\frac{n-1}{n}\right)^n = \lim_{n\to\infty}\frac{1}{\left(\frac{n}{n-1}\right)^n} = \lim_{n\to\infty}\frac{1}{\left(1+\frac{1}{n-1}\right)^{n-1}\left(1+\frac{1}{n-1}\right)} = \frac{1}{\mathrm{e}}.$$

(b) $$\begin{aligned}\lim_{n\to\infty}\left(\frac{3n+1}{3n-1}\right)^{4n} &= \lim_{n\to\infty}\left(\frac{3n+1}{3n}\right)^{4n}\cdot\lim_{n\to\infty}\left(\frac{3n}{3n-1}\right)^{4n} \\ &= \lim_{n\to\infty}\left(\left(1+\frac{1}{3n}\right)^{3n}\right)^{\frac{4}{3}}\cdot\lim_{n\to\infty}\left(\left(1+\frac{1}{3n-1}\right)^{3n}\right)^{\frac{4}{3}} = \mathrm{e}^{\frac{8}{3}}.\end{aligned}$$

Proposition 2.25 e *is irrational.*

Proof. Suppose e is rational, say $\mathrm{e} = p/q$ with positive integers p and q. By (2.19)

$$0 < q!(\mathrm{e} - s_q) < \frac{1}{q}. \tag{2.21}$$

By our assumption, $q!\mathrm{e}$ is an integer. Since

$$q!s_q = q!\left(1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{q!}\right)$$

is also an integer, we see that $q!(\mathrm{e} - s_q)$ is an integer. Since $q \geq 1$, (2.21) implies the existence of an integer between 0 and 1 which is absurd. ∎

2.3.5 The Root and the Ratio Tests

Theorem 2.26 (Root Test) *Given* $\sum a_n$, *put* $\alpha = \overline{\lim_{n\to\infty}} \sqrt[n]{|\, a_n \,|}$.
Then

(a) *if* $\alpha < 1$, $\sum a_n$ *converges;*
(b) *if* $\alpha > 1$, $\sum a_n$ *diverges;*
(c) *if* $\alpha = 1$, *the test gives no information.*

Proof. (a) If $\alpha < 1$ choose β such that $\alpha < \beta < 1$, and an integer n_0 such that

$$\sqrt[n]{|\, a_n \,|} < \beta$$

for $n \geq n_0$ (such n_0 exists since α is the supremum of the limit set of $(\sqrt[n]{|\, a_n \,|})$). That is, $n \geq n_0$ implies

$$|\, a_n \,| < \beta^n.$$

Since $0 < \beta < 1$, $\sum \beta^n$ converges. Convergence of $\sum a_n$ now follows from the comparison test.
(b) If $\alpha > 1$ there is a subsequence (a_{n_k}) such that $\sqrt[n_k]{|\, a_{n_k} \,|} \to \alpha$. Hence $|\, a_n \,| > 1$ for infinitely many n, so that the necessary condition for convergence, $a_n \to 0$, fails.
To prove (c) consider the series $\sum \dfrac{1}{n}$ and $\sum \dfrac{1}{n^2}$. For each of the series $\alpha = 1$, but the first diverges, the second converges. ∎

Remark. (a) $\sum a_n$ converges, if there exists $q < 1$ such that $\sqrt[n]{|\, a_n \,|} \leq q$ for almost all n.
(b) $\sum a_n$ diverges if $|\, a_n \,| \geq 1$ for infinitely many n.

Theorem 2.27 (Ratio Test) *The series* $\sum a_n$

(a) *converges if* $\overline{\lim_{n\to\infty}} \left| \dfrac{a_{n+1}}{a_n} \right| < 1$,

(b) *diverges if* $\left| \dfrac{a_{n+1}}{a_n} \right| \geq 1$ *for all but finitely many* n.

In place of (b) one can also use the (weaker) statement
(b') $\sum a_n$ diverges if $\underline{\lim}_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| > 1$.

Indeed, if (b') is satisfied, almost all elements of the sequence $\left| \frac{a_{n+1}}{a_n} \right|$ are ≥ 1.

Corollary 2.28 *The series* $\sum a_n$

(a) *converges if* $\lim \left| \frac{a_{n+1}}{a_n} \right| < 1$,

(b) *diverges if* $\lim \left| \frac{a_{n+1}}{a_n} \right| > 1$.

Proof of Theorem 2.27. If condition (a) holds, we can find $\beta < 1$ and an integer m such that $n \geq m$ implies

$$\left| \frac{a_{n+1}}{a_n} \right| < \beta.$$

In particular,

$$\begin{aligned} &| a_{m+1} | < \beta | a_m |, \\ &| a_{m+2} | < \beta | a_{m+1} | < \beta^2 | a_m |, \\ &\qquad \cdots \\ &| a_{m+p} | < \beta^p | a_m |. \end{aligned}$$

That is,

$$| a_n | < \frac{| a_m |}{\beta^m} \beta^n$$

for $n \geq m$, and (a) follows from the comparison test, since $\sum \beta^n$ converges. If $| a_{n+1} | \geq | a_n |$ for $n \geq n_0$, it is seen that the condition $a_n \to 0$ does not hold, and (b) follows. ∎

Remark 2.2 Homework 7.5 shows that in (b) "all but finitely many" cannot be replaced by the weaker assumption "infinitely many."

Example 2.12 (a) The series $\sum_{n=0}^{\infty} n^2/2^n$ converges since, if $n \geq 3$,

$$\left| \frac{a_{n+1}}{a_n} \right| = \frac{(n+1)^2 2^n}{2^{n+1} n^2} = \frac{1}{2} \left(1 + \frac{1}{n} \right)^2 \leq \frac{1}{2} \left(1 + \frac{1}{3} \right)^2 = \frac{8}{9} < 1.$$

(b) Consider the series

$$\frac{1}{2} + 1 + \frac{1}{8} + \frac{1}{4} + \frac{1}{32} + \frac{1}{16} + \frac{1}{128} + \frac{1}{64} + \cdots$$
$$= \frac{1}{2^1} + \frac{1}{2^0} + \frac{1}{2^3} + \frac{1}{2^2} + \frac{1}{2^5} + \frac{1}{2^4} + \cdots$$

where $\underline{\lim}_{n\to\infty} \dfrac{a_{n+1}}{a_n} = \dfrac{1}{8}$, $\overline{\lim}_{n\to\infty} \dfrac{a_{n+1}}{a_n} = 2$, but $\lim \sqrt[n]{a_n} = \frac{1}{2}$. Indeed, $a_{2n} = 1/2^{2n-2}$ and $a_{2n+1} = 1/2^{2n+1}$ yields

$$\frac{a_{2n+1}}{a_{2n}} = \frac{1}{8}, \quad \frac{a_{2n}}{a_{2n-1}} = 2.$$

The root test indicates convergence; the ratio test does not apply.
(c) For $\sum \dfrac{1}{n}$ and $\sum \dfrac{1}{n^2}$ both the ratio and the root test do not apply since both (a_{n+1}/a_n) and $(\sqrt[n]{a_n})$ converge to 1.

The ratio test is frequently easier to apply than the root test. However, the root test has wider scope.

Remark 2.3 For any sequence (c_n) of positive real numbers,

$$\underline{\lim}_{n\to\infty} \frac{c_{n+1}}{c_n} \le \underline{\lim}_{n\to\infty} \sqrt[n]{c_n} \le \overline{\lim}_{n\to\infty} \sqrt[n]{c_n} \le \overline{\lim}_{n\to\infty} \frac{c_{n+1}}{c_n}.$$

For the proof, see [Rud76, 3.37 Theorem]. In particular, if $\lim \dfrac{c_{n+1}}{c_n}$ exists, then $\lim \sqrt[n]{c_n}$ also exists and both limits coincide.

Proposition 2.29 (Leibniz criterion) *Let $\sum b_n$ be an alternating serie, that is $\sum b_n = \sum(-1)^{n+1}a_n$ with a decreasing sequence of positive numbers $a_1 \ge a_2 \ge \cdots \ge 0$. If $\lim a_n = 0$ then $\sum b_n$ converges.*

Proof. The proof is quite the same as in Example 2.8 (b). We find for the partial sums s_n of $\sum b_n$

$$| s_n - s_m | \le a_{m+1}$$

if $n \ge m$. Since (a_n) tends to 0, the Cauchy criterion applies to (s_n). Hence, $\sum b_n$ is convergent. ∎

2.3.6 Absolute Convergence

The series $\sum a_n$ is said to *converge absolutely* if the series $\sum | a_n |$ converges.

Proposition 2.30 *If $\sum a_n$ converges absolutely, then $\sum a_n$ converges.*

Proof. The assertion follows from the inequality

$$\left| \sum_{k=m}^{n} a_k \right| \le \sum_{k=m}^{n} | a_k |$$

plus the Cauchy criterion. ∎

Remarks 2.4 For series with positive terms, absolute convergence is the same as convergence. If $\sum a_n$ converges but $\sum |\, a_n \,|$ diverges, we say that $\sum a_n$ *converges nonabsolutely*. For instance $\sum (-1)^{n+1}/n$ converges nonabsolutely. The comparison test as well as the root and the ratio tests, is really a test for absolute convergence and cannot give any information about nonabsolutely convergent series.
We shall see that we may operate with absolutely convergent series very much as with finite sums. We may multiply them, we may change the order in which the additions are carried out without effecting the sum of the series. But for nonabsolutely convergent sequences this is no longer true and more care has to be taken when dealing with them.

Without proof we mention the fact that one can multiply absolutely convergent series; for the proof, see [Rud76, Theorem 3.50].

Proposition 2.31 *If $\sum a_n$ converges absolutely with $\sum_{n=0}^{\infty} a_n = A$, $\sum b_n$ converges, $\sum_{n=0}^{\infty} b_n = B$, $c_n = \sum_{k=0}^{n} a_k b_{n-k}$, $n \in \mathbb{Z}_+$. Then $\sum_{n=0}^{\infty} c_n = AB$.*

2.3.7 Decimal Expansion of Real Numbers

Proposition 2.32 (a) *Let α be a real number with $0 \le \alpha < 1$. Then there exists a sequence (a_n), $a_n \in \{0, 1, 2, \ldots, 9\}$ such that*

$$\alpha = \sum_{n=1}^{\infty} a_n \, 10^{-n}. \tag{2.22}$$

The sequence (a_n) is called a decimal expansion *of α.*
(b) Given a sequence (a_k), $a_k \in \{0, 1, \ldots, 9\}$, then there exists a real number $\alpha \in [0, 1]$ such that

$$\alpha = \sum_{n=1}^{\infty} a_n \, 10^{-n}.$$

Proof. (b) Comparison with the geometric series yields

$$\sum_{n=1}^{\infty} a_n 10^{-n} \le 9 \sum_{n=1}^{\infty} 10^{-n} = \frac{9}{10} \cdot \frac{1}{1 - 1/10} = 1.$$

Hence the series $\sum_{n=1}^{\infty} a_n 10^{-n}$ converges to some $\alpha \in [0, 1]$.
(a) Given $\alpha \in [0, 1)$ we use induction to construct a sequence (a_n) with (2.22) and

$$s_n \le \alpha < s_n + 10^{-n}, \quad \text{where} \quad s_n = \sum_{k=1}^{n} a_k \, 10^{-k}.$$

First, cut $[0, 1]$ into 10 pieces $I_j := [j/10, (j + 1)/10)$, $j = 0, \ldots, 9$, of equal length. If $\alpha \in I_j$, put $a_1 := j$. Then,

$$s_1 = \frac{a_1}{10} \le \alpha < s_1 + \frac{1}{10}.$$

Suppose $a_1, \ldots, a_n$ are already constructed and

$$s_n \le \alpha < s_n + 10^{-n}.$$

Consider the intervals $I_j := [s_n + j/10^{n+1}, s_n + (j+1)/10^{n+1})$, $j = 0, \ldots, 9$. There is exactly one j such that $\alpha \in I_j$. Put $a_{n+1} := j$, then

$$s_n + \frac{a_{n+1}}{10^{n+1}} \le \alpha < s_n + \frac{a_{n+1}+1}{10^{n+1}}$$
$$s_{n+1} \le \alpha < s_{n+1} + 10^{-n-1}.$$

The induction step is complete. By construction $|\, \alpha - s_n \,| < 10^{-n}$, that is, $\lim s_n = \alpha$. ∎

Remarks 2.5 (a) The proof shows that any real number $\alpha \in [0, 1)$ can be approximated by rational numbers.
(b) The construction avoids decimal expansion of the form $\alpha = \ldots a9999\ldots$, $a < 9$, and gives instead $\alpha = \ldots (a+1)000\ldots$. It gives a bijective correspondence between the real numbers of the interval $[0, 1)$ and the sequences (a_n), $a_n \in \{0, 1, \ldots, 9\}$, not ending with nines. However, the sequence $(a_n) = (0, 1, 9, 9, \cdots)$ corresponds to the real number 0.02.
(c) It is not difficult to see that $\alpha \in [0, 1)$ is rational if and only if there exist positive integers n_0 and p such that $n \ge n_0$ implies $a_n = a_{n+p}$—the decimal expansion is *periodic* from n_0 on.

2.3.8 Complex Sequences and Series

Almost all notions and theorems carry over from real sequences to complex sequences. For example

> A sequence (z_n) of complex numbers *converges to* z if for every (real) $\varepsilon > 0$ there exists a positive integer $n_0 \in \mathbb{N}$ such that $n \ge n_0$ implies
>
> $$|\, z - z_n \,| < \varepsilon.$$

The following proposition shows that convergence of a complex sequence can be reduced to the convergence of two real sequences.

Proposition 2.33 *The complex sequence (z_n) converges to some complex number z if and only if the real sequences $(\operatorname{Re} z_n)$ converges to $\operatorname{Re} z$ and the real sequence $(\operatorname{Im} z_n)$ converges to $\operatorname{Im} z$.*

Proof. Using the (complex) limit law $\lim(z_n + c) = c + \lim z_n$ it is easy to see that we can restrict ourselves to the case $z = 0$. Suppose first $z_n \to 0$. Proposition 1.20 (d) gives $|\operatorname{Re} z_n| \le |z_n|$. Hence $\operatorname{Re} z_n$ tends to 0 as $n \to \infty$. Similarly, $|\operatorname{Im} z_n| \le |z_n|$ and therefore $\operatorname{Im} z_n \to 0$.
Suppose now $x_n := \operatorname{Re} z_n \to 0$ and $y_n := \operatorname{Im} z_n \to 0$ as n goes to infinity. Since

$|\,z_n\,|^2 = x_n^2 + y_n^2$, $|\,z_n\,|^2 \to 0$ as $n \to \infty$; this implies $z_n \to 0$. ∎

Since the complex field $\mathbb{C}$ is not an ordered field, all notions and propositions where the *order* is involved do not make sense for complex series or they need modifications. The sandwich theorem does not hold; there is no notion of monotonic sequences, upper and lower limits. But still there are bounded sequences ($|\,z_n\,| \le C$), limit points, subsequences, Cauchy sequences, series, and absolute convergence. The following theorems are true for complex sequences, too:

Proposition/Lemma/Theorem 1, 2, 3, 9, 10, 12, 15, 17, 18

The Bolzano–Weierstraß Theorem for bounded complex sequences (z_n) can be proved by considering the real and the imaginary sequences $(\,\mathrm{Re}\,z_n)$ and $(\,\mathrm{Im}\,z_n)$ separately.
The comparison test for series now reads:

(a) If $|\,a_n\,| \le C\,|\,b_n\,|$ for some $C > 0$ and for almost all $n \in \mathbb{N}$, and if $\sum |\,b_n\,|$ converges, then $\sum a_n$ converges.
(b) If $|\,a_n\,| \ge C\,|\,d_n\,|$ for some $C > 0$ and for almost all n, and if $\sum |\,d_n\,|$ diverges, then $\sum a_n$ diverges.

The Cauchy criterion, the root, and the ratio tests are true for complex series as well. Propositions 19, 20, 26, 27, 28, 30, 31 are true for complex series.

2.3.9 Power Series

Definition 2.10 Given a sequence (c_n) of complex numbers, the series

$$\sum_{n=0}^{\infty} c_n\, z^n \tag{2.23}$$

is called a *power series*. The numbers c_n are called the *coefficients* of the series; z is a complex number.

In general, the series will converge or diverge, depending on the choice of z. More precisely, with every power series there is associated a circle with center 0, the *circle of convergence*, such that (2.23) converges if z is in the interior of the circle and diverges if z is in the exterior. The radius R of this disc of convergence is called the *radius of convergence*.
On the disc of convergence, a power series defines a function since it associates to each z with $|\,z\,| < R$ a complex number, namely the sum of the numerical series $\sum_n c_n z^n$. For example, $\sum_{n=0}^{\infty} z^n$ defines the function $f(z) = \frac{1}{1-z}$ for $|\,z\,| < 1$. If almost all coefficients c_n are 0, say $c_n = 0$ for all $n \ge m+1$, the power series is a finite sum and the corresponding function is a polynomial: $\sum_{n=0}^{\infty} c_n z^n = \sum_{n=0}^{m} c_n z^n = c_0 + c_1 z + c_2 z^2 + \cdots + c_m z^m$.

Theorem 2.34 *Given a power series $\sum c_n z^n$, put*

$$\alpha = \overline{\lim_{n\to\infty}} \sqrt[n]{|\,c_n\,|}, \quad R = \frac{1}{\alpha}. \tag{2.24}$$

If $\alpha = 0$, $R = +\infty$; if $\alpha = +\infty$, $R = 0$. Then $\sum c_n z^n$ converges if $|\,z\,| < R$, and diverges if $|\,z\,| > R$.

The behavior on the circle of convergence cannot be described so simple.
Proof. Put $a_n = c_n z^n$ and apply the root test:

$$\overline{\lim_{n\to\infty}} \sqrt[n]{|\, a_n \,|} = |\, z \,| \overline{\lim_{n\to\infty}} \sqrt[n]{|\, c_n \,|} = \frac{|\, z \,|}{R}.$$

This gives convergence if $|\, z \,| < R$ and divergence if $|\, z \,| > R$. ∎

The nonnegative number R is called the *radius of convergence*.

Example 2.13 (a) The series $\sum_{n=0}^{m} c_n z^m$ has $c_n = 0$ for almost all n. Hence $\alpha = \lim_{n\to\infty} \sqrt[n]{|\, c_n \,|} = \lim_{n\to\infty} 0 = 0$ and $R = +\infty$.
(b) The series $\sum n^n z^n$ has $R = 0$.
(c) The series $\displaystyle\sum \frac{z^n}{n!}$ has $R = +\infty$. (In this case the ratio test is easier to apply than the root test. Indeed,

$$\alpha = \lim_{n\to\infty} \left| \frac{c_{n+1}}{c_n} \right| = \lim_{n\to\infty} \frac{n!}{(n+1)!} = \lim_{n\to\infty} \frac{1}{n+1} = 0,$$

and therefore $R = +\infty$.)
(d) The series $\sum z^n$ has $R = 1$. If $|\, z \,| = 1$ diverges since (z^n) does not tend to 0. This generalizes the geometric series; formula (2.9) still holds if $|\, q \,| < 1$:

$$\sum_{n=2}^{\infty} 2 \left(\frac{\mathrm{i}}{3} \right)^n = \frac{2(\mathrm{i}/3)^2}{1 - \mathrm{i}/3} = -\frac{3 + \mathrm{i}}{15}.$$

(e) The series $\sum z^n/n$ has $R = 1$. It diverges if $z = 1$. It converges for all other z with $|\, z \,| = 1$ (without proof).
(f) The series $\sum z^n/n^2$ has $R = 1$. It converges for all z with $|\, z \,| = 1$ by the comparison test, since $|\, z^n/n^2 \,| = 1/n^2$.

2.3.10 Rearrangements

The generalized associative law for finite sums says that we can insert brackets without effecting the sum, for example, $((a_1 + a_2) + (a_3 + a_4)) = (a_1 + (a_2 + (a_3 + a_4)))$. We will see that a similar statement holds for series:
Suppose that $\sum_k a_k$ is a converging series and $\sum_l b_l$ is a sum obtained from $\sum_k a_k$ by "inserting brackets", for example

$$b_1 + b_2 + b_3 + \cdots = \underbrace{(a_1 + a_2)}_{b_1} + \underbrace{(a_3 + \cdots + a_{10})}_{b_2} + \underbrace{(a_{11} + a_{12})}_{b_3} + \cdots$$

Then $\sum_l b_l$ converges and the sum is the same. If $\sum_k a_k$ diverges to $+\infty$, the same is true for $\sum_l b_l$. However, divergence of $\sum a_k$ does not imply divergence of $\sum b_l$ in general, since $1 - 1 + 1 - 1 + 1 - 1 + \cdots$ diverges but $(1 - 1) + (1 - 1) + \cdots$ converges. For the proof let $s_n = \sum_{k=1}^{n}$ and $t_m = \sum_{l=1}^{m} b_l$. By construction, $t_m = s_{n_m}$ for a suitable subsequence (s_{n_m})

of the partial sums of $\sum_k a_k$. Convergence (proper or improper) of (s_n) implies convergence (proper or improper) of any subsequence. Hence, $\sum_l b_l$ converges.
For finite sums, the generalized commutative law holds:

$$a_1 + a_2 + a_3 + a_4 = a_2 + a_4 + a_1 + a_3;$$

that is, any rearrangement of the summands does not effect the sum. We will see in Example 2.14 below that this is not true for arbitrary series but for absolutely converging ones, (see Proposition 2.36 below).

Definition 2.11 Let $\sigma\colon \mathbb{N} \to \mathbb{N}$ be a bijective mapping, that is in the sequence $(\sigma(1), \sigma(2), \dots)$ every positive integer appears once and only once. Putting

$$a'_n = a_{\sigma(n)}, \quad (n = 1, 2, \dots),$$

we say that $\sum a'_n$ is a *rearrangement* of $\sum a_n$.

If (s_n) and (s'_n) are the partial sums of $\sum a_n$ and a rearrangement $\sum a'_n$ of $\sum a_n$, it is easily seen that, in general, these two sequences consist of entirely different numbers. We are led to the problem of determining under what conditions all rearrangements of a convergent series will converge and whether the sums are necessarily the same.

Example 2.14 (a) Consider the convergent series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - + \cdots \tag{2.25}$$

and one of its rearrangements

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \cdots . \tag{2.26}$$

If s is the sum of (2.25) then $s > 0$ since

$$\left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots > 0.$$

We will show that (2.26) converges to $s' = s/2$. Namely

$$\begin{aligned} s' = \sum a'_n &= \left(1 - \frac{1}{2}\right) - \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{6}\right) - \frac{1}{8} + \left(\frac{1}{5} - \frac{1}{10}\right) - \frac{1}{12} + \cdots \\ &= \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \frac{1}{12} + \cdots \\ &= \frac{1}{2}\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + - \cdots\right) = \frac{1}{2}s \end{aligned}$$

Since $s \neq 0$, $s' \neq s$. Hence, there exist rearrangements which converge; however to a different limit.

(b) Consider the following rearrangement of the series (2.25)

$$\begin{aligned}\sum a'_n = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \\ + \left(\frac{1}{5} + \frac{1}{7}\right) - \frac{1}{6} + \\ + \left(\frac{1}{9} + \frac{1}{11} + \frac{1}{13} + \frac{1}{15}\right) - \frac{1}{8} \\ + \cdots \\ + \left(\frac{1}{2^n+1} + \frac{1}{2^n+3} + \cdots + \frac{1}{2^{n+1}-1}\right) - \frac{1}{2n+2} + \cdots\end{aligned}$$

Since for every positive integer $n \geq 10$

$$\left(\frac{1}{2^n+1} + \frac{1}{2^n+3} + \cdots + \frac{1}{2^{n+1}-1}\right) - \frac{1}{2n+2} > 2^{n-1} \cdot \frac{1}{2^{n+1}} - \frac{1}{2n+2} > \frac{1}{4} - \frac{1}{2n+2} > \frac{1}{5}$$

the rearranged series diverges to $+\infty$.

Without proof (see [Rud76, 3.54 Theorem]) we remark the following surprising theorem. It shows (together with the Proposition 2.36) that the absolute convergence of a series is necessary and sufficient for every rearrangement to be convergent (to the same limit).

Proposition 2.35 *Let $\sum a_n$ be a series of real numbers which converges, but not absolutely. Suppose $-\infty \leq \alpha \leq \beta \leq +\infty$. Then there exists a rearrangement $\sum a'_n$ with partial sums s'_n such that*

$$\underline{\lim_{n\to\infty}}\, s'_n = \alpha, \quad \overline{\lim_{n\to\infty}}\, s'_n = \beta.$$

Proposition 2.36 *If $\sum a_n$ is a series of complex numbers which converges absolutely, then every rearrangement of $\sum a_n$ converges, and they all converge to the same sum.*

Proof. Let $\sum a'_n$ be a rearrangement with partial sums s'_n. Given $\varepsilon > 0$, by the Cauchy criterion for the series $\sum |\, a_n \,|$ there exists $n_0 \in \mathbb{N}$ such that $n \geq m \geq n_0$ implies

$$\sum_{k=m}^{n} |\, a_k \,| < \varepsilon. \tag{2.27}$$

Now choose p such that the integers $1, 2, \ldots, n_0$ are all contained in the set $\sigma(1), \sigma(2), \ldots, \sigma(p)$.

$$\{1, 2, \ldots, n_0\} \subseteq \{\sigma(1), \sigma(2), \ldots, \sigma(p)\}.$$

Then, if $n \geq p$, the numbers $a_1,\ a_2, \ldots, a_{n_0}$ will cancel in the difference $s_n - s'_n$, so that

$$|\, s_n - s'_n \,| = \left| \sum_{k=1}^{n} a_k - \sum_{k=1}^{n} a_{\sigma(k)} \right| \leq \left| \sum_{k=n_0+1}^{n} \pm a_k \right| \leq \sum_{k=n_0+1}^{n} |\, a_k \,| < \varepsilon,$$

by (2.27). Hence (s'_n) converges to the same sum as (s_n).
The same argument shows that $\sum a'_n$ also absolutely converges. ∎

2.3.11 Products of Series

If we multiply two finite sums $a_1 + a_2 + \cdots + a_n$ and $b_1 + b_2 + \cdots + b_m$ by the distributive law, we form all products $a_i b_j$ put them into a sequence $p_0, p_1, \cdots, p_s$, $s = mn$, and add up $p_0 + p_1 + p_2 + \cdots + p_s$. This method can be generalized to series $a_0 + a_1 + \cdots$ and $b_0 + b_1 + \cdots$. Surely, we can form all products $a_i b_j$, we can arrange them in a sequence $p_0, p_1, p_2, \cdots$ and form the *product series* $p_0 + p_1 + \cdots$. For example, consider the table

$$\begin{array}{cccc} a_0b_0 & a_0b_1 & a_0b_2 & \cdots \\ a_1b_0 & a_1b_1 & a_1b_2 & \cdots \\ a_2b_0 & a_2b_1 & a_2b_2 & \cdots \\ \vdots & \vdots & \vdots & \end{array} \qquad \begin{array}{cccc} p_0 & p_1 & p_3 & \cdots \\ p_2 & p_4 & p_7 & \cdots \\ p_5 & p_8 & p_{12} & \cdots \\ \vdots & \vdots & \vdots & \end{array}$$

and the diagonal enumeration of the products. The question is: under which conditions on $\sum a_n$ and $\sum b_n$ the product series converges and its sum does not depend on the arrangement of the products $a_i b_k$.

Proposition 2.37 *If both series $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=0}^{\infty} b_k$ converge absolutely with $A = \sum_{k=0}^{\infty} a_k$ and $B = \sum_{k=0}^{\infty} b_k$, then any of their product series $\sum p_k$ converges absolutely and $\sum_{k=0}^{\infty} p_k = AB$.*

Proof. For the nth partial sum of any product series $\sum_{k=0}^{n} |\, p_k \,|$ we have

$$|\, p_0 \,| + |\, p_1 \,| + \cdots + |\, p_n \,| \le (|\, a_0 \,| + \cdots + |\, a_m \,|) \cdot (|\, b_0 \,| + \cdots + |\, b_m \,|),$$

if m is sufficiently large. More than ever,

$$|\, p_0 \,| + |\, p_1 \,| + \cdots + |\, p_n \,| \le \sum_{k=0}^{\infty} |\, a_k \,| \cdot \sum_{k=0}^{\infty} |\, b_k \,|.$$

That is, any series $\sum_{k=0}^{\infty} |\, p_k \,|$ is bounded and hence convergent by Lemma 2.19 (c). By Proposition 2.36 *all* product series converge to the same sum $s = \sum_{k=0}^{\infty} p_k$. Consider now the very special product series $\sum_{k=1}^{\infty} q_n$ with partial sums consisting of the sum of the elements in the upper left square. Then

$$q_1 + q_2 + \cdots + q_{(n+1)^2} = (a_0 + a_1 + \cdots + a_n)(b_0 + \cdots + b_n).$$

converges to $s = AB$. ∎

Arranging the elements $a_i b_j$ as above in a diagonal array and summing up the elements on the nth diagonal $c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$, we obtain the *Cauchy product*

$$\sum_{n=0}^{\infty} c_n = \sum_{n=0}^{\infty} (a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0).$$

Corollary 2.38 *If both series $\sum_{k=0}^{\infty} a_k$ and $\sum_{k=0}^{\infty} b_k$ converge absolutely with $A = \sum_{k=0}^{\infty} a_k$ and $B = \sum_{k=0}^{\infty} b_k$, their Cauchy product $\sum_{k=0}^{\infty} c_k$ converges absolutely and $\sum_{k=0}^{\infty} c_k = AB$.*

Example 2.15 We compute the Cauchy product of two geometric series:

$$\begin{aligned}(1+p+p^2+\cdots)(1+q+q^2+\cdots) &= 1+(p+q)+(p^2+pq+q^2)+\\ &\quad +(p^3+p^2q+pq^2+q^2)+\cdots\\ &= \frac{p-q}{p-q}+\frac{p^2-q^2}{p-q}+\frac{p^3-q^3}{p-q}+\cdots = \frac{1}{p-q}\sum_{n=1}^{\infty}(p^n-q^n)\\ &\underset{|p|<1,|q|<1}{=} \frac{1}{p-q}\frac{p}{1-p}-\frac{q}{1-q} = \frac{1}{p-q}\frac{p(1-q)-q(1-p)}{(1-p)(1-q)} = \frac{1}{1-p}\cdot\frac{1}{1-q}.\end{aligned}$$

Cauchy Product of Power Series

In case of power series the Cauchy product is appropriate since it is again a power series (which is not the case for other types of product series). Indeed, the Cauchy product of $\sum_{k=0}^{\infty} a_k z^k$ and $\sum_{k=0}^{\infty} b_k z^k$ is given by the general element

$$\sum_{k=0}^{n} a_k z^k b_{n-k} z^{n-k} = z^n \sum_{k=0}^{n} a_k b_{n-k},$$

such that

$$\sum_{k=0}^{\infty} a_k z^k \cdot \sum_{k=0}^{\infty} b_k z^k = \sum_{n=0}^{\infty}(a_0 b_n + \cdots + a_n b_0) z^n.$$

Corollary 2.39 *Suppose that $\sum_n a_n z^n$ and $\sum_n b_n z^n$ are power series with positive radius of convergence R_1 and R_2, respectively. Let $R = \min\{R_1, R_2\}$. Then the Cauchy product $\sum_{n=0}^{\infty} c_n z^n$, $c_n = a_0 b_n + \cdots + a_n b_0$, converges absolutely for $|z| < R$ and*

$$\sum_{n=0}^{\infty} a_n z^n \cdot \sum_{n=0}^{\infty} b_n z^n = \sum_{n=0}^{\infty} c_n z^n, \quad |z| < R.$$

This follows from the previous corollary and the fact that both series converge absolutely for $|z| < R$.

Example 2.16 (a)

$$\sum_{n=0}^{\infty}(n+1)z^n = \frac{1}{(1-z)^2}, \quad |z| < 1.$$

Indeed, consider the Cauchy product of $\sum_{n=0}^{\infty} z^n = \frac{1}{1-z}$, $|z| < 1$ with itself. Since $a_n = b_n = 1$, $c_n = \sum_{k=0}^{n} a_k b_{n-k} = \sum_{k=0}^{n} 1 \cdot 1 = n+1$, the claim follows.
(b)

$$\begin{aligned}(z+z^2)\sum_{n=0}^{\infty} z^n &= \sum_{n=0}^{\infty}(z^{n+1}+z^{n+2}) = \sum_{n=1}^{\infty} z^n + \sum_{n=2}^{\infty} z^n =\\ &= z + 2\sum_{n=2}^{\infty} z^n = z + 2z^2 + 2z^3 + 2z^4 + \cdots = z + \frac{2z^2}{1-z} = \frac{z+z^2}{1-z}.\end{aligned}$$

Chapter 3

Functions and Continuity

This chapter is devoted to another central notion in analysis—the notion of a continuous function. We will see that sums, product, quotients, and compositions of continuous functions are continuous. If nothing is specified otherwise D will denote a finite union of intervals.

Definition 3.1 Let $D \subset \mathbb{R}$ be a subset of $\mathbb{R}$. A *function* is a map $f\colon D \to \mathbb{R}$.
(a) The set D is called the *domain* of f; we write $D = D(f)$.
(b) If $A \subseteq D$, $f(A) := \{f(x) \mid x \in A\}$ is called the *image* of A under f.
The function $f\restriction A\colon A \to \mathbb{R}$ given by $f\restriction A(a) = f(a)$, $a \in A$, is called the *restriction* of f to A.
(c) If $B \subset \mathbb{R}$, we call $f^{-1}(B) := \{x \in D \mid f(x) \in B\}$ the *preimage* of B under f.
(d) The *graph* of f is the set $\operatorname{graph}(f) := \{(x, f(x)) \mid x \in D\}$.

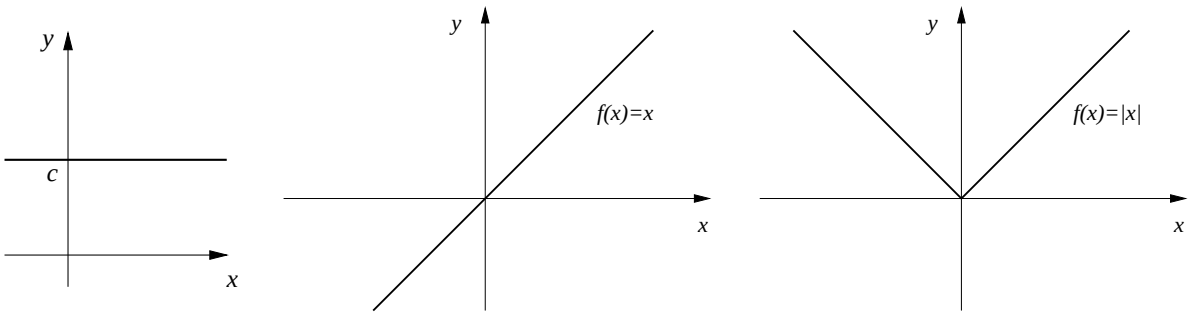
Later we will consider functions in a wider sense: From the complex numbers into complex numbers and from $\mathbb{F}^n$ into $\mathbb{F}^m$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$.

We say that a function $f\colon D \to \mathbb{R}$ is *bounded*, if $f(D) \subset \mathbb{R}$ is a bounded set of real numbers, i. e. there is a $C > 0$ such that $|\, f(x)\,| \leq C$ for all $x \in D$. We say that f is *bounded above* (resp. *bounded below*) if there exists $C \in R$ such that $f(x) < C$ (resp. $f(x) > C$) for all x in the domain of f.

Example 3.1 (a) Power series (with radius of convergence $R > 0$), polynomials and rational functions are the most important examples of functions.
Let $c \in \mathbb{R}$. Then $f(x) = c$, $f\colon \mathbb{R} \to \mathbb{R}$, is called the *constant* function.

(b) Properties of the functions change drastically if we change the domain or the image set. Let $f\colon \mathbb{R} \to \mathbb{R}$, $g\colon \mathbb{R} \to \mathbb{R}_+$, $k\colon \mathbb{R}_+ \to \mathbb{R}$, $h\colon \mathbb{R}_+ \to \mathbb{R}_+$ function given by $x \mapsto x^2$. g is surjective, k is injective, h is bijective, f is neither injective nor surjective. Obviously, $f\restriction\mathbb{R}_+ = k$ and $g\restriction\mathbb{R}_+ = h$.

(c) Let $f(x) = \sum_{n=0}^{\infty} x^n$, $f\colon (-1,1) \to \mathbb{R}$ and $h(x) = \frac{1}{1-x}$, $h\colon \mathbb{R}\setminus\{1\} \to \mathbb{R}$. Then $h\restriction(-1,1) = f$.



The graphs of the constant, the identity, and absolute value functions.

3.1 Limits of a Function

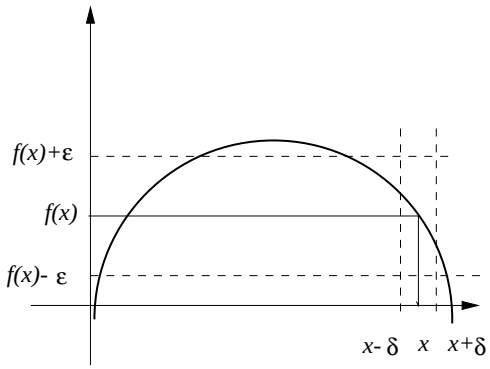
Definition 3.2 (ε-δ-definition) Let (a, b) a finite or infinite interval and $x_0 \in (a, b)$. Let $f\colon (a, b) \setminus \{x_0\} \to \mathbb{R}$ be a real-valued function. We call $A \in \mathbb{R}$ the *limit of f in x_0* ("The limit of $f(x)$ is A as x approaches x_0"; "f approaches A near x_0") if the following is satisfied

> For any $\varepsilon > 0$ there exists $\delta > 0$ such that $x \in (a, b)$ and $0 < |\, x - x_0 \,| < \delta$ imply $|\, f(x) - A \,| < \varepsilon$.

We write

$$\lim_{x \to x_0} f(x) = A.$$

Roughly speaking, if x is close to x_0, $f(x)$ must be closed to A.



Using quantifiers $\lim_{x \to x_0} f(x) = A$ reads as

$$\forall\, \varepsilon > 0 \;\; \exists\, \delta > 0 \;\; \forall\, x \in (a, b) : \quad 0 < |\, x - x_0 \,| < \delta \implies |\, f(x) - A \,| < \varepsilon.$$

Note that the formal negation of $\lim_{x \to x_0} f(x) = A$ is

$$\exists\, \varepsilon > 0 \;\; \forall\, \delta > 0 \;\; \exists\, x \in (a, b) : \quad 0 < |\, x - x_0 \,| < \delta \quad \text{and} \quad |\, f(x) - A \,| \geq \varepsilon.$$

Proposition 3.1 (sequences definition) *Let f and x_0 be as above. Then $\lim\limits_{x\to x_0} f(x) = A$ if and only if for every sequence (x_n) with $x_n \in (a,b)$, $x_n \neq x_0$ for all n, and $\lim\limits_{n\to\infty} x_n = x_0$ we have $\lim\limits_{n\to\infty} f(x_n) = A$.*

Proof. Suppose $\lim_{x\to x_0} f(x) = A$, and $x_n \to x_0$ where $x_n \neq x_0$ for all n. Given $\varepsilon > 0$ we find $\delta > 0$ such that $|\, f(x) - A \,| < \varepsilon$ if $0 < |\, x - x_0 \,| < \delta$. Since $x_n \to x_0$, there is a positive integer n_0 such that $n \geq n_0$ implies $|\, x_n - x_0 \,| < \delta$. Therefore $n \geq n_0$ implies $|\, f(x_n) - A \,| < \varepsilon$. That is, $\lim_{n\to\infty} f(x_n) = A$.
Suppose to the contrary that the condition of the proposition is fulfilled but $\lim_{x\to x_0} f(x) \neq A$. Then there is some $\varepsilon > 0$ such that for all $\delta = 1/n$, $n \in \mathbb{N}$, there is an $x_n \in (a,b)$ such that $0 < |\, x_n - x_0 \,| < 1/n$, but $|\, f(x_n) - A \,| \geq \varepsilon$. We have constructed a sequence (x_n), $x_n \neq x_0$ and $x_n \to x_0$ as $n \to \infty$ such that $\lim_{n\to\infty} f(x_n) \neq A$ which contradicts our assumption. Hence $\lim\limits_{x\to x_0} f(x) = A$. ∎

Example. $\lim_{x\to 1} x + 3 = 4$. Indeed, given $\varepsilon > 0$ choose $\delta = \varepsilon$. Then $|\, x - 1 \,| < \delta$ implies $|\, (x+3) - 4 \,| < \delta = \varepsilon$.

3.1.1 One-sided Limits, Infinite Limits, and Limits at Infinity

Definition 3.3 (a) We are writing

$$\lim_{x\to x_0+0} f(x) = A$$

if for all sequences (x_n) with $x_n > x_0$ and $\lim\limits_{n\to\infty} x_n = x_0$, we have $\lim\limits_{n\to\infty} f(x_n) = A$. Sometimes we use the notation $f(x_0 + 0)$ in place of $\lim\limits_{x\to x_0+0} f(x)$. We call $f(x_0 + 0)$ the *right-hand limit* of f at x_0 or we say "A is the limit of f as x approaches x_0 from above (from the right)."
Similarly one defines the *left-hand limit* of f at x_0, $\lim\limits_{x\to x_0-0} f(x) = A$ with $x_n < x_0$ in place of $x_n > x_0$. Sometimes we use the notation $f(x_0 - 0)$.
(b) We are writing

$$\lim_{x\to+\infty} f(x) = A$$

if for all sequences (x_n) with $\lim\limits_{n\to\infty} x_n = +\infty$ we have $\lim\limits_{n\to\infty} f(x_n) = A$. Sometimes we use the notation $f(+\infty)$. In a similar way we define $\lim\limits_{x\to-\infty} f(x) = A$.
(c) Finally, the notions of (a), (b), and Definition 3.2 still make sense in case $A = +\infty$ and $A = -\infty$. For example,

$$\lim_{x\to x_0-0} f(x) = -\infty$$

if for all sequences (x_n) with $x_n < x_0$ and $\lim\limits_{n\to\infty} x_n = x_0$ we have $\lim\limits_{n\to\infty} f(x_n) = -\infty$.

Remark 3.1 All notions in the above definition can be given in ε-δ or ε-D or E-δ or E-D languages using inequalities. For example, $\lim\limits_{x\to x_0-0} f(x) = -\infty$ if and only if

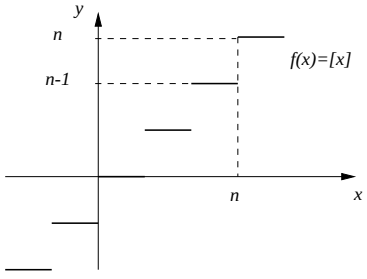
$$\forall\, E > 0\ \ \exists\, \delta > 0\ \ \forall\, x \in D(f) : 0 < x_0 - x < \delta \implies f(x) < -E.$$

For example, we show that $\lim_{x\to 0-0} \frac{1}{x} = -\infty$. To $E > 0$ choose $\delta = \frac{1}{E}$. Then $0 < -x < \delta = \frac{1}{E}$ implies $0 < E < -\frac{1}{x}$ and hence $f(x) < -E$. This proves the claim.
Similarly, $\lim\limits_{x\to+\infty} f(x) = +\infty$ if and only if

$$\forall\, E > 0\ \exists\, D > 0\ \forall\, x \in D(f) : x > D \implies f(x) > E.$$

The proves of equivalence of ε-δ definitions and sequence definitions are along the lines of Proposition 3.1.
For example, $\lim_{x\to+\infty} x^2 = +\infty$. To $E > 0$ choose $D = \sqrt{E}$. Then $x > D$ implies $x > \sqrt{E}$; thus $x^2 > E$.

Example 3.2 (a) $\lim\limits_{x\to+\infty} \dfrac{1}{x} = 0$. For, let $\varepsilon > 0$; choose $D = 1/\varepsilon$. Then $x > D = 1/\varepsilon$ implies $0 < 1/x < \varepsilon$. This proves the claim.
(b) Consider the entier function $f(x) = [x]$, defined in Example 2.6 (b). If $n \in \mathbb{Z}$, $\lim\limits_{x\to n-0} f(x) = n - 1$ whereas $\lim\limits_{x\to n+0} f(x) = n$.



Proof. We use the ε-δ definition of the one-sided limits to prove the first claim. Let $\varepsilon > 0$. Choose $\delta = \frac{1}{2}$ then $0 < n - x < \frac{1}{2}$ implies $n - \frac{1}{2} < x < n$ and therefore $f(x) = n - 1$. In particular $|\, f(x) - (n-1)\,| = 0 < \varepsilon$. Similarly one proves $\lim\limits_{x\to n+0} f(x) = n$. ∎

Since the one-sided limits are different, $\lim\limits_{x\to n} f(x)$ does not exist.

Definition 3.4 Suppose we are given two functions f and g, both defined on $(a, b) \setminus \{x_0\}$. By $f + g$ we mean the function which assigns to each point $x \neq x_0$ of (a, b) the number $f(x) + g(x)$. Similarly, we define the difference $f - g$, the product fg, and the quotient f/g, with the understanding that the quotient is defined only at those points x at which $g(x) \neq 0$.

Proposition 3.2 *Suppose that f and g are functions defined on $(a, b) \setminus \{x_0\}$, $a < x_0 < b$, and $\lim_{x\to x_0} f(x) = A$, $\lim_{x\to x_0} g(x) = B$, $\alpha, \beta \in \mathbb{R}$. Then*
(a) $\lim\limits_{x\to x_0} f(x) = A'$ *implies* $A' = A$.
(b) $\lim\limits_{x\to x_0} (\alpha f + \beta g)(x) = \alpha A + \beta B$;
(c) $\lim\limits_{x\to x_0} (fg)(x) = AB$;
(d) $\lim\limits_{x\to x_0} \dfrac{f}{g}(x) = \dfrac{A}{B}$, *if* $B \neq 0$.
(e) $\lim\limits_{x\to x_0} |\, f(x)\,| = |\, A\,|$.

Proof. In view of Proposition 3.1, all these assertions follow immediately from the analogous properties of sequences, see Proposition 2.3. As an example, we show (c). Let (x_n) , $x_n \neq x_0$, be a sequence tending to x_0. By assumption, $\lim_{n\to\infty} f(x_n) = A$ and $\lim_{n\to\infty} g(x_n) = B$. By the Propostition 2.3 $\lim_{n\to\infty} f(x_n)g(x_n) = AB$, that is, $\lim_{n\to\infty}(fg)(x_n) = AB$. By

Proposition 3.1, $\lim_{x\to x_0} fg(x) = AB$. ∎

Remark 3.2 The proposition remains true if we replace (at the same time in all places) $x \to x_0$ by $x \to x_0 + 0$, $x \to x_0 - 0$, $x \to +\infty$, or $x \to -\infty$. Moreover we can replace A or B by $+\infty$ or by $-\infty$ provided the right members of (b), (c), (d) and (e) are defined.
Note that $+\infty + (-\infty)$, $0 \cdot \infty$, ∞/∞, and $A/0$ are not defined.
The *extended real number system* consists of the real field $\mathbb{R}$ and two symbols, $+\infty$ and $-\infty$. We preserve the original order in $\mathbb{R}$ and define

$$-\infty < x < +\infty$$

for every $x \in \mathbb{R}$.
It is the clear that $+\infty$ is an upper bound of every subset of the extended real number system, and every nonempty subset has a least upper bound. If, for example, E is a set of real numbers which is not bounded above in $\mathbb{R}$, then $\sup E = +\infty$ in the extended real system. Exactly the same remarks apply to lower bounds.
The extended real system does not form a field, but it is customary to make the following conventions:

(a) If x is real then

$$x + \infty = +\infty, \quad x - \infty = -\infty, \quad \frac{x}{+\infty} = \frac{x}{-\infty} = 0.$$

(b) If $x > 0$ then $x \cdot (+\infty) = +\infty$ and $x \cdot (-\infty) = -\infty$.
(c) If $x < 0$ then $x \cdot (+\infty) = -\infty$ and $x \cdot (-\infty) = +\infty$.

When it is desired to make the distinction between the real numbers on the one hand and the symbols $+\infty$ and $-\infty$ on the other hand quite explicit, the real numbers are called *finite*.
In Homework 9.2 (a) and (b) you are invited to give explicit proves in two special cases.

Example 3.3 (a) Let $p(x)$ and $q(x)$ be polynomials and $a \in \mathbb{R}$. Then

$$\lim_{x\to a} p(x) = p(a).$$

This immediately follows from $\lim_{x\to a} x = a$, $\lim_{x\to a} c = c$ and Proposition 3.2. Indeed, by (b) and (c), for $p(x) = 3x^3 - 4x + 7$ we have $\lim_{x\to a}(3x^2 - 4x + 7) = 3\left(\lim_{x\to a} x\right)^3 - 4\lim_{x\to a} x + 7 = 3a^2 - 4a + 7 = p(a)$. This works for arbitrary polynomials . Suppose moreover that $q(a) \neq 0$. Then by (d),

$$\lim_{x\to a} \frac{p(x)}{q(x)} = \frac{p(a)}{q(a)}.$$

Hence, the limit of a rational function $f(x)$ as x approaches a point a of the domain of f is $f(a)$.

(b) Let $f(x) = \frac{p(x)}{q(x)}$ be a rational function with polynomials $p(x) = \sum_{k=0}^{r} a_k x^k$ and $q(x) = \sum_{k=0}^{s} b_k x^k$ with real coefficients a_k and b_k and of degree r and s, respectively. Then

$$\lim_{x\to+\infty} f(x) = \begin{cases} 0, & \text{if} \quad r < s, \\ \frac{a_r}{b_s}, & \text{if} \quad r = s, \\ +\infty, & \text{if} \quad r > s \quad \text{and} \quad \frac{a_r}{b_s} > 0, \\ -\infty, & \text{if} \quad r > s \quad \text{and} \quad \frac{a_r}{b_s} < 0. \end{cases}$$

The first two statements ($r \geq s$) follow from Example 3.2 (b) together with Proposition 3.2. Namely, $a_k x^{k-r} \to 0$ as $x \to +\infty$ provided $0 \leq k < r$. The statements for $r > s$ follow from $x^{r-s} \to +\infty$ as $x \to +\infty$ and the above remark.
Note that

$$\lim_{x\to-\infty} f(x) = (-1)^{r+s} \lim_{x\to+\infty} f(x)$$

since

$$\frac{p(-x)}{q(-x)} = \frac{(-1)^r a_r x^r + \dots}{(-1)^s b_s x^s + \dots} = (-1)^{r+s} \frac{a_r x^r + \dots}{b_s x^s + \dots}.$$

3.2 Continuous Functions

Definition 3.5 Let f be a function and $x_0 \in D(f)$. We say that f *is continuous at* x_0 if

$$\forall\, \varepsilon > 0 \;\; \exists\, \delta > 0 \;\; \forall\, x \in D(f) : \; |\, x - x_0 \,| < \delta \implies |\, f(x) - f(x_0) \,| < \varepsilon. \tag{3.1}$$

We say that f is continuous in $A \subset D(f)$ if f is continuous at all points $x_0 \in A$.

Proposition 3.1 shows that the above definition of continuity in x_0 is equivalent to: For all sequences (x_n), $x_n \in D(f)$, with $\lim\limits_{n\to\infty} x_n = x_0$, $\lim\limits_{n\to\infty} f(x_n) = f(x_0)$. In other words, f is continuous at x_0 if $\lim\limits_{x\to x_0} f(x) = f(x_0)$.

Example 3.4 (a) In example 3.2 we have seen that every polynomial is continuous in $\mathbb{R}$ and every rational functions f is continuous in their domain $D(f)$.
$f(x) = |\, x \,|$ is continuous in $\mathbb{R}$.
(b) Continuity is a *local* property: If two functions $f, g\colon D \to \mathbb{R}$ coincide in a neighborhood $U_\varepsilon(x_0) \subset D$ of some point x_0, then f is continuous at x_0 if and only if g is continuous at x_0.
(c) $f(x) = [x]$ is continuous in $\mathbb{R} \setminus \mathbb{Z}$. If x_0 is not an integer, then $n < x_0 < n+1$ for some $n \in \mathbb{N}$ and $f(x) = n$ coincides with a constant function in a neighborhood $x \in U_\varepsilon(x_0)$. By (b), f is continuous at x_0. If $x_0 = n \in \mathbb{Z}$, $\lim_{x\to n}[x]$ does not exist; hence f is not continuous at n.
(d) $f(x) = \dfrac{x^2-1}{x-1}$ if $x \neq 1$ and $f(1) = 1$. Then f is not continuous at $x_0 = 1$ since

$$\lim_{x\to1} \frac{x^2-1}{x-1} = \lim_{x\to1}(x+1) = 2 \neq 1 = f(1).$$

There are two reasons for a function not being continuous at x_0. First, $\lim_{x\to x_0} f(x)$ does not exist. Secondly, f has a limit at x_0 but $\lim_{x\to x_0} f(x) \neq f(x_0)$.

Proposition 3.3 *Suppose $f, g\colon D \to \mathbb{R}$ are continuous at $x_0 \in D$. Then $f+g$ and fg are also continuous at x_0. If $g(x_0) \neq 0$, then f/g is continuous at x_0.*

The proof is obvious from Proposition 3.2.
The set $\mathrm{C}(D)$ of continuous function on $D \subset \mathbb{R}$ form a commutative algebra with 1.

Proposition 3.4 *Let $f\colon D \to \mathbb{R}$ and $g\colon E \to \mathbb{R}$ functions with $f(D) \subset E$. Suppose f is continuous at $a \in D$, and g is continuous at $b = f(a) \in E$. Then the composite function $g\circ f\colon D \to \mathbb{R}$ is continuous at a.*

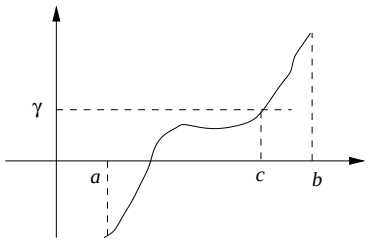
Proof. Let (x_n) be a sequence with $x_n \in D$ and $\lim_{n\to\infty} x_n = a$. Since f is continuous at a, $\lim_{n\to\infty} f(x_n) = b$. Since g is continuous at b, $\lim_{n\to\infty} g(f(x_n)) = g(b)$; hence $g\circ f(x_n) \to g\circ f(a)$. This completes the proof. ∎

Example 3.5 $f(x) = \frac{1}{x}$ is continuous for $x \neq 0$, $g(x) = \sin x$ is continuous (see below), hence, $(g\circ f)(x) = \sin\frac{1}{x}$ is continuous on $\mathbb{R}\setminus\{0\}$.

3.2.1 The Intermediate Value Theorem

In this paragraph, $[a,b] \subset \mathbb{R}$ is a closed, bounded interval, $a, b \in \mathbb{R}$.
The intermediate value theorem is the basis for several existence theorems in analysis. It is again equivalent to the order completeness of $\mathbb{R}$.

Theorem 3.5 (Intermediate Value Theorem) *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function and γ a real number between $f(a)$ and $f(b)$. Then there exists $c \in [a,b]$ such that $f(c) = \gamma$.*



The statement is clear from the graphical presentation. Nevertheless, it needs a proof since pictures do not prove anything. The statement is wrong for rational numbers. For example, let $D = \{x \in \mathbb{Q} \mid 1 \leq x \leq 2\}$ and $f(x) = x^2 - 2$. Then $f(1) = -1$ and $f(2) = 2$ but there is no $p \in D$ with $f(p) = 0$ since 2 has no rational square root.

Proof. Without loss of generality suppose $f(a) \leq f(b)$. Starting with $[a_1, b_1] = [a,b]$, we successively construct a nested sequence of intervals $[a_n, b_n]$ such that $f(a_n) \leq \gamma \leq f(b_n)$. As in the proof of Proposition 2.12, the $[a_n, b_n]$ is one of the two halfintervals $[a_{n-1}, m]$ and $[m, b_{n-1}]$ where $m = (a_{n-1} + b_{n-1})/2$ is the midpoint of the $(n-1)$st interval. By Proposition 2.11 the monotonic sequences (a_n) and (b_n) both converge to a common point c. Since f is continuous,

$$\lim_{n\to\infty} f(a_n) = f(\lim_{n\to\infty} a_n) = f(c) = f(\lim_{n\to\infty} b_n) = \lim_{n\to\infty} f(b_n).$$

By Proposition 2.14, $f(a_n) \leq \gamma \leq f(b_n)$ implies

$$\lim_{n\to\infty} f(a_n) \leq \gamma \leq \lim_{n\to\infty} f(b_n);$$

Hence, $\gamma = f(c)$. ∎

Example 3.6 (a) We again show the existence of the nth root of a positive real number $a > 0$, $n \in \mathbb{N}$. By Example 3.2, the polynomial $p(x) = x^n - a$ is continuous in $\mathbb{R}$. We find $p(0) = -a < 0$ and by Bernoulli's inequality

$$p(1+a) = (1+a)^n - a \geq 1 + (n-1)a \geq 1 > 0.$$

Theorem 3.5 shows that p has a root in the interval $(0, 1+a)$.
(b) A polynomial p of odd degree with real coefficients has a real zero. Namely, by Example 3.3, if the leading coefficient a_r of p is positive, $\lim\limits_{x\to-\infty} p(x) = -\infty$ and $\lim\limits_{x\to\infty} p(x) = +\infty$. Hence there are a and b with $a < b$ and $p(a) < 0 < p(b)$. Therefore, there is a $c \in (a, b)$ such that $p(c) = 0$.

There are polynomials of even degree having no real zeros. For example $f(x) = x^{2k} + 1$.

Remark 3.3 Theorem 3.5 is not true for continuous functions $f\colon \mathbb{Q} \to \mathbb{R}$. For example, $f(x) = x^2 - 2$ is continuous, $f(0) = -2 < 0 < 2 = f(2)$. However, there is no $r \in \mathbb{Q}$ with $f(r) = 0$.

3.2.2 Continuous Functions on Bounded and Closed Intervals—The Theorem about Maximum and Minimum

We say that $f\colon [a, b] \to \mathbb{R}$ is continuous, if f is continuous on (a, b) and $f(a+0) = f(a)$ and $f(b-0) = f(b)$.

Theorem 3.6 (Theorem about Maximum and Minimum) *Let $f\colon [a, b] \to \mathbb{R}$ be continuous. Then f is bounded and attains its maximum and its minimum, that is, there exists $C > 0$ with $|\, f(x)\,| \leq C$ for all $x \in [a, b]$ and there exist $p, q \in [a, b]$ with $\sup\limits_{a\leq x\leq b} f(x) = \max\limits_{a\leq x\leq b} f(x) = f(p)$ and $\inf\limits_{a\leq x\leq b} f(x) = \min\limits_{a\leq x\leq b} f(x) = f(q)$.*

Remarks 3.4 (a) The theorem is not true in case of open, half-open or infinite intervals. For example, $f\colon (0, 1] \to \mathbb{R}$, $f(x) = \frac{1}{x}$ is continuous but not bounded. The function $f\colon (0, 1) \to \mathbb{R}$, $f(x) = x$ is continuous and bounded. However, it doesn't attain maximum and minimum. Finally, $f(x) = x^2$ on $\mathbb{R}_+$ is continuous but not bounded.
(b) Put $M := \max\limits_{x\in K} f(x)$ and $m := \min\limits_{x\in K} f(x)$. By the Theorem about maximum and minimum and the intermediate value theorem, for all $\gamma \in \mathbb{R}$ with $m \leq \gamma \leq M$ there exists $c \in [a, b]$ such that $f(c) = \gamma$; that is, f attains all values between m and M.

Proof. We give the proof in case of the maximum. Replacing f by $-f$ yields the proof for the minimum. Let

$$A = \sup_{a\leq x\leq b} f(x) \in \mathbb{R} \cup \{+\infty\}.$$

(Note that $A = +\infty$ is equivalent to f is not bounded above.) Then there exists a sequence $(x_n) \in [a,b]$ such that $\lim_{n\to\infty} f(x_n) = A$. Since (x_n) is bounded, by the Theoremm of Weierstraß there exists a convergent subsequence (x_{n_k}) with $p = \lim\limits_k x_{n_k}$ and $a \le p \le b$. Since f is continuous,

$$A = \lim_{k\to\infty} f(x_{n_k}) = f(p).$$

In particular, A is a *finite* real number; that is, f is bounded above by A and f attains its maximum A at point $p \in [a,b]$. ∎

3.3 Uniform Continuity

Let D be a finite or infinite interval.

Definition 3.6 A function $f\colon D \to \mathbb{R}$ is called *uniformly continuous* if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, x' \in D \,|\, |x - x'| < \delta$ implies $|f(x) - f(x')| < \varepsilon$.

f is uniformly continuous on $[a,b]$ if and only if

$$\forall\, \varepsilon > 0 \;\; \exists\, \delta > 0 \;\; \forall\, x, y \in [a,b] : |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon. \tag{3.2}$$

Remark 3.5 If f is uniformly continuous on D then f is continuous on D. However, the converse direction is not true.
Consider, for example, $f\colon (0,1) \to \mathbb{R}$, $f(x) = \frac{1}{x}$ which is continuous. Suppose to the contrary that f is uniformly continuous. Then to $\varepsilon = 1$ there exists $\delta > 0$ with (3.2). By the Archimedian property there exists $n \in \mathbb{N}$ such that $\frac{1}{2n} < \delta$. Consider $x_n = \frac{1}{n}$ and $y_n = \frac{1}{2n}$. Then $|x_n - y_n| = \frac{1}{2n} < \delta$. However,

$$|f(x_n) - f(y_n)| = 2n - n = n \ge 1.$$

A contradiction! Hence, f is not uniformly continuous on $(0,1)$.

Let us consider the differences between the concepts of continuity and uniform continuity. First, uniform continuity is a property of a function on a set, whereas continuity can be defined in a single point. To ask whether or not a given function is uniformly continuous at a certain point is meaningless. Secondly, if f is continuous on D, then it is possible to find, for each $\varepsilon > 0$ and for each point $x_0 \in D$, a number $\delta = \delta(x_0, \varepsilon) > 0$ having the property specified in Definition 3.5. This δ depends on ε *and* on x_0. If f is, however, uniformly continuous on X, then it is possible, for each $\varepsilon > 0$ to find *one* $\delta = \delta(\varepsilon) > 0$ which will do for all possible points x_0 of X.
That the two concepts are equivalent on bounded and closed intervals follows from the next proposition.

Proposition 3.7 *Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on a bounded and closed interval. Then f is uniformly continuous on $[a,b]$.*

Proof. Suppose to the contrary that f is not uniformly continuous. Then there exists $\varepsilon_0 > 0$ without matching $\delta > 0$; for every positive integer $n \in \mathbb{N}$ there exists a pair of points x_n, x'_n with $|\, x_n - x'_n \,| < 1/n$ but $|\, f(x_n) - f(x'_n) \,| \geq \varepsilon_0$. Since $[a, b]$ is bounded and closed, (x_n) has a subsequence converging to some point $p \in [a, b]$. Since $|\, x_n - x'_n \,| < 1/n$, the sequence (x'_n) also converges to p. Hence

$$\lim_{k\to\infty} \big(f(x_{n_k}) - f(x'_{n_k})\big) = f(p) - f(p) = 0$$

which contradicts $\big|\, f(x_{n_k}) - f(x'_{n_k}) \,\big| \geq \varepsilon_0$ for all k. ∎

There exists an example of a *bounded* continuous function $f\colon [0, 1) \to \mathbb{R}$ which is not uniformly continuous, see [Kön90, p. 91].

Discontinuities

If x is a point in the domain of a function f at which f is not continuous, we say f is *discontinuous* at x or f has a *discontinuity* at x. It is customary to divide discontinuities into two types.

Definition 3.7 Let $f\colon (a, b) \to \mathbb{R}$ be a function which is discontinuous at a point x_0. If the one-sided limits $\lim_{x\to x_0+0} f(x)$ and $\lim_{x\to x_0-0} f(x)$ exist, then f is said to have a *simple* discontinuity or a discontinuity of the *first kind*. Otherwise the discontinuity is said to be of the *second kind*.

Example 3.7 (a) $f(x) = \operatorname{sign}(x)$ is continuous on $\mathbb{R} \setminus \{0\}$ since it is locally constant. Moreover, $f(0 + 0) = 1$ and $f(0 - 0) = -1$. Hence, $\operatorname{sign}(x)$ has a simple discontinuity at $x_0 = 0$.
(b) Define $f(x) = 0$ if x is rational, and $f(x) = 1$ if x is irrational. Then f has a discontinuity of the second kind at every point x since neither $f(x + 0)$ nor $f(x - 0)$ exists.
(c) Define

$$f(x) = \begin{cases} \sin \frac{1}{x}, & \text{if} \quad x \neq 0; \\ 0, & \text{if} \quad x = 0. \end{cases}$$

Consider the two sequences

$$x_n = \frac{1}{\frac{\pi}{2} + n\pi} \quad \text{and} \quad y_n = \frac{1}{n\pi},$$

Then both sequences (x_n) and (y_n) approach 0 from above but $\lim_{n\to\infty} f(x_n) = 1$ and $\lim_{n\to\infty} f(y_n) = 0$; hence $f(0 + 0)$ does not exist. Therefore f has a discontinuity of the second kind at $x = 0$. We have not yet shown that $\sin x$ is a continuous function. This will be done in Section 3.5.2.

3.4 Monotonic Functions

Definition 3.8 Let f be a real function on the interval (a, b). Then f is said to be *monotonically increasing* on (a, b) if $a < x < y < b$ implies $f(x) \leq f(y)$. If the last inequality is reversed, we obtain the definition of a *monotonically decreasing* function. The class of *monotonic functions* consists of both the increasing and the decreasing functions.
If $a < x < y < b$ implies $f(x) < f(y)$, the function is said to be *strictly increasing*. Similarly, *strictly decreasing* functions are defined.

Theorem 3.8 *Let f be a monotonically increasing function on (a, b). Then the one-sided limits $f(x + 0)$ and $f(x - 0)$ exist at every point x of (a, b). More precisely,*

$$\sup_{t\in(a,x)} f(t) = f(x - 0) \leq f(x) \leq f(x + 0) = \inf_{t\in(x,b)} f(t). \tag{3.3}$$

Furthermore, if $a < x < y < b$, then

$$f(x + 0) \leq f(y - 0). \tag{3.4}$$

Analogous results evidently hold for monotonically decreasing functions.

Proof. See Appendix B to this chapter. ∎

Proposition 3.9 *Let $f\colon [a, b] \to \mathbb{R}$ be a strictly monotonically increasing continuous function and $A = f(a)$ and $B = f(b)$. Then f maps $[a, b]$ bijectively onto $[A, B]$ and the inverse function*

$$f^{-1}\colon [A, B] \to \mathbb{R}$$

is again strictly monotonically increasing and continuous.

Note that the inverse function $f^{-1}\colon [A, B] \to [a, b]$ is defined by $f(y_0) = x_0$, $y_0 \in [A, B]$, where x_0 is the unique element of $[a, b]$ with $f(x_0) = y_0$. However, we can think of f^{-1} as a function into $\mathbb{R}$. A similar statement is true for strictly decreasing functions.
Proof. By Remark 3.4, f maps $[a, b]$ onto the whole closed interval $[A, B]$ (intermediate value theorem). Since $x < y$ implies $f(x) < f(y)$, f is injective and hence bijective. Hence, the inverse mapping $f^{-1}\colon [A, B] \to [a, b]$ exists and is again strictly increasing ($u < v$ implies $f^{-1}(u) = x < y = f^{-1}(v)$ otherwise, $x \geq y$ implies $u \geq v$).
We show that $g = f^{-1}$ is continuous. Suppose (u_n) is a sequence in $[A, B]$ with $u_n \to u$ and $u_n = f(x_n)$ and $u = f(x)$. We have to show that (x_n) converges to x. Suppose to the contrary that there exists $\varepsilon_0 > 0$ such that $|\, x_n - x \,| \geq \varepsilon_0$ for infinitely many n. Since $(x_n) \subseteq [a, b]$ is bounded, there exists a converging subsequence (x_{n_k}), say, $x_{n_k} \to c$ as $k \to \infty$. The above inequality is true for the limit c, too, that is $|\, c - x \,| \geq \varepsilon_0$. By continuity of f, $x_{n_k} \to c$ implies $f(x_{n_k}) \to f(c)$. That is $u_{n_k} \to f(c)$. Since $u_n \to u = f(x)$ and the limit of a converging sequence is unique, $f(c) = f(x)$. Since f is bijective, $x = c$; this contradicts $|\, c - x \,| \geq \varepsilon_0$. Hence, g is continuous at u. ∎

Example 3.8 The function $f\colon \mathbb{R}_+ \to \mathbb{R}_+$, $f(x) = x^n$, is continuous and strictly increasing. Hence $x = g(y) = \sqrt[n]{y}$ is continuous, too. This gives an alternative proof of homework 5.5.

3.5 Exponential, Trigonometric, and Hyperbolic Functions and their Inverses

3.5.1 Exponential and Logarithm Functions

In this section we are dealing with the exponential function which is one of the most important in analysis. We use the exponential series to define the function. We will see that this definition is consistent with the definition e^x for rational $x \in \mathbb{Q}$ as defined in Chapter 1.

Definition 3.9 For $z \in \mathbb{C}$ put

$$E(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \cdots . \tag{3.5}$$

Note that $E(0) = 1$ and $E(1) = \mathrm{e}$ by the definition at page 61. The radius of convergence of the exponential series (3.5) is $R = +\infty$, i. e. the series converges absolutely for all $z \in \mathbb{C}$, see Example 2.13 (c).
Applying Proposition 2.31 (Cauchy product) on multiplication of absolutely convergent series, we obtain

$$\begin{aligned} E(z)E(w) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{m=0}^{\infty} \frac{w^m}{m!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{z^k\, w^{n-k}}{k!\,(n-k)!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} z^k w^{n-k} = \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!}, \end{aligned}$$

which gives us the important addition formula

$$E(z+w) = E(z)E(w), \quad z, w \in \mathbb{C}. \tag{3.6}$$

One consequence is that

$$E(z)E(-z) = E(0) = 1, \quad z \in \mathbb{C}. \tag{3.7}$$

This shows that $E(z) \neq 0$ for all z. By (3.5), $E(x) > 0$ if $x > 0$; hence (3.7) shows $E(x) > 0$ for all real x.
Iteration of (3.6) gives

$$E(z_1 + \cdots + z_n) = E(z_1) \cdots E(z_n). \tag{3.8}$$

Let us take $z_1 = \cdots = z_n = 1$. Since $E(1) = \mathrm{e}$ by (2.15), we obtain

$$E(n) = \mathrm{e}^n, \quad n \in \mathbb{N}. \tag{3.9}$$

If $p = m/n$, where m, n are positive integers, then

$$E(p)^n = E(pn) = E(m) = \mathrm{e}^m, \tag{3.10}$$

so that

$$E(p) = \mathrm{e}^p, \quad p \in \mathbb{Q}_+. \tag{3.11}$$

It follows from (3.7) that $E(-p) = \mathrm{e}^{-p}$ if p is positive and rational. Thus (3.11) holds for all rational p. This justifies the redefinition

$$\mathrm{e}^x := E(x), \quad x \in \mathbb{C}.$$

The notation $\exp(x)$ is often used in place of e^x.

Proposition 3.10 *We can estimate the remainder term $r_n := \sum_{k=n}^{\infty} z^k/k!$ as follows*

$$|\, r_n(z) \,| \le \frac{2\,|\, z \,|^n}{n!} \quad \textit{if} \quad |\, z \,| \le \frac{n+1}{2}. \tag{3.12}$$

Proof. We have

$$\begin{aligned} |\, r_n(z) \,| &\le \sum_{k=n}^{\infty} \left| \frac{z^k}{k!} \right| = \frac{|\, z \,|^n}{n!} \left(1 + \frac{|\, z \,|}{n+1} + \frac{|\, z \,|^2}{(n+1)(n+2)} + \cdots + \frac{|\, z \,|^k}{(n+1)\cdots(n+k)} + \cdots \right) \\ &\le \frac{|\, z \,|^n}{n!} \left(1 + \frac{|\, z \,|}{n+1} + \frac{|\, z \,|^2}{(n+1)^2} + \cdots + \frac{|\, z \,|^k}{(n+1)^k} + \cdots \right). \end{aligned}$$

$|\, z \,| \le (n+1)/2$ implies,

$$|\, r_n(z) \,| \le \frac{|\, z \,|^n}{n!} \left(1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^k} + \cdots \right) \le \frac{2\,|\, z \,|^n}{n!}.$$

■

Example 3.9 (a) Inserting $n = 1$ gives

$$|\, E(z) - 1 \,| = |\, r_1(z) \,| \le 2\,|\, z \,|, \quad |\, z \,| \le 1.$$

In particular, $E(z)$ is continuous at $z_0 = 0$. Indeed, to $\varepsilon > 0$ choose $\delta = \varepsilon/2$ then $|\, z \,| < \delta$ implies $|\, E(z) - 1 \,| \le 2\,|\, z \,| \le \varepsilon$; hence $\lim_{z \to 0} E(z) = E(0) = 1$ and E is continuous at 0.
(b) Inserting $n = 2$ gives

$$|\, \mathrm{e}^z - 1 - z \,| = |\, r_2(z) \,| \le |\, z \,|^2, \quad |\, z \,| \le \frac{3}{2}.$$

This implies

$$\left| \frac{\mathrm{e}^z - 1}{z} - 1 \right| \le |\, z \,|, \quad |\, z \,| < \frac{3}{2}.$$

The sandwich theorem gives $\lim_{z \to 0} \frac{\mathrm{e}^z - 1}{z} = 1$.

By (3.5), $\lim_{x\to\infty} E(x) = +\infty$; hence (3.7) shows that $\lim_{x\to-\infty} E(x) = 0$. By (3.5), $0 < x < y$ implies that $E(x) < E(y)$; by (3.7), it follows that $E(-y) < E(-x)$; hence, E is strictly increasing on the whole real axis.
The addition formula also shows that

$$\lim_{h\to 0}(E(z+h) - E(z)) = E(z)\lim_{h\to 0}(E(h) - 1) = E(z)\cdot 0 = 0, \tag{3.13}$$

where $\lim_{h\to 0} E(h) = 1$ directly follows from Example 3.9. Hence, $E(z)$ is continuous for all z.

Proposition 3.11 *Let* e^x *be defined on* $\mathbb{R}$ *by the power series* (3.5). *Then*

(a) e^x *is continuous for all* x.
(b) e^x *is a strictly increasing function and* $\mathrm{e}^x > 0$.
(c) $\mathrm{e}^{x+y} = \mathrm{e}^x\mathrm{e}^y$.
(d) $\lim\limits_{x\to+\infty} \mathrm{e}^x = +\infty$, $\lim\limits_{x\to-\infty} \mathrm{e}^x = 0$.
(e) $\lim\limits_{x\to+\infty} \dfrac{x^n}{\mathrm{e}^x} = 0$ *for every* $n \in \mathbb{N}$.

Proof. We have already proved (a) to (d); (3.5) shows that

$$\mathrm{e}^x > \frac{x^{n+1}}{(n+1)!}$$

for $x > 0$, so that

$$\frac{x^n}{\mathrm{e}^x} < \frac{(n+1)!}{x},$$

and (e) follows. Part (e) shows that e^x tends faster to $+\infty$ than any power of x, as $x \to +\infty$. ∎

Since e^x, $x \in \mathbb{R}$, is a strictly increasing continuous function, by Proposition 3.9 e^x has an strictly increasing continuous inverse function $\log y$, $\log\colon (0,+\infty) \to \mathbb{R}$. The function $\log$ is defined by

$$\mathrm{e}^{\log y} = y, \quad y > 0, \tag{3.14}$$

or, equivalently, by

$$\log(\mathrm{e}^x) = x, \quad x \in \mathbb{R}. \tag{3.15}$$

Writing $u = \mathrm{e}^x$ and $v = \mathrm{e}^y$, (3.6) gives

$$\log(uv) = \log(\mathrm{e}^x\mathrm{e}^y) = \log(\mathrm{e}^{x+y}) = x + y,$$

such that

$$\log(uv) = \log u + \log v, \quad u > 0, v > 0. \tag{3.16}$$

This shows that $\log$ has the familiar property which makes the logarithm useful for computations. Another customary notation for $\log x$ is $\ln x$. Proposition 3.11 shows that

$$\lim_{x\to+\infty} \log x = +\infty, \quad \lim_{x\to 0+0} \log x = -\infty.$$

We summarize what we have proved so far.

Proposition 3.12 *Let the logarithm* $\log\colon (0,+\infty) \to \mathbb{R}$ *be the inverse function to the exponential function* e^x*. Then*

(a) $\log$ *is continuous on* $(0,+\infty)$.
(b) $\log$ *is strictly increasing.*
(c) $\log(uv) = \log u + \log v$ *for* $u, v > 0$.
(d) $\lim\limits_{x\to+\infty} \log x = +\infty, \quad \lim\limits_{x\to 0+0} \log x = -\infty.$

It is seen from (3.14) that

$$x = \mathrm{e}^{\log x} \Longrightarrow x^n = \mathrm{e}^{n\,\log x} \tag{3.17}$$

if $x > 0$ and n is an integer. Similarly, if m is a positive integer, we have

$$x^{\frac{1}{m}} = \mathrm{e}^{\frac{\log x}{m}} \tag{3.18}$$

Combining (3.17) and (3.18), we obtain

$$x^\alpha = \mathrm{e}^{\alpha\,\log x}. \tag{3.19}$$

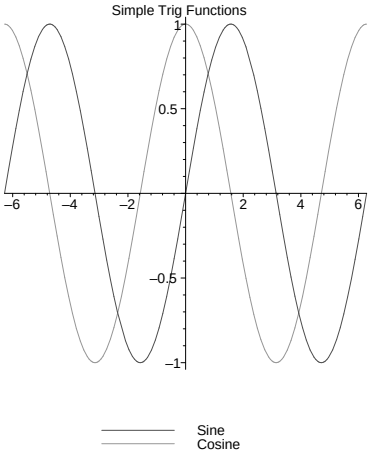
for any rational α. We now define x^α for any real α and $x > 0$, by (3.19). In the same way, we redefine the exponential function

$$a^x = \mathrm{e}^{x\log a}, \quad a > 0, \quad x \in \mathbb{R}.$$

It turns out that in case $a \neq 1$, $f(x) = a^x$ is strict monotonic and continuous since e^x is so. Hence, f has a stict monotonic continuous inverse function $\log_a\colon (0,+\infty) \to \mathbb{R}$ defined by

$$\log_a(a^x) = x, \quad x \in \mathbb{R}, \quad a^{\log_a x} = x, \quad x > 0.$$

3.5.2 Trigonometric Functions and their Inverses



In this section we redefine the trigonometric functions using the exponential function e^z. We will see that the new definitions coincide with the old ones.

Definition 3.10 For $z \in \mathbb{C}$ define

$$\cos z = \frac{1}{2}\left(\mathrm{e}^{\mathrm{i}\,z} + \mathrm{e}^{-\mathrm{i}\,z}\right), \quad \sin z = \frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}\,z} - \mathrm{e}^{-\mathrm{i}\,z}\right) \tag{3.20}$$

such that

$$\mathrm{e}^{\mathrm{i}\,z} = \cos z + \mathrm{i}\sin z \quad \text{(Euler formula)} \tag{3.21}$$

Proposition 3.13 (a) *The functions* $\sin z$ *and* $\cos z$ *can be written as power series which converge absolutely for all* $z \in \mathbb{C}$*:*

$$\begin{aligned} \cos z &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}\, z^{2n} = 1 - \frac{1}{2}z^2 + \frac{1}{4!}z^4 - \frac{1}{6!}z^6 + - \cdots \\ \sin z &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}\, z^{2n+1} = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - + \cdots . \end{aligned} \tag{3.22}$$

(b) $\sin x$ *and* $\cos x$ *are real valued and continuous on* $\mathbb{R}$*, where* $\cos x$ *is an even and* $\sin x$ *is an odd function, i. e.* $\cos(-x) = \cos x$*,* $\sin(-x) = -\sin x$*. We have*

$$\sin^2 x + \cos^2 x = 1; \tag{3.23}$$

$$\begin{aligned} \cos(x+y) &= \cos x \cos y - \sin x \sin y; \\ \sin(x+y) &= \sin x \cos y + \cos x \sin y. \end{aligned} \tag{3.24}$$

Proof. (a) Inserting $\mathrm{i}z$ into (3.5) in place of z and using $(\mathrm{i}^n) = (\mathrm{i}, -1, -\mathrm{i}, 1, \mathrm{i}, -1, \dots)$, we have

$$\mathrm{e}^{\mathrm{i}\,z} = \sum_{n=0}^{\infty} \mathrm{i}^n \frac{z^n}{n!} = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{(2k)!} + \mathrm{i} \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k+1}}{(2k+1)!}.$$

Inserting $-\mathrm{i}z$ into (3.5) in place of z we have

$$\mathrm{e}^{-\mathrm{i}\,z} = \sum_{n=0}^{\infty} \mathrm{i}^n \frac{z^n}{n!} = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{(2k)!} - \mathrm{i} \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k+1}}{(2k+1)!}.$$

Inserting this into (3.20) proves (a).

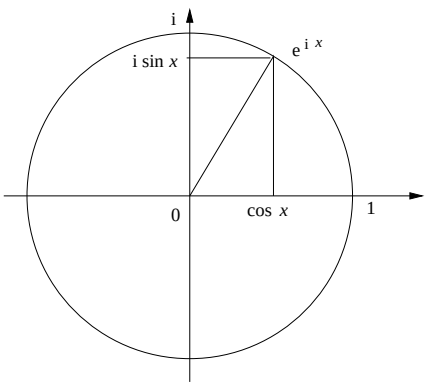
(b) Since the exponential function is continuous on $\mathbb{C}$, $\sin z$ and $\cos z$ are also continuous on $\mathbb{C}$. In particular, their restrictions to $\mathbb{R}$ are continuous. Now let $x \in \mathbb{R}$, then $\overline{\mathrm{i}\,x} = -\mathrm{i}\,x$. By Homework 11.3 and (3.20) we obtain

$$\cos x = \frac{1}{2}\left(\mathrm{e}^{\mathrm{i}\,x} + \mathrm{e}^{\overline{\mathrm{i}\,x}}\right) = \frac{1}{2}\left(\mathrm{e}^{\mathrm{i}\,x} + \overline{\mathrm{e}^{\mathrm{i}\,x}}\right) = \frac{1}{2}\left(\mathrm{e}^{\mathrm{i}\,x} + \mathrm{e}^{\overline{\mathrm{i}\,x}}\right) = \mathrm{Re}\,\left(\mathrm{e}^{\mathrm{i}\,x}\right)$$

and similarly

$$\sin x = \mathrm{Im}\,\left(\mathrm{e}^{\mathrm{i}\,x}\right).$$

Hence, $\sin x$ and $\cos x$ are real for real x.



For $x \in \mathbb{R}$ we have $\left|\,\mathrm{e}^{\mathrm{i}\,x}\,\right| = 1$. Namely by (3.7) and Homework 11.3

$$\left|\,\mathrm{e}^{\mathrm{i}x}\,\right|^2 = \mathrm{e}^{\mathrm{i}x}\,\overline{\mathrm{e}^{\mathrm{i}x}} = \mathrm{e}^{\mathrm{i}x}\mathrm{e}^{-\mathrm{i}\,x} = \mathrm{e}^0 = 1,$$

so that for $x \in \mathbb{R}$

$$\left|\,\mathrm{e}^{\mathrm{i}\,x}\,\right| = 1. \tag{3.25}$$

On the other hand, the Euler formula and the fact that $\cos x$ and $\sin x$ are real give

$$1 = \left|\,\mathrm{e}^{\mathrm{i}\,x}\,\right| = |\cos x + \mathrm{i}\sin x\,| = \cos^2 x + \sin^2 x.$$

Hence, $e^{ix} = \cos x + i \sin x$ is a point on the unit circle in the complex plane, and $\cos x$ and $\sin x$ are its coordinates. This establishes the equivalence between the old definition of $\cos x$ as the length of the adjacent side in a rectangular triangle with hypothenuse 1 and angle $x\frac{180^\circ}{\pi}$ and the power series definition of $\cos x$. The only missing link is: the length of the arc from 1 to e^{ix} is x.
It follows directly from the definition that $\cos(-z) = \cos z$ and $\sin(-z) = -\sin z$ for all $z \in \mathbb{C}$. The addition laws for $\sin x$ and $\cos x$ follow from (3.6) applied to $e^{i(x+y)}$. This completes the proof of (b). ∎

Lemma 3.14 *There exists a unique number $\tau \in (0, 2)$ such that $\cos \tau = 0$. We define the number π by*

$$\pi = 2\tau. \tag{3.26}$$

The proof is based on the following Lemma.

Lemma 3.15

$$\text{(a)} \quad 0 < x < \sqrt{6} \quad \textit{implies} \quad x - \frac{x^3}{6} < \sin x < x. \tag{3.27}$$

$$\text{(b)} \quad 0 < x < \sqrt{2} \quad \textit{implies} \quad 0 < \cos x, \tag{3.28}$$

$$0 < \sin x < x < \frac{\sin x}{\cos x}, \tag{3.29}$$

$$\cos^2 x < \frac{1}{1 + x^2}. \tag{3.30}$$

(c) *$\cos x$ is strictly decreasing on $[0, \pi]$; whereas $\sin x$ is strictly increasing on $[-\pi/2, \pi/2]$.*

In particular, the sandwich theorem applied to statement (a), $1 - \frac{x^2}{6} < \frac{\sin x}{x} < 1$ as $x \to 0 + 0$ gives $\lim_{x \to 0+0} \frac{\sin x}{x} = 1$. Since $\frac{\sin x}{x}$ is an even function, this implies $\lim_{x \to 0} \frac{\sin x}{x} = 1$.
The proof of the lemma is in the Appendix B to this chapter.
Proof of Lemma 3.14. $\cos 0 = 1$. By the Lemma 3.15, $\cos^2 1 < 1/2$. By the double angle formula for cosine, $\cos 2 = 2\cos^2 1 - 1 < 0$. By continuity of $\cos x$ and Theorem 3.5, cos has a zero τ in the interval $(0, 2)$.
By addition laws,

$$\cos x - \cos y = -2 \sin\left(\frac{x+y}{2}\right) \sin\left(\frac{x-y}{2}\right).$$

So that by Lemma 3.15 $0 < x < y < 2$ implies $0 < \sin((x + y)/2)$ and $\sin((x - y)/2) < 0$; therefore $\cos x > \cos y$. Hence, $\cos x$ is strictly decreasing on $(0, 2)$. The zero τ is therefore unique. ∎

By definition, $\cos\left(\frac{\pi}{2}\right) = 0$; and (3.23) shows $\sin(\pi/2) = \pm 1$. By (3.27), $\sin \pi/2 = 1$. Thus $\mathrm{e}^{\mathrm{i}\pi/2} = \mathrm{i}$, and the addition formula for e^z gives

$$\mathrm{e}^{\pi\,\mathrm{i}} = -1, \quad \mathrm{e}^{2\pi\,\mathrm{i}} = 1; \tag{3.31}$$

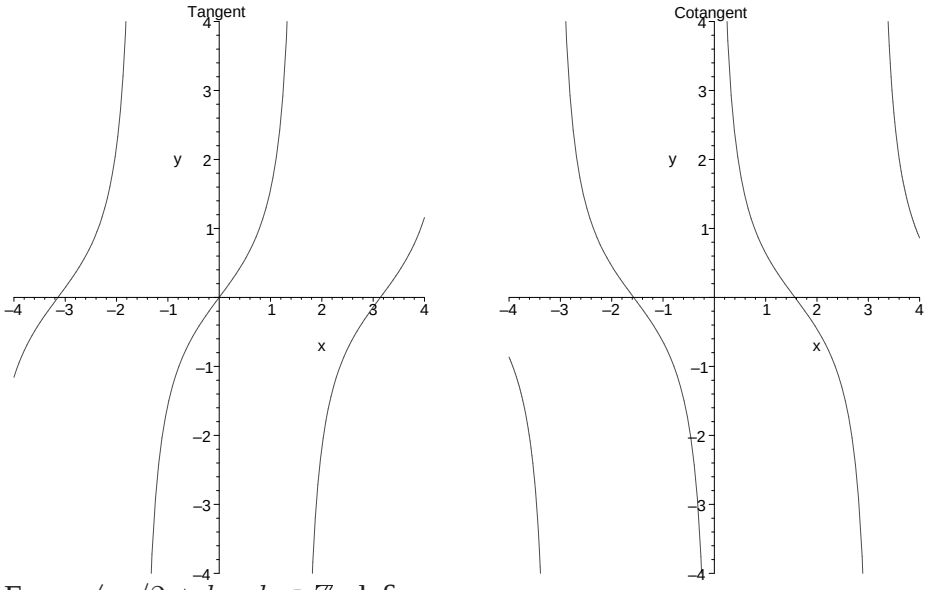
hence,

$$\mathrm{e}^{z+2\pi\,\mathrm{i}} = \mathrm{e}^z, \quad z \in \mathbb{C}. \tag{3.32}$$

Proposition 3.16 (a) *The function* e^z *is periodic with period* $2\pi\mathrm{i}$.
We have $\mathrm{e}^{\mathrm{i}x} = 1$, $x \in \mathbb{R}$, *if and only if* $x = 2k\pi$, $k \in \mathbb{Z}$.
(b) *The functions* $\sin z$ *and* $\cos z$ *are periodic with period* 2π.
The real zeros of the sine and cosine functions are $\{k\pi \mid k \in \mathbb{Z}\}$ *and* $\{\pi/2 + k\pi \mid k \in \mathbb{Z}\}$, *respectively.*

Proof. We have already proved (a). (b) follows from (a) and (3.20). ∎

Tangent and Cotangent Functions



For $x \neq \pi/2 + k\pi$, $k \in \mathbb{Z}$, define

$$\tan x = \frac{\sin x}{\cos x}. \tag{3.33}$$

For $x \neq k\pi$, $k \in \mathbb{Z}$, define

$$\cot x = \frac{\cos x}{\sin x}. \tag{3.34}$$

Lemma 3.17 (a) $\tan x$ *is continuous at* $x \in \mathbb{R} \setminus \{\pi/2 + k\pi \mid k \in \mathbb{Z}\}$, *and* $\tan(x + \pi) = \tan x$;
(b) $\lim\limits_{x \to \frac{\pi}{2} - 0} \tan x = +\infty$, $\lim\limits_{x \to \frac{\pi}{2} + 0} \tan x = -\infty$;
(c) $\tan x$ *is strictly increasing on* $(-\pi/2, \pi/2)$;

Proof. (a) is clear by Proposition 3.3 since $\sin x$ and $\cos x$ are continuous. We show only (c) and let (b) as an exercise. Let $0 < x < y < \pi/2$. Then $0 < \sin x < \sin y$ and $\cos x > \cos y > 0$. Therefore

$$\tan x = \frac{\sin x}{\cos x} < \frac{\sin y}{\cos y} = \tan y.$$

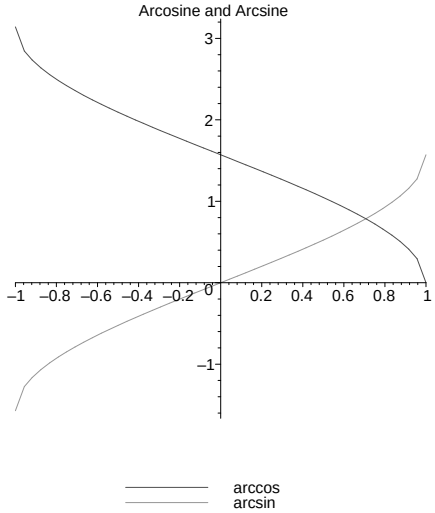
Hence, $\tan$ is strictly increasing on $(0, \pi/2)$. Since $\tan(-x) = -\tan(x)$, $\tan$ is strictly increasing on the whole interval $(-\pi/2, \pi/2)$. ∎

Similarly as Lemma 3.17 one proves the next lemma.

Lemma 3.18 (a) $\cot x$ *is continuous at* $x \in \mathbb{R} \setminus \{k\pi \mid k \in \mathbb{Z}\}$, *and* $\cot(x + \pi) = \cot x$;
(b) $\lim_{x \to 0-0} \cot x = -\infty$, $\lim_{x \to 0+0} \cot x = +\infty$;
(c) $\cot x$ *is strictly decreasing on* $(0, \pi)$.

Inverse Trigonometric Functions

We have seen in Lemma 3.15 that $\cos x$ is strictly decreasing on $[0, \pi]$ and $\sin x$ is strictly increasing on $[-\pi/2, \pi/2]$. Obviously, the images are $\cos[0, \pi] = \sin[-\pi/2, \pi/2] = [-1, 1]$. Using Proposition 3.9 we obtain that the inverse functions exists and they are monotonic and continuous.



Proposition 3.19 (and Definition) *There exists the inverse function to* $\cos$

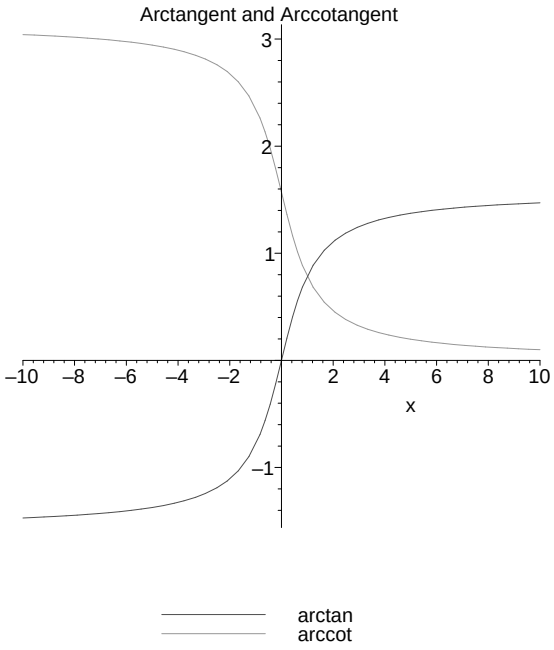
$$\arccos\colon [-1, 1] \to [0, \pi] \tag{3.35}$$

given by $\arccos(\cos x) = x$, $x \in [0, \pi]$ *or* $\cos(\arccos y) = y$, $y \in [-1, 1]$. *The function* $\arccos x$ *is strictly decreasing and continuous.*
There exists the inverse function to $\sin$

$$\arcsin\colon [-1, 1] \to [-\pi/2, \pi/2] \tag{3.36}$$

given by $\arcsin(\sin x) = x$, $x \in [-\pi/2, \pi/2]$ *or* $\sin(\arcsin y) = y$, $y \in [-1, 1]$. *The function* $\arcsin x$ *is strictly increasing and continuous.*

Note that $\arcsin x + \arccos x = \pi/2$ if $x \in [-1, 1]$. Indeed, let $y = \arcsin x$; then $x = \sin y = \cos(\pi/2 - y)$. Since $y \in [0, \pi]$, $\pi/2 - y \in [-\pi/2, \pi/2]$, and we have $\arccos x = \pi/2 - y$. Therefore $y + \arccos x = \pi/2$.



By Lemma 3.17, $\tan x$ is strictly increasing on $(-\pi/2, \pi/2)$. Therefore, there exists the inverse function on the image $\tan(-\pi/2, \pi/2) = \mathbb{R}$.

Proposition 3.20 (and Definition) *There exists the inverse function to* $\tan$

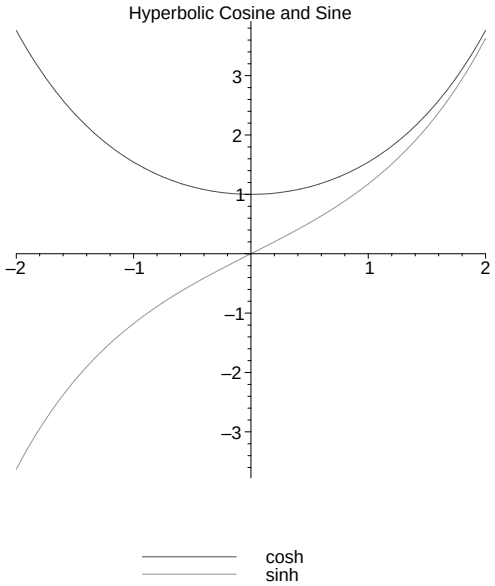
$$\arctan\colon \mathbb{R} \to (-\pi/2, \pi/2) \tag{3.37}$$

given by $\arctan(\tan x) = x$*,* $x \in (-\pi/2, \pi/2)$ *or* $\tan(\arctan y) = y$*,* $y \in \mathbb{R}$*. The function* $\arctan x$ *is strictly increasing and continuous. There exists the inverse function to* $\cot x$

$$\operatorname{arccot} : \mathbb{R} \to (0, \pi) \tag{3.38}$$

given by $\operatorname{arccot}(\cot x) = x$*,* $x \in (0, \pi)$ *or* $\cot(\operatorname{arccot} y) = y$*,* $y \in \mathbb{R}$*. The function* $\operatorname{arccot} x$ *is strictly decreasing and continuous.*

3.5.3 Hyperbolic Functions and their Inverses



The functions

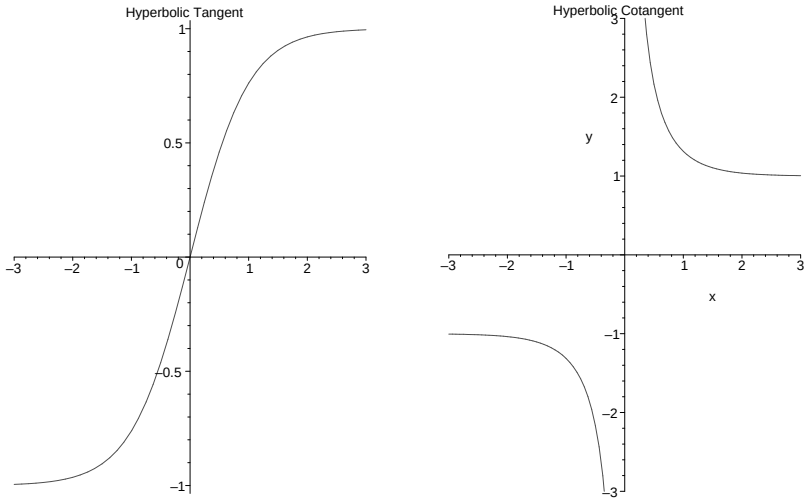
$$\sinh x = \frac{\mathrm{e}^x - \mathrm{e}^{-x}}{2}, \tag{3.39}$$

$$\cosh x = \frac{\mathrm{e}^x + \mathrm{e}^{-x}}{2}, \tag{3.40}$$

$$\tanh x = \frac{\mathrm{e}^x - \mathrm{e}^{-x}}{\mathrm{e}^x + \mathrm{e}^{-x}} = \frac{\sinh x}{\cosh x} \tag{3.41}$$

$$\coth x = \frac{\mathrm{e}^x + \mathrm{e}^{-x}}{\mathrm{e}^x - \mathrm{e}^{-x}} = \frac{\cosh x}{\sinh x} \tag{3.42}$$

are called *hyperbolic sine, hyperbolic cosine, hyperbolic tangent,* and *hyperbolic cotangent,* respectively. There are many analogies between these functions and their ordinary trigonometric counterparts.



The functions $\sinh x$ and $\tanh x$ are strictly increasing with $\sinh(\mathbb{R}) = \mathbb{R}$ and $\tanh(\mathbb{R}) = (-1, 1)$. Hence, their inverse functions are defined on $\mathbb{R}$ and on $(-1, 1)$, respectively, and are also strictly increasing and continuous. The function

$$\operatorname{arsinh} : \mathbb{R} \to \mathbb{R} \tag{3.43}$$

is given by $\operatorname{arsinh}(\sinh(x)) = x$, $x \in \mathbb{R}$ or $\sinh(\operatorname{arsinh}(y)) = y$, $y \in \mathbb{R}$.
The function

$$\operatorname{artanh} : (-1, 1) \to \mathbb{R} \tag{3.44}$$

is defined by $\operatorname{artanh}(\tanh(x)) = x$, $x \in \mathbb{R}$ or $\tanh(\operatorname{artanh}(y)) = y$, $y \in (-1, 1)$.
The function $\cosh$ is strictly increasing on the half line $\mathbb{R}_+$ with $\cosh(\mathbb{R}_+) = [1, \infty)$. Hence, the inverse function is defined on $[1, \infty)$ taking values in $\mathbb{R}_+$. It is also strictly increasing and continuous.

$$\operatorname{arcosh} : [1, \infty) \to \mathbb{R}_+ \tag{3.45}$$

is defined via $\operatorname{arcosh}(\cosh(x)) = x$, $x \geq 0$ or by $\cosh(\operatorname{arcosh}(y)) = y$, $y \geq 1$.
The function $\coth$ is strictly decreasing on the $x < 0$ and on $x > 0$ with $\coth(\mathbb{R} \setminus 0) = \mathbb{R} \setminus [-1, 1]$. Hence, the inverse function is defined on $\mathbb{R} \setminus [-1, 1]$ taking values in $\mathbb{R} \setminus 0$. It is also strictly decreasing and continuous.

$$\operatorname{arcoth} : \mathbb{R} \setminus [-1, 1] \to \mathbb{R} \tag{3.46}$$

is defined via $\operatorname{arcoth}(\coth(x)) = x$, $x \neq 0$ or by $\coth(\operatorname{arcoth}(y)) = y$, $y < -1$ or $y > 1$.

3.6 Appendix B

3.6.1 Monotonic Functions have One-Sided Limits

Proof of Theorem 3.8. By hypothesis, the set $\{f(t) \mid a < t < x\}$ is bounded above by $f(x)$, and therefore has a least upper bound which we shall denote by A. Evidently $A \leq f(x)$. We

have to show that $A = f(x-0)$.
Let $\varepsilon > 0$ be given. It follows from the definition of A as a least upper bound that there exists $\delta > 0$ such that $a < x - \delta < x$ and

$$A - \varepsilon < f(x-\delta) \le A. \tag{3.47}$$

Since f is monotonic, we have

$$f(x-\delta) < f(t) \le A, \quad \text{if} \quad x - \delta < t < x. \tag{3.48}$$

Combining (3.47) and (3.48), we see that

$$|\, f(t) - A \,| < \varepsilon \quad \text{if} \quad x - \delta < t < x.$$

Hence $f(x-0) = A$.
The second half of (3.3) is proved in precisely the same way. Next, if $a < x < y < b$, we see from (3.3) that

$$f(x+0) = \inf_{x<t<b} f(t) = \inf_{x<t<y} f(t). \tag{3.49}$$

The last equality is obtained by applying (3.3) to (a, y) instead of (a, b). Similarly,

$$f(y-0) = \sup_{a<t<y} f(t) = \sup_{x<t<y} f(t). \tag{3.50}$$

Comparison of the (3.49) and (3.50) gives (3.4). ∎

3.6.2 Proofs for $\sin x$ and $\cos x$ inequalities

Proof of Lemma 3.15. (a) By (3.22)

$$\cos x = \left(1 - \frac{1}{2!}x^2\right) + x^4\left(\frac{1}{4!} - \frac{1}{6!}x^2\right) + \cdots.$$

$0 < x < \sqrt{2}$ implies $1 - x^2/2 > 0$ and, moreover $1/(2n)! - x^2/(2n+2)! > 0$ for all $n \in \mathbb{N}$; hence $C(x) > 0$.
By (3.22),

$$\sin x = x\left(1 - \frac{1}{3!}x^2\right) + x^5\left(\frac{1}{5!} - \frac{1}{7!}x^2\right) + \cdots.$$

Now,

$$1 - \frac{1}{3!}x^2 > 0 \iff x < \sqrt{6}, \quad \frac{1}{5!} - \frac{1}{7!}x^2 > 0 \iff x < \sqrt{42}, \ldots.$$

Hence, $S(x) > 0$ if $0 < x < \sqrt{6}$. This gives (3.27). Similarly,

$$x - \sin x = x^3\left(\frac{1}{3!} - \frac{1}{5!}x^2\right) + x^7\left(\frac{1}{7!} - \frac{1}{9!}x^2\right) + \cdots,$$

and we obtain $\sin x < x$ if $0 < x < \sqrt{20}$. Finally we have to check whether $\sin x - x\cos x > 0$; equivalently

$$0 \overset{?}{<} x^3\left(\frac{1}{2!}-\frac{1}{3!}\right) - x^5\left(\frac{1}{4!}-\frac{1}{5!}\right) + x^7\left(\frac{1}{6!}-\frac{1}{7!}\right) - + \cdots$$
$$0 \overset{?}{<} x^3\left(\frac{2}{3!} - x^2\frac{4}{5!}\right) + x^7\left(\frac{6}{7!} - x^2\frac{8}{9!}\right) + \cdots$$

Now $\sqrt{10} > x > 0$ implies

$$\frac{2n}{(2n+1)!} - \frac{2n+2}{(2n+3)!}x^2 > 0$$

for all $n \in \mathbb{N}$. This completes the proof of (a)
(b) Using (3.23), we get

$$0 < x\cos x < \sin x \Longrightarrow 0 < x^2\cos^2 x < \sin^2 x$$
$$\Longrightarrow x^2\cos^2 x + \cos^2 x < 1 \Longrightarrow \cos^2 x < \frac{1}{1+x^2}.$$

(c) In the proof of Lemma 3.14 we have seen that $\cos x$ is strictly decreasing in $(0, \pi/2)$. By (3.23), $\sin x = \sqrt{1-\cos^2 x}$ is strictly increasing. Since $\sin x$ is an odd function, $\sin x$ is strictly increasing on $[-\pi/2, \pi/2]$. Since $\cos x = -\sin(x - \pi/2)$, the statement for $\cos x$ follows. ∎

3.6.3 Estimates for π

Proposition 3.21 *For real x we have*

$$\cos x = \sum_{k=0}^{n}(-1)^k\frac{x^{2k}}{(2k)!} + r_{2n+2}(x) \tag{3.51}$$
$$\sin x = \sum_{k=0}^{n}(-1)^k\frac{x^{2k+1}}{(2k+1)!} + r_{2n+3}(x), \tag{3.52}$$

where

$$|\, r_{2n+2}(x)\,| \le \frac{|\,x\,|^{2n+2}}{(2n+2)!} \quad \textit{if} \quad |\,x\,| \le 2n+3, \tag{3.53}$$
$$|\, r_{2n+3}(x)\,| \le \frac{|\,x\,|^{2n+3}}{(2n+3)!} \quad \textit{if} \quad |\,x\,| \le 2n+4. \tag{3.54}$$

Proof. Let

$$r_{2n+2}(x) = \pm\frac{x^{2n+2}}{(2n+2)!}\left(1 - \frac{x^2}{(2n+3)(2n+4)} \pm \cdots\right).$$

Put

$$a_k := \frac{x^{2k}}{(2n+3)(2n+4)\cdots(2n+2(k+1))}.$$

Then we have, by definition

$$r_{2n+2}(x) = \pm\frac{x^{2n+2}}{(2n+2)!}\,(1 - a_1 + a_2 - + \cdots)\,.$$

Since

$$a_k = a_{k-1}\frac{x^2}{(2n+2k+1)(2n+2k+2)},$$

$|\,x\,| \le 2n+3$ implies

$$1 > a_1 > a_2 > \cdots > 0$$

and finally as in the proof of the Leibniz criterion

$$0 \le 1 - a_1 + a_2 - a_3 + - \cdots \le 1.$$

Hence, $|\,r_{2n+2}(x)\,| \le |\,x\,|^{2n+2}/(2n+2)!$. The estimate for the remainder of the sine series is similar. ∎

This is an application of Proposition 3.21. For numerical calculations it is convenient to use the following order of operations

$$\cos x = \left(\cdots\left(\left(\left(\frac{-x^2}{2n(2n-1)}+1\right)\frac{-x^2}{(2n-2)(2n-3)}+1\right)\frac{-x^2}{(2n-4)(2n-5)}+1\right)\cdots \right.$$
$$\left. \cdots\right)\frac{-x^2}{2}+1+r_{2n+2}(x).$$

First we compute $\cos 1.5$ and $\cos 1.6$. Choosing $n = 7$ we obtain

$$\cos x = \left(\left(\left(\left(\left(\left(\frac{-x^2}{182}+1\right)\frac{-x^2}{132}+1\right)\frac{-x^2}{90}+1\right)\frac{-x^2}{56}+1\right)\frac{-x^2}{30}+1\right)\frac{-x^2}{12}+1\right)\frac{-x^2}{2}+$$
$$+1+r_{16}(x).$$

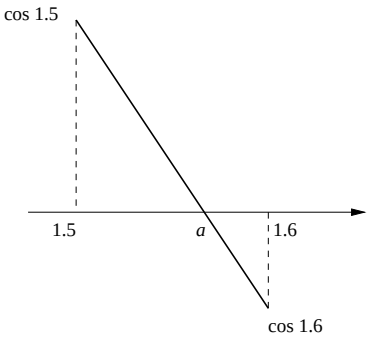
By Proposition 3.21

$$|\,r_{16}(x)\,| \le \frac{|\,x\,|^{16}}{16!} \le 0.9\cdot 10^{-10} \quad \text{if} \quad |\,x\,| \le 1.6.$$

The calculations give

$$\cos 1.5 = 0.07073720163 \pm 20\cdot 10^{-11} > 0, \cos 1.6 \;\; = -0.02919952239 \pm 20\cdot 10^{-11} < 0.$$

By the itermediate value theorem, $1.5 < \pi/2 < 1.6$.



Now we compute $\cos x$ for two values of x which are close to the linear interpolation

$$a = 1.5 + 0.1\frac{\cos 1.5}{\cos 1.5 - \cos 1.6} = 1.57078\ldots$$

$$\cos 1.5707 = 0.000096326273 \pm 20 \cdot 10^{-11} > 0,$$
$$\cos 1.5708 = -0.00000367326 \pm 20 \cdot 10^{-11} < 0.$$

Hence, $1.5707 < \pi/2 < 1.5708$.

The next linear interpolation gives

$$b = 1.5707 + 0.00001\frac{\cos 1.5707}{\cos 1.707 - \cos 1.708} = 1.570796326\ldots$$
$$\cos 1.570796326 = 0.00000000073 \pm 20 \cdot 10^{-11} > 0,$$
$$\cos 1.570796327 = -0.00000000027 \pm 20 \cdot 10^{-11} < 0.$$

Therefore $1.570796326 < \pi/2 < 1.570796327$ so that

$$\pi = 3.141592653 \pm 10^{-9}.$$

Chapter 4

Differentiation

4.1 The Derivative of a Function

We define the derivative of a function and prove the main properties like product, quotient and chain rule. We relate the derivative of a function with the derivative of its inverse function. We prove the mean value theorem and consider local extrema. Taylor's theorem will be formulated.

Definition 4.1 Let $f\colon (a,b) \to \mathbb{R}$ be a function and $x_0 \in (a,b)$. If the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{4.1}$$

exists, we call f *differentiable* at x_0. The limit is denoted by $f'(x_0)$. We say f *is differentiable* if f is differentiable at every point $x \in (a,b)$. We thus have associated to every function f a function f' whose domain is the set of points x_0 where the limit (4.1) exists; f' is called the *derivative* of f.

Sometimes the Leibniz notation is used to denote the derivative of f

$$f'(x_0) = \frac{\mathrm{d}f(x_0)}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x} f(x_0).$$

Remarks 4.1 (a) Replacing $x - x_0$ by h , we see that $f'(x_0) = \lim\limits_{h \to 0} \dfrac{f(x_0+h) - f(x_0)}{h}$.
(b) The limits

$$\lim_{h \to 0-0} \frac{f(x_0+h) - f(x_0)}{h}, \quad \lim_{h \to 0+0} \frac{f(x_0+h) - f(x_0)}{h}$$

are called *left-hand and right-hand derivatives of f in x_0*, respectively. In particular for $f\colon [a,b] \to \mathbb{R}$, we can consider the right-hand derivative at a and the left-hand derivative at b.

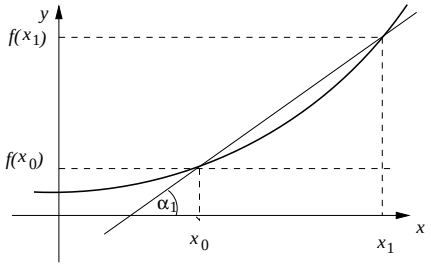
Example 4.1 (a) For $f(x) = c$ the constant function

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \to x_0} \frac{c - c}{x - x_0} = 0.$$

(b) For $f(x) = x$,

$$f'(x_0) = \lim_{x \to x_0} \frac{x - x_0}{x - x_0} = 1.$$

(c) The slope of the tangent line. Given a function $f\colon (a,b) \to \mathbb{R}$ which is differentiable at x_0. Then $f'(x_0)$ is the slope of the tangent line to the graph of f through the point $(x_0, f(x_0))$.



The slope of the secant line through $(x_0, f(x_0))$ and $(x_1, f(x_1))$ is

$$m = \tan \alpha_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

One can see: If x_1 approaches x_0, the secant line through $(x_0, f(x_0))$ and $(x_1, f(x_1))$ approaches the tangent line through $(x_0, f(x_0))$. Hence, the slope of the tangent line is the limit of the slopes of the secant lines if x approaches x_0:

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

Proposition 4.1 *If f is differentiable at $x_0 \in (a,b)$, then f is continuous at x_0.*

Proof. By Proposition 3.2 we have

$$\lim_{x \to x_0} (f(x) - f(x_0)) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}(x - x_0) = f'(x_0) \lim_{x \to x_0} (x - x_0) = f'(x_0) \cdot 0 = 0.$$

■

The converse of this proposition is not true. For example $f(x) = |\,x\,|$ is continuous in $\mathbb{R}$ but differentiable in $\mathbb{R} \setminus \{0\}$ since $\lim\limits_{h \to 0+0} \dfrac{|\,h\,|}{h} = 1$ whereas $\lim\limits_{h \to 0-0} \dfrac{|\,h\,|}{h} = -1$. Later we will become aquainted with a function which is continuous on the whole line without being differentiable at any point!

Proposition 4.2 *Let $f\colon (r,s) \to \mathbb{R}$ be a function and $a \in (r,s)$. Then f is differentiable at a if and only if there exists a number $c \in \mathbb{R}$ and a function φ defined in a neighborhood of a such that*

$$f(x) = f(a) + (x - a)c + \varphi(x), \tag{4.2}$$

where

$$\lim_{x \to a} \frac{\varphi(x)}{x - a} = 0. \tag{4.3}$$

In this case $f'(a) = c$.

The proposition says that a function f differentiable at a can be approximated by a linear function, in our case by

$$y = f(a) + (x - a)f'(a).$$

The graph of this linear function is the tangent line to the graph of f at the point $(a, f(a))$. Later we will use this point of view to define differentiability of functions $f\colon \mathbb{R}^n \to \mathbb{R}^m$.

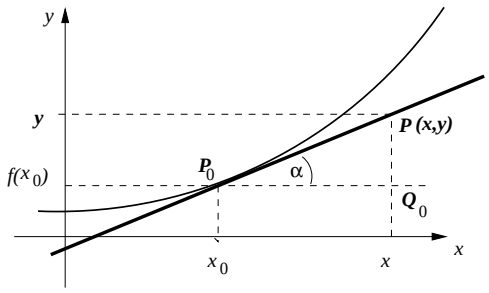
Proof. Suppose first f satisfies (4.2) and (4.3). Then

$$\lim_{x\to a}\frac{f(x)-f(a)}{x-a}=\lim_{x\to a}\left(c+\frac{\varphi(x)}{x-a}\right)=c.$$

Hence, f is differentiable at a with $f'(a)=c$.
Now, let f be differentiable at a with $f'(a)=c$. Put $\varphi(x)=f(x)-f(a)-(x-a)f'(a)$. Then

$$\lim_{x\to a}\frac{\varphi(x)}{x-a}=\lim_{x\to a}\frac{f(x)-f(a)}{x-a}-f'(a)=0.$$

■



Let us compute the linear function whose graph is the tangent line through $(x_0, f(x_0))$. Consider the rectangular triangle PP_0Q_0. By Example 4.1 (c) we have

$$f'(x_0)=\tan\alpha=\frac{y-y_0}{x-x_0},$$

such that the tangent line has the equation

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

This function is called the *linearization of f at x_0*. It is also the Taylor polynomial of degree 1 of f at x_0, see Section4.5 below.

Proposition 4.3 *Suppose f and g are defined on (a,b) and are differentible at a point $x\in(a,b)$. Then $f+g$, fg, and f/g are differentiable at x and*

(a) $(f+g)'(x)=f'(x)+g'(x)$;
(b) $(fg)'(x)=f'(x)g(x)+f(x)g'(x)$;
(c) $\left(\frac{f}{g}\right)'(x)=\frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}$.

In (c), *we assume that $g(x)\neq 0$.*

Proof. (a) Since

$$\frac{(f+g)(x+h)-(f+g)(x)}{h}=\frac{f(x+h)-f(x)}{h}+\frac{g(x+h)-g(x)}{h},$$

the claim follows from Proposition 3.2.
Let $h=fg$ and t be variable. Then

$$\begin{aligned}h(t)-h(x)&=f(t)(g(t)-g(x))+g(x)(f(t)-f(x))\\ \frac{h(t)-h(x)}{t-x}&=f(t)\frac{g(t)-g(x)}{t-x}+g(x)\frac{f(t)-f(x)}{t-x}.\end{aligned}$$

Noting that $f(t) \to f(x)$ as $t \to x$, (b) follows.
Next let $h = f/g$. Then

$$\begin{aligned}\frac{h(t)-h(x)}{t-x} &= \frac{\frac{f(t)}{g(t)} - \frac{f(x)}{g(x)}}{t-x} = \frac{f(t)g(x) - f(x)g(t)}{g(x)g(t)(t-x)} \\ &= \frac{1}{g(t)g(x)} \frac{f(t)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(t)}{t-x} \\ &= \frac{1}{g(t)g(x)} \left(g(x)\frac{f(t)-f(x)}{t-x} - f(x)\frac{g(t)-g(x)}{t-x} \right).\end{aligned}$$

Letting $t \to x$, and applying Propositions 3.2 and 4.1, we obtain (c). ∎

Example 4.2 (a) $f(x) = x^n$, $n \in \mathbb{Z}$. We will prove $f'(x) = nx^{n-1}$ by induction on $n \in \mathbb{N}$. The cases $n = 0, 1$ are OK by Example 4.1. Suppose the statement is true for some fixed n. We will show that $(x^{n+1})' = (n+1)x^n$.
By the product rule and the induction hypothesis

$$(x^{n+1})' = (x^n \cdot x)' = (x^n)'x + x^n(x') = nx^{n-1}\, x + x^n = (n+1)x^n.$$

This proves the claim for positive integers n. For negative n consider $f(x) = 1/x^{-n}$ and use the quotient rule.
(b) $(\mathrm{e}^x)' = \mathrm{e}^x$.

$$(\mathrm{e}^x)' = \lim_{h\to 0} \frac{\mathrm{e}^{x+h} - \mathrm{e}^x}{h} = \lim_{h\to 0} \frac{\mathrm{e}^x \mathrm{e}^h - \mathrm{e}^x}{h} = \mathrm{e}^x \lim_{h\to 0} \frac{\mathrm{e}^h - 1}{h} = \mathrm{e}^x; \tag{4.4}$$

the last equation simply follows from Example 3.9 (b)
(c) $(\sin x)' = \cos x$, $\quad(\cos x)' = -\sin x$. Using $\sin(x+y) - \sin(x-y) = 2\cos(x)\sin(y)$ we have

$$\begin{aligned}(\sin x)' &= \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} = \lim_{h\to 0} \frac{2\cos\frac{2x+h}{2}\sin\frac{h}{2}}{h} \\ &= \lim_{h\to 0} \cos\left(x + \frac{h}{2}\right) \lim_{h\to 0} \frac{\sin\frac{h}{2}}{\frac{h}{2}}.\end{aligned}$$

Since $\cos x$ is continuous and $\lim\limits_{h\to 0} \dfrac{\sin h}{h} = 1$ (by the argument after Lemma 3.15), we obtain $(\sin x)' = \cos x$. The proof for $\cos x$ is analogous.
(d) $(\tan x)' = \dfrac{1}{\cos^2 x}$. Using the quotiont rule for the function $\tan x = \sin x/\cos x$ we have

$$(\tan x)' = \frac{(\sin x)'\cos x - \sin x(\cos x)'}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

The next proposition deals with composite functions and is probably the most important statement about derivatives.

Proposition 4.4 (Chain rule) *Let $g\colon (\alpha,\beta) \to \mathbb{R}$ be differentiable at $x_0 \in (\alpha,\beta)$ and let $f\colon (a,b) \to \mathbb{R}$ be differentiable at $y_0 = g(x_0) \in (a,b)$. Then $h = f\circ g$ is differentiable at x_0, and*

$$h'(x_0) = f'(y_0)g'(x_0). \tag{4.5}$$

Proof. We have

$$\begin{aligned}\frac{f(g(x)) - f(g(x_0))}{x - x_0} &= \frac{f(g(x)) - f(g(x_0))}{g(x) - g(x_0)}\,\frac{g(x) - g(x_0)}{x - x_0}\\ &\underset{x\to x_0}{\longrightarrow} \lim_{y\to y_0}\frac{f(y) - f(y_0)}{y - y_0}\cdot g'(x_0) = f'(y_0)g'(x_0).\end{aligned}$$

Here we used that $y = g(x)$ tends to $y_0 = g(x_0)$ as $x \to x_0$, since g is continuous at x_0. ∎

Proposition 4.5 *Let $f\colon (a,b) \to \mathbb{R}$ be strictly monotonic and continuous. Suppose f is differentiable at x. Then the inverse function $g = f^{-1}\colon f((a,b)) \to \mathbb{R}$ is differentiable at $y = f(x)$ with*

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{f'(g(y))}. \tag{4.6}$$

Proof. Let $(y_n) \subset f((a,b))$ be a sequence with $y_n \to y$ and $y_n \neq y$ for all n. Put $x_n = g(y_n)$. Since g is continuous (by Corollary 3.9), $\lim_{n\to\infty} x_n = x$. Since g is injective, $x_n \neq x$ for all n. We have

$$\lim_{n\to\infty}\frac{g(y_n) - g(y)}{y_n - y} = \lim_{n\to\infty}\frac{x_n - x}{f(x_n) - f(x)} = \lim_{n\to\infty}\frac{1}{\frac{f(x_n)-f(x)}{x_n - x}} = \frac{1}{f'(x)}.$$

Hence $g'(y) = 1/f'(x) = 1/f'(g(y))$. ∎

We give some applications of the last two propositions.

Example 4.3 (a) Let $f\colon \mathbb{R} \to \mathbb{R}$ be differentiable; define $F\colon \mathbb{R} \to \mathbb{R}$ by $F(x) := f(ax + b)$ with some $a, b \in \mathbb{R}$. Then

$$F'(x) = af'(ax + b).$$

(b) In what follows f is the original function (with known derivative) and g is the inverse function to f. We fix the notion $y = f(x)$ and $x = g(y)$.
$\log\colon \mathbb{R}_+ \setminus \{0\} \to \mathbb{R}$ is the inverse function to $f(x) = \mathrm{e}^x$. By the above proposition

$$(\log y)' = \frac{1}{(\mathrm{e}^x)'} = \frac{1}{\mathrm{e}^x} = \frac{1}{y}.$$

(c) $x^\alpha = \mathrm{e}^{\alpha\log x}$. Hence, $(x^\alpha)' = (\mathrm{e}^{\alpha\log x})' = \mathrm{e}^{\alpha\log x}\alpha\dfrac{1}{x} = \alpha x^{\alpha-1}$.
(d) Suppose $f > 0$ and $g = \log f$. Then $g' = f'\dfrac{1}{f}$; hence $f' = f\,g'$.

(e) $\arcsin\colon [-1,1] \to \mathbb{R}$ is the inverse function to $y = f(x) = \sin x$. If $x \in (-1,1)$ then

$$(\arcsin(y))' = \frac{1}{(\sin x)'} = \frac{1}{\cos x}.$$

Since $y \in [-1,1]$ implies $x = \arcsin y \in [-\pi/2, \pi/2]$, $\cos x \geq 0$. Therefore, $\cos x = \sqrt{1-\sin^2 x} = \sqrt{1-y^2}$. Hence

$$(\arcsin y)' = \frac{1}{\sqrt{1-y^2}}, \quad -1 < y < 1.$$

Note that the derivative is not defined at the endpoints $y = -1$ and $y = 1$.
(f)

$$(\arctan y)' = \frac{1}{(\tan x)'} = \frac{1}{\frac{1}{\cos^2 x}} = \cos^2 x.$$

Since $y = \tan x$ we have

$$y^2 = \tan^2 x = \frac{\sin^2 x}{\cos^2 x} = \frac{1-\cos^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} - 1$$
$$\cos^2 x = \frac{1}{1+y^2}$$
$$(\arctan y)' = \frac{1}{1+y^2}.$$

4.2 The Derivatives of Elementary Functions

function	derivative
const.	0
$x^n\,(n \in \mathbb{N})$	nx^{n-1}
$x^\alpha\,(\alpha \in \mathbb{R}, x > 0)$	$\alpha x^{\alpha-1}$
e^x	e^x
$a^x,\ (a > 0)$	$a^x \log a$
$\log x$	$\dfrac{1}{x}$
$\log_a x$	$\dfrac{1}{x \log a}$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\dfrac{1}{\cos^2 x}$
$\cot x$	$-\dfrac{1}{\sin^2 x}$
$\sinh x$	$\cosh x$
$\cosh x$	$\sinh x$
$\tanh x$	$\dfrac{1}{\cosh^2 x}$
$\coth x$	$-\dfrac{1}{\sinh^2 x}$
$\arcsin x$	$\dfrac{1}{\sqrt{1-x^2}}$
$\arccos x$	$-\dfrac{1}{\sqrt{1-x^2}}$
$\arctan x$	$\dfrac{1}{1+x^2}$
$\operatorname{arccot} x$	$-\dfrac{1}{1+x^2}$
$\operatorname{arsinh} x$	$\dfrac{1}{\sqrt{x^2+1}}$
$\operatorname{arcosh} x$	$\dfrac{1}{\sqrt{x^2-1}}$
$\operatorname{artanh} x$	$\dfrac{1}{1-x^2}$
$\operatorname{arcoth} x$	$\dfrac{1}{1-x^2}$

4.2.1 Derivatives of Higher Order

Let $f\colon D \to \mathbb{R}$ be differentiable. If the derivative $f'\colon D \to \mathbb{R}$ is differentiable at $x \in D$, then

$$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} = f''(x) = (f')'(x)$$

is called the *second derivative* of f at x. Similarly, one defines inductively higher order derivatives. Continuing in this manner, we obtain functions

$$f,\ f',\ f'',\ f^{(3)},\ \ldots,\ f^{(k)}$$

each of which is the derivative of the preceding one. $f^{(n)}$ is called the *nth derivative* of f or the derivative of order n of f. We also use the Leibniz notation

$$f^{(k)}(x) = \frac{\mathrm{d}^k f(x)}{\mathrm{d}x^k} = \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^k f(x).$$

Definition 4.2 Let $D \subset \mathbb{R}$ and $k \in \mathbb{N}$ a positive integer. We denote by $\mathrm{C}^k(D)$ the set of all functions $f\colon D \to \mathbb{R}$ such that $f^{(k)}(x)$ exists for all $x \in D$ and $f^{(k)}(x)$ is continuous. Obviously $\mathrm{C}(D) \supset \mathrm{C}^1(D) \supset \mathrm{C}^2(D) \supset \cdots$. Further, we set

$$\mathrm{C}^\infty(D) = \bigcap_{k\in\mathbb{N}} \mathrm{C}^k(D) = \{f\colon D \to \mathbb{R} \mid f^{(k)}(x) \quad \text{exists} \quad \forall\, k \in \mathbb{N}, x \in D\}. \tag{4.7}$$

$f \in \mathrm{C}^k(D)$ is called *k times continuously differentiable*. $\mathrm{C}(D) = \mathrm{C}^0(D)$ is the vector space of continuous functions on D.

Using induction over n, one proves the following proposition.

Proposition 4.6 (Leibniz formula) *Let f and g be n times differentiable. Then fg is n times differentiable with*

$$(f(x)g(x))^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(k)}(x) g^{(n-k)}(x). \tag{4.8}$$

4.3 Local Extrema and the Mean Value Theorem

Many properties of a function f like monotony, convexity, and existence of local extrema can be studied using the derivative f'. From estimates for f' we obtain estimates for the growth of f.

Definition 4.3 Let $f\colon [a,b] \to \mathbb{R}$ be a function. We say that f has a *local maximum* at the point ξ, $\xi \in (a,b)$, if there exists $\delta > 0$ such that $f(x) \le f(\xi)$ for all $x \in [a,b]$ with $|\,x - \xi\,| < \delta$. *Local minima* are defined likewise.

We say that ξ is a *local extremum* if it is either a local maximum or a local minimum.

Proposition 4.7 *Let f be defined on $[a,b]$. If f has a local extremum at a point $\xi \in (a,b)$, and if $f'(\xi)$ exists, then $f'(\xi) = 0$.*

Proof. Suppose f has a local maximum at ξ. According with the definition choose $\delta > 0$ such that

$$a < \xi - \delta < \xi < \xi + \delta < b.$$

If $\xi - \delta < x < \xi$, then

$$\frac{f(x) - f(\xi)}{x - \xi} \geq 0.$$

Letting $x \to \xi$, we see that $f'(\xi) \geq 0$.
If $\xi < x < \xi + \delta$, then

$$\frac{f(x) - f(\xi)}{x - \xi} \leq 0.$$

Letting $x \to \xi$, we see that $f'(\xi) \leq 0$. Hence, $f'(\xi) = 0$. ∎

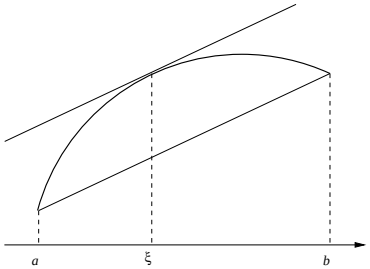
Remarks 4.2 (a) $f'(x) = 0$ is a necessary but not a sufficient condition for a local extremum in x. For example $f(x) = x^3$ has $f'(x) = 0$, but x^3 has no local extremum.
(b) If f attains its local extrema at the boundary, like $f(x) = x$ on $[0,1]$, we do not have $f'(\xi) = 0$.

Theorem 4.8 (Rolle's Theorem) *Let $f\colon [a,b] \to \mathbb{R}$ be continuous with $f(a) = f(b)$ and let f be differentiable in (a,b). Then there exists a point $\xi \in (a,b)$ with $f'(\xi) = 0$.*

In particular, between two zeros of a differentiable function there is a zero of its derivative.
Proof. If f is the constant function, the theorem is trivial since $f'(x) \equiv 0$ on (a,b). Otherwise, there exists $x_0 \in (a,b)$ such that $f(x_0) > f(a)$ or $f(x_0) < f(a)$. Then f attains its maximum or minimum, respectively, at a point $\xi \in (a,b)$. By Proposition 4.7, $f'(\xi) = 0$. ∎

Theorem 4.9 (Mean Value Theorem) *Let $f\colon [a,b] \to \mathbb{R}$ be continuous and differentiable in (a,b). Then there exists a point $\xi \in (a,b)$ such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a} \tag{4.9}$$



Geometrically, the mean value theorem states that there exists a tangent line through some point $(\xi, f(\xi))$ which is parallel to the secant line AB, $A = (a, f(a))$, $B = (b, f(b))$.

Theorem 4.10 (Generalized Mean Value Theorem) *Let f and g be continuous functions on $[a,b]$ which are differentiable on (a,b). Then there exists a point $\xi \in (a,b)$ such that*

$$(f(b)-f(a))g'(\xi) = (g(b)-g(a))f'(\xi).$$

Proof. Put

$$h(t) = (f(b)-f(a))g(t) - (g(b)-g(a))f(t).$$

Then h is continuous in $[a,b]$ and differentiable in (a,b) and

$$h(a) = f(b)g(a) - f(a)g(b) = h(b).$$

Rolle's theorem shows that there exists $\xi \in (a,b)$ such that

$$h'(\xi) = f(b)-f(a))h'(\xi) - (g(b)-g(a))f'(\xi) = 0.$$

The theorem follows.
In case that g' is nonzero on (a,b) and $g(b)-g(a) \neq 0$, the generalized MVT states the existence of some $\xi \in (a,b)$ such that

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(\xi)}{g'(\xi)}.$$

This is in particular true for $g(x) = x$ and $g' = 1$ which gives the assertion of the Mean Value Theorem. ∎

Remark 4.3 Note that the MVT fails if f is complex-valued, continuous on $[a,b]$, and differentiable on (a,b). Indeed, $f(x) = \mathrm{e}^{\mathrm{i}x}$ on $[0,2\pi]$ is a counter example. f is continuous on $[0,2\pi]$, differentiable on $(0,2\pi)$ and $f(0) = f(2\pi) = 1$. However, there is no $\xi \in (0,2\pi)$ such that $0 = \frac{f(2\pi)-f(0)}{2\pi} = f'(\xi) = \mathrm{i}\mathrm{e}^{\mathrm{i}\xi}$ since the exponential function has no zero, see (3.7) ($\mathrm{e}^z \cdot \mathrm{e}^{-z} = 1$) in Subsection 3.5.1.

Corollary 4.11 *Suppose f is differentiable on (a,b).*

If $f'(x) \geq 0$ for all $x \in (a,b)$, then f in monotonically increasing.
If $f'(x) = 0$ for all $x \in (a,b)$, then f is constant.
If $f'(x) \leq 0$ for all x in (a,b), then f is monotonically decreasing.

Proof. All conclusions can be read off from the equality

$$f(x) - f(t) = (x-t)f'(\xi)$$

which is valid for each pair x,t, $a < t < x < b$ and for some $\xi \in (t,x)$. ∎

4.3.1 Local Extrema and Convexity

Proposition 4.12 *Let $f\colon (a,b) \to \mathbb{R}$ be differentiable and suppose $f''(\xi)$ exists at a point $\xi \in (a,b)$. If*

$$f'(\xi) = 0 \quad \textit{and} \quad f''(\xi) > 0,$$

then f has a local minimum at ξ. Similarly, if

$$f'(\xi) = 0 \quad \textit{and} \quad f''(\xi) < 0,$$

f has a local maximum at ξ.

Remark 4.4 The condition of Proposition 4.12 is sufficient but not necessary for the existence of a local extremum. For example, $f(x) = x^4$ has a local minimum at $x = 0$, but $f''(0) = 0$.

Proof. We consider the case $f''(\xi) > 0$; the proof of the other case is analogous. Since

$$f''(\xi) = \lim_{x \to \xi} \frac{f'(x) - f'(\xi)}{x - \xi} > 0.$$

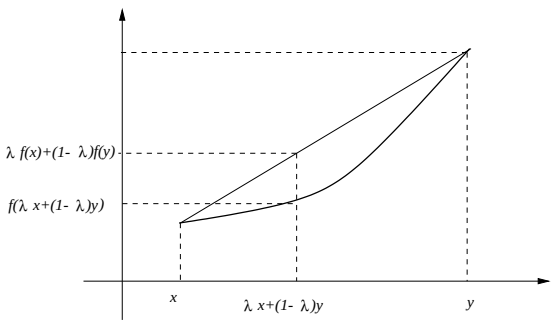
By Homework 10.4 there exists $\delta > 0$ such that

$$\frac{f'(x) - f'(\xi)}{x - \xi} > \frac{|\, f''(\xi)\,|}{2} > 0, \quad \text{for all } x \text{ with } \quad 0 < |\, x - \xi\,| < \delta.$$

Since $f'(\xi) = 0$ it follows that

$$\begin{aligned} f'(x) &< 0 \quad \text{if} \quad \xi - \delta < x < \xi, \\ f'(x) &> 0 \quad \text{if} \quad \xi < x < \xi + \delta. \end{aligned}$$

Hence, by Corollary 4.11, f is decreasing in $(\xi - \delta, \xi)$ and increasing in $(\xi, \xi + \delta)$. Therefore, f has a local minimum at ξ. ∎



Definition 4.4 A function $f\colon (a,b) \to \mathbb{R}$ is said to be *convex* if for all $x, y \in (a,b)$ and all $\lambda \in [0,1]$

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y). \tag{4.10}$$

A function f is said to be *concave* if $-f$ is convex.

Proposition 4.13 *(a) Convex functions are continuous.*
(b) Suppose $f\colon (a,b) \to \mathbb{R}$ is twice differentiable. Then f is convex if and only if $f''(x) \ge 0$ for all $x \in (a,b)$.

Proof. The proof is in Appendix C to this chapter. ∎

4.4 L'Hospital's Rule

Theorem 4.14 (L'Hospital's Rule) *Suppose f and g are differentiable in (a, b) and $g(x) \neq 0$ for all $x \in (a, b)$, where $-\infty \leq a < b \leq +\infty$. Suppose*

$$\lim_{x \to a+0} \frac{f'(x)}{g'(x)} = A. \tag{4.11}$$

If

$$\text{(a)} \quad \lim_{x \to a+0} f(x) = \lim_{x \to a+0} g(x) = 0 \quad \textit{or} \tag{4.12}$$

$$\text{(b)} \quad \lim_{x \to a+0} f(x) = \lim_{x \to a+0} g(x) = +\infty, \tag{4.13}$$

then

$$\lim_{x \to a+0} \frac{f(x)}{g(x)} = A. \tag{4.14}$$

The analogous statements are of course also true if $x \to b - 0$, or if $g(x) \to -\infty$.

Proof. First we consider the case of finite $a \in \mathbb{R}$. (a) One can extend the definition of f and g via $f(a) = g(a) = 0$. Then f and g are continuous at a. By the generalized mean value theorem, for every $x \in (a, b)$ there exists a $\xi \in (a, x)$ such that

$$\frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f(x)}{g(x)} = \frac{f'(\xi)}{g'(\xi)}.$$

If x approaches a then ξ also approaches a, and (a) follows.
(b) Now let $f(a + 0) = g(a + 0) = +\infty$. Given $\varepsilon > 0$ choose $\delta > 0$ such that

$$\left| \frac{f'(t)}{g'(t)} - A \right| < \varepsilon$$

if $t \in (a, a + \delta)$. By the generalized mean value theorem for any $x, y \in (a, a + \delta)$ with $x \neq y$,

$$\left| \frac{f(x) - f(y)}{g(x) - g(y)} - A \right| < \varepsilon.$$

We have

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(y)}{g(x) - g(y)} \frac{1 - \frac{g(y)}{g(x)}}{1 - \frac{f(y)}{f(x)}}.$$

The right factor tends to 1 as x approaches a, in particular there exists $\delta_1 > 0$ with $\delta_1 < \delta$ such that $x \in (a, a + \delta_1)$ implies

$$\left| \frac{f(x)}{g(x)} - \frac{f(x) - f(y)}{g(x) - g(y)} \right| < \varepsilon.$$

Further, the triangle inequality gives

$$\left| \frac{f(x)}{g(x)} - A \right| < 2\varepsilon.$$

This proves (b).
The case $x \to +\infty$ can be reduced to the limit process $y \to 0+0$ using the substitution $y = 1/x$. ∎

L'Hospital's rule applies also applies in the cases $A = +\infty$ and $A = -\infty$.

Example 4.4 (a) $\lim\limits_{x\to 0} \dfrac{\sin x}{x} = \lim\limits_{x\to 0} \dfrac{\cos x}{1} = 1.$
(b) $\lim\limits_{x\to 0+0} \dfrac{\sqrt{x}}{1-\cos x} = \lim\limits_{x\to 0+0} \dfrac{\frac{1}{2\sqrt{x}}}{\sin x} = \lim\limits_{x\to 0+0} \dfrac{1}{2\sqrt{x}\sin x} = +\infty.$
(c)

$$\lim_{x\to 0+0} x\log x = \lim_{x\to 0+0} \frac{\log x}{\frac{1}{x}} = \lim_{x\to 0+0} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = \lim_{x\to 0+0} -x = 0.$$

Remark 4.5 It is easy to transform other indefinite expressions to $\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$ of l'Hospital's rule.

$$\begin{aligned} 0\cdot\infty: &\quad f\cdot g = \frac{f}{\frac{1}{g}} \\ \infty-\infty: &\quad f-g = \frac{\frac{1}{g}-\frac{1}{f}}{\frac{1}{fg}}; \\ 0^0: &\quad f^g = \mathrm{e}^{g\log f}. \end{aligned}$$

Similarly, expressions of the form 1^∞ and ∞^0 can be transformed.

4.5 Taylor's Theorem

The aim of this section is to show how n times differentiable functions can be approximated by polynomials of degree n.
First consider a polynomial $p(x) = a_n x^n + \cdots + a_1 x + a_0$. We compute

$$\begin{aligned} p'(x) &= na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1, \\ p''(x) &= n(n-1)a_n x^{n-2} + (n-1)(n-2)a_{n-1}x^{n-2} + \cdots + 2a_2, \\ &\vdots \\ p^{(n)}(x) &= n!\, a_n. \end{aligned}$$

Inserting $x = 0$ gives $p(0) = a_0$, $p'(0) = a_1$, $p''(0) = 2a_2$,..., $p^{(n)}(0) = n!a_n$. Hence,

$$p(x) = p(0) + \frac{p''(0)}{1!}x + \frac{p''(0)}{2!}x^2 + \cdots + \frac{p^{(n)}(0)}{n!}x^n. \tag{4.15}$$

Now, fix $a \in \mathbb{R}$ and let $q(x) = p(x+a)$. Since $q^{(k)}(0) = p^{(k)}(a)$, (4.15) gives

$$p(x+a) = q(x) = \sum_{k=0}^{n} \frac{q^{(k)}(0)}{k!}\, x^k,$$

$$p(x+a) = \sum_{k=0}^{n} \frac{p^{(k)}(a)}{k!}\, x^k.$$

Replacing in the above equation $x + a$ by x yields

$$p(x) = \sum_{k=0}^{n} \frac{p^{(k)}(a)}{k!} (x-a)^k. \tag{4.16}$$

Theorem 4.15 (Taylor's Theorem) *Suppose f is a real function on $[r, s]$, $n \in \mathbb{N}$, $f^{(n)}$ is continuous on $[r, s]$, $f^{(n+1)}(t)$ exists for all $t \in (r, s)$. Let a and x be distinct points of $[r, s]$ and define*

$$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x-a)^k. \tag{4.17}$$

Then there exists a point ξ between x and a such that

$$f(x) = P_n(x) + \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-a)^{n+1}. \tag{4.18}$$

For $n = 0$, this is just the mean value theorem. $P_n(x)$ is called the *nth Taylor polynomial of f at $x = a$*, and the second summand of (4.18)

$$R_{n+1}(x, a) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-a)^{n+1}$$

is called the *Lagrangian remainder term*.
In general, the theorem shows that f can be approximated by a polynomial of degree n, and that (4.18) allows to estimate the error, if we know the bounds of $\left| f^{(n+1)}(x) \right|$.
Proof. Consider a and x to be fixed; let M be the number defined by

$$f(x) = P_n(x) + M(x-a)^{n+1}$$

and put

$$g(t) = f(t) - P_n(t) - M(t-a)^{n+1}, \quad \text{for} \quad r \le t \le s. \tag{4.19}$$

We have to show that $(n+1)!M = f^{(n+1)}(\xi)$ for some ξ between a and x. By (4.17) and (4.19),

$$g^{(n+1)}(t) = f^{(n+1)}(t) - (n+1)!M, \quad \text{for} \quad r < t < s. \tag{4.20}$$

Hence the proof will be complete if we can show that $g^{(n+1)}(\xi) = 0$ for some ξ between a and x.
Since $P_n^{(k)}(a) = f^{(k)}(a)$ for $k = 0, 1, \ldots, n$, we have

$$g(a) = g'(a) = \cdots = g^{(n)}(a) = 0.$$

Our choice of M shows that $g(x) = 0$, so that $g'(\xi_1) = 0$ for some ξ_1 between a and x, by Rolle's theorem. Since $g'(a) = 0$ we conclude similarly that $g''(\xi_2) = 0$ for some ξ_2 between a and ξ_1. After $n + 1$ steps we arrive at the conclusion that $g^{(n+1)}(\xi_{n+1}) = 0$ for some ξ_{n+1} between a and ξ_n, that is, between a and x. ∎

Definition 4.5 Suppose that f is a real function defined on $[r, s]$ such that $f^{(n)}(t)$ exists for all $t \in (r, s)$ and all $n \in \mathbb{N}$. Let x and a points of $[r, s]$. Then

$$T_f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k \tag{4.21}$$

is called the *Taylor series* of f at a.

Remarks 4.6 (a) The radius r of convergence of a Taylor series can be 0.
(b) If T_f converges, it may happen that $T_f(x) \neq f(x)$. If $T_f(x)$ at a point a converges to $f(x)$ in a certain neighborhood $U_r(a)$, $r > 0$, f is called to be *analytic* at a.

Example 4.5 We give an example for (b). Define $f\colon \mathbb{R} \to \mathbb{R}$ via

$$f(x) = \begin{cases} \mathrm{e}^{-1/x^2}, & \text{if} \quad x \neq 0, \\ 0, & \text{if} \quad x = 0. \end{cases}$$

We will show that $f \in \mathrm{C}^\infty(\mathbb{R})$ with $f^{(k)}(0) = 0$. For we will prove by induction on n that there exists a polynomial p_n such that

$$f^{(n)}(x) = p_n\left(\frac{1}{x}\right) \mathrm{e}^{-1/x^2}, \quad x \neq 0$$

and $f^{(n)}(0) = 0$. For $n = 0$ the statement is clear taking $p_0(x) \equiv 1$. Suppose the statement is true for n. First, let $x \neq 0$ then

$$f^{(n+1)}(x) = \left(p_n\left(\frac{1}{x}\right) \mathrm{e}^{-1/x^2}\right)' = \left(-\frac{1}{x^2} p_n'\left(\frac{1}{x}\right) + \frac{2}{x^3} p_n\left(\frac{1}{x}\right)\right) \mathrm{e}^{-1/x^2}.$$

Choose $p_{n+1}(t) = -p_n'(t)t^2 + 2p_n(t)t^3$.
Secondly,

$$f^{(n+1)}(0) = \lim_{h\to 0} \frac{f^{(n)}(h) - f^{(n)}(0)}{h} = \lim_{h\to 0} \frac{p_n\left(\frac{1}{h}\right) \mathrm{e}^{-1/h^2}}{h} = \lim_{x\to\pm\infty} x p_n(x) \mathrm{e}^{-x^2} = 0,$$

where we used Proposition 2.7 in the last equality.
Hence $T_f \equiv 0$ at 0—the Taylor series is identically 0—and $T_f(x)$ does not converge to $f(x)$ in a neigborhood of 0.

4.5.1 Examples of Taylor Series

(a) Power series coincide with their Taylor series.

$$\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad x \in \mathbb{R}, \quad \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}, \quad x \in (-1, 1).$$

(b) $f(x) = \log(1 + x)$, see Homework 13.4.

(c) $f(x) = (1+x)^\alpha$, $\alpha \in \mathbb{R}$, $a = 0$. We have

$$f^{(k)}(x) = \alpha(\alpha-1)\cdots(\alpha-k+1)(1+x)^{\alpha-k}, \quad \text{in particular} \quad f^{(k)}(0) = \alpha(\alpha-1)\cdots(\alpha-k+1).$$

Therefore,

$$(1+x)^\alpha = \sum_{k=1}^{n} \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}\, x^k + R_n(x) \tag{4.22}$$

The quotient test shows that the corresponding power series converges for $|\,x\,| < 1$. Consider the Lagrangian remainder term with $0 < \xi < x < 1$ and $n+1 > \alpha$. Then

$$|\,R_{n+1}(x)\,| = \left| \binom{\alpha}{n+1}(1+\xi)^{\alpha-n-1}x^{n+1} \right| \le \left| \binom{\alpha}{n+1}x^{n+1} \right| \le \left| \binom{\alpha}{n+1} \right| \longrightarrow 0$$

as $n \to \infty$. Hence,

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n, \quad 0 < x < 1. \tag{4.23}$$

(4.23) is called the *binomial series*. Its radius of convergence is $R = 1$. Looking at other forms of the remainder term gives that (4.23) holds for $-1 < x < 1$.
(d) $y = f(x) = \arctan x$. Since $y' = 1/(1+x^2)$ and $y'' = -2x/(1+x^2)^2$ we see that

$$y'(1+x^2) = 1.$$

Differentiating this n times and using Leibniz's formula, Proposition 4.6 we have

$$\sum_{k=0}^{n} (y')^{(k)}(1+x^2)^{(n-k)} \binom{n}{k} = 0.$$

$$\Longrightarrow \binom{n}{n} y^{(n+1)}\,(1+x^2) + \binom{n}{n-1} y^{(n)}\, 2x + \binom{n}{n-2} y^{(n-1)}\, 2 = 0;$$

$$x = 0: \quad y^{(n+1)} + n(n-1)y^{(n-1)} = 0.$$

This yields

$$y^{(n)}(0) = \begin{cases} 0, & \text{if} \quad n = 2k, \\ (-1)^k(2k)!, & \text{if} \quad n = 2k+1. \end{cases}$$

Therefore,

$$\arctan x = \sum_{k=0}^{n} \frac{(-1)^k}{2k+1}\, x^{2k+1} + R_{2n+2}(x). \tag{4.24}$$

One can prove that $-1 < x \le 1$ implies $R_{2n+2}(x) \to 0$ as $n \to \infty$. In particular, $x = 1$ gives

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - + \cdots.$$

4.6 Appendix C

Corollary 4.16 (to the mean value theorem) *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a differentiable function with*

$$f'(x) = cf(x) \quad \textit{for all } x \in \mathbb{R}, \tag{4.25}$$

where $c \in \mathbb{R}$ is a fixed number. Let $A = f(0)$. Then

$$f(x) = A\mathrm{e}^{cx} \quad \textit{for all } x \in \mathbb{R}. \tag{4.26}$$

Proof. Consider $F(x) = f(x)\mathrm{e}^{-cx}$. Using the product rule for derivatives and (4.25) we obtain

$$F'(x) = f'(x)\mathrm{e}^{-cx} + f(x)(-c)\mathrm{e}^{-cx} = (f'(x) - cf'(x))\,\mathrm{e}^{-cx} = 0.$$

By Corollary 4.11, $F(x)$ is constant. Since $F(0) = f(0) = A$, $F(x) = A$ for all $x \in \mathbb{R}$; the statement follows. ∎

The Continuity of derivatives

We have seen that there exist derivatives f' which are not continuous at some point. However, not every function is a derivative. In particular, derivatives which exist at every point of an interval have one important property: The intermediate value theorem holds. The precise statement follows.

Proposition 4.17 *Suppose f is differentiable on $[a, b]$ and suppose $f'(a) < \lambda < f'(b)$. Then there is a point $x \in (a, b)$ such that $f'(x) = \lambda$.*

Proof. Put $g(t) = f(t) - \lambda t$. Then g is differentiable and $g'(a) < 0$. Therefore, $g(t_1) < g(a)$ for some $t_1 \in (a, b)$. Similarly, $g'(b) > 0$, so that $g(t_2) < g(b)$ for some $t_2 \in (a, b)$. Hence, g attains its minimum in the *open* interval (a, b) in some point $x \in (a, b)$. By Proposition 4.7, $g'(x) = 0$. Hence, $f'(x) = \lambda$. ∎

Corollary 4.18 *If f is differentiable on $[a, b]$, then f' cannot have discontinuities of the first kind.*

Proof of Proposition 4.13. (a) Suppose first that $f'' \geq 0$ for all x. By Corollary 4.11, f' is increasing. Let $a < x < y < b$ and $\lambda \in [0, 1]$. Put $t = \lambda x + (1 - \lambda)y$. Then $x < t < y$ and by the mean value theorem there exist $\xi_1 \in (x, t)$ and $\xi_2 \in (t, y)$ such that

$$\frac{f(t) - f(x)}{t - x} = f'(\xi_1) \leq f'(\xi_2) = \frac{f(y) - f(t)}{y - t}.$$

Since $t - x = (1 - \lambda)(y - x)$ and $y - t = \lambda(y - x)$ it follows that

$$\begin{aligned} &\frac{f(t) - f(x)}{1 - \lambda} \leq \frac{f(y) - f(t)}{\lambda} \\ &\quad \Longrightarrow f(t) \leq \lambda f(x) + (1 - \lambda)f(y). \end{aligned}$$

Hence, f is convex.
(b) Let $f\colon (a,b) \to \mathbb{R}$ be convex and twice differentiable. Suppose to the contrary $f''(x_0) < 0$ for some $x_0 \in (a,b)$. Let $c = f'(x_0)$; put

$$\varphi(x) = f(x) - (x - x_0)c.$$

Then $\varphi\colon (a,b) \to \mathbb{R}$ is twice differentiable with $\varphi'(x_0) = 0$ and $\varphi''(x_0) < 0$. Hence, by Proposition 4.12, φ has a local maximum in x_0. By definition, there is a $\delta > 0$ such that $U_\delta(x_0) \subset (a,b)$ and

$$\varphi(x_0 - \delta) < \varphi(x_0), \quad \varphi(x_0 + \delta) < \varphi(x_0).$$

It follows that

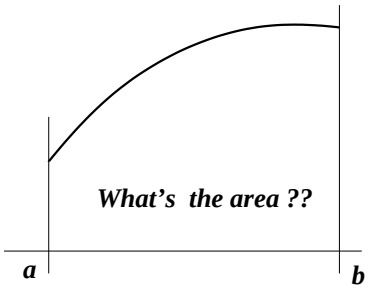
$$f(x_0) = \varphi(x_0) > \frac{1}{2}\left(\varphi(x_0 - \delta) + \varphi(x_0 + \delta)\right) = \frac{1}{2}\left(f(x_0 - \delta) + f(x_0 + \delta)\right).$$

This contradicts the convexity of f if we set $x = x_0 - \delta$, $y = x_0 + \delta$, and $\lambda = 1/2$. ∎

Chapter 5

Integration

In the first section of this chapter derivatives will not appear! Roughly speaking, integration generalizes "addition". The formula $\text{distance} = \text{velocity} \times \text{time}$ is only valid for constant velocity. The right formula is $s = \int_{t_0}^{t_1} v(t)\,\mathrm{d}t$. We need integrals to compute length of curves, areas of surfaces, and volumes.
The study of integrals requires a long preparation, but once this preliminary work has been completed, integrals will be an invaluable tool for creating new functions, and the derivative will reappear more powerful than ever. The relation between the integral and derivatives is given in the Fundamental Theorem of Calculus.



The integral formalizes a simple intuitive concept—that of area. It is not a surprise that to learn the definition of an intuitive concept can present great difficulties—"area" is certainly not an exception.

5.1 The Riemann–Stieltjes Integral

In this section we will only define the area of some very special regions—those which are bounded by the horizontal axis, the vertical lines through $(a, 0)$ and $(b, 0)$ and the graph of a function f such that $f(x) \geq 0$ for all x in $[a, b]$. If f is negative on a subinterval of $[a, b]$, the integral will represent the difference of the areas above and below the x-axis.
All intervals $[a, b]$ are finite intervals.

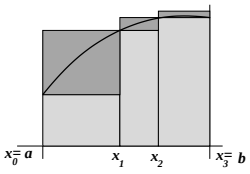
Definition 5.1 Let $[a, b]$ be an interval. By a *partition* of $[a, b]$ we mean a finite set of points $x_0, x_1, \ldots, x_n$, where

$$a = x_0 \leq x_1 \leq \cdots \leq x_n = b.$$

We write

$$\Delta x_i = x_i - x_{i-1}, \quad i = 1, \ldots, n.$$

Now suppose f is a bounded real function defined on $[a, b]$. Corresponding to each partition P of $[a, b]$ we put



$$M_i = \sup\{f(x) \mid x \in [x_{i-1}, x_i]\} \tag{5.1}$$

$$m_i = \inf\{f(x) \mid x \in [x_{i-1}, x_i]\} \tag{5.2}$$

$$U(P, f) = \sum_{i=1}^{n} M_i \Delta x_i, \quad L(P, f) = \sum_{i=1}^{n} m_i \Delta x_i, \tag{5.3}$$

and finally

$$\overline{\int_a^b} f \,\mathrm{d}x = \inf U(P, f), \tag{5.4}$$

$$\underline{\int_a^b} f \,\mathrm{d}x = \sup L(P, f), \tag{5.5}$$

where the infimum and supremum are taken over all partitions P of $[a, b]$. The left members of (5.4) and (5.5) are called the *upper* and *lower Riemann integrals* of f over $[a, b]$, respectively. If the upper and lower integrals are equal, we say that f Riemann-integrable on $[a, b]$ and we write $f \in \mathcal{R}$ (that is $\mathcal{R}$ denotes the Riemann-integrable functions), and we denote the common value of (5.4) and (5.5) by

$$\int_a^b f \,\mathrm{d}x \quad \text{or by} \quad \int_a^b f(x) \,\mathrm{d}x. \tag{5.6}$$

This is the *Riemann integral* of f over $[a, b]$.

Since f is bounded, there exist two numbers m and M such that $m \leq f(x) \leq M$ for all $x \in [a, b]$. Hence for every partition P

$$m(b - a) \leq L(P, f) \leq U(P, f) \leq M(b - a),$$

so that the numbers $L(P, f)$ and $U(P, f)$ form a bounded set. This shows that the upper and the lower integrals are defined for *every* bounded function f. The question of their equality, and hence the question of the integrability of f, is a more delicate one. Instead of investigating it separately for the Riemann integral, we shall immediately consider a more general situation.

Definition 5.2 Let α be a monotonically increasing function on $[a, b]$ (since $\alpha(a)$ and $\alpha(b)$ are finite, it follows that α is bounded on $[a, b]$). Corresponding to each partition P of $[a, b]$, we write

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}).$$

It is clear that $\Delta\alpha_i \geq 0$. For any real function f which is bounded on $[a, b]$ we put

$$U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta\alpha_i, \tag{5.7}$$

$$L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta\alpha_i, \tag{5.8}$$

where M_i and m_i have the same meaning as in Definition 5.1, and we define

$$\overline{\int_a^b} f \, \mathrm{d}\alpha = \inf U(P, f, \alpha), \tag{5.9}$$

$$\underline{\int_a^b} f \, \mathrm{d}\alpha = \sup U(P, f, \alpha), \tag{5.10}$$

where the infimum and the supremum are taken over all partitions P.
If the left members of (5.9) and (5.10) are equal, we denote their common value by

$$\int_a^b f \, \mathrm{d}\alpha \quad \text{or sometimes by} \quad \int_a^b f(x) \, \mathrm{d}\alpha(x). \tag{5.11}$$

This is the *Riemann–Stieltjes integral* (or simply the *Stieltjes integral*) of f with respect to α, over $[a, b]$. If (5.11) exists, we say that f is integrable with respect to α in the Riemann sense, and write $f \in \mathcal{R}(\alpha)$.

By taking $\alpha(x) = x$, the Riemann integral is seen to be a special case of the Riemann–Stieltjes integral. Let us mention explicitely, that in the general case, α need not even be continuous.
We shall now investigate the existence of the integral (5.11). Without saying so every time, f will be assumed real and bounded, and α increasing on $[a, b]$.

Definition 5.3 We say that a partition P^* is a *refinement* of the partition P if $P^* \supset P$ (that is, every point of P is a point of P^*). Given two partitions, P_1 and P_2, we say that P^* is their *common refinement* if $P^* = P_1 \cup P_2$.

Lemma 5.1 *If P^* is a refinement of P, then*

$$L(P, f, \alpha) \le L(P^*, f, \alpha) \quad \textit{and} \quad U(P, f, \alpha) \ge U(P^*, f, \alpha). \tag{5.12}$$

Proof. We only prove the first inequality of (5.12); the proof of the second one is analogous. Suppose first that P^* contains just one point more than P. Let this extra point be x^*, and suppose $x_{i-1} \le x^* < x_i$, where x_{i-1} and x_i are two consecutive points of P. Put

$$w_1 = \inf\{f(x) \mid x \in [x_{i-1}, x^*]\}, \quad w_2 = \inf\{f(x) \mid x \in [x^*, x_i]\}.$$

Clearly, $w_1 \ge m_i$ and $w_2 \ge m_i$ (since $\inf M \ge \inf N$ if $M \subset N$, see homework 1.4 (b)), where, as before, $m_i = \inf\{f(x) \mid x \in [x_{i-1}, x_i]\}$. Hence

$$\begin{aligned} L(P^*, f, \alpha) - L(P, f, \alpha) &= w_1(\alpha(x^*) - \alpha(x_{i-1})) + w_2(\alpha(x_i) - \alpha(x^*)) - m_i(\alpha(x_i) - \alpha(x_{i-1})) \\ &= (w_1 - m_i)(\alpha(x^*) - \alpha(x_{i-1})) + (w_2 - m_i)(\alpha(x_i) - \alpha(x^*)) \ge 0. \end{aligned}$$

If P^* contains k points more than P, we repeat this reasoning k times, and arrive at (5.12). ∎

Proposition 5.2

$$\underline{\int_a^b} f \, \mathrm{d}\alpha \leq \overline{\int_a^b} f \, \mathrm{d}\alpha.$$

Proof. Let P^* be the common refinement of two partitions P_1 and P_2. By Lemma 5.1

$$L(P_1, f, \alpha) \leq L(P^*, f, \alpha) \leq U(P^*, f, \alpha) \leq U(P_2, f, \alpha).$$

Hence

$$L(P_1, f, \alpha) \leq U(P_2, f, \alpha). \tag{5.13}$$

If P_2 is fixed and the supremum is taken over all P_1, (5.13) gives

$$\underline{\int_a^b} f \, \mathrm{d}\alpha \leq U(P_2, f, \alpha). \tag{5.14}$$

The proposition follows by taking the infimum over all P_2 in (5.14). ∎

Proposition 5.3 (Riemann Criterion) *$f \in \mathcal{R}(\alpha)$ on $[a, b]$ if and only if for every $\varepsilon > 0$ there exists a partition P such that*

$$U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon. \tag{RC}$$

Proof. For every P we have

$$L(P, f, \alpha) \leq \underline{\int_a^b} f \, \mathrm{d}\alpha \leq \overline{\int_a^b} f \, \mathrm{d}\alpha \leq U(P, f, \alpha).$$

Thus (RC) implies

$$0 \leq \overline{\int_a^b} f \, \mathrm{d}\alpha - \underline{\int_a^b} f \, \mathrm{d}\alpha < \varepsilon.$$

since the above inequality can be satisfied for every $\varepsilon > 0$, we have

$$\overline{\int_a^b} f \, \mathrm{d}\alpha = \underline{\int_a^b} f \, \mathrm{d}\alpha,$$

that is $f \in \mathcal{R}(\alpha)$.
Conversely, suppose $f \in \mathcal{R}(\alpha)$, and let $\varepsilon > 0$ be given. Then there exist partitions P_1 and P_2 such that

$$U(P_2, f, \alpha) - \int_a^b f \, \mathrm{d}\alpha < \frac{\varepsilon}{2}, \quad \int_a^b f \, \mathrm{d}\alpha - L(P_1, f, \alpha) < \frac{\varepsilon}{2}. \tag{5.15}$$

We choose P to be the common refinement of P_1 and P_2. Then Lemma 5.1, together with (5.15), shows that

$$U(P, f, \alpha) \leq U(P_2, f, \alpha) < \int_a^b f \, \mathrm{d}\alpha + \frac{\varepsilon}{2} < L(P_1, f, \alpha) + \varepsilon \leq L(P, f, \alpha) + \varepsilon,$$

so that (RC) holds for this partition P. ∎

Proposition 5.3 furnishes a convenient criterion for integrability. Before we apply it, we state some closely related facts.

Lemma 5.4 (a) *If* (RC) *holds for P and some ε, then* (RC) *holds with the same ε for every refinement of P.*
(b) *If* (RC) *holds for $P = \{x_0, \ldots, x_n\}$ and if s_i, t_i are arbitrary points in $[x_{i-1}, x_i]$, then*

$$\sum_{i=1}^{n} | f(s_i) - f(t_i) | \Delta\alpha_i < \varepsilon.$$

(c) *If $f \in \mathcal{R}(\alpha)$ and* (RC) *holds as in* (b)*, then*

$$\left| \sum_{i=1}^{n} f(t_i)\Delta\alpha_i - \int_a^b f \, \mathrm{d}\alpha \right| < \varepsilon.$$

Proof. Lemma 5.1 implies (a). Under the assumptions made in (b), both $f(s_i)$ and $f(t_i)$ lie in $[m_i, M_i]$, so that $| f(s_i) - f(t_i) | \leq M_i - m_i$. Thus

$$\sum_{i=1}^{n} | f(t_i) - f(s_i) | \Delta\alpha_i \leq U(P, f, \alpha) - L(P, f, \alpha),$$

which proves (b). The obvious inequalities

$$L(P, f, \alpha) \leq \sum_i f(t_i)\Delta\alpha_i \leq U(P, f, \alpha)$$

and

$$L(P, f, \alpha) \leq \int_a^b f \, \mathrm{d}\alpha \leq U(P, f, \alpha)$$

prove (c). ∎

Theorem 5.5 *If f is continuous on $[a, b]$ then $f \in \mathcal{R}(\alpha)$ on $[a, b]$.*

Proof. Let $\varepsilon > 0$ be given. Choose $\eta > 0$ so that

$$(\alpha(b) - \alpha(a))\eta < \varepsilon.$$

Since f is uniformly continuous on $[a, b]$ (Proposition 3.7), there exists a $\delta > 0$ such that

$$| f(x) - f(t) | < \eta \tag{5.16}$$

if $x, t \in [a, b]$ and $| x - t | < \delta$. If P is any partition of $[a, b]$ such that $\Delta x_i < \delta$ for all i, then (5.16) implies that

$$M_i - m_i \leq \eta, \quad i = 1, \ldots, n \tag{5.17}$$

and therefore

$$U(P,f,\alpha)-L(P,f,\alpha)=\sum_{i=1}^{n}(M_i-m_i)\Delta\alpha_i\le\eta\sum_{i=1}^{n}\Delta\alpha_i=\eta(\alpha(b)-\alpha(a))<\varepsilon.$$

By Proposition 5.3, $f\in\mathcal{R}(\alpha)$. ∎

Example 5.1 (a) The proof of Theorem 5.5 together with Lemma 5.4 shows that

$$\left|\sum_{i=1}^{n}f(t_i)\Delta\alpha_i-\int_a^b f\,\mathrm{d}\alpha\right|<\varepsilon$$

if $\Delta x_i<\delta$.
We compute $I=\int_a^b\sin x\,\mathrm{d}x$. Let $\varepsilon>0$. Since $\sin x$ is continuous, $f\in\mathcal{R}$. There exists $\delta>0$ such that $|\,x-t\,|<\delta$ implies

$$|\sin x-\sin t\,|<\frac{\varepsilon}{b-a}. \tag{5.18}$$

In this case (RC) is satisfied and consequently

$$\left|\sum_{i=1}^{n}\sin(t_i)\Delta x_i-\int_a^b\sin x\,\mathrm{d}x\right|<\varepsilon$$

for every partition P with $\Delta x_i<\delta$, $i=1,\dots,n$.
For we choose an equidistant partition of $[a,b]$, $x_i=a+(b-a)i/n$, $i=0,\dots,n$. Then $h=\Delta x_i=(b-a)/n$ and the condition (5.18) is satisfied provided $n>\dfrac{(b-a)^2}{\varepsilon}$. We have, by addition the formula $\cos(x-y)-\cos(x+y)=2\sin x\sin y$

$$\begin{aligned}\sum_{i=1}^{n}\sin x_i\,\Delta x_i=\sum_{i=1}^{n}\sin(a+ih)h&=\frac{h}{2\sin h/2}\sum_{i=1}^{n}2\,\sin h/2\,\sin(a+ih)\\&=\frac{h}{2\sin h/2}\sum_{i=1}^{n}\left(\cos(a+(i-1/2)h)-\cos(a+(i+1/2)h)\right)\\&=\frac{h}{2\sin h/2}\left(\cos(a+h/2)-\cos(a+(n+1/2)h)\right)\\&=\frac{h/2}{\sin h/2}\left(\cos(a+h/2)-\cos(b+h/2)\right)\end{aligned}$$

Since $\lim_{h\to0}\sin h/h=1$ and $\cos x$ is continuous, we find that the above expression tends to $\cos a-\cos b$. Hence $\int_a^b\sin x\,\mathrm{d}x=\cos a-\cos b$.
(b) For $x\in[a,b]$ define

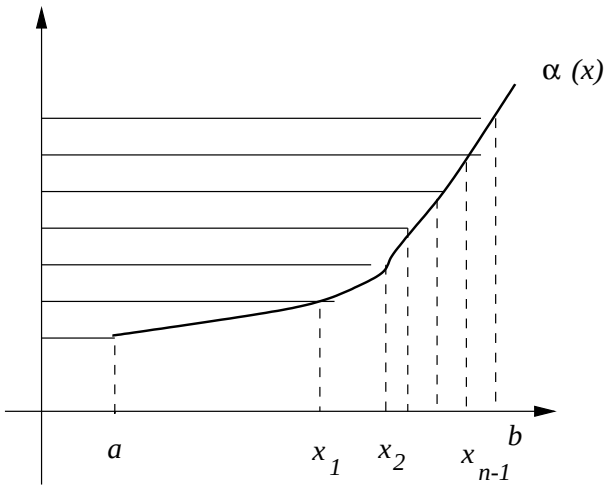
$$f(x)=\begin{cases}1, & x\in\mathbb{Q},\\ 0, & x\notin\mathbb{Q}.\end{cases}$$

We will show $f\notin\mathcal{R}$. Let P be any partition of $[a,b]$. Since any interval contains rational as well as irrational points, $m_i=0$ and $M_i=1$ for all i. Hence $L(P,f)=0$ whereas

$U(P,f) = \sum_{i=1}^{n} \Delta x_i = b - a$. We conclude that the upper and lower Riemann integrals don't coincide; $f \notin \mathcal{R}$. A similar reasoning shows $f \notin \mathcal{R}(\alpha)$ if $\alpha(b) > \alpha(a)$ since $L(P,f,\alpha) = 0 < U(P,f,\alpha) = \alpha(b) - \alpha(a)$.

Proposition 5.6 *If f is monotonic on $[a,b]$, and α is continuous on $[a,b]$, then $f \in \mathcal{R}(\alpha)$.*

Proof.



Let $\varepsilon > 0$ be given. For any positive integer n, choose a partition such that

$$\Delta\alpha_i = \frac{\alpha(b) - \alpha(a)}{n}, \quad i = 1, \ldots, n.$$

This is possible by the intermediate value theorem (Theorem 3.5) since α is continuous.

We suppose that f is monotonically increasing (the proof is analogous in the other case). Then

$$M_i = f(x_i), \quad m_i = f(x_{i-1}), \quad i = 1, \ldots, n,$$

so that

$$\begin{aligned} U(P,f,\alpha) - L(P,f,\alpha) &= \frac{\alpha(b) - \alpha(a)}{n} \sum_{i=1}^{n} (f(x_i) - f(x_{i-1})) \\ &= \frac{\alpha(b) - \alpha(a)}{n} (f(b) - f(a)) < \varepsilon \end{aligned}$$

if n is taken large enough. By Proposition 5.3, $f \in \mathcal{R}(\alpha)$. ∎

Without proofs which can be found in [Rud76, pp. 126 –128] we note the following facts.

Proposition 5.7 *If f is bounded on $[a,b]$, f has finitely many points of discontinuity on $[a,b]$, and α is continuous at every point at which f is discontinuous. Then $f \in \mathcal{R}(\alpha)$.*

Proof. We give an idea of the proof in case of the Riemann integral ($\alpha(x) = x$) and one single discontinuity at c, $a < c < b$. For, let $\varepsilon > 0$ be given and $m \le f(x) \le M$ for all $x \in [a,b]$ and put $C = M - m$. First choose point a' and b' with $a < a' < c < b' < b$ and $C(b' - a') < \varepsilon$. Let f_j, $j = 1, 2$, denote the restriction of f to the subintervals $I_1 = [a, a']$ and $I_2 = [b, b']$, respectively. Since f_j is continuous on I_j, $f_j \in \mathcal{R}$ over I_j and therefore, by the Riemann criterion, there exist partitions P_j, $j = 1, 2$, of I_j such that $U(P_j, f_j) - L(P_j, f_j) < \varepsilon$, $j = 1, 2$. Let $P = P_1 \cup P_2$ be a partition of $[a,b]$. Then

$$\begin{aligned} U(P,f) - L(P,f) &= U(P_1,f_1) - L(P_1,f) + U(P_2,f) - L(P_2,f) + (M_0 - m_0)(b' - a') \\ &\le \varepsilon + \varepsilon + C(b' - a') < 3\varepsilon, \end{aligned}$$

where M_0 and m_0 are the supremum and infimum of $f(x)$ on $[a', b']$. The Riemann criterion is satisfied for f on $[a, b]$, $f \in \mathcal{R}$. ∎

Proposition 5.8 *If $f \in \mathcal{R}(\alpha)$ on $[a, b]$, $m \leq f(x) \leq M$, φ is continuous on $[m, M]$, and $h(x) = \varphi(f(x))$ on $[a, b]$. Then $h \in \mathcal{R}(\alpha)$ on $[a, b]$.*

Remark 5.1 (a) A bounded function f is Riemann-integrable on $[a, b]$ if and only if f is continuous almost everywhere on $[a, b]$. (The proof of this fact can be found in [Rud76, Theorem 11.33]).
"Almost everywhere" means that the discontinuities form a set of (Lebesgue) measure 0. A set $M \subset \mathbb{R}$ has measure 0 if for given $\varepsilon > 0$ there exist intervals I_n, $n \in \mathbb{N}$ such that $M \subset \bigcup_{n\in\mathbb{N}} I_n$ and $\sum_{n\in\mathbb{N}} | I_n | < \varepsilon$. Here, $| I |$ denotes the length of the interval. Examples of sets of measure 0 are finite sets, countable sets, and the Cantor set (which is uncountable).
(b) Note that such a "chaotic" function (at point 0) as

$$f(x) = \begin{cases} \cos \frac{1}{x}, & x \neq 0, \\ 0, & x = 0, \end{cases}$$

is integrable on $[-\pi, \pi]$ since there is only one single discontinuity at 0.

5.1.1 Properties of the Integral

Proposition 5.9 (a) *If $f_1, f_2 \in \mathcal{R}(\alpha)$ on $[a, b]$ then $f_1 + f_2 \in \mathcal{R}(\alpha)$, $cf \in \mathcal{R}(\alpha)$ for every constant c and*

$$\int_a^b (f_1 + f_2)\,\mathrm{d}\alpha = \int_a^b f_1\,\mathrm{d}\alpha + \int_a^b f_2\,\mathrm{d}\alpha, \quad \int_a^b cf\,\mathrm{d}\alpha = c\int_a^b f\,\mathrm{d}\alpha.$$

(b) *If $f_1, f_2 \in \mathcal{R}(\alpha)$ and $f_1(x) \leq f_2(x)$ on $[a, b]$, then*

$$\int_a^b f_1\,\mathrm{d}\alpha \leq \int_a^b f_2\,\mathrm{d}\alpha.$$

(c) *If $f \in \mathcal{R}(\alpha)$ on $[a, b]$ and if $a < c < b$, then $f \in \mathcal{R}(\alpha)$ on $[a, c]$ and on $[c, b]$, and*

$$\int_a^b f\,\mathrm{d}\alpha = \int_a^c f\,\mathrm{d}\alpha + \int_c^b f\,\mathrm{d}\alpha.$$

(d) *If $f \in \mathcal{R}(\alpha)$ on $[a, b]$ and $| f(x) | \leq M$ on $[a, b]$, then*

$$\left| \int_a^b f\,\mathrm{d}\alpha \right| \leq M(\alpha(b) - \alpha(a)).$$

(e) *If $f \in \mathcal{R}(\alpha_1)$ and $f \in \mathcal{R}(\alpha_2)$, then $f \in \mathcal{R}(\alpha_1 + \alpha_2)$ and*

$$\int_a^b f\mathrm{d}(\alpha_1 + \alpha_2) = \int_a^b f\mathrm{d}\alpha_1 + \int_a^b f\mathrm{d}\alpha_2;$$

if $f \in \mathcal{R}(\alpha)$ and c is a positive constant, then $f \in \mathcal{R}(c\alpha)$ and

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

Proof. If $f = f_1 + f_2$ and P is any partition of $[a, b]$, we have

$$L(P, f_1, \alpha) + L(P, f_2, \alpha) \leq L(P, f, \alpha) \leq U(P, f, \alpha) \leq U(P, f_1, \alpha) + U(P, f_2, \alpha) \tag{5.19}$$

since $\inf_{I_i} f_1 + \inf_{I_i} f_2 \leq \inf_{I_i}(f_1 + f_2)$ and $\sup_{I_i} f_1 + \sup_{I_i} f_2 \geq \sup_{I_i}(f_1 + f_2)$.
If $f_1 \in \mathcal{R}(\alpha)$ and $f_2 \in \mathcal{R}(\alpha)$, let $\varepsilon > 0$ be given. There are partitons P_j, $j = 1, 2$, such that

$$U(P_j, f_j, \alpha) - L(P_j, f_j, \alpha) < \varepsilon.$$

These inequalities persist if P_1 and P_2 are replaced by their common refinement P. Then (5.19) implies

$$U(P, f, \alpha) - L(P, f, \alpha) < 2\varepsilon$$

which proves that $f \in \mathcal{R}(\alpha)$. With the same P we have

$$U(P, f_j, \alpha) < \int_a^b f_j \, \mathrm{d}\alpha + \varepsilon, \quad j = 1, 2;$$

since $L(P, f, \alpha) < \int_a^b f \, \mathrm{d}\alpha < U(P, f, \alpha)$; hence (5.19) implies

$$\int_a^b f \, \mathrm{d}\alpha \leq U(P, f, \alpha) < \int_a^b f_1 \, \mathrm{d}\alpha + \int_a^b f_2 \, \mathrm{d}\alpha + 2\varepsilon.$$

Since ε was arbitrary, we conclude that

$$\int_a^b f \, \mathrm{d}\alpha \leq \int_a^b f_1 \, \mathrm{d}\alpha + \int_a^b f_2 \, \mathrm{d}\alpha. \tag{5.20}$$

If we replace f_1 and f_2 in (5.20) by $-f_1$ and $-f_2$, respectively, the inequality is reversed, and the equality is proved.
(b) Put $f = f_2 - f_1$. It suffices to prove that $\int_a^b f \, \mathrm{d}\alpha \geq 0$. For every partition P we have $m_i \geq 0$ since $f \geq 0$. Hence

$$\int_a^b f \, \mathrm{d}\alpha \geq L(P, f, \alpha) = \sum_{i=1}^n m_i \Delta\alpha_i \geq 0$$

since in addition $\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}) \geq 0$ (α is increasing).
The proofs of the other assertions are so similar that we omit the details. In part (c) the point is that (by passing to refinements) we may restrict ourselves to partitions which contain the point c, in approximating $\int_a^b f \, \mathrm{d}\alpha$, cf. Homework 14.5. ∎

Note that in (c), $f \in \mathcal{R}(\alpha)$ on $[a, c]$ and on $[c, b]$ in general does not imply that $f \in \mathcal{R}(\alpha)$ on $[a, b]$. For example consider the interval $[-1, 1]$ with

$$f(x) = \alpha(x) = \begin{cases} 0, & -1 \leq x < 0, \\ 1, & 0 \leq x \leq 1. \end{cases}$$

Then $\int_0^1 f\,\mathrm{d}\alpha = 0$. The integral vanishes since α is constant on $[0,1]$. However, $f \notin \mathcal{R}(\alpha)$ on $[-1,1]$ since for any partition P including the point 0, we have $U(P,f,\alpha) = 1$ and $L(P,f,\alpha) = 0$.

Proposition 5.10 *If $f, g \in \mathcal{R}(\alpha)$ on $[a,b]$, then*
(a) $fg \in \mathcal{R}(\alpha)$*;*
(b) $|f| \in \mathcal{R}(\alpha)$ *and* $\left|\int_a^b f\,\mathrm{d}\alpha\right| \le \int_a^b |f|\,\mathrm{d}\alpha$.

Proof. If we take $\varphi(t) = t^2$, Proposition 5.8 shows that $f^2 \in \mathcal{R}(\alpha)$ if $f \in \mathcal{R}(\alpha)$. The identity

$$4fg = (f+g)^2 - (f-g)^2$$

completes the proof of (a).
If we take $\varphi(t) = |t|$, Proposition 5.8 shows that $|f| \in \mathcal{R}(\alpha)$. Choose $c = \pm 1$ so that $c\int f\,\mathrm{d}\alpha \ge 0$. Then

$$\left|\int f\,\mathrm{d}\alpha\right| = c\int f\,\mathrm{d}\alpha = \int cf\,\mathrm{d}\alpha \le \int |f|\,\mathrm{d}\alpha,$$

since $\pm f \le |f|$. ∎

The *unit step function* or *Heaviside function* $H(x)$ is defined by $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \ge 0$.

Example 5.2 (a) If $a < s < b$, f is bounded on $[a,b]$, f is continuous at s, and $\alpha(x) = H(x-s)$, then

$$\int_a^b f\,\mathrm{d}\alpha = f(s).$$

For the proof, consider the partition P with $n = 3$; $a = x_0 < x_1 < s = x_2 < x_3 = b$. Then $\Delta\alpha_1 = \Delta\alpha_3 = 0$, $\Delta\alpha_2 = 1$, and

$$U(P,f,\alpha) = M_2, \quad L(P,f,\alpha) = m_2.$$

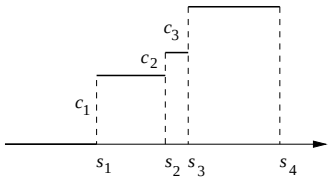
Since f is continuous at s, we see that M_2 and m_2 converge to $f(s)$ as $x \to s$.
(b) Suppose $c_n \ge 0$ for all $n = 1,\dots,N$ and (s_n), $n = 1,\dots,N$, is a strictly increasing finite sequence of distinct points in (a,b). Further, $\alpha(x) = \sum_{n=1}^N c_n H(x - s_n)$. Then

$$\int_a^b f\ \mathrm{d}\alpha = \sum_{n=1}^N c_n f(s_n).$$

This follows from (a) and Proposition 5.9 (e).

Proposition 5.11 *Suppose* $c_n \geq 0$ *for all positive integers* $n \in \mathbb{N}$, $\sum_{n=1}^{\infty} c_n$ *converges,* (s_n) *is a strictly increasing sequence of distinct points in* (a, b)*, and*

$$\alpha(x) = \sum_{n=1}^{\infty} c_n H(x - s_n). \tag{5.21}$$



Let f *be continuous on* $[a, b]$*. Then*

$$\int_a^b f \, \mathrm{d}\alpha = \sum_{n=1}^{\infty} c_n f(s_n). \tag{5.22}$$

Proof. The comparison test shows that the series (5.21) converges for every x. Its sum α is evidently an increasing function with $\alpha(a) = 0$ and $\alpha(b) = \sum c_n$. Let $\varepsilon > 0$ be given, choose N so that

$$\sum_{n=N+1}^{\infty} c_n < \varepsilon.$$

Put

$$\alpha_1(x) = \sum_{n=1}^{N} c_n H(x - s_n), \quad \alpha_2(x) = \sum_{n=N+1}^{\infty} c_n H(x - s_n).$$

By Proposition 5.9 and Example 5.2

$$\int_a^b f \, \mathrm{d}\alpha_1 = \sum_{n=1}^{N} c_n f(s_n).$$

Since $\alpha_2(b) - \alpha_2(a) < \varepsilon$, by Proposition 5.9 (d),

$$\left| \int_a^b f \, \mathrm{d}\alpha_2 \right| \leq M\varepsilon,$$

where $M = \sup | f(x) |$. Since $\alpha = \alpha_1 + \alpha_2$ it follows that

$$\left| \int_a^b f \, \mathrm{d}\alpha - \sum_{n=1}^{N} c_n f(s_n) \right| \leq M\varepsilon.$$

If we let $N \to \infty$ we obtain (5.22). ∎

Proposition 5.12 *Assume that* α *is increasing and* $\alpha' \in \mathcal{R}$ *on* $[a, b]$*. Let* f *be a bounded real function on* $[a, b]$*.*
Then $f \in \mathcal{R}(\alpha)$ *if and only if* $f\alpha' \in \mathcal{R}$*. In that case*

$$\int_a^b f \, \mathrm{d}\alpha = \int_a^b f(x)\alpha'(x) \, \mathrm{d}x. \tag{5.23}$$

The statement remains true if α *is continuous on* $[a, b]$ *and differentiable up to finitely many points* $c_1, c_2, \ldots, c_n$*.*

Proof. Let $\varepsilon > 0$ be given and apply the Riemann criterion Proposition 5.3 to α': There is a partition $P = \{x_0, \dots, x_n\}$ of $[a, b]$ such that

$$U(P, \alpha') - L(P, \alpha') < \varepsilon. \tag{5.24}$$

The mean value theorem furnishes points $t_i \in [x_{i-1}, x_i]$ such that

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}) = \alpha'(t_i)(x_i - x_{i-1}) = \alpha'(t_i)\Delta x_i, \quad \text{for} \quad i = 1, \dots, n.$$

If $s_i \in [x_{i-1}, x_i]$, then

$$\sum_{i=1}^{n} |\, \alpha'(s_i) - \alpha'(t_i) \,|\, \Delta x_i < \varepsilon \tag{5.25}$$

by (5.24) and Lemma 5.4 (b). Put $M = \sup |\, f(x) \,|$. Since

$$\sum_{i=1}^{n} f(s_i)\Delta\alpha_i = \sum_{i=1}^{n} f(s_i)\alpha'(t_i)\Delta x_i$$

it follows from (5.25) that

$$\left| \sum_{i=1}^{n} f(s_i)\Delta\alpha_i - \sum_{i=1}^{n} f(s_i)\alpha'(s_i)\Delta x_i \right| \leq M\varepsilon. \tag{5.26}$$

In particular,

$$\sum_{i=1}^{n} f(s_i)\Delta\alpha_i \leq U(P, f\alpha') + M\varepsilon,$$

for all choices of $s_i \in [x_{i-1}, x_i]$, so that

$$U(P, f, \alpha) \leq U(P, f\alpha') + M\varepsilon.$$

The same argument leads from (5.26) to

$$U(P, f\alpha') \leq U(P, f, \alpha) + M\varepsilon.$$

Thus

$$|\, U(P, f, \alpha) - U(P, f\alpha) \,| \leq M\varepsilon. \tag{5.27}$$

Now (5.25) remains true if P is replaced by any refinement. Hence (5.26) also remains true. We conclude that

$$\left| \overline{\int_a^b} f \, \mathrm{d}\alpha - \overline{\int_a^b} f(x)\alpha'(x) \, \mathrm{d}x \right| \leq M\varepsilon.$$

But ε is arbitrary. Hence

$$\overline{\int_a^b} f \, \mathrm{d}\alpha = \overline{\int_a^b} f(x)\alpha'(x) \, \mathrm{d}x,$$

for *any* bounded f. The equality for the lower integrals follows from (5.26) in exactly the same way. The proposition follows. ∎

We now summarize the two cases.

Proposition 5.13 *Let f be continuous on $[a,b]$. Except for finitely many points $c_0, c_1, \ldots, c_n$ with $c_0 = a$ and $c_n = b$ there exists $\alpha'(x)$ which is continuous and bounded on $[a,b] \setminus \{c_0, \ldots, c_n\}$.*
Then $f \in \mathcal{R}(\alpha)$ and

$$\int_a^b f \,\mathrm{d}\alpha = \int_a^b f(x)\alpha'(x)\,\mathrm{d}x + \sum_{i=1}^{n-1} f(c_i)(\alpha(c_i+0) - \alpha(c_i-0)) + f(a)(\alpha(a+0) - \alpha(a)) + f(b)(\alpha(b) - \alpha(b-0)).$$

Proof (Sketch of proof). (a) Note that $A_i^+ = \alpha(c_i+0) - \alpha(c_i)$ and $A_i^- = \alpha(c_i) - \alpha(c_i-0)$ exist by Theorem 3.8. Define

$$\alpha_1(x) = \sum_{i=0}^{n-1} A_i^+ H(x - c_i) + \sum_{i=1}^{k} -A_i^- H(c_i - x).$$

(b) Then $\alpha_2 = \alpha - \alpha_1$ is continuous.
(c) Since α_1 is piecewise constant, $\alpha_1'(x) = 0$ for $x \neq c_k$. Hence $\alpha_2'(x) = \alpha'(x)$. for $x \neq c_i$. Applying Proposition 5.12 gives

$$\int_a^b f \mathrm{d}\alpha_2 = \int_a^b f\alpha_2' \,\mathrm{d}x = \int_a^b f\alpha' \,\mathrm{d}x.$$

Further,

$$\int_a^b f \,\mathrm{d}\alpha = \int_a^b f \mathrm{d}(\alpha_1 + \alpha_2) = \int_a^b f\alpha' \,\mathrm{d}x + \int_a^b f \mathrm{d}\alpha_1.$$

By Proposition 5.11

$$\int_a^b f \mathrm{d}\alpha_1 = \sum_{i=1}^{n} A_i^+ f(c_i) - \sum_{i=1}^{n-1} A_i^- (-f(c_i)).$$

■

Example 5.3 (a) The Fundamental Theorem of Calculus, see Theorem 5.15 yields

$$\int_0^2 x \,\mathrm{d}x^3 = \int_0^2 x \cdot 3x^2 \,\mathrm{d}x = 3\frac{x^4}{4}\bigg|_0^2 = 12.$$

(b) $f(x) = x^2$.

$$\alpha(x) = \begin{cases} x, & 0 \le x < 1, \\ 7, & x = 1, \\ x^2 + 10, & 1 < x < 2, \\ 64, & x = 2. \end{cases}$$

$$\begin{aligned}\int_0^2 f\,\mathrm{d}\alpha &= \int_0^2 f\alpha'\,\mathrm{d}x + f(1)(\alpha(1+0)-\alpha(1-0)) + f(2)(\alpha(2)-\alpha(2-0))\\ &= \int_0^1 x^2\cdot 1\,\mathrm{d}x + \int_1^2 x^2\cdot 2x\,\mathrm{d}x + 1(11-1) + 4(64-14)\\ &= \frac{x^3}{3}\Big|_0^1 + \frac{x^4}{2}\Big|_1^2 + 10 + 200 = \frac{1}{3} + 8 - \frac{1}{2} + 210 = 217\frac{5}{6}.\end{aligned}$$

Remark 5.2 The three preceding proposition show the flexibility of the Stieltjes process of integration. If α is a pure step function, the integral reduces to an infinite series. If α has an initegrable derivative, the integral reduces to the ordinary Riemann integral. This makes it possible to study series and integral simultaneously, rather than separately.

5.2 Integration and Differentiation

We shall see that integration and differentiation are, in a certain sense, inverse operations.

Theorem 5.14 *Let $f \in \mathcal{R}$ on $[a,b]$. For $a \le x \le b$ put*

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

Then F is continuous on $[a,b]$; furthermore, if f is continuous at $x_0 \in [a,b]$ then F is differentiable at x_0 and

$$F'(x_0) = f(x_0).$$

Proof. Since $f \in \mathcal{R}$, f is bounded. Suppose $|\,f(t)\,| \le M$ on $[a,b]$. If $a \le x < y \le b$, then

$$|\,F(y) - F(x)\,| = \left|\int_x^y f(t)\,\mathrm{d}t\right| \le M(y-x),$$

by Proposition 5.9 (c) and (d). Given $\varepsilon > 0$, we see that

$$|\,F(y) - F(x)\,| < \varepsilon,$$

provided that $|\,y - x\,| < \varepsilon/M$. This proves continuity (and, in fact, uniform continuity) of F. Now suppose that f is continuous at x_0. Given $\varepsilon > 0$, choose $\delta > 0$ such that

$$|\,f(t) - f(x_0)\,| < \varepsilon$$

if $|\,t - x_0\,| < \delta$, $t \in [a,b]$. Hence, if

$$x_0 - \delta < s \le x_0 \le t < x_0 + \delta, \quad \text{and} \quad a \le s < t \le b,$$

we have by Proposition 5.9 (d) as before

$$\begin{aligned}\left|\frac{F(t)-F(s)}{t-s} - f(x_0)\right| &= \left|\frac{1}{t-s}\int_s^t f(r)\,\mathrm{d}r - \frac{1}{t-s}\int_s^t f(x_0)\,\mathrm{d}r\right|\\ &= \frac{1}{t-s}\left|\int_s^t (f(u)-f(x_0))\,\mathrm{d}u\right| < \varepsilon.\end{aligned}$$

This in particular holds for $s = x_0$, that is

$$\left| \frac{F(t) - F(x_0)}{t - x_0} - f(x_0) \right| < \varepsilon.$$

It follows that $F'(x_0) = f(x_0)$. ∎

Definition 5.4 A function $F\colon [a,b] \to \mathbb{R}$ is called an *antiderivative* or a *primitive* of a function $f\colon [a,b] \to \mathbb{R}$ if F is differentiable and $F' = f$.

Remarks 5.3 (a) There exist functions f not having an antiderivative, for example the Heaviside function $H(x)$ has a simple discontinuity at 0; but by Corollary 4.18 derivatives cannot have simple discontinuities.
(b) The antiderivative F of a function f (if it exists) is unique up to an additive constant. More precisely, if F is a antiderivative on $[a,b]$, then $F_1(x) = F(x) + c$ is also a antiderivative of f. If F and G are antiderivatives of f on $[a,b]$, then there is a constant c so that $F(x) - G(x) = c$. The first part is obvious since $F_1'(x) = F'(x) + c' = f(x)$. Suppose F and G are antiderivatives of f. Put $H(x) = F(x) - G(x)$; then $H'(x) = 0$ and $H(x)$ is constant by Corollary 4.11.

Notation for the antiderivative:

$$F(x) = \int f(x)\,\mathrm{d}x = \int f\,\mathrm{d}x.$$

The function f is called the *integrand*. Integration and differentiation are inverse to each other:

$$\frac{\mathrm{d}}{\mathrm{d}x} \int f(x)\,\mathrm{d}x = f(x), \qquad \int f'(x)\,\mathrm{d}x = f(x).$$

Theorem 5.15 (Fundamental Theorem of Calculus) *Let $f\colon [a,b] \to \mathbb{R}$ be continuous.* (a) *If*

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

Then $F(x)$ is an antiderivative of $f(x)$ on $[a,b]$.
(b) *If $G(x)$ is an antiderivative of $f(x)$ then*

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

Proof. (a) By Theorem 5.14 $F(x) = \int_a^x f(x)\,\mathrm{d}x$ is differentiable at any point $x_0 \in [a,b]$ with $F'(x) = f(x)$.
(b) By the above remark, the antiderivative is unique up to a constant, hence $F(x) - G(x) = C$. Since $F(a) = \int_a^a f(x)\,\mathrm{d}x = 0$ we obtain

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

■

Note that the FTC is also true if $f \in \mathcal{R}$ and G is an antiderivative of f on $[a,b]$. Indeed, let $\varepsilon > 0$ be given. By the Riemann criterion, Proposition 5.3 there exists a partition $P = \{x_0, \ldots, x_n\}$ of $[a,b]$ such that $U(P,f) - L(P,f) < \varepsilon$. By the mean value theorem, there exist points $t_i \in [x_{i-1}, x_i]$ such that

$$F(x_i) - F(x_{i-1}) = f(t_i)(x_i - x_{i-1}), \quad i = 1, \ldots, n.$$

Thus

$$F(b) - F(a) = \sum_{i=1}^{n} f(t_i)\Delta x_i.$$

It follows from Lemma 5.4 (c) and the above equation that

$$\left| \sum_{i=1}^{n} f(t_i)\Delta x_i - \int_a^b f(x)\,\mathrm{d}x \right| = \left| F(b) - F(a) - \int_a^b f(x)\,\mathrm{d}x \right| < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the proof is complete.

5.2.1 Table of Antiderivatives

By differentiating the right hand side one gets the left hand side of the table.

function	domain	antiderivative
x^α	$\alpha \in \mathbb{R} \setminus \{-1\},\ x > 0$	$\frac{1}{\alpha+1} x^{\alpha+1}$
$\frac{1}{x}$	$x < 0 \quad \text{or} \quad x > 0$	$\log \lvert x \rvert$
e^x	$\mathbb{R}$	e^x
a^x	$a > 0,\ a \neq 1,\ x \in \mathbb{R}$	$\frac{a^x}{\log a}$
$\sin x$	$\mathbb{R}$	$-\cos x$
$\cos x$	$\mathbb{R}$	$\sin x$
$\frac{1}{\sin^2 x}$	$\mathbb{R} \setminus \{k\pi \mid k \in \mathbb{Z}\}$	$-\cot x$
$\frac{1}{\cos^2 x}$	$\mathbb{R} \setminus \left\{\frac{\pi}{2} + k\pi \mid k \in \mathbb{Z}\right\}$	$\tan x$
$\frac{1}{1+x^2}$	$\mathbb{R}$	$\arctan x$
$\frac{1}{\sqrt{1+x^2}}$	$\mathbb{R}$	$\operatorname{arsinh} x = \log(x + \sqrt{x^2+1})$
$\frac{1}{\sqrt{1-x^2}}$	$-1 < x < 1$	$\arcsin x$
$\frac{1}{\sqrt{x^2-1}}$	$x < -1 \quad \text{or} \quad x > 1$	$\log(x + \sqrt{x^2-1})$

5.2.2 Integration Rules

The aim of this subsection is to calculate antiderivatives of composed functions using antiderivatives of (already known) simpler functions.
Notation:

$$f(x)|_a^b := f(b) - f(a).$$

Proposition 5.16 (a) *Let f and g be functions with antiderivatives F and G, respectively. Then $af(x) + bg(x)$, $a, b \in \mathbb{R}$, has the antiderivative $aF(x) + bG(x)$.*

$$\int (af + bg)\,\mathrm{d}x = a\int f\,\mathrm{d}x + b\int g\,\mathrm{d}x \quad \text{(Linearity.)}$$

(b) *If f and g are differentiable, and $f(x)g'(x)$ has a antiderivative then $f'(x)g(x)$ has a antiderivative, too:*

$$\int f'g\,\mathrm{d}x = fg - \int fg'\,\mathrm{d}x, \quad \text{(Integration by parts.)} \tag{5.28}$$

If f and g are continuously differentiable on $[a, b]$ then

$$\int_a^b f'g\,\mathrm{d}x = f(x)g(x)|_a^b - \int_a^b fg'\,\mathrm{d}x. \tag{5.29}$$

(c) *If $\varphi\colon D \to \mathbb{R}$ is continuously differentiable with $\varphi(D) \subset I$, and $f\colon I \to \mathbb{R}$ has a antiderivative F, then*

$$\int f(\varphi(x))\varphi'(x)\,\mathrm{d}x = F(\varphi(x)), \quad \text{(Change of variable.)} \tag{5.30}$$

If $\varphi\colon [a, b] \to \mathbb{R}$ is continuously differentiable with $\varphi([a, b]) \subset I$ and $f\colon I \to \mathbb{R}$ is continuous, then

$$\int_a^b f(\varphi(t))\varphi'(t)\,\mathrm{d}t = \int_{\varphi(a)}^{\varphi(b)} f(x)\,\mathrm{d}x.$$

Proof. Since differentiation is linear, (a) follows.
(b) Differentiating the right hand side, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}(fg - \int fg'\,\mathrm{d}x) = f'g + fg' - fg' = f'g$$

which proves the statement.
(c) By the chain rule $F(\varphi(x))$ is differentiable with

$$\frac{\mathrm{d}}{\mathrm{d}x}F(\varphi(x)) = F'(\varphi(x))\varphi'(x) = f(\varphi(x))\varphi'(x),$$

and (c) follows.
The statements about the Riemann integrals follow from the statements about antiderivatives

using the fundamental theorem of calculus. For example, we show the second part of (c). By the above part, $F(\varphi(t))$ is an antiderivative of $f(\varphi(t))\varphi'(t)$. By the FTC we have

$$\int_a^b f(\varphi(t))\varphi'(t)\,\mathrm{d}t = F(\varphi(t))|_a^b = F(\varphi(b)) - F(\varphi(a)).$$

On the other hand, again by the FTC,

$$\int_{\varphi(a)}^{\varphi(b)} f(x)\,\mathrm{d}x = F(x)|_{\varphi(a)}^{\varphi(b)} = F(\varphi(b)) - F(\varphi(a)).$$

This completes the proof of (c). ∎

Corollary 5.17 *Suppose F is the antiderivative of f.*

$$\int f(ax+b)\,\mathrm{d}x = \frac{1}{a}F(ax+b), \quad a \neq 0; \tag{5.31}$$

$$\int \frac{g'(x)}{g(x)}\,\mathrm{d}x = \log|\,g(x)\,|, \quad \textit{(g differentiable and } g(x) \neq 0\textit{)}. \tag{5.32}$$

Example 5.4 (a) The antiderivative of a polymnomial. If $p(x) = \sum_{k=0}^n a_k x^k$, then $\int p(x)\,\mathrm{d}x = \sum_{k=0}^n \frac{a_k}{k+1}x^{k+1}$.
(b) Put $f'(x) = \mathrm{e}^x$ and $g(x) = x$, then $f(x) = \mathrm{e}^x$ and $g'(x) = 1$ and we obtain

$$\int x\mathrm{e}^x\,\mathrm{d}x = x\mathrm{e}^x - \int 1\cdot\mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x(x-1).$$

(c) $I = (0,\infty)$. $\int \log x\,\mathrm{d}x = \int 1\cdot\log x\,\mathrm{d}x = x\log x - \int x\frac{1}{x}\,\mathrm{d}x = x\log x - x$.
(d)

$$\begin{aligned}\int \arctan x\,\mathrm{d}x &= \int 1\cdot\arctan x\,\mathrm{d}x = x\arctan x - \int x\frac{1}{1+x^2}\,\mathrm{d}x\\ &= x\arctan x - \frac{1}{2}\int\frac{(1+x^2)'}{1+x^2}\,\mathrm{d}x = x\arctan x - \frac{1}{2}\log(1+x^2).\end{aligned}$$

In the last equation we made use of (5.32).
(e) Recurrent computation of integrals.

$$I_n := \int \frac{\mathrm{d}x}{(1+x^2)^n}, \quad n\in\mathbb{N}.$$

$I_1 = \arctan x$.

$$I_n = \int\frac{(1+x^2)-x^2}{(1+x^2)^n} = I_{n-1} - \int\frac{x^2\,\mathrm{d}x}{(1+x^2)^n}.$$

Put $u = x$, $v' = \dfrac{x}{(1+x^2)^n}$. Then $U' = 1$ and

$$v = \int\frac{x\,\mathrm{d}x}{(1+x^2)^n} = \frac{1}{2}\frac{(1+x^2)^{1-n}}{1-n}.$$

Hence,

$$I_n = I_{n-1} - \frac{1}{2}\frac{x(1+x^2)^{1-n}}{1-n} - \frac{1}{2(1-n)}\int (1+x^2)^{1-n}\,\mathrm{d}x$$
$$I_n = \frac{x}{(2n-2)(1+x^2)^{n-1}} + \frac{2n-3}{2n-2}\,I_{n-1}.$$

In particular, $I_2 = \frac{x}{2(1+x^2)} + \frac{1}{2}\arctan x$ and $I_3 = \frac{x}{4(1+x^2)^2} + \frac{3}{4}I_2$.

Proposition 5.18 (Mean Value Theorem of Integration) *Let $f, \varphi\colon [a,b] \to \mathbb{R}$ be continuous functions and $\varphi \geq 0$. Then there exists $\xi \in [a,b]$ such that*

$$\int_a^b f(x)\varphi(x)\,\mathrm{d}x = f(\xi)\int_a^b \varphi(x)\,\mathrm{d}x. \tag{5.33}$$

In particular, in case $\varphi = 1$ we have

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

for some $\xi \in [a,b]$.

Proof. Put $m = \inf\{f(x) \mid x \in [a,b]\}$ and $M = \sup\{f(x) \mid x \in [a,b]\}$. Since $\varphi \geq 0$ we obtain $m\varphi(x) \leq f(x)\varphi(x) \leq M\varphi(x)$. By Proposition 5.9 (a) and (b) we have

$$m\int_a^b \varphi(x)\,\mathrm{d}x \leq \int_a^b f(x)\varphi(x)\,\mathrm{d}x \leq M\int_a^b \varphi(x)\,\mathrm{d}x.$$

Hence there is a $\mu \in [m,M]$ such that

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

Since f is continuous on $[a,b]$ the intermediate value theorem Theorem 3.5 ensures that there is a ξ with $\mu = f(\xi)$. The claim follows. ∎

Example 5.5 The trapezoid rule. Let $f\colon [0,1] \to \mathbb{R}$ be twice continuously differentiable. Then there exists $\xi \in [0,1]$ such that

$$\int_0^1 f(x)\,\mathrm{d}x = \frac{1}{2}\left(f(0)+f(1)\right) - \frac{1}{12}f''(\xi). \tag{5.34}$$

Proof. Let $\varphi(x) = \frac{1}{2}x(1-x)$ such that $\varphi(x) \geq 0$ for $x \in [0,1]$, $\varphi'(x) = \frac{1}{2} - x$, and $\varphi''(x) = -1$. Using integration by parts twice as well as Theorem 5.18 we find

$$\begin{aligned}
\int_0^1 f(x)\,\mathrm{d}x &= -\int_0^1 \varphi''(x)f(x)\,\mathrm{d}x = -\varphi'(x)f(x)\big|_0^1 + \int_0^1 \varphi'(x)f'(x)\,\mathrm{d}x \\
&= \frac{1}{2}\left(f(0)+f(1)\right) + \varphi(x)f'(x)\big|_0^1 - \int_0^1 \varphi(x)f''(x)\,\mathrm{d}x \\
&= \frac{1}{2}\left(f(0)+f(1)\right) - f''(\xi)\int_0^1 \varphi(x)\,\mathrm{d}x \\
&= \frac{1}{2}\left(f(0)+f(1)\right) - \frac{1}{12}f''(\xi).
\end{aligned}$$

Indeed, $\int_0^1 \left(\frac{1}{2}x - \frac{1}{2}x^2\right)\,\mathrm{d}x = \frac{1}{4}x^2 - \frac{1}{6}x^3\big|_0^1 = \frac{1}{4} - \frac{1}{6} = \frac{1}{12}$. ■

5.2.3 Integration of Rational Functions

We will give a useful method to compute antiderivatives of an arbitrary rational function. Consider a rational function p/q where p and q are polinomials. We will assume that $\deg p < \deg q$; for otherwise we can express p/q as a polynomial function plus a rational function which *is* of this form, for eample

$$\frac{x^2}{x-1} = x + 1 + \frac{1}{x-1}.$$

Polynomials

We need some preliminary facts on polynomials which are stated here without proof. Recall that a non-zero constant polynomial has degree zero, $\deg c = 0$ if $c \neq 0$. By definition, the zero polynomial has degree $-\infty$, $\deg 0 = -\infty$.

Theorem 5.19 (Fundamental Theorem of Algebra) *Every polynomial p of positive degree with complex coefficients has a complex root, i. e. there exists a complex number z such that $p(z) = 0$.*

Lemma 5.20 (Long Division) *Let p and q be polynomials, then there exist unique polynomials r and s such that*

$$p = qs + r, \quad \deg r < \deg q.$$

Lemma 5.21 *Let p be a complex polynomial of degree $n \geq 1$ and leading coefficient a_n. Then there exist n uniquely determined numbers $z_1, \ldots, z_n$ (which may be equal) such that*

$$p(z) = a_n(z - z_1)(z - z_2)\cdots(z - z_n).$$

Proof. We use induction over n and the two preceding statements. In case $n = 1$ the linear polynomial $p(z) = az + b$ can be written in the desired form

$$p(z) = a\left(z - \frac{-b}{a}\right) \quad \text{with the unique root} \quad z_1 = -\frac{b}{a}.$$

Suppose the statement is true for all polynomials of degree $n-1$. We will show it for degree n polynomials. For, let z_n be a complex root of p which exists by Theorem 5.19; $p(z_n) = 0$. Using long division of p by the linear polynomial $q(z) = z - z_n$ we obtain a quotient polynomial $p_1(z)$ and a remainder polynomial $r(z)$ of degree 0 (a constant polynomial) such that

$$p(z) = (z - z_n)p_1(z) + r(z).$$

Inserting $z = z_n$ gives $p(z_n) = 0 = r(z_n)$; hence the constant r vanishes and we have

$$p(z) = (z - z_n)p_1(z)$$

with a polynomial $p_1(z)$ of degree $n-1$. Applying the induction hypothesis to p_1 the statement follows. ∎

A root α of p is said to be a *root of multiplicity* k, $k \in \mathbb{N}$, if α appears exactly k times among the zeros $z_1, z_2, \ldots, z_n$. In that case $(z-\alpha)^k$ divides $p(z)$ but $(z-\alpha)^{k+1}$ not.
If p is a real polynomial, i. e. a polynomial with real coefficients, and α is a root of multiplicity k of p then $\overline{\alpha}$ is also a root of multiplicity k of p. Indeed, taking the complex conjugation of the equation

$$p(z) = (z-\alpha)^k q(z)$$

we have since $\overline{p(z)} = \overline{p}(\overline{z}) = p(\overline{z})$

$$p(\overline{z}) = (\overline{z}-\overline{\alpha})^k \overline{q}(\overline{z}) \underset{z:=\overline{z}}{\Longrightarrow} p(z) = (z-\overline{\alpha})^k \overline{q}(z).$$

Note that the product of the two complex linear factors $z-\alpha$ and $z-\overline{\alpha}$ yield a real quadratic factor

$$(z-\alpha)(z-\overline{\alpha}) = z^2 - (\alpha+\overline{\alpha})z + \alpha\overline{\alpha} = z^2 - 2\operatorname{Re}\alpha + |\,\alpha\,|^2 .$$

Using this fact, the real version of Lemma 5.21 is as follows.

Lemma 5.22 *Let q be a real polynomial of degree n with leading coefficient a_n. Then there exist real numbers $\alpha_i, \beta_j, \gamma_j$ and multiplicities $r_i, s_j \in \mathbb{N}$, $i = 1, \ldots, k$, $j = 1, \ldots, l$ such that*

$$q(x) = a_n \prod_{i=1}^{k} (x-\alpha_i)^{r_i} \prod_{j=1}^{l} (x^2 - 2\beta_j x + \gamma_j)^{s_j}.$$

We assume that the quadratic factors cannot be factored further; this means

$$\beta_j^2 - \gamma_j < 0, \quad j = 1, \ldots, l.$$

Of course, $\deg q = \sum_i r_i + \sum_j 2s_j = n$.

Example 5.6 (a) $x^4 - 4 = (x^2+2)(x^2-2) = (x-\sqrt{2})(x+\sqrt{2})(x-\mathrm{i}\sqrt{2})(x+\mathrm{i}\sqrt{2}) = (x-\sqrt{2})(x+\sqrt{2})(x^2+2)$
(b) $x^3 + x - 2$. One can guess the first zero $x_1 = 1$. Using long division one gets

$$\begin{array}{rrrr}
x^3 & & +x & -2 \\
\underline{-(x^3} & \underline{-x^2)} & & \\
 & x^2 & +x & -2 \\
 & \underline{-(x^2} & \underline{-x} & \underline{\quad)} \\
 & & 2x & -2 \\
 & & \underline{-(2x} & \underline{-2)} \\
 & & & 0
\end{array} \qquad = (x-1)(x^2+x+2)$$

There are no further real zeros of $x^2 + x + 2$.

5.2.4 Partial Fraction Decomposition

Proposition 5.23 *Let $p(x)$ and $q(x)$ be real polynomials with $\deg p < \deg q$. There exist real numbers A_{ir}, B_{js}, and C_{js} such that*

$$\frac{p(x)}{q(x)} = \sum_{i=1}^{k} \left(\sum_{r=1}^{r_i} \frac{A_{ir}}{(x-\alpha_i)^r} \right) + \sum_{j=1}^{l} \left(\sum_{s=1}^{s_j} \frac{B_{js}x + C_{js}}{(x^2 - 2\beta_j x + \gamma_j)^s} \right) \tag{5.35}$$

where the α_i, β_j, γ_j, r_i, and s_j have the same meaning as in Lemma 5.22.

Example 5.7 (a) Compute $\displaystyle\int f(x)\,\mathrm{d}x = \int \frac{x^4}{x^3-1}\,\mathrm{d}x$. We use long division to obtain a rational function p/q with $\deg p < \deg q$, $f(x) = x + \frac{x}{x^3-1}$. To obtain the partial fraction decomposition (PFD), we need the factorization of the denominator polynomial $q(x) = x^3 - 1$. One can guess the first real zero $x_1 = 1$ and divide q by $x-1$; $q(x) = (x-1)(x^2+x+1)$.
The PFD then reads

$$\frac{x}{x^3-1} = \frac{a}{x-1} + \frac{bx+c}{x^2+x+1}.$$

We have to determine a, b, c. Multiplication by $x^3 - 1$ gives

$$0 \cdot x^2 + 1 \cdot x + 0 = a(x^2+x+1) + (bx+c)(x-1) = (a+b)x^2 + (a-b+c)x + a - c.$$

The two polynomials on the left and on the right must coincide, that is, there coefficients must be equal:

$$0 = a - c, \quad 1 = a - b + c, \quad 0 = a + b;$$

which gives $a = -b = c = \frac{1}{3}$. Hence,

$$\frac{x}{x^3-1} = \frac{1}{3}\,\frac{1}{x-1} - \frac{1}{3}\,\frac{x-1}{x^2+x+1}.$$

We can integrate the first two terms but we have to rewrite the last one

$$\frac{x-1}{x^2+x+1} = \frac{1}{2}\,\frac{2x+1}{x^2+x+1} - \frac{3}{2}\frac{1}{\left(x+\frac{1}{2}\right)^2 + \frac{3}{4}}.$$

Recall that

$$\int \frac{2x-2\beta}{x^2-2\beta x+\gamma}\,\mathrm{d}x = \log\left|\, x^2 - 2\beta x + \gamma \,\right|, \quad \int \frac{\mathrm{d}x}{(x+b)^2+a^2} = \frac{1}{a}\arctan\frac{x+b}{a}.$$

Therefore,

$$\int \frac{x^4}{x^3-1}\,\mathrm{d}x = \frac{1}{2}x^2 + \frac{1}{2}\log|\,x-1\,| - \frac{1}{6}\log(x^2+x+1) + \frac{1}{\sqrt{3}}\arctan\frac{2x+1}{\sqrt{3}}.$$

(b) If $q(x) = (x-1)^3(x+2)(x^2+2)^2(x^2+1)$ and $p(x)$ is any polynomial with $\deg p < \deg q = 10$, then the partial fraction decomposition reads as

$$\frac{p(x)}{q(x)} = \frac{A_{11}}{x-1} + \frac{A_{12}}{(x-1)^2} + \frac{A_{13}}{(x-1)^3} + \frac{A_{21}}{x+2} + \frac{B_{11}x+C_{11}}{x^2+2} + \frac{B_{12}x+C_{12}}{(x^2+2)^2} + \frac{B_{21}+C_{21}}{x^2+1}. \tag{5.36}$$

Suppose now that $p(x) \equiv 1$. One can immediately compute A_{13} and A_{21}. Multiplying (5.36) by $(x-1)^3$ yields

$$\frac{1}{(x+2)(x^2+2)^2(x^2+1)} = A_{13} + (x-1)p_1(x)$$

with a rational function p_1 not having $(x-1)$ in the denominator. Inserting $x = 1$ gives $A_{13} = \frac{1}{3^2 \cdot 3 \cdot 2} = \frac{1}{54}$. Similarly,

$$A_{21} = \left.\frac{1}{(x-1)^3(x^2+2)^2(x^2+1)}\right|_{x=-2} = \frac{1}{(-3)^3 \cdot 6 \cdot 5}.$$

5.2.5 Other Classes of Elementary Integrable Functions

An *elementary function* is the compositions of rational, exponential, trigonometric functions and their inverse functions, for example

$$f(x) = \frac{\mathrm{e}^{\sin(\sqrt{x}-1)}}{x + \log x}.$$

A function is called *elementary integrable* if it has an elementary antiderivative. Rational functions are elementary integrable. "Most" functions are not elementary integrable as

$$\mathrm{e}^{-x^2}, \quad \frac{\mathrm{e}^x}{x}, \quad \frac{1}{\log x}, \quad \frac{\sin x}{x}.$$

They define "new" functions

$$\begin{aligned}
W(x) &:= \int_0^x \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t, && \text{(Gaussian integral)},\\
\mathrm{li}(x) &:= \int_0^x \frac{\mathrm{d}t}{\log t} && \text{(integral logarithm)}\\
\mathrm{F}(\varphi,k) &:= \int_0^\varphi \frac{\mathrm{d}x}{\sqrt{1-k^2\sin^2 x}} && \text{(elliptic integral of the first kind)},\\
\mathrm{E}(\varphi,k) &:= \int_0^\varphi \sqrt{1-k^2\sin^2 x}\,\mathrm{d}x && \text{(elliptic integral of the second kind)}.
\end{aligned}$$

$\int \boldsymbol{R}(\cos \boldsymbol{x}, \sin \boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$

Let $R(u,v)$ be a rational function in two variables u and v, that is $R(u,v) = \frac{p(u,v)}{q(u,v)}$ with polinomials p and q in two variables. We substitute $t = \tan\dfrac{x}{2}$. Then

$$\sin x = \frac{2t}{1+t^2}, \quad \cos x = \frac{1-t^2}{1+t^2}, \quad \mathrm{d}x = \frac{2\mathrm{d}t}{1+t^2}.$$

Hence

$$\int R(\cos x, \sin x)\,\mathrm{d}x = \int R\left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right)\frac{2\mathrm{d}t}{1+t^2} = \int R_1(t)\,\mathrm{d}t$$

with another rational function $R_1(t)$.
There are 3 special cases where another substitution is apropriate.

(a) $R(-u,v) = -R(u,v)$, R is odd in u. Substitute $t = \sin x$.

(b) $R(u,-v) = -R(u,v)$, R is odd in v. Substitute $t = \cos x$.

(c) $R(-u,-v) = R(u,v)$. Substitute $t = \tan x$.

Example 5.8 (1) $\int \sin^3 x\,\mathrm{d}x$. Here, $R(u,v) = v^3$ is an odd function in v, such that (b) applies; $t = \cos x,\ \mathrm{d}t = -\sin x\,\mathrm{d}x$, $\sin^2 x = 1 - \cos^2 x = 1 - t^2$. This yields

$$\int \sin^3 x\,\mathrm{d}x = -\int \sin^2 x \cdot (-\sin x\,\mathrm{d}x) = -\int (1-t^2)\,\mathrm{d}t = -t + \frac{t^3}{3} + \text{ const.}$$
$$= -\cos x + \frac{\cos^3}{3} + \text{ const.}.$$

(2) $\int \tan x\,\mathrm{d}x$. Here, $R(u,v) = \frac{v}{u}$. All (a), (b), and (c) apply to this situation. For example, let $t = \sin x$. Then $\cos^2 x = 1 - t^2,\ \mathrm{d}t = \cos x\,\mathrm{d}x$ and

$$\int \tan x\,\mathrm{d}x = \int \frac{\sin x \cdot \cos x\,\mathrm{d}x}{\cos^2 x} = \int \frac{t\,\mathrm{d}t}{1-t^2} = -\frac{1}{2}\int \frac{\mathrm{d}(1-t^2)}{1-t^2} =$$
$$= -\frac{1}{2}\log(1-t^2) = -\log|\cos x|\,.$$

$\int R(x, \sqrt[n]{ax+b})\,\mathrm{d}x$

The substitution

$$t = \sqrt[n]{ax+b}$$

yields $x = (t^n - b)/a,\ \mathrm{d}x = nt^{n-1}\,\mathrm{d}t/a$, and therefore

$$\int R(x, \sqrt[n]{ax+b})\,\mathrm{d}x = \frac{n}{a}\int R\left(\frac{t^n-b}{a}, t\right) t^{n-1}\,\mathrm{d}t.$$

$\int R(x, \sqrt{ax^2+2bx+c})\,\mathrm{d}x$

Using the method of complete squares the above integral can be written in one of the three basic forms

$$\int R(t, \sqrt{t^2+1})\,\mathrm{d}t, \quad \int R(t, \sqrt{t^2-1})\,\mathrm{d}t, \quad \int R(t, \sqrt{1-t^2})\,\mathrm{d}t.$$

Further substitutions

$$\begin{array}{lll} t = \sinh u, & \sqrt{t^2+1} = \cosh u, & \mathrm{d}t = \cosh u\,\mathrm{d}u, \\ t = \pm\cosh u, & \sqrt{t^2-1} = \sinh u, & \mathrm{d}t = \pm\sinh u\,\mathrm{d}u, \\ t = \pm\cos u, & \sqrt{1-t^2} = \sin u, & \mathrm{d}t = \mp\sin u\,\mathrm{d}u \end{array}$$

reduce the integral to already known integrals.

Example 5.9 Compute $I = \int \frac{\mathrm{d}x}{\sqrt{x^2+6x+5}}$. Hint: $t = \sqrt{x^2+6x+5} - x$.
Then $(x+t)^2 = x^2+2tx+t^2 = x^2+6x+5$ such that $t^2+2tx = 6x+5$ and therefore $x = \frac{t^2-5}{6-2t}$ and

$$\mathrm{d}x = \frac{2t(6-2t)+2(t^2-5)}{(6-2t)^2}\,\mathrm{d}t = \frac{-2t^2+12t-10}{(6-2t)^2}\,\mathrm{d}t.$$

Hence, using $t + x = t + \frac{t^2-5}{6-2t} = \frac{-t^2+6t-5}{6-2t}$,

$$I = \int \frac{(-2t^2+12t-10)\,\mathrm{d}t}{(6-2t)^2}\,\frac{1}{t+x} = \int \frac{2(6-2t)(-t^2+6t-5)}{(-t^2+6t-5)(6-2t)^2}\,\mathrm{d}t$$
$$= 2\int \frac{\mathrm{d}t}{6-2t} = -\log|\,6-2t\,| + \text{ const.} = \log\left|\,6-2\sqrt{x^2+6x+5}+2x\,\right| + \text{ const.}$$

5.3 Improper Integrals

The notion of the Riemann integral defined so far is apparently too tight for some applications: we can integrate only over finite intervals and the functions are necessarily bounded. If the integration interval is unbounded or the function to integrate is unbounded we speak about *improper* integrals. We consider three cases: one limit of the integral is infinite; the function is not defined at one of the end points a or b of the interval; both a and b are critical points (either infinity or the function is not defined there).

5.3.1 Integrals on unbounded intervals

Definition 5.5 Suppose $f \in \mathcal{R}$ on $[a,b]$ for all $b > a$ where a is fixed. Define

$$\int_a^\infty f(x)\,\mathrm{d}x = \lim_{b\to+\infty}\int_a^b f(x)\,\mathrm{d}x \tag{5.37}$$

if this limit exists (and is finite). In that case, we say that the integral on the left *converges*. If it also converges if f has been replaced by $|\,f\,|$, it is said to *converge absolutely*.

If an integral converges absolutely, then it converges, see Example 5.11 below, where

$$\left|\int_a^\infty f\,\mathrm{d}x\right| \le \int_a^\infty |\,f\,|\,\mathrm{d}x.$$

Similarly, one defines $\int_{-\infty}^b f(x)\,\mathrm{d}x$. Moreover,

$$\int_{-\infty}^\infty f\,\mathrm{d}x := \int_{-\infty}^0 f\,\mathrm{d}x + \int_0^\infty f\,\mathrm{d}x$$

if both integrals on the right side converge.

Example 5.10 (a) The integral $\displaystyle\int_1^\infty \frac{\mathrm{d}x}{x^s}$ converges for $s > 1$ and diverges for $0 < s \leq 1$. Indeed,

$$\int_1^R \frac{\mathrm{d}x}{x^s} = \frac{1}{1-s} \cdot \frac{1}{x^{s-1}}\bigg|_1^R = \frac{1}{s-1}\left(1 - \frac{1}{R^{s-1}}\right).$$

Since

$$\lim_{R\to+\infty} \frac{1}{R^{s-1}} = \begin{cases} 0, & \text{if} \quad s > 1, \\ +\infty, & \text{if} \quad 0 < s < 1, \end{cases}$$

it follows that

$$\int_0^\infty \frac{\mathrm{d}x}{x^s} = \frac{1}{s-1}, \quad \text{if} \quad s > 1.$$

(b)

$$\int_0^R \mathrm{e}^{-x}\,\mathrm{d}x = -\mathrm{e}^{-x}\big|_0^R = 1 - \frac{1}{\mathrm{e}^R}.$$

Hence $\displaystyle\int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}x = 1$.

Proposition 5.24 (Cauchy criterion) *The improper integral* $\displaystyle\int_a^\infty f\,\mathrm{d}x$ *converges if and only if for every* $\varepsilon > 0$ *there exists some* $b > a$ *such that for all* $c, d > b$

$$\left| \int_c^d f\,\mathrm{d}x \right| < \varepsilon.$$

Proof. The following *Cauchy criterion for limits of functions* is easily proved using sequences:

The limit $\lim\limits_{x\to\infty} F(x)$ exists if and only if

$$\forall\,\varepsilon > 0\ \ \exists\, R > 0\ \ \forall\, x, y > R : |\,F(x) - F(y)\,| < \varepsilon. \tag{5.38}$$

Indeed, suppose that (x_n) is any sequence converging to $+\infty$ as $n \to \infty$. We will show that $(F(x_n))$ converges if (5.38) is satisfied. Let $\varepsilon > 0$. By assumption, there exists $R > 0$ with the above property. Since $x_n \underset{n\to\infty}{\longrightarrow} +\infty$ there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies $x_n > R$. Hence, $|\,F(x_n) - F(x_m)\,| < \varepsilon$ as $m, n \geq n_0$. Thus, $(F(x_n))$ is a Cauchy sequence and therefore convergent. This proves one direction of the above criterion. The inverse direction is even simpler: Suppose that $\lim\limits_{x\to+\infty} F(x) = A$ exists (and is finite!). We will show that the above criterion is satisfied.Let $\varepsilon > 0$. By definition of the limit there exists $R > 0$ such that $x, y > R$ imply $|\,F(x) - A\,| < \varepsilon/2$ and $|\,F(y) - A\,| < \varepsilon/2$. By the triangle inequality,

$$|\,F(x) - F(y)\,| = |\,F(x) - A - (F(y) - A)\,| \leq |\,F(x) - A\,| + |\,F(y) - A\,| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

as $x, y > R$ which completes the proof of this Cauchy criterion.
Applying this criterion to the function $F(t) = \int_a^t f\,\mathrm{d}x$ noting that $|\,F(d) - F(c)\,| = \left|\int_c^d f\,\mathrm{d}x\right|$, the limit $\lim_{t\to\infty} F(t)$ exists. ∎

Example 5.11 (a) If $\int_a^\infty f\,\mathrm{d}x$ converges absolutely, then $\int_a^\infty f\,\mathrm{d}x$ converges. Indeed, let $\varepsilon > 0$ and $\int_a^\infty |\,f\,|\,\mathrm{d}x$ converges. By the Cauchy Criterion for the later integral and by the triangle inequality, Proposition 5.10, there exists $b > 0$ such that for all $c, d > b$

$$\left| \int_c^d f\,\mathrm{d}x \right| \le \int_c^d |\,f\,|\,\mathrm{d}x < \varepsilon. \tag{5.39}$$

Hence, the Cauchy criterion is satisfied for f if it holds for $|\,f\,|$. Thus, $\int_a^\infty f\,\mathrm{d}x$ converges.
(b) $\int_1^\infty \frac{\sin x}{x}\,\mathrm{d}x$. Partial integration with $u = \frac{1}{x}$ and $v' = \sin x$ yields $u' = -\frac{1}{x^2}$, $v = -\cos x$ and

$$\begin{aligned}
\int_c^d \frac{\sin x}{x}\,\mathrm{d}x &= -\frac{1}{x}\cos x\Big|_c^d - \int_c^d \frac{\cos x}{x^2}\,\mathrm{d}x \\
\left| \int_c^d \frac{\sin x}{x}\,\mathrm{d}x \right| &\le \left| -\frac{1}{d}\cos d + \frac{1}{c}\cos c \right| + \left| \int_c^d \frac{\mathrm{d}x}{x^2} \right| \\
&\le \frac{1}{c} + \frac{1}{d} + \left| \frac{1}{d} - \frac{1}{c} \right| \le 2\left(\frac{1}{c} + \frac{1}{d}\right) < \varepsilon
\end{aligned}$$

if c and d are sufficiently large. Hence, $\int_1^\infty \frac{\sin x}{x}\,\mathrm{d}x$ converges.
The integral does not converge absolutely. For non-negative integers $n \in \mathbb{Z}_+$ we have

$$\int_{n\pi}^{(n+1)\pi} \left| \frac{\sin x}{x} \right| \mathrm{d}x \ge \frac{1}{(n+1)\pi} \int_{n\pi}^{(n+1)\pi} |\sin x|\,\mathrm{d}x = \frac{2}{(n+1)\pi};$$

hence

$$\int_1^{(n+1)\pi} \left| \frac{\sin x}{x} \right| \mathrm{d}x \ge \frac{2}{\pi} \sum_{k=1}^{n} \frac{1}{k+1}.$$

Since the harmonic series diverges, so does the integral $\int_\pi^\infty \left| \frac{\sin x}{x} \right| \mathrm{d}x$.

Proposition 5.25 *Suppose $f \in \mathcal{R}$ is nonnegative, $f \ge 0$. Then $\int_a^\infty f\,\mathrm{d}x$ converges if there exists $C > 0$ such that*

$$\int_a^b f\,\mathrm{d}x < C, \quad \textit{for all} \quad b > a.$$

The proof is similar to the proof of Lemma 2.19 (c); we omit it. Analogous propositions are true for integrals $\int_{-\infty}^a f\,\mathrm{d}x$.

Proposition 5.26 (Integral criterion for series) *Assume that $f \in \mathcal{R}$ is nonnegative $f \ge 0$ and decreasing on $[1, +\infty)$. Then $\int_1^\infty f\,\mathrm{d}x$ converges if and only if the series $\sum_{n=1}^\infty f(n)$ converges.*

Proof. Since $f(n) \le f(x) \le f(n-1)$ for $n - 1 \le x \le n$,

$$f(n) \le \int_{n-1}^n f\,\mathrm{d}x \le f(n-1).$$

Summation over $n = 2, 3, \ldots, N$ yields

$$\sum_{n=2}^N f(n) \le \int_1^N f\,\mathrm{d}x \le \sum_{n=1}^{N-1} f(n).$$

If $\int_1^\infty f\,\mathrm{d}x$ converges the series $\sum_{n=1}^\infty f(n)$ is bounded and therefore convergent.
Conversely, if $\sum_{n=1}^\infty f(n)$ converges, the integral $\int_1^R f\,\mathrm{d}x \le \sum_{n=1}^\infty f(n)$ is bounded as $R\to\infty$, hence convergent by Proposition 5.25. ∎

Example 5.12 $\sum_{n=2}^\infty \frac{1}{n(\log n)^\alpha}$ converges if and only if $\int_2^\infty \frac{\mathrm{d}x}{x(\log x)^\alpha}$ converges. The substitution $y=\log x$, $\mathrm{d}y=\frac{\mathrm{d}x}{x}$ gives

$$\int_2^\infty \frac{\mathrm{d}x}{x(\log x)^\alpha} = \int_{\log 2}^\infty \frac{\mathrm{d}y}{y^\alpha}$$

which converges if and only if $\alpha>1$ (see Example 5.10).

5.3.2 Integrals of Unbounded Functions

Definition 5.6 Suppose f is a real function on $[a,b)$ and $f\in\mathcal{R}$ on $[a,t]$ for every t, $a<t<b$. Define

$$\int_a^b f\,\mathrm{d}x = \lim_{t\to b-0}\int_a^t f\,\mathrm{d}x$$

if the limit on the right exists. Similarly, one defines

$$\int_a^b f\,\mathrm{d}x = \lim_{t\to a+0}\int_t^b f\,\mathrm{d}x$$

if f is unbounded at a and integrable on $[t,b]$ for all t with $a<t<b$.

In both cases we say that $\int_a^b f\,\mathrm{d}x$ *converges*.

Example 5.13 (a)

$$\int_0^1 \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \lim_{t\to1-0}\int_0^t \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \lim_{t\to1-0}\arcsin x\Big|_0^t = \lim_{t\to1-0}\arcsin t = \arcsin 1 = \frac{\pi}{2}.$$

(b)

$$\int_0^1 \frac{\mathrm{d}x}{x^\alpha} = \lim_{t\to0+0}\int_t^1 \frac{\mathrm{d}x}{x^\alpha} = \lim_{t\to0+0}\begin{cases}\frac{1}{1-\alpha}x^{1-\alpha}\big|_t^1, & \alpha\neq1\\ \log x\big|_t^1, & \alpha=1\end{cases} = \begin{cases}\frac{1}{1-\alpha}, & \alpha<1,\\ +\infty, & \alpha\ge1.\end{cases}$$

Remarks 5.4 (a) The analogous statements to Proposition 5.24 and Proposition 5.25 are true for improper integrals $\int_a^b f\,\mathrm{d}x$.
For example, $\int_0^1 \frac{\mathrm{d}x}{x(1-x)}$ diverges since both improper integrals $\int_0^{\frac12} f\,\mathrm{d}x$ and $\int_{\frac12}^1 f\,\mathrm{d}x$ diverge, $\int_0^1 \frac{\mathrm{d}x}{\sqrt{x}(1-x)}$ diverges since it diverges at $x=1$, finally $I=\int_0^1 \frac{\mathrm{d}x}{\sqrt{x(1-x)}}$ converges. Indeed, the substitution $x=\sin^2 t$ gives $I=\pi$.
(b) If f is unbounded both at a and at b we define the improper integral

$$\int_a^b f\,\mathrm{d}x = \int_a^c f\,\mathrm{d}x + \int_c^b f\,\mathrm{d}x$$

if c is between a and b and both improper integrals on the right side exist.
(c) Also, if f is unbounded at a define

$$\int_a^\infty f\,\mathrm{d}x = \int_a^b f\,\mathrm{d}x + \int_b^\infty f\,\mathrm{d}x$$

if the two improper integrals on the right side exist.
(d) If f is unbounded in the interior of the interval $[a,b]$, say at c, we define the improper integral

$$\int_a^b f\,\mathrm{d}x = \int_a^c f\,\mathrm{d}x + \int_c^b f\,\mathrm{d}x$$

if the two improper integrals on the right side exist. For example,

$$\begin{aligned}\int_{-1}^1 \frac{\mathrm{d}x}{\sqrt{|\,x\,|}} &= \int_{-1}^0 \frac{\mathrm{d}x}{\sqrt{|\,x\,|}} + \int_0^1 \frac{\mathrm{d}x}{\sqrt{|\,x\,|}} = \lim_{t\to 0-0}\int_{-1}^t \frac{\mathrm{d}x}{\sqrt{|\,x\,|}} + \lim_{t\to 0+0}\int_t^1 \frac{\mathrm{d}x}{\sqrt{|\,x\,|}}\\ &= \lim_{t\to 0-0} -2\sqrt{-x}\Big|_{-1}^t + \lim_{t\to 0+0} 2\sqrt{x}\Big|_t^1 = 4.\end{aligned}$$

5.3.3 The Gamma function

For $x>0$ set

$$\Gamma(x) = \int_0^\infty t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t. \tag{5.40}$$

By Example 5.13, $\Gamma_1(x) = \int_0^1 t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t$ converges since for every $t>0$

$$t^{x-1}\mathrm{e}^{-t} \le \frac{1}{t^{1-x}}.$$

By Example 5.10, $\Gamma_2(x) = \int_1^\infty t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t$ converges since for every $t \ge t_0$

$$t^{x-1}\mathrm{e}^{-t} \le \frac{1}{t^2}.$$

Note that $\lim_{t\to\infty} t^{x+1}\mathrm{e}^{-t} = 0$ by Proposition 3.11. Hence, $\Gamma(x)$ is defined for every $x>0$.

Proposition 5.27 *For every positive x*

$$x\Gamma(x) = \Gamma(x+1). \tag{5.41}$$

In particular, for $n\in\mathbb{N}$ we have $\Gamma(n+1) = n!$.

Proof. Using integration by parts,

$$\int_\varepsilon^R t^x\mathrm{e}^{-t}\,\mathrm{d}t = -t^x\mathrm{e}^{-t}\Big|_\varepsilon^R + x\int_\varepsilon^R t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t.$$

Taking the limits $\varepsilon\to 0+0$ and $R\to+\infty$ one has $\Gamma(x+1) = x\Gamma(x)$. Since by Example 5.10

$$\Gamma(1) = \int_0^\infty \mathrm{e}^{-t}\,\mathrm{d}t = 1,$$

it follows from (5.41) that

$$\Gamma(n+1) = n\Gamma(n) = \cdots = n(n-1)(n-2)\cdots\Gamma(1) = n!$$

■

The Gamma function interpolates the factorial function $n!$ which is defined only for positive integers n. However, this property alone is not sufficient for a complete characterization of the Gamma function. We need another property. This will be done more in detail in the appendix to this chapter.

5.4 Integration of Vector-Valued Functions

A mapping $\gamma\colon [a,b] \to \mathbb{R}^k$, $\gamma(t) = (\gamma_1(t), \ldots, \gamma_k(t))$ is said to be continuous if all the mappings γ_i, $i = 1, \ldots, k$, are continuous. Moreover, if all the γ_i are differentiable, we write $\gamma'(t) = (\gamma_1'(t), \ldots, \gamma_k'(t))$.

Definition 5.7 Let $f_1, \ldots, f_k$ be real functions on $[a,b]$ and let $f = (f_1, \ldots, f_k)$ be the correponding mapping from $[a,b]$ into $\mathbb{R}^k$. If α increases on $[a,b]$, to say that $f \in \mathcal{R}(\alpha)$ means that $f_j \in \mathcal{R}(\alpha)$ for $j = 1, \ldots, k$. In this case we define

$$\int_a^b f \,\mathrm{d}\alpha = \left(\int_a^b f_1 \,\mathrm{d}\alpha, \ldots, \int_a^b f_k \,\mathrm{d}\alpha \right).$$

In other words $\int_a^b f \,\mathrm{d}\alpha$ is the point in $\mathbb{R}^k$ whose jth coordinate is $\int_a^b f_j \,\mathrm{d}\alpha$. It is clear that parts (a), (c), and (e) of Proposition 5.9 are valid for these vector valued integrals; we simply apply the earlier results to each coordinate. The same is true for Proposition 5.12, Theorem 5.14, and Theorem 5.15. To illustrate this, we state the analog of the fundamental theorem of calculus.

Theorem 5.28 *If $f = (f_1, \ldots, f_k) \in \mathcal{R}$ on $[a,b]$ and if $F = (F_1, \ldots, F_k)$ is an antiderivative of f on $[a,b]$, then*

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

The analog of Proposition 5.10 (b) offers some new features. Let $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ be any vector in $\mathbb{R}^k$. We denote its *Euclidean norm* by $\|x\| = \sqrt{x_1^2 + \cdots + x_k^2}$.

Proposition 5.29 *If $f = (f_1, \ldots, f_k) \in \mathcal{R}(\alpha)$ on $[a,b]$ then $\|f\| \in \mathcal{R}(\alpha)$ and*

$$\left\| \int_a^b f \,\mathrm{d}\alpha \right\| \leq \int_a^b \|f\| \,\mathrm{d}\alpha. \tag{5.42}$$

Proof. By the definition of the norm,

$$\|f\| = \left(f_1^2 + f_2^2 + \cdots + f_k^2\right)^{\frac{1}{2}}.$$

By Proposition 5.10 (a) each of the functions f_i^2 belong to $\mathcal{R}(\alpha)$; hence so does their sum $f_1^2 + f_2^2 + \cdots + f_k^2$. Note that the square-root is a continuous function on the positive half line. If we

apply Proposition 5.8 we see $\|f\| \in \mathcal{R}(\alpha)$.
To prove (5.42), put $y = (y_1, \dots, y_k)$ with $y_j = \int_a^b f_j \, \mathrm{d}\alpha$. Then we have $y = \int_a^b f \, \mathrm{d}\alpha$, and

$$\|y\|^2 = \sum_{j=1}^{k} y_j^2 = \sum_{j=1}^{k} y_j \int_a^b f_j \, \mathrm{d}\alpha = \int_a^b \sum_{j=1}^{k} (y_j f_j) \, \mathrm{d}\alpha.$$

By the Cauchy–Schwarz inequality, Proposition 1.25,

$$\sum_{j=1}^{k} y_j f_j(t) \le \|y\| \, \|f(t)\|, \quad t \in [a,b].$$

Inserting this into the preceding equation, the monotony of the integral gives

$$\|y\|^2 \le \|y\| \int_a^b \|f\| \, \mathrm{d}\alpha.$$

If $y = 0$, (5.42) is trivial. If $y \neq 0$, division by $\|y\|$ gives (5.42). ∎

Integration of Complex Valued Functions

This is a special case of the above arguments with $k = 2$. Let $\varphi\colon [a,b] \to \mathbb{C}$ be a complex-valued function. Let $u, v\colon [a,b] \to \mathbb{R}$ be the real and imaginary parts of φ, respectively; $u = \operatorname{Re}\varphi$ and $v = \operatorname{Im}\varphi$.
The function $\varphi = u + \mathrm{i}v$ is said to be *integrable* if $u, v \in \mathcal{R}$ on $[a,b]$ and we set

$$\int_a^b \varphi \, \mathrm{d}x = \int_a^b u \, \mathrm{d}x + \mathrm{i} \int_a^b v \, \mathrm{d}x.$$

The fundamental theorem of calculus holds: If the complex function φ is Riemann integrable, $\varphi \in \mathcal{R}$ on $[a,b]$ and $F(x)$ is an antiderivative of φ, then

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

Similarly, if u and v are both continuous, $F(x) = \int_a^x \varphi(t) \, \mathrm{d}t$ is an antiderivative of $\varphi(x)$.
Proof. Let $F = U + \mathrm{i}V$ be the antiderivative of φ where $U' = u$ and $V' = v$. By the fundamental theorem of calculus

$$\int_a^b \varphi \, \mathrm{d}x = \int_a^b u \, \mathrm{d}x + \mathrm{i} \int_a^b v \, \mathrm{d}x = U(b) - U(a) + \mathrm{i}\,(V(b) - V(a)) = F(b) - F(a).$$

∎

Example:

$$\int_a^b \mathrm{e}^{\alpha t} \, \mathrm{d}t = \frac{1}{\alpha} \mathrm{e}^{\alpha t} \Big|_a^b, \quad \alpha \in \mathbb{C}.$$

5.5 Inequalities

Besides the triangle inequality $\left| \int_a^b f \, \mathrm{d}\alpha \right| \leq \int_a^b |\, f \,| \, \mathrm{d}\alpha$ which was shown in Proposition 5.10 we can formulate Hölder's, Minkowski's, and the Cauchy–Schwarz inequalities for Riemann–Stieltjes integrals. For, let $p > 0$ be a fixed positive real number and α an increasing function on $[a, b]$. For $f \in \mathcal{R}(\alpha)$ define the L^p-norm

$$\|f\|_p = \left(\int_a^b |\, f \,|^p \, \mathrm{d}\alpha \right)^{\frac{1}{p}}. \tag{5.43}$$

Cauchy–Schwarz Inequality

Proposition 5.30 *Let $f, g \colon [a, b] \to \mathbb{C}$ be complex valued functions and $f, g \in \mathcal{R}$ on $[a, b]$. Then*

$$\left(\int_a^b |\, fg \,| \, \mathrm{d}x \right)^2 \leq \int_a^b |\, f \,|^2 \, \mathrm{d}x \cdot \int_a^b |\, g \,|^2 \, \mathrm{d}x. \tag{5.44}$$

Proof. Letting $f = |\, f \,|$ and $g = |\, g \,|$, it suffices to show $(\int fg \, \mathrm{d}x)^2 \leq \int f^2 \, \mathrm{d}x \cdot \int g^2 \, \mathrm{d}x$. For, put $A = \int_a^b g^2 \, \mathrm{d}x$, $B = \int_a^b fg \, \mathrm{d}x$, and $C = \int_a^b f^2 \, \mathrm{d}x$. Let $\lambda \in \mathbb{C}$ be arbitrary. By the positivity and linearity of the integal,

$$0 \leq \int_a^b (f + \lambda g)^2 \, \mathrm{d}x = \int_a^b f^2 \, \mathrm{d}x + 2\lambda \int_a^b fg \, \mathrm{d}x + \lambda^2 \int_a^b g^2 \, \mathrm{d}x = C + 2B\lambda + A\lambda^2 =: h(\lambda).$$

Thus, h is non-negative for all complex values λ.
Case 1. $A = 0$. Inserting this, we get $2B\lambda + C \geq 0$ for all $\lambda \in \mathbb{C}$. This implies $B = 0$ and $C \geq 0$; the inequality is satisfied.
Case 2. $A > 0$. Dividing the above inequality by A, we have

$$0 \leq \lambda^2 + \frac{2B}{A}\lambda + \frac{C}{A} \leq \left(\lambda + \frac{B}{A} \right)^2 - \left(\frac{B}{A} \right)^2 + \frac{C}{A}.$$

This is satisfied for all λ if and only if

$$\left(\frac{B}{A} \right)^2 \leq \frac{C}{A} \quad \text{and, finally} \quad B^2 \leq AC.$$

This completes the proof. ∎

Proposition 5.31 (a) Cauchy–Schwarz inequality. *Suppose $f, g \in \mathcal{R}(\alpha)$, then*

$$\left| \int_a^b fg \, \mathrm{d}\alpha \right| \leq \int_a^b |\, fg \,| \, \mathrm{d}\alpha \leq \sqrt{\int_a^b |\, f \,|^2 \, \mathrm{d}\alpha} \sqrt{\int_a^b |\, g \,|^2 \, \mathrm{d}\alpha} \quad \textit{or} \tag{5.45}$$

$$\int_a^b |\, fg \,| \, \mathrm{d}\alpha \leq \|f\|_2 \, \|g\|_2 \,. \tag{5.46}$$

(b) Hölder's inequality. *Let p and q be positive real numbers such that $\frac{1}{p}+\frac{1}{q}=1$. If $f, g \in \mathcal{R}(\alpha)$, then*

$$\left| \int_a^b fg \, \mathrm{d}\alpha \right| \leq \int_a^b |\, fg \,| \, \mathrm{d}\alpha \leq \|f\|_p \, \|g\|_q \,. \tag{5.47}$$

(c) Minkowski's inequality. *Let $p \geq 1$ and $f, g \in \mathcal{R}(\alpha)$, then*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p \,. \tag{5.48}$$

5.6 Appendix D

The composition of an integrable and a continuous function is integrable

Proof of Proposition 5.8. Let $\varepsilon > 0$. Since φ is uniformly continuous on $[m, M]$, there exists $\delta > 0$ such that $\delta < \varepsilon$ and $|\, \varphi(s) - \varphi(t) \,| < \varepsilon$ if $|\, s - t \,| < \delta$ and $[s, t \in [m, M]$.
Since $f \in \mathcal{R}(\alpha)$, there exists a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ such that

$$U(P, f, \alpha) - L(P, f, \alpha) < \delta^2. \tag{5.49}$$

Let M_i and m_i have the same meaning as in Definition 5.1, and let M_i^* and m_i^* the analogous numbers for h. Divide the numbers $1, 2, \ldots, n$ into two classes: $i \in A$ if $M_i - m_i < \delta$ and $i \in B$ if $M_i - m_i > \delta$. For $i \in A$ our choice of δ shows that $M_i^* - m_i^* \leq \varepsilon$. For $i \in B$, $M_i^* - m_i^* \leq 2K$ where $K = \sup\{|\, \varphi(t) \,| \mid m \leq t \leq M\}$. By (5.49), we have

$$\delta \sum_{i \in B} \Delta\alpha_i \leq \sum_{i \in B} (M_i - m_i) \Delta\alpha_i < \delta^2 \tag{5.50}$$

so that $\sum_{i \in B} \Delta\alpha_i < \delta$. It follows that

$$\begin{aligned} U(P, h, \alpha) - L(P, h, \alpha) = \sum_{i \in A} (M_i^* - m_i^*) \Delta\alpha_i + \sum_{i \in B} (M_i^* - m_i^*) \Delta\alpha_i \leq \\ \varepsilon(\alpha(b) - \alpha(a)) + 2K\delta < \varepsilon(\alpha(b) - \alpha(a) + 2K). \end{aligned}$$

Since ε was arbitrary, Proposition 5.3 implies that $h \in \mathcal{R}(\alpha)$. ∎

Convex Functions are Continuous

Proposition 5.32 *Every convex function $f : (a, b) \to \mathbb{R}$, $-\infty \leq a < b \leq +\infty$, is continuous.*

Proof. There is a very nice geometric proof in Rudin's book "Real and Complex Analysis", see [Rud66, 3.2 Theorem]. We give another proof here.
Let $x \in (a, b)$; choose a finite subinterval (x_1, x_2) with $a < x_1 < x < x_2 < b$. Since $f(x) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)$, $\lambda \in [0, 1]$, f is bounded above on $[x_1, x_2]$. Chosing x_3 with $x_1 < x_3 < x$ the convexity of f implies

$$\frac{f(x_3) - f(x_1)}{x_3 - x_1} \leq \frac{f(x) - f(x_1)}{x - x_1} \Longrightarrow f(x) \geq \frac{f(x_3) - f(x_1)}{x_3 - x_1} (x - x_1).$$

This means that f is bounded below on $[x_3, x_2]$ by a linear function; hence f is bounded on $[x_3, x_2]$, say $|\, f(x)\,| \le C$ on $[x_3, x_2]$.
The convexity implies

$$f\left(\frac{1}{2}(x+h)+\frac{1}{2}(x-h)\right) \le \frac{1}{2}\left(f(x+h)+f(x-h)\right)$$
$$\Longrightarrow f(x)-f(x-h) \le f(x+h)-f(x).$$

Iteration yields

$$f(x-(\nu-1)h)-f(x-\nu h) \le f(x+h)-f(x) \le f(x+\nu h)-f(x+(\nu-1)h).$$

Summing up over $\nu = 1, \ldots, n$ we have

$$f(x)-f(x-nh) \le n\left(f(x+h)-f(x)\right) \le f(x+nh)-f(x)$$
$$\Longrightarrow \frac{1}{n}\left(f(x)-f(x-nh)\right) \le f(x+h)-f(x) \le \frac{1}{n}\left(f(x+nh)-f(x)\right).$$

Let $\varepsilon > 0$ be given; choose $n \in \mathbb{N}$ such that $2C/n < \varepsilon$ and choose h such that $x_3 < x - nh < x < x + nh < x_2$. The above inequality then implies

$$|\, f(x+h)-f(x)\,| \le \frac{2C}{n} < \varepsilon.$$

This shows continuity of f at x. ∎

If g is an increasing convex function and f is a convex function, then $g \circ f$ is convex since $f(\lambda x + \mu y) \le \lambda f(x) + \mu f(y)$, $\lambda + \mu = 1$, $\lambda, \mu \ge 0$, implies

$$g(f(\lambda x+\mu y)) \le g(\lambda f(x)+\mu g(x)) \le \lambda g(f(x))+\mu g(f(y)).$$

5.6.1 More on the Gamma Function

Let $I \subset \mathbb{R}$ be an interval. A positive function $F\colon I \to \mathbb{R}$ is called *logarithmic convex* if $\log F\colon I \to \mathbb{R}$ is convex, i. e. for every $x, y \in I$ and every λ, $0 \le \lambda \le 1$ we have

$$F(\lambda x+(1-\lambda)y) \le F(x)^{\lambda}\, F(y)^{1-\lambda}.$$

Proposition 5.33 *The Gamma function is logarithmic convex.*

Proof. Let $x, y > 0$ and $0 < \lambda < 1$ be given. Set $p = 1/\lambda$ and $q = 1/(1-\lambda)$. Then $1/p + 1/q = 1$ and we apply Hölder's inequality to the functions

$$f(t) = t^{\frac{x-1}{p}} \mathrm{e}^{-\frac{t}{p}}, \quad g(t) = t^{\frac{y-1}{q}} \mathrm{e}^{-\frac{t}{q}}$$

and obtain

$$\int_{\varepsilon}^{R} f(t)g(t)\,\mathrm{d}t \le \left(\int_{\varepsilon}^{R} f(t)^p\,\mathrm{d}t\right)^{\frac{1}{p}} \left(\int_{\varepsilon}^{R} g(t)^q\,\mathrm{d}t\right)^{\frac{1}{q}}.$$

Note that

$$f(t)g(t) = t^{\frac{x}{p}+\frac{y}{q}-1}\mathrm{e}^{-t}, \quad f(t)^p = t^{x-1}\mathrm{e}^{-t}, \quad g(t)^q = t^{y-1}\mathrm{e}^{-t}.$$

Taking the limts $\varepsilon \to 0+0$ and $R \to +\infty$ we obtain

$$\Gamma\left(\frac{x}{p}+\frac{y}{q}\right) \le \Gamma(x)^{\frac{1}{p}}\Gamma(y)^{\frac{1}{q}}.$$

■

Remark 5.5 One can prove that a convex function (see Definition 4.4) is continuous, see Proposition 5.32. Also, an increasing convex function of a convex function f is convex, for example e^f is convex if f is. We conclude that $\Gamma(x)$ is continuous for $x > 0$.

Theorem 5.34 *Let $F\colon (0,+\infty) \to (0,+\infty)$ be a function with*

(a) $F(1) = 1$,
(b) $F(x+1) = xF(x)$,
(c) F *is logarithmic convex.*

Then $F(x) = \Gamma(x)$ for all $x > 0$.

Proof. Since $\Gamma(x)$ has the properties (a), (b), and (c) it suffices to prove that F is uniquely determined by (a), (b), and (c). By (b),

$$F(x+n) = F(x)x(x+1)\cdots(x+n)$$

for every positive x and every positive integer n. In particular $F(n+1) = n!$ and it suffices to show that $F(x)$ is uniquely determined for every x with $x \in (0,1)$. Since $n+x = (1-x)n + x(n+1)$ from (c) it follows

$$F(n+x) \le F(n)^{1-x}F(n+1)^x = F(n)^{1-x}F(n)^x\, n^x = (n-1)!n^x.$$

Similarly, from $n+1 = x(n+x) + (1-x)((n+1+x)$ it follows

$$n! = F(n+1) \le F(n+x)^x F(n+1+x)^{1-x} = F(n+x)(n+x)^{1-x}.$$

Combining both inequalities,

$$n!(n+x)^{x-1} \le F(n+x) \le (n-1)!n^x$$

and moreover

$$a_n(x) := \frac{n!(n+x)^{x-1}}{x(x+1)\cdots(x+n-1)} \le F(x) \le \frac{(n-1)!n^x}{x(x+1)\cdots(x+n-1)} =: b_n(x).$$

Since $\frac{b_n(x)}{a_n(x)} = \frac{(n+x)n^x}{n(n+x)^x}$ converges to 1 as $n \to \infty$,

$$F(x) = \lim_{n\to\infty} \frac{(n-1)!n^x}{x(x+1)\cdots(x+n)}.$$

Hence F is uniquely determined. ■

Stirling's Formula

We give an asymptotic formula for $n!$ as $n \to \infty$. We call two sequences (a_n) and (b_n) to be *asymptotically equal* if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$, and we write $a_n \sim b_n$.

Proposition 5.35 (Stirling's Formula) *The asymptotical behavior of $n!$ is*

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n.$$

Proof. Using the trapezoid rule (5.34) with $f(x) = \log x$, $f''(x) = -1/x^2$ we have

$$\int_k^{k+1} \log x \,\mathrm{d}x = \frac{1}{2}\left(\log k + \log(k+1)\right) + \frac{1}{12\xi_k^2}$$

with $k \le \xi_k \le k+1$. Summation over $k = 1, \dots, n-1$ gives

$$\int_1^n \log x \,\mathrm{d}x = \sum_{k=1}^{n} \log k - \frac{1}{2}\log n + \frac{1}{12}\sum_{k=1}^{n-1} \frac{1}{\xi_k^2}.$$

Since $\int \log x \,\mathrm{d}x = x\log x - x$ (integration by parts), we have

$$n\log n - n + 1 = \sum_{k=1}^{n} \log k - \frac{1}{2}\log n + \frac{1}{12}\sum_{k=1}^{n-1} \frac{1}{\xi_k^2}$$
$$\sum_{k=1}^{n} \log k = \left(n + \frac{1}{2}\right)\log n - n + \gamma_n,$$

where $\gamma_n = 1 - \frac{1}{12}\sum_{k=1}^{n-1} \frac{1}{\xi_k^2}$. Exponentiating both sides of the equation we find with $c_n = \mathrm{e}^{\gamma_n}$

$$n! = n^{n+\frac{1}{2}}\,\mathrm{e}^{-n}\,c_n. \tag{5.51}$$

Since $0 < 1/\xi_k^2 \le 1/k^2$, the limit

$$\gamma = \lim_{n\to\infty} \gamma_n = 1 - \sum_{k=1}^{\infty} \frac{1}{\xi_k^2}$$

exists, and so the limit $c = \lim_{n\to\infty} c_n = \mathrm{e}^{\gamma}$.
Proof of $c_n \to \sqrt{2\pi}$. Using (5.51) we have

$$\frac{c_n^2}{c_{2n}} = \frac{(n!)^2\sqrt{2n}(2n)^{2n}}{n^{2n+1}(2n)!} = \sqrt{2}\,\frac{2^{2n}(n!)^2}{\sqrt{n}(2n)!}$$

and $\lim_{n\to\infty} \frac{c_n^2}{c_{2n}} = \frac{c^2}{c} = c$. Using Wallis's product formula for π

$$\pi = 2\prod_{k=1}^{\infty} \frac{4k^2}{4k^2-1} = \lim_{n\to\infty} 2\frac{2{\cdot}2{\cdot}4{\cdot}4\cdots 2n{\cdot}2n}{1{\cdot}3{\cdot}3{\cdot}5\cdots(2n-1)(2n+1)} \tag{5.52}$$

we have

$$\left(2\prod_{k=1}^{n}\frac{4k^2}{4k^2-1}\right)^{\frac{1}{2}} = \sqrt{2}\frac{2\cdot4\cdots2n}{3\cdot5\cdots(2n-1)\sqrt{2n+1}} = \frac{1}{\sqrt{n+\frac{1}{2}}}\cdot\frac{2^2\cdot4^2\cdots(2n)^2}{2\cdot3\cdot4\cdots(2n-1)(2n)}$$
$$= \frac{1}{\sqrt{n+\frac{1}{2}}}\cdot\frac{2^{2n}(n!)^2}{(2n)!},$$

such that

$$\sqrt{\pi} = \lim_{n\to\infty}\frac{2^{2n}(n!)^2}{\sqrt{n}(2n)!}.$$

Consequently, $c = \sqrt{2\pi}$ which completes the proof.

■

Proof of Hölder's Inequality

Proof of Proposition 5.31. We prove (b). The other two statements are consequences, their proofs are along the lines in Section 1.3. The main idea is to approximate the integral on the left by Riemann sums and use Hölder's inequality (1.22). Let $\varepsilon > 0$; without loss of generality, let $f, g \geq 0$. By Proposition 5.10 fg, f^p, $g^q \in \mathcal{R}(\alpha)$ and by Proposition 5.3 there exist partitions P_1, P_2, and P_3 of $[a,b]$ such that $U(fg,P_1,\alpha)-L(fg,P_1,\alpha)<\varepsilon$, $U(f^p,P_2,\alpha)-L(f^p,P_2,\alpha)<\varepsilon$, and $U(g^q,P_3,\alpha)-L(g^q,P_3,\alpha)<\varepsilon$. Let $P=\{x_0,x_1,\ldots,x_n\}$ be the common refinement of P_1, P_2, and P_3. By Lemma 5.4 (a) and (c)

$$\int_a^b fg\,\mathrm{d}\alpha < \sum_{i=1}^{n}(fg)(t_i)\Delta\alpha_i + \varepsilon, \tag{5.53}$$

$$\sum_{i=1}^{n} f(t_i)^p\Delta\alpha_i < \int_a^b f^p\,\mathrm{d}\alpha + \varepsilon, \tag{5.54}$$

$$\sum_{i=1}^{n} g(t_i)^q\Delta\alpha_i < \int_a^b g^q\,\mathrm{d}\alpha + \varepsilon, \tag{5.55}$$

for any $t_i \in [x_{i-1},x_i]$. Using the two preceding inequalities and Hölder's inequality (1.22) we have

$$\sum_{i=1}^{n} f(t_i)\Delta\alpha_i^{\frac{1}{p}} g(t_i)\Delta\alpha_i^{\frac{1}{q}} \leq \left(\sum_{i=1}^{n} f(t_i)^p\Delta\alpha_i\right)^{\frac{1}{p}}\left(\sum_{i=1}^{n} g(t_i)^q\Delta\alpha_i\right)^{\frac{1}{q}}$$
$$< \left(\int_a^b f^p\,\mathrm{d}\alpha+\varepsilon\right)^{\frac{1}{p}}\left(\int_a^b g^q\,\mathrm{d}\alpha+\varepsilon\right)^{\frac{1}{q}}.$$

By (5.53),

$$\int_a^b fg\,\mathrm{d}\alpha < \sum_{i=1}^{n}(fg)(t_i)\Delta\alpha_i+\varepsilon < \left(\int_a^b f^p\,\mathrm{d}\alpha+\varepsilon\right)^{\frac{1}{p}}\left(\int_a^b g^q\,\mathrm{d}\alpha+\varepsilon\right)^{\frac{1}{q}}+\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the claim follows. ∎

Chapter 6

Sequences of Functions and Basic Topology

In the present chapter we draw our attention to complex-valued functions (including the real-valued), although many of the theorems and proofs which follow extend to vector-valued functions without difficulty and even to mappings into more general spaces. We stay within this simple framework in order to focus attention on the most important aspects of the problem that arise when **limit processes are interchanged.**

6.1 Discussion of the Main Problem

Definition 6.1 Suppose (f_n), $n \in \mathbb{N}$, is a sequence of functions defined on a set E, and suppose that the sequence of numbers $(f_n(x))$ converges for every $x \in E$. We can then define a function f by

$$f(x) = \lim_{n\to\infty} f_n(x), \quad x \in E. \tag{6.1}$$

Under these circumstances we say that (f_n) *converges* on E and f is the *limit* (or the *limit function*) of (f_n). Sometimes we say that "(f_n) *converges pointwise* to f on E" if (6.1) holds. Similarly, if $\sum_{n=1}^{\infty} f_n(x)$ converges for every $x \in E$, and if we define

$$f(x) = \sum_{n=1}^{\infty} f_n(x), \quad x \in E, \tag{6.2}$$

the function f is called the *sum* of the series $\sum_{n=1}^{\infty} f_n$.

The main problem which arises is to determine whether important properties of the functions f_n are preserved under the limit operations (6.1) and (6.2). For instance, if the functions f_n are continuous, or differentiable, or integrable, is the same true of the limit function? What are the relations between f_n' and f', say, or between the integrals of f_n and that of f? To say that f is continuous at x means

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

Hence, to ask whether the limit of a sequence of continuous functions is continuous is the same as to ask whether

$$\lim_{t\to x}\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty}\lim_{t\to x} f_n(t) \tag{6.3}$$

i. e. whether the order in which limit processes are carried out is immaterial. We shall now show by means of several examples that limit processes cannot in general be interchanged without affecting the result. Afterwards, we shall prove that under certain conditions the order in which limit operations are carried out is inessential.

Example 6.1 (a) Our first example, and the simplest one, concerns a "double sequence." For positive integers $m, n \in \mathbb{N}$ let

$$s_{mn} = \frac{m}{m+n}.$$

Then, for fixed n

$$\lim_{m\to\infty} s_{mn} = 1,$$

so that $\lim\limits_{n\to\infty}\lim\limits_{m\to\infty} s_{mn} = 1$. On the other hand, for every fixed m,

$$\lim_{n\to\infty} s_{mn} = 0,$$

so that $\lim\limits_{m\to\infty}\lim\limits_{n\to\infty} s_{mn} = 0$. The two limits cannot be interchanged.

(b) Let $f_n(x) = x^n$ on $[0,1]$. Then $f(x) = \begin{cases} 0, & 0 \le x < 1, \\ 1, & x = 1. \end{cases}$ All functions $f_n(x)$ are continous on $[0,1]$; however, the limit $f(x)$ is discontinuous at $x = 1$; that is $\lim\limits_{x\to 1-0}\lim\limits_{n\to\infty} t^n = 0 \neq 1 = \lim\limits_{n\to\infty}\lim\limits_{t\to 1-0} t^n$. The limits cannot be interchanged.

After these examples, which show what can go wrong if limit processes are interchanged carelessly, we now define a new notion of convergence, stronger than pointwise convergence as defined in Definition 6.1, which will enable us to arrive at positive results.

6.2 Uniform Convergence

6.2.1 Definitions and Example

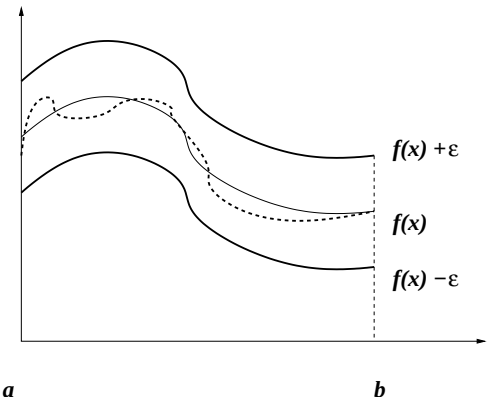
Definition 6.2 A sequence of functions (f_n) converges *uniformly* on E to a function f if for every $\varepsilon > 0$ there is a positive integer n_0 such that $n \ge n_0$ implies

$$|\, f_n(x) - f(x)\,| \le \varepsilon \tag{6.4}$$

for all $x \in E$. We write $f_n \rightrightarrows f$ on E.

As a formula, $f_n \rightrightarrows f$ on E if

$$\forall\, \varepsilon > 0 \;\; \exists\, n_0 \in \mathbb{N} \;\; \forall\, n \ge n_0 \;\; \forall\, x \in E : |\, f_n(x) - f(x)\,| \le \varepsilon.$$



Uniform convergence of f_n to f on $[a, b]$ means that f_n is in the ε-tube of f for sufficiently large n

It is clear that every uniformly convergent sequence is pointwise convergent (to the same function). Quite explicitly, the difference between the two concepts is this: If (f_n) converges pointwise on E to a function f, for every $\varepsilon > 0$ and for every $x \in E$, there exists an integer n_0 depending on both ε *and* $x \in E$ such that (6.4) holds if $n \geq n_0$. If (f_n) converges uniformly on E it is possible, for each $\varepsilon > 0$ to find *one* integer n_0 which will do for *all* $x \in E$.
We say that the series $\sum_{k=1}^{\infty} f_k(x)$ converges *uniformly* on E if the sequence $(s_n(x))$ of partial sums defined by

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

converges uniformly on E.

Proposition 6.1 (Cauchy criterion) (a) *The sequence of functions (f_n) defined on E converges uniformly on E if and only if for every $\varepsilon > 0$ there is an integer n_0 such that $n, m \geq n_0$ and $x \in E$ imply*

$$| f_n(x) - f_m(x) | \leq \varepsilon. \tag{6.5}$$

(b) The series of functions $\sum_{k=1}^{\infty} g_k(x)$ defined on E converges uniformly on E if and only if for every $\varepsilon > 0$ there is an integer n_0 such that $n, m \geq n_0$ and $x \in E$ imply

$$\left| \sum_{k=m}^{n} g_k(x) \right| \leq \varepsilon.$$

Proof. Suppose (f_n) converges uniformly on E and let f be the limit function. Then there is an integer n_0 such that $n \geq n_0$, $x \in E$ implies

$$| f_n(x) - f(x) | \leq \frac{\varepsilon}{2},$$

so that

$$| f_n(x) - f_m(x) | \leq | f_n(x) - f(x) | + | f_m(x) - f(x) | \leq \varepsilon$$

if $m, n \geq n_0$, $x \in E$.
Conversely, suppose the Cauchy condition holds. By Proposition 2.18, the sequence $(f_n(x))$ converges for every x to a limit which may we call $f(x)$. Thus the sequence (f_n) converges

pointwise on E to f. We have to prove that the convergence is uniform. Let $\varepsilon > 0$ be given, choose n_0 such that (6.5) holds. Fix n and let $m \to \infty$ in (6.5). Since $f_m(x) \to f(x)$ as $m \to \infty$ this gives

$$| f_n(x) - f(x) | \leq \varepsilon$$

for every $n \geq n_0$ and $x \in E$.
(b) immediately follows from (a) with $f_n(x) = \sum_{k=1}^{n} g_k(x)$. ∎

Remark 6.1 Suppose

$$\lim_{n\to\infty} f_n(x) = f(x), \quad x \in E.$$

Put

$$M_n = \sup_{x\in E} | f_n(x) - f(x) | .$$

Then $f_n \rightrightarrows f$ uniformly on E if and only if $M_n \to 0$ as $n \to \infty$. **(prove!)**

The following comparison test of a function series with a numerical series gives a sufficient criterion for uniform convergence.

Theorem 6.2 (Weierstraß) *Suppose (f_n) is a sequence of functions defined on E, and suppose*

$$| f_n(x) | \leq M_n, \quad x \in E,\ n \in \mathbb{N}. \tag{6.6}$$

Then $\sum_{n=1}^{\infty} f_n$ converges uniformly on E if $\sum_{n=1}^{\infty} M_n$ converges.

Proof. If $\sum M_n$ converges, then, for arbitrary $\varepsilon > 0$ there exists n_0 such that $m, n \geq n_0$ implies $\sum_{i=m}^{n} M_i \leq \varepsilon$. Hence,

$$\left| \sum_{i=m}^{n} f_i(x) \right| \underset{\text{tr.In.}}{\leq} \sum_{i=m}^{n} | f_i(x) | \underset{(6.6)}{\leq} \sum_{i=m}^{n} M_i \leq \varepsilon, \qquad \forall\, x \in E.$$

Uniform convergence now follows from Proposition 6.1. ∎

Proposition 6.3 (Comparison Test) *If $\sum_{n=1}^{\infty} g_n(x)$ converges uniformly on E and $| f_n(x) | \leq g_n(x)$ for all sufficiently large n and all $x \in E$ then $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on E.*

Proof. Apply the Cauchy criterion. Note that

$$\left| \sum_{n=k}^{m} f_n(x) \right| \leq \sum_{n=k}^{m} | f_n(x) | \leq \sum_{n=k}^{m} g_n(x) < \varepsilon.$$

∎

Application of Weierstraß' Theorem to Power Series and Fourier Series

Proposition 6.4 *Let*

$$\sum_{n=0}^{\infty} a_n z^n, \quad a_n \in \mathbb{C}, \tag{6.7}$$

be a power series with radius of convergence $R > 0$.
Then (6.7) *converges uniformly on the closed disc* $\{z \mid |z| \le r\}$ *for every* r *with* $0 \le r < R$.

Proof. We apply Weierstraß' theorem to $f_n(z) = a_n z^n$. Note that

$$|f_n(z)| = |a_n|\,|z|^n \le |a_n|\, r^n.$$

Since $r < R$, r belongs to the disc of convergence, the series $\sum_{n=0}^{\infty} |a_n|\, r^n$ converges by Theorem 2.34. By Theorem 6.2, the series $\sum_{n=0}^{\infty} a_n z^n$ converges uniformly on $\{z \mid |z| \le r\}$. ∎

Remark 6.2 (a) The power series

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

has the same radius of convergence R as the series (6.7) and hence also converges uniformly on the closed disc $\{z \mid |z| \le r\}$.
Indeed, this simply follows from the fact that

$$\overline{\lim_{n\to\infty}} \sqrt[n]{(n+1)\,|a_{n+1}|} = \lim_{n\to\infty} \sqrt[n]{n+1}\, \overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} = \frac{1}{R}.$$

(b) Note that the power series in general does *not* converge uniformly on the whole open disc of convergence $|z| < R$. As an example, consider the geometric series

$$f(z) = \frac{1}{1-z} = \sum_{k=0}^{\infty} z^k, \quad |z| < 1.$$

Note that the condition

$$\exists\, \varepsilon_0 > 0 \;\; \forall\, n \in \mathbb{N} \;\; \exists\, x_n \in E\colon \; |f_n(x_n) - f(x_n)| \ge \varepsilon_0$$

implies that (f_n) *does not* converge uniformly to f on E. **Prove!**
To $\varepsilon = 1$ and every $n \in \mathbb{N}$ choose $z_n = \frac{n}{n+1}$ and we obtain, using Bernoulli's inequality,

$$z_n^n = \left(1 - \frac{1}{n+1}\right)^n \ge 1 - n\frac{1}{n+1} = 1 - z_n, \quad \text{hence} \quad \frac{z_n^n}{1-z_n} \ge 1. \tag{6.8}$$

so that

$$|s_{n-1}(z_n) - f(z_n)| = \left|\sum_{k=0}^{n-1} z_n^k - \frac{1}{1-z_n}\right| = \left|\sum_{k=n}^{\infty} z_n^k\right| = \frac{z_n^n}{1-z_n} \underset{(6.8)}{\ge} 1.$$

The geometric series doesn't converge uniformly on the whole open unit disc.

Example 6.2 (a) A series of the form

$$\sum_{n=0}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx), \quad a_n,\, b_n, x \in \mathbb{R}, \tag{6.9}$$

is called a *Fourier series* (see Section 6.3 below). If both $\sum_{n=0}^{\infty} |\, a_n \,|$ and $\sum_{n=0}^{\infty} |\, b_n \,|$ converge then the series (6.9) converges uniformly on $\mathbb{R}$ to a function $F(x)$.
Indeed, since $|\, a_n \cos(nx) \,| \le |\, a_n \,|$ and $|\, b_n \sin(nx) \,| \le |\, b_n \,|$, by Theorem 6.2, the series (6.9) converges uniformly on $\mathbb{R}$.

(b) Let $f\colon \mathbb{R} \to \mathbb{R}$ be the sum of the Fourier series

$$f(x) = \sum_{n=1}^{\infty} \frac{\sin nx}{n} \tag{6.10}$$

Note that (a) does not apply since $\sum_n |\, b_n \,| = \sum_n \frac{1}{n}$ diverges.
If $f(x)$ exists, so does $f(x+2\pi) = f(x)$, and $f(0) = 0$. We will show that the series converges uniformly on $[\delta, 2\pi - \delta]$ for every $\delta > 0$. For, put

$$s_n(x) = \sum_{k=1}^{n} \sin kx = \operatorname{Im} \left(\sum_{k=1}^{n} \mathrm{e}^{\mathrm{i}kx} \right).$$

If $\delta \le x \le 2\pi - \delta$ we have

$$|\, s_n(x) \,| \le \left| \sum_{k=1}^{n} \mathrm{e}^{\mathrm{i}kx} \right| = \left| \frac{\mathrm{e}^{\mathrm{i}(n+1)x} - \mathrm{e}^{\mathrm{i}x}}{\mathrm{e}^{\mathrm{i}x} - 1} \right| \le \frac{2}{|\, \mathrm{e}^{\mathrm{i}x/2} - \mathrm{e}^{-\mathrm{i}x/2} \,|} = \frac{1}{\sin \frac{x}{2}} \le \frac{1}{\sin \frac{\delta}{2}}.$$

Note that $|\operatorname{Im} z| \le |z|$ and $\left|\, \mathrm{e}^{\mathrm{i}x} \,\right| = 1$. Since $\sin \frac{x}{2} \ge \sin \frac{\delta}{2}$ for $\delta/2 \le x/2 \le \pi - \delta/2$ we have for $0 < m < n$

$$\begin{aligned}
\left| \sum_{k=m}^{n} \frac{\sin kx}{k} \right| &= \left| \sum_{k=m}^{n} \frac{s_k(x) - s_{k-1}(x)}{k} \right| \\
&= \left| \sum_{k=m}^{n} s_k(x) \left(\frac{1}{k} - \frac{1}{k+1} \right) + \frac{s_n(x)}{n+1} - \frac{s_{m-1}(x)}{m} \right| \\
&\le \frac{1}{\sin \frac{\delta}{2}} \left(\left| \sum_{k=m}^{n} \left(\frac{1}{k} - \frac{1}{k+1} \right) + \frac{1}{n+1} \right| + \left| \frac{1}{m} \right| \right) \\
&\le \frac{1}{\sin \frac{\delta}{2}} \left(\frac{1}{m} - \frac{1}{n+1} + \frac{1}{n+1} + \frac{1}{m} \right) \le \frac{2}{m \sin \frac{\delta}{2}}
\end{aligned}$$

The right side becomes arbitrarily small as $m \to \infty$. Using Proposition 6.1 (b) uniform convergence of (6.10) on $[\delta, 2\pi - \delta]$ follows.

6.2.2 Uniform Convergence and Continuity

Theorem 6.5 *Let $E \subset \mathbb{R}$ be a subset and $f_n\colon E \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of continuous functions on E uniformly converging to some function $f\colon E \to \mathbb{R}$.*
Then f is continuous on E.

Proof. Let $a \in E$ and $\varepsilon > 0$ be given. Since $f_n \rightrightarrows f$ there is an $r \in \mathbb{N}$ such that

$$| f_r(x) - f(x) | \leq \varepsilon/3 \quad \text{for all} \quad x \in E.$$

Since f_r is continuous at a, there exists $\delta > 0$ such that $| x - a | < \delta$ implies

$$| f_r(x) - f(a) | \leq \varepsilon/3.$$

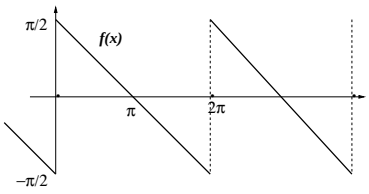
Hence $| x - a | < \delta$ implies

$$| f(x) - f(a) | \leq | f(x) - f_r(x) | + | f_r(x) - f_r(a) | + | f_r(a) - f(a) | \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

This proves the assertion. ∎

The same is true for functions $f\colon E \to \mathbb{C}$ where $E \subset \mathbb{C}$.

Example 6.3 (Example 6.2 continued) (a) A power series defines a continuous functions on the disc of convergence $\{z \mid | z | < R\}$.
Indeed, let $| z_0 | < R$. Then there exists $r \in \mathbb{R}$ such that $| z_0 | < r < R$. By Proposition 6.4, $\sum_n a_n x^n$ converges uniformly on $\{x \mid | x | \leq r\}$. By Theorem 6.5, the function, defined by the sum of the power series is continuous on $[-r, r]$.
(b) The sum of the Fourier series (6.9) is a continuous function on $\mathbb{R}$ if both $\sum_n | a_n |$ and $\sum_n | b_n |$ converge.
(c) The sum of the Fourier series $f(x) = \sum_n \frac{\sin(nx)}{n}$ is continuous on $[\delta, 2\pi - \delta]$ for all δ with $0 < \delta < \pi$ by the above theorem. Clearly, f is 2π-periodic since all partial sums are.



Later (see the section on Fourier series) we will show that

$$f(x) = \begin{cases} 0, & x = 0, \\ \frac{\pi - x}{2}, & x \in (0, 2\pi), \end{cases}$$

Since f is discontinuous at $x_0 = 2\pi n$, the Fourier series does not converge uniformly on $\mathbb{R}$.

Also, Example 6.1 (b) shows that the continuity of the $f_n(x) = x^n$ alone is not sufficient for the continuity of the limit function. On the other hand, the sequence of continuous functions (x^n) on $(0, 1)$ converges to the continuous function 0. However, the convergence is not uniform. **Prove!**

6.2.3 Uniform Convergence and Integration

Example 6.4 Let $f_n(x) = 2n^2 x \mathrm{e}^{-n^2 x^2}$; clearly $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in \mathbb{R}$. Further

$$\int_0^1 f_n(x)\,\mathrm{d}x = -\mathrm{e}^{-n^2x^2}\Big|_0^1 = \left(1 - \mathrm{e}^{-n^2}\right) \longrightarrow 1.$$

On the other hand $\int_0^1 \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = \int_0^1 0\,\mathrm{d}x = 0$. Thus, $\lim_{n\to\infty}$ and integration cannot be interchanged. The reason, (f_n) converges pointwise to 0 but not uniformly. Indeed,

$$f_n\left(\frac{1}{n}\right) = \frac{2n^2}{n}\mathrm{e}^{-1} = \frac{2n}{\mathrm{e}} \underset{n\to\infty}{\longrightarrow} +\infty.$$

Theorem 6.6 *Let α be an increasing function on $[a,b]$. Suppose $f_n \in \mathcal{R}(\alpha)$ on $[a,b]$ for all $n \in \mathbb{N}$ and suppose $f_n \to f$ uniformly on $[a,b]$. Then $f \in \mathcal{R}(\alpha)$ on $[a,b]$ and*

$$\int_a^b f\,\mathrm{d}\alpha = \lim_{n\to\infty} \int_a^b f_n\,\mathrm{d}\alpha. \tag{6.11}$$

Proof. Put

$$\varepsilon_n = \sup_{x\in[a,b]} |\, f_n(x) - f(x)\,|\,.$$

Then

$$f_n - \varepsilon_n \le f \le f_n + \varepsilon_n,$$

so that the upper and the lower integrals of f satisfy

$$\int_a^b (f_n - \varepsilon_n)\,\mathrm{d}\alpha \le \underline{\int_a^b} f\,\mathrm{d}\alpha \le \overline{\int_a^b} f\,\mathrm{d}\alpha \le \int_a^b (f_n + \varepsilon_n)\,\mathrm{d}\alpha. \tag{6.12}$$

Hence,

$$0 \le \overline{\int} f\,\mathrm{d}\alpha - \underline{\int} f\,\mathrm{d}\alpha \le 2\varepsilon_n(\alpha(b) - \alpha(a)).$$

Since $\varepsilon_n \to 0$ as $n \to \infty$ (Remark 6.1), the upper and the lower integrals of f are equal. Thus $f \in \mathcal{R}(\alpha)$. Another application of (6.12) yields

$$\int_a^b (f_n - \varepsilon_n)\,\mathrm{d}\alpha \le \int_a^b f\,\mathrm{d}\alpha \le \int_a^b (f_n + \varepsilon_n)\,\mathrm{d}\alpha$$
$$\left| \int_a^b f\,\mathrm{d}\alpha - \int_a^b f_n\,\mathrm{d}\alpha \right| \le \varepsilon_n((\alpha(b) - \alpha(a)).$$

This implies (6.11). ∎

Corollary 6.7 *If $f_n \in \mathcal{R}(\alpha)$ on $[a,b]$ and if the series*

$$f(x) = \sum_{n=1}^{\infty} f_n(x), \quad a \le x \le b$$

converges uniformly on $[a,b]$, then

$$\int_a^b \left(\sum_{n=1}^{\infty} f_n\right) \mathrm{d}\alpha = \sum_{n=1}^{\infty} \left(\int_a^b f_n\,\mathrm{d}\alpha\right).$$

In other words, the series may be integrated term by term.

Corollary 6.8 *Let $f_n\colon [a,b] \to \mathbb{R}$ be a sequence of continuous functions uniformly converging on $[a,b]$ to f. Let $x_0 \in [a,b]$.*
Then the sequence $F_n(x) = \int_{x_0}^x f_n(t)\,\mathrm{d}t$ converges uniformly to $F(x) = \int_{x_0}^x f(t)\,\mathrm{d}t$.

Proof. The pointwise convergence of F_n follows from the above theorem with $\alpha(t) = t$ and a and b replaced by x_0 and x.
We show uniform convergence: Let $\varepsilon > 0$. Since $f_n \rightrightarrows f$ on $[a,b]$, there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies $|\, f_n(t) - f(t)\,| \leq \frac{\varepsilon}{b-a}$ for all $t \in [a,b]$. For $n \geq n_0$ and all $x \in [a,b]$ we thus have

$$|\, F_n(x) - F(x)\,| = \left| \int_{x_0}^x (f_n(t) - f(t))\,\mathrm{d}t \right| \leq \int_{x_0}^x |\, f_n(t) - f(t)\,|\,\mathrm{d}t \leq \frac{\varepsilon}{b-a}\,(b-a) = \varepsilon.$$

Hence, $F_n \rightrightarrows F$ on $[a,b]$. ∎

Example 6.5 (a) For every real $t \in (-1,1)$ we have

$$\log(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} \mp \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} t^n. \tag{6.13}$$

Proof. In Homework 13.5 (a) there was computed the Taylor series

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

of $\log(1+x)$ and it was shown that $T(x) = \log(1+x)$ if $x \in (0,1)$.
By Proposition 6.4 the geometric series $\sum_{n=0}^{\infty} (-1)^n x^n$ converges uniformly to the function $\frac{1}{1+x}$ on $[-r,r]$ for all $0 < r < 1$. By Corollary 6.7 we have for all $t \in [-r,r]$

$$\begin{aligned} \log(1+t) = \log(1+x)\big|_0^t = \int_0^t \frac{\mathrm{d}x}{1+x} &= \int_0^t \sum_{n=0}^{\infty} (-1)^n x^n\,\mathrm{d}x \\ \underset{\text{Cor 6.7}}{=} \sum_{n=0}^{\infty} \int_0^t (-1)^n x^n\,\mathrm{d}x &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^{n+1}\Big|_0^t = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} t^n \end{aligned}$$

∎

(b) For $|\, t\,| < 1$ we have

$$\arctan t = t - \frac{t^3}{3} + \frac{t^5}{5} \mp \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{t^{2n+1}}{2n+1} \tag{6.14}$$

As in the previous example we use the uniform convergence of the geometric series on $[-r,r]$ for every $0 < r < 1$ that allows to exchange integration and summation

$$\arctan t = \int_0^t \frac{\mathrm{d}x}{1+x^2} = \int_0^t \sum_{n=0}^{\infty} (-1)^n x^{2n}\,\mathrm{d}x = \sum_{n=0}^{\infty} (-1)^n \int_0^t x^{2n}\,\mathrm{d}x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} t^{2n+1}.$$

Note that you are, in general, not allowed to insert $t = 1$ into the equations (6.13) and (6.14). However, the following proposition (the proof is in the appendix to this chapter) fills this gap.

Proposition 6.9 (Abel's Limit Theorem) *Let $\sum_{n=0}^{\infty} a_n$ a convergent series of real numbers. Then the power series*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

converges for $x \in [0,1]$ and is continuous on $[0,1]$.

As a consequence of the above proposition we have

$$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \pm \cdots = \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{n},$$
$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} \pm \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}.$$

Example 6.6 We have $f_n(x) = \dfrac{1}{n}\,\mathrm{e}^{-\frac{x}{n}} \rightrightarrows f(x) \equiv 0$ on $[0, +\infty)$. Indeed, $|\, f_n(x) - 0\,| \le \frac{1}{n} < \varepsilon$ if $n \ge \frac{1}{\varepsilon}$ and for all $x \in \mathbb{R}_+$. However,

$$\int_0^{\infty} f_n(t)\,\mathrm{d}t = -\mathrm{e}^{-\frac{t}{n}}\Big|_0^{+\infty} = \lim_{t\to+\infty}\left(1 - \mathrm{e}^{-\frac{t}{n}}\right) = 1.$$

Hence

$$\lim_{n\to\infty} \int_0^{\infty} f_n(t)\,\mathrm{d}t = 1 \neq 0 = \int_0^{\infty} f(t)\,\mathrm{d}t.$$

That is, Theorem 6.6 fails in case of improper integrals.

6.2.4 Uniform Convergence and Differentiation

Example 6.7 Let

$$f_n(x) = \frac{\sin(nx)}{\sqrt{n}}, \quad x \in \mathbb{R},\, n \in \mathbb{N}, \tag{6.15}$$

$f(x) = \lim\limits_{n\to\infty} f_n(x) = 0$. Then $f'(x) = 0$, and

$$f_n'(x) = \sqrt{n}\cos(nx),$$

so that (f_n') does not converge to f'. For instance $f_n'(0) = \sqrt{n} \underset{n\to\infty}{\longrightarrow} +\infty$ as $n \to \infty$, whereas $f'(0) = 0$. Note that (f_n) converges uniformly to 0 on $\mathbb{R}$ since $|\sin(nx)/\sqrt{n}\,| \le 1/\sqrt{n}$ becomes small, independently on $x \in \mathbb{R}$.
Consequently, uniform convergence of (f_n) implies nothing about the sequence (f_n'). Thus, stronger hypothesis are required for the assertion that $f_n \to f$ implies $f_n' \to f'$

Theorem 6.10 *Suppose (f_n) is a sequence of continuously differentiable functions on $[a,b]$ pointwise converging to some function f. Suppose further that (f_n') converges uniformly on $[a,b]$.*
Then f_n converges uniformly to f on $[a,b]$, f is continuously differentiable and on $[a,b]$, and

$$f'(x) = \lim_{n\to\infty} f_n'(x), \quad a \le x \le b. \tag{6.16}$$

Proof. Put $g(x) = \lim_{n\to\infty} f_n'(x)$, then g is continuous by Theorem 6.5. By the Fundamental Theorem of Calculus, Theorem 5.14,

$$f_n(x) = f_n(a) + \int_a^x f_n'(t)\,\mathrm{d}t.$$

By assumption on (f_n') and by Corollary 6.8 the sequence

$$\left(\int_a^x f_n'(t)\,\mathrm{d}t\right)_{n\in\mathbb{N}}$$

converges uniformly on $[a,b]$ to $\int_a^x g(t)\,\mathrm{d}t$. Taking the limit $n\to\infty$ in the above equation, we thus obtain

$$f(x) = f(a) + \int_a^x g(t)\,\mathrm{d}t.$$

Since g is continuous, the right hand side defines a differentiable function, namely the antiderivative of $g(x)$, by the FTC. Hence, $f'(x) = g(x)$; since g is continuous the proof is now complete. ∎

For a more general result (without the additional assumption of continuity of f_n') see [Rud76, 7.17 Theorem].

Corollary 6.11 *Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence R.*
(a) *Then f is differentiable on $(-R,R)$ and we have*

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}, \quad x \in (-R,R). \tag{6.17}$$

(b) *The function f is infinitely often differentiable on $(-R,R)$ and we have*

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1) a_n x^{n-k}, \tag{6.18}$$

$$a_n = \frac{1}{n!} f^{(n)}(0), \quad n \in \mathbb{N}_0. \tag{6.19}$$

In particular, f coincides with its Taylor series.

Proof. (a) By Remark 6.2 (a), the power series $\sum_{n=0}^{\infty}(a_n x^n)'$ has the same radius of convergence and converges uniformly on every closed subinterval $[-r,r]$ of $(-R,R)$. By Theorem 6.10, $f(x)$ is differentiable and differentiation and summation can be interchanged.

(b) Iterated application of (a) yields that $f^{(k-1)}$ is differentiable on $(-R, R)$ with (6.18). In particular, inserting $x = 0$ into (6.18) we find

$$f^{(k)}(0) = k!a_k, \quad \Longrightarrow \quad a_k = \frac{f^{(k)}(0)}{k!}.$$

These are exactly the Taylor coefficients of f hat $a = 0$. Hence, f coincides with its Taylor series. ∎

Example 6.8 For $x \in (-1, 1)$ we have

$$\sum_{n=1}^{\infty} nx^n = \frac{x}{(1-x)^2}.$$

Since the geometric series $f(x) = \sum_{n=0}^{\infty} x^n$ equals $1/(1-x)$ on $(-1, 1)$ by Corollary 6.11 we have

$$\frac{1}{(1-x)^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-x}\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{n=0}^{\infty} x^n\right) = \sum_{n=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}x}\left(x^n\right) = \sum_{n=1}^{\infty} nx^{n-1}.$$

Multiplying the preceding equation by x gives the result.

6.3 Fourier Series

In this section we consider basic notions and results of the theory of Fourier series. The question is to write a periodic function as a series of $\cos kx$ and $\sin kx$, $k \in \mathbb{N}$. In contrast to Taylor expansions the periodic function need not to be infinitely often differentiable. Two Fourier series may have the same behavior in one interval, but may behave in different ways in some other interval. We have here a very striking contrast between Fourier series and power series.
In this section a *periodic* function is meant to be a 2π-periodic complex valued function on $\mathbb{R}$, that is $f\colon \mathbb{R} \to \mathbb{C}$ satisfies $f(x + 2\pi) = f(x)$ for all $x \in \mathbb{R}$. Special periodic functions are the trigonometric polynomials.

Definition 6.3 A function $f\colon \mathbb{R} \to \mathbb{R}$ is called *trigonometric polynomial* if there are real numbers a_k, b_k, $k = 0, \ldots, n$ with

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kx + b_k \sin kx. \tag{6.20}$$

The coefficients a_k and b_k are uniquely determined by f since

$$\begin{aligned} a_k &= \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,\mathrm{d}x, \quad k = 0, 1, \ldots, n, \\ b_k &= \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,\mathrm{d}x, \quad k = 1, \ldots, n. \end{aligned} \tag{6.21}$$

This is immediate from

$$\begin{aligned}
&\int_0^{2\pi} \cos kx \, \sin mx \, \mathrm{d}x = 0, \\
&\int_0^{2\pi} \cos kx \, \cos mx \, \mathrm{d}x = \pi\delta_{km}, \quad k, m \in \mathbb{N}, \\
&\int_0^{2\pi} \sin kx \, \sin mx \, \mathrm{d}x = \pi\delta_{km},
\end{aligned} \tag{6.22}$$

where $\delta_{km} = 1$ if $k = m$ and $\delta_{km} = 0$ if $k \neq m$ is the so called *Kronecker symbol*, see Homework 19.2. For example, if $m \geq 1$ we have

$$\begin{aligned}
\frac{1}{\pi}\int_0^{2\pi} f(x)\cos mx \, \mathrm{d}x &= \frac{1}{\pi}\int_0^{2\pi} \left(\frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kx + b_k \sin kx\right) \cos mx \, \mathrm{d}x \\
&= \frac{1}{\pi}\left(\sum_{k=1}^{n} \int_0^{2\pi} \left(a_k \cos kx \cos mx + b_k \sin kx \cos mx\right) \, \mathrm{d}x\right) \\
&= \frac{1}{\pi}\left(\sum_{k=1}^{n} a_k \pi \delta_{km}\right) = a_m.
\end{aligned}$$

Sometimes it is useful to consider complex trigonometric polynomials. Using the formulas expressing $\cos x$ and $\sin x$ in terms of $\mathrm{e}^{\mathrm{i}x}$ and $\mathrm{e}^{-\mathrm{i}x}$ we can write the above polynomial (6.20) as

$$f(x) = \sum_{k=-n}^{n} c_k \mathrm{e}^{\mathrm{i}kx}, \tag{6.23}$$

where $c_0 = a_0/2$ and

$$c_k = \frac{1}{2}\left(a_k - \mathrm{i}b_k\right), \quad c_{-k} = \frac{1}{2}\left(a_k + \mathrm{i}b_k\right), \quad k \geq 1.$$

To obtain the coefficients c_k using integration we need the notion of an integral of a complex-valued function, see Section 5.5. If $m \neq 0$ we have

$$\int_a^b \mathrm{e}^{\mathrm{i}mx} \, \mathrm{d}x = \frac{1}{\mathrm{i}m}\mathrm{e}^{\mathrm{i}mx}\Big|_a^b.$$

If $a = 0$ and $b = 2\pi$ and $m \in \mathbb{Z}$ we obtain

$$\int_0^{2\pi} \mathrm{e}^{\mathrm{i}mx} \, \mathrm{d}x = \begin{cases} 0, & m \in \mathbb{Z} \setminus \{0\}, \\ 2\pi, & m = 0. \end{cases} \tag{6.24}$$

We conclude,

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} f(x)\,\mathrm{e}^{-\mathrm{i}kx} \, \mathrm{d}x, \quad k = 0, \pm 1, \ldots, \pm n.$$

Definition 6.4 Let $f\colon \mathbb{R} \to \mathbb{C}$ be a periodic function with $f \in \mathcal{R}$ on $[0, 2\pi]$. We call

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) \mathrm{e}^{-\mathrm{i}kx} \, \mathrm{d}x, \quad k \in \mathbb{Z} \tag{6.25}$$

the *Fourier coefficients* of f, and the series

$$\sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{\mathrm{i}kx}, \tag{6.26}$$

i. e. the sequence of partial sums

$$s_n = \sum_{k=-n}^{n} c_k \mathrm{e}^{\mathrm{i}kx}, \quad n \in \mathbb{N},$$

the *Fourier series* of f.

The Fourier series can also be written as

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx. \tag{6.27}$$

where a_k and b_k are given by (6.21). One can ask whether the Fourier series of a function converges to the function itself. It is easy to see: If the function f is the *uniform* limit of a series of trigonometric polynomials

$$f(x) = \sum_{k=-\infty}^{\infty} \gamma_k \mathrm{e}^{\mathrm{i}kx} \tag{6.28}$$

then f coincides with its Fourier series. Indeed, since the series (6.28) converges uniformly, by Proposition 6.6 we can change the order of summation and integration and obtain

$$\begin{aligned} c_k &= \frac{1}{2\pi} \int_0^{2\pi} \left(\sum_{m=-\infty}^{\infty} \gamma_m \mathrm{e}^{\mathrm{i}mx} \right) \mathrm{e}^{-\mathrm{i}kx} \, \mathrm{d}x \\ &= \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \int_0^{2\pi} \gamma_m \mathrm{e}^{\mathrm{i}(m-k)x} \, \mathrm{d}x = \gamma_k. \end{aligned}$$

In general, the Fourier series of f neither converges uniformly nor pointwise to f. For Fourier series convergence with respect to the L^2-norm

$$\|f\|_2 = \left(\frac{1}{2\pi} \int_0^{2\pi} |\, f \,|^2 \, \mathrm{d}x \right)^{\frac{1}{2}} \tag{6.29}$$

is the appropriate notion.

6.3.1 An Inner Product on the Periodic Functions

Let V be the linear space of periodic functions $f\colon \mathbb{R} \to \mathbb{C}$, $f \in \mathcal{R}$ on $[0, 2\pi]$. We introduce an inner product on V by

$$f \cdot g = \frac{1}{2\pi} \int_0^{2\pi} f(x)\, \overline{g(x)}\, \mathrm{d}x, \quad f, g \in V.$$

One easily checks the following properties for $f, g, h \in V$, $\lambda, \mu \in \mathbb{C}$.

$$f + g \cdot h = f \cdot h + g \cdot h,$$

$$f \cdot g + h = f \cdot g + f \cdot h,$$

$$\lambda f \cdot \mu g = \lambda\, \overline{\mu}\, f \cdot g,$$

$$f \cdot g = \overline{g \cdot f}.$$

For every $f \in V$ we have $f \cdot f = 1/(2\pi) \int_0^{2\pi} |\, f \,|^2 \, \mathrm{d}x \geq 0$. However, $f \cdot f = 0$ does not imply $f = 0$ (you can change f at finitely many points without any impact on $f \cdot f$). If $f \in V$ is continuous, then $f \cdot f = 0$ implies $f = 0$, see Homework 14.3. Put $\|f\|_2 = \sqrt{f \cdot f}$.
Note that in the physical literature the inner product in $\mathrm{L}^2(X)$ is often *linear* in the second component and antilinear in the first component. Define for $k \in \mathbb{Z}$ the periodic function $\mathrm{e}_k\colon \mathbb{R} \to \mathbb{C}$ by $\mathrm{e}_k(x) = \mathrm{e}^{\mathrm{i}kx}$, the Fourier coefficients of $f \in V$ take the form

$$c_k = f \cdot \mathrm{e}_k, \quad k \in \mathbb{Z}.$$

From (6.24) it follows that the functions e_k, $k \in \mathbb{Z}$, satisfy

$$\mathrm{e}_k \cdot \mathrm{e}_l = \delta_{kl}. \tag{6.30}$$

Any such subset $\{\mathrm{e}_k \mid k \in \mathbb{N}\}$ of an inner product space V satisfying (6.30) is called an *orthonormal system (ONS)*. Using $\mathrm{e}_k(x) = \cos kx + \mathrm{i} \sin kx$ the real orthogonality relations (6.22) immediately follow from (6.30).
The next lemma shows that the Fourier series of f is the best L^2-approximation of a periodic function $f \in V$ by trigonometric polynomials.

Lemma 6.12 (Least Square Approximation) *Suppose $f \in V$ has the Fourier coefficients c_k, $k \in \mathbb{Z}$ and let $\gamma_k \in \mathbb{C}$ be arbitrary. Then*

$$\left\| f - \sum_{k=-n}^{n} c_k \mathrm{e}_k \right\|_2^2 \leq \left\| f - \sum_{k=-n}^{n} \gamma_k \mathrm{e}_k \right\|_2^2, \tag{6.31}$$

and equality holds if and only if $c_k = \gamma_k$ for all k. Further,

$$\left\| f - \sum_{k=-n}^{n} c_k \mathrm{e}_k \right\|_2^2 = \|f\|_2^2 - \sum_{k=-n}^{n} |\, c_k \,|^2 . \tag{6.32}$$

Proof. Let $\sum$ always denote $\sum_{k=-n}^{n}$. Put $g_n = \sum \gamma_k \mathrm{e}_k$. Then

$$f \cdot g_n = f \cdot \sum \gamma_k \, \mathrm{e}_k = \sum \overline{\gamma_k} f \cdot \mathrm{e}_k = \sum c_k \overline{\gamma_k}$$

and $g_n \cdot \mathrm{e}_k = \gamma_k$ such that

$$g_n \cdot g_n = \sum |\, \gamma_k \,|^2 .$$

Noting that $|\, a - b \,|^2 = (a-b)(\overline{a} - \overline{b}) = |\, a \,|^2 + |\, b \,|^2 - \overline{a}b - a\overline{b}$, it follows that

$$\begin{aligned} \|f - g_n\|_2^2 = f - g_n \cdot f - g_n &= f \cdot f - f \cdot g_n - g_n \cdot f + g_n \cdot g_n \\ &= \|f\|_2^2 - \sum \overline{c_k} \gamma_k - \sum c_k \overline{\gamma}_k + \sum |\, \gamma_k \,|^2 \\ &= \|f\|_2^2 - \sum |\, c_k \,|^2 + \sum |\, \gamma_k - c_k \,|^2 \end{aligned} \tag{6.33}$$

which is evidently minimized if and only if $\gamma_k = c_k$. Inserting this into (6.33), equation (6.32) follows. ∎

Corollary 6.13 (Bessel's Inequality) *Under the assumptions of the above lemma we have*

$$\sum_{k=-\infty}^{\infty} |\, c_k \,|^2 \leq \|f\|_2^2 . \tag{6.34}$$

Proof. By equation (6.32), for every $n \in \mathbb{N}$ we have

$$\sum_{k=-n}^{n} |\, c_k \,|^2 \leq \|f\|_2^2 .$$

Taking the limit $n \to \infty$ or $\sup_{n \in \mathbb{N}}$ shows the assertion. ∎

An ONS $\{e_k \mid k \in \mathbb{Z}\}$ is said to be *complete* if instead of Bessel's inequality, equality holds for all $f \in V$.

Definition 6.5 Let $f_n, f \in V$, we say that (f_n) *converges to* f *in* L^2 (denoted by $f_n \underset{\|\cdot\|_2}{\longrightarrow} f$) if

$$\lim_{n \to \infty} \|f_n - f\|_2 = 0.$$

Explicitly

$$\int_0^{2\pi} |\, f_n(x) - f(x) \,|^2 \, \mathrm{d}x \underset{n \to \infty}{\longrightarrow} 0.$$

Remarks 6.3 (a) Note that the L^2-limit in V is not unique; changing $f(x)$ at finitely many points of $[0, 2\pi]$ does not change the integral $\int_0^{2\pi} |\, f - f_n \,|^2 \, \mathrm{d}x$.
(b) If $f_n \rightrightarrows f$ on $\mathbb{R}$ then $f_n \underset{\|\cdot\|_2}{\longrightarrow} f$. Indeed, let $\varepsilon > 0$. Then there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies $\sup_{x\in\mathbb{R}} |\, f_n(x) - f(x) \,| \leq \varepsilon$. Hence

$$\int_0^{2\pi} |\, f_n - f \,|^2 \, \mathrm{d}x \leq \int_0^{2\pi} \varepsilon^2 \, \mathrm{d}x = 2\pi\varepsilon^2.$$

This shows $f_n - f \underset{\|\cdot\|_2}{\longrightarrow} 0$.
(c) The above Lemma, in particular (6.32), shows that the Fourier series converges in L^2 to f if and only if

$$\|f\|_2^2 = \sum_{k=-\infty}^{\infty} |\, c_k \,|^2 \,. \tag{6.35}$$

This is called *Parseval's Completeness Relation*. We will see that it holds for all $f \in V$.

Let use write

$$f(x) \sim \sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{\mathrm{i}kx}$$

to express the fact that (c_k) are the (complex) Fourier coefficients of f. Further

$$s_n(f) = s_n(f; x) = \sum_{k=-n}^{n} c_k \mathrm{e}^{\mathrm{i}kx} \tag{6.36}$$

denotes the nth partial sum.

Theorem 6.14 (Parseval's Completeness Theorem) *The ONS $\{\mathrm{e}_k \mid k \in \mathbb{Z}\}$ is complete. More precisely, if $f, g \in V$ with*

$$f \sim \sum_{k=-\infty}^{\infty} c_k \mathrm{e}_k, \quad g \sim \sum_{k=-\infty}^{\infty} \gamma_k \mathrm{e}_k,$$

then

$$\text{(i)} \quad \lim_{n\to\infty} \frac{1}{2\pi} \int_0^{2\pi} |\, f - s_n(f) \,|^2 \, \mathrm{d}x = 0, \tag{6.37}$$

$$\text{(ii)} \quad \frac{1}{2\pi} \int_0^{2\pi} f \, \overline{g} \, \mathrm{d}x = \sum_{k=-\infty}^{\infty} c_k \, \overline{\gamma_k}, \tag{6.38}$$

$$\text{(iii)} \quad \frac{1}{2\pi} \int_0^{2\pi} |\, f \,|^2 \, \mathrm{d}x = \sum_{k=-\infty}^{\infty} |\, c_k \,|^2 = \frac{a_0^2}{4} + \frac{1}{2} \sum_{k=1}^{\infty} (a_k^2 + b_k^2) \quad \textit{Parseval's formula}. \tag{6.39}$$

The proof is in Rudin's book, [Rud76, 8.16, p.191]. It uses Stone–Weierstraß theorem about the uniform approximation of a continuoous function by polynomials. An elementary proof is in Forster's book [For01, §23].

Example 6.9 (a) Consider the periodic function $f \in V$ given by

$$f(x) = \begin{cases} 1, & 0 \le x < \pi \\ -1, & \pi \le x < 2\pi. \end{cases}$$

Since f is an odd function the coefficients a_k vanish. We compute the Fourier coefficients b_k.

$$b_k = \frac{2}{\pi}\int_0^\pi \sin kx \,\mathrm{d}x = -\frac{2}{k\pi} - \cos kx\Big|_0^\pi = \frac{2}{k\pi}\left((-1)^{k+1}+1\right) = \begin{cases} 0, & \text{if } k \text{ is even,} \\ \frac{4}{k\pi}, & \text{if } k \text{ is odd..} \end{cases}$$

The Fourier series of f reads

$$f \sim \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{\sin(2n+1)x}{2n+1}.$$

Noting that

$$\sum_{k\in\mathbb{Z}} |\, c_k \,|^2 = \frac{a_0^2}{4} + \frac{1}{2}\sum_{n\in\mathbb{N}}(a_n^2 + b_n^2)$$

Parseval's formula gives

$$\|f\|_2^2 = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}x = 1 = \frac{1}{2}\sum_{n\in\mathbb{N}} b_n^2 = \frac{8}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2} =: \frac{8}{\pi^2}s_1 \Longrightarrow s_1 = \frac{\pi^2}{8}.$$

Now we can compute $s = \sum_{n=1}^{\infty}\frac{1}{n^2}$. Since this series converges absolutely we are allowed to rearrange the elements in such a way that we first add all the odd terms, which gives s_1 and then all the even terms which gives s_0. Using $s_1 = \pi^2/8$ we find

$$\begin{aligned} s &= s_1 + s_0 = s_1 + \frac{1}{2^2} + \frac{1}{4^2} + \frac{1}{6^2} + \cdots \\ s &= s_0 + \frac{1}{2^2}\left(\frac{1}{1^2} + \frac{1}{2^2} + \cdots\right) = s_1 + \frac{s}{4} \\ s &= \frac{4}{3}s_1 = \frac{\pi^2}{6}. \end{aligned}$$

(b) Fix $a \in [0, 2\pi]$ and consider $f \in V$ with

$$f(x) = \begin{cases} 1, & 0 \le x \le a, \\ 0, & a < x < 2\pi \end{cases}$$

The Fourier coefficients of f are $c_0 = \frac{1}{2\pi}\int_0^a \mathrm{d}x = \frac{a}{2\pi}$ and

$$c_k = f\cdot \mathrm{e}_k = \frac{1}{2\pi}\int_0^a \mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x = \frac{\mathrm{i}}{2\pi k}\left(\mathrm{e}^{-\mathrm{i}ka} - 1\right), \quad k \neq 0.$$

If $k \neq 0$,

$$|\, c_k \,|^2 = \frac{1}{4\pi^2k^2}\left(1 - \mathrm{e}^{\mathrm{i}ka}\right)\left(1 - \mathrm{e}^{-\mathrm{i}ka}\right) = \frac{1-\cos ka}{2\pi^2k^2},$$

hence Parseval's formula gives

$$\begin{aligned}\sum_{k=-\infty}^{\infty} |\, c_k \,|^2 &= \frac{a^2}{4\pi^2} + \sum_{k=1}^{\infty} \frac{1-\cos ak}{\pi^2 k^2} \\ &= \frac{a^2}{4\pi^2} + \frac{1}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2} - \frac{1}{\pi^2}\sum_{k=1}^{\infty}\frac{\cos ak}{k^2} \\ &= \frac{a^2}{4\pi^2} + \frac{1}{\pi^2}\left(s - \sum_{k=1}^{\infty}\frac{\cos ak}{k^2}\right),\end{aligned}$$

where $s = \sum 1/k^2$. On the other hand

$$\|f\|_2^2 = \frac{1}{2\pi}\int_0^a \mathrm{d}x = \frac{a}{2\pi}.$$

Hence, (6.37) reads

$$\begin{aligned}\frac{a^2}{4\pi^2} + \frac{1}{\pi^2}\left(s - \sum_{k=1}^{\infty}\frac{\cos ka}{k^2}\right) &= \frac{a}{2\pi} \\ \sum_{k=1}^{\infty}\frac{\cos ka}{k^2} = \frac{a^2}{4} - \frac{a\pi}{2} + \frac{\pi^2}{6} &= \frac{(a-\pi)^2}{4} - \frac{\pi^2}{12}. \qquad (6.40)\end{aligned}$$

Since the series

$$\sum_{k=1}^{\infty}\frac{\cos kx}{k^2} \qquad (6.41)$$

converges uniformly on $\mathbb{R}$ (use Theorem 6.2 and $\sum_k 1/k^2$ is an upper bound) (6.41) is the Fourier series of the function

$$\frac{(x-\pi)^2}{4} - \frac{\pi^2}{12}, \quad x \in [0, 2\pi]$$

and the Fourier series converges uniformly on $\mathbb{R}$ to the above function. Since the term by term differentiated series converges uniformly on $[\delta, 2\pi - \delta]$, see Example 6.2, we obtain

$$-\sum_{k=1}^{\infty}\frac{\sin kx}{k} = \sum_{k=1}^{\infty}\left(\frac{\cos kx}{k^2}\right)' = \left(\frac{(x-\pi)^2}{4} - \frac{\pi^2}{12}\right)' = \frac{x-\pi}{2}$$

which is true for $x \in (0, 2\pi)$.
We can also integrate the Fourier series and obtain by Corollary 6.7

$$\begin{aligned}\int_0^x \sum_{k=1}^{\infty}\frac{\cos kt}{k^2}\,\mathrm{d}t &= \sum_{k=1}^{\infty}\frac{1}{k^2}\int_0^x \cos kt\,\mathrm{d}t = \sum_{k=1}^{\infty}\frac{1}{k^3}\sin kt\Big|_0^x \\ &= \sum_{k=1}^{\infty}\frac{\sin kx}{k^3}.\end{aligned}$$

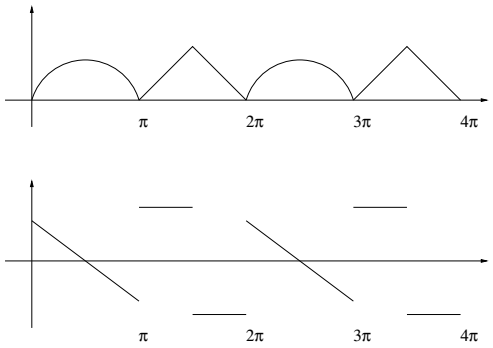
On the other hand,

$$\int_0^x \left(\frac{(t-\pi)^2}{4} - \frac{\pi^2}{12} \right) \mathrm{d}t = \frac{(x-\pi)^3}{12} - \frac{\pi^2}{12}x + \frac{\pi^3}{12}.$$

By homework 19.5

$$f(x) = \sum_{k=1}^{\infty} \frac{\sin kx}{k^3} = \frac{(x-\pi)^3}{12} - \frac{\pi^2}{12}x + \frac{\pi^3}{12}$$

defines a continuously differentiable periodic function on $\mathbb{R}$.



Theorem 6.15 *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous periodic function which is piecewise continuously differentiable, i. e. there exists a partition $\{t_0, \ldots, t_r\}$ of $[0, 2\pi]$ such that $f|[t_{i-1}, t_i]$ is continuously differentiable.*
Then the Fourier series of f converges uniformly to f.

Proof. Let $\varphi_i\colon [t_{i-1}, t_i] \to \mathbb{R}$ denote the continuous derivative of $f|[t_{i-1}, t_i]$ and $\varphi\colon \mathbb{R} \to \mathbb{R}$ the periodic function that coincides with φ_i on $[t_{i-1}, t_i]$. By Bessel's inequality, the Fourier coefficients γ_k of φ satisfy

$$\sum_{k=-\infty}^{\infty} |\,\gamma_k\,|^2 \leq \|\varphi\|_2^2 < \infty.$$

If $k \neq 0$ the Fourier coefficients c_k of f can be found using integration by parts from the Fourier coefficients of γ_k.

$$\int_{t_{i-1}}^{t_i} f(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x = \frac{\mathrm{i}}{k}\left(f(x)\mathrm{e}^{-\mathrm{i}kx}\Big|_{t_{i-1}}^{t_i} - \int_{t_{i-1}}^{t_i} \varphi(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x \right).$$

Hence summation over $i = 1, \ldots, r$ yields,

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} f(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x = \frac{1}{2\pi}\sum_{i=1}^{r}\int_{t_{i-1}}^{t_i} f(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x$$
$$c_k = \frac{-\mathrm{i}}{2\pi k}\int_0^{2\pi} \varphi(x)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}x = \frac{-\mathrm{i}\gamma_k}{k}.$$

Note that the term

$$\sum_{i=1}^{r} f(x)\mathrm{e}^{-\mathrm{i}kx}\Big|_{t_{i-1}}^{t_i}$$

vanishes since f is continuous and $f(2\pi) = f(0)$. Since for $\alpha, \beta \in \mathbb{C}$ we have $|\,\alpha\beta\,| \leq \frac{1}{2}(|\,\alpha\,|^2 + |\,\beta\,|^2)$, we obtain

$$|\,c_k\,| \leq \frac{1}{2}\left(\frac{1}{|\,k\,|^2} + |\,\gamma_k\,|^2 \right).$$

Since both $\sum_{k=1}^{\infty} \frac{1}{k^2}$ and $\sum_{k=-\infty}^{\infty} |\gamma_k|^2$ converge,

$$\sum_{k=-\infty}^{\infty} |c_k| < \infty.$$

Thus, the Fourier series converges uniformly to a continuous function g (see Theorem 6.5). Since the Fourier series converges both to f and to g in the L^2 norm, $\|f-g\|_2 = 0$. Since both f and g are continuous, they coincide. This completes the proof. ∎

Note that for any $f \in V$, the series $\sum_{k\in\mathbb{Z}} |c_k|^2$ converges while the series $\sum_{k\in\mathbb{Z}} |c_k|$ converges only if the Fourier series converges uniformly to f.

6.4 Basic Topology

In the study of functions of several variables we need some topological notions like neighborhood, open set, closed set, and compactness.

6.4.1 Finite, Countable, and Uncountable Sets

Definition 6.6 If there exists a 1-1 mapping of the set A onto the the B (a bijection), we say that A and B have the same *cardinality* or that A and B are *equivalent*; we write $A \sim B$.

Definition 6.7 For any nonnegative integer $n \in \mathbb{N}_0$ let N_n be the set $\{1, 2, \ldots, n\}$. For any set A we say that:

(a) A is *finite* if $A \sim N_n$ for some n. The empty set $\varnothing$ is also considered to be finite.
(b) A is *infinite* if A is not finite.
(c) A is *countable* if $A \sim \mathbb{N}$.
(d) A is *uncountable* if A is neither finite nor countable.
(e) A is *at most countable* if A is finite or countable.

For finite sets A and B we evidently have $A \sim B$ if A and B have the same number of elements. For infinite sets, however, the idea of "having the same number of elements" becomes quite vague, whereas the notion of 1-1 correspondence retains its clarity.

Example 6.10 (a) $\mathbb{Z}$ is countable. Indeed, the arrangement

$$0, 1, -1, 2, -2, 3, -3, \ldots$$

gives a bijection between $\mathbb{N}$ and $\mathbb{Z}$. An infinite set $\mathbb{Z}$ can be equivalent to one of its proper subsets N.

(b) Countable sets represent the "smallest" infinite cardinality: No uncountable set can be a subset of a countable set. Any countable set can be arranged in a sequence. In particular, $\mathbb{Q}$ is contable, see Example 2.6 (c).
(c) The countable union of countable sets is a countable set; this is Cantor's First Diagonal Process:

$$\begin{array}{ccccccccc}
x_{11} & \to & x_{12} & & x_{13} & \to & x_{14} & & \dots \\
 & \swarrow & & \nearrow & & \swarrow & & \nearrow & \\
x_{21} & & x_{22} & & x_{23} & & x_{24} & & \dots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & & \\
x_{31} & & x_{32} & & x_{33} & & x_{34} & & \dots \\
 & \swarrow & & \nearrow & & \swarrow & & & \\
x_{41} & & x_{42} & & x_{43} & & x_{44} & & \dots \\
\downarrow & \nearrow & & & & & & & \\
x_{51} & & \dots & & & & & &
\end{array}$$

(d) Let $A = \{(x_n) \mid x_n \in \{0,1\} \ \forall\, n \in \mathbb{N}\}$ be the set of all sequences whose elements are 0 and 1. This set A is uncountable. In particular, $\mathbb{R}$ is uncountable.
Proof. Suppose to the contrary that A is countable and arrange the elements of A in a sequence $(s_n)_{n\in\mathbb{N}}$ of distinct elements of A. We construct a sequence s as follows. If the nth element in s_n is 1 we let the nth digit of s be 0, and vice versa. Then the sequence s differs from every member $s_1, s_2, \dots$ at least in one place; hence $s \notin A$—a contradiction since s is indeed an element of A. This proves, A is uncountable. ∎

6.4.2 Metric Spaces and Normed Spaces

Definition 6.8 A set X is said to be a *metric space* if for any two points $x, y \in X$ there is associated a real number $d(x, y)$, called the *distance* of x and y such that

(a) $d(x, x) = 0$ and $d(x, y) > 0$ for all $x, y \in X$ with $x \neq y$ (positive definiteness);
(b) $d(x, y) = d(y, x)$ (symmetry);
(c) $d(x, y) \leq d(x, z) + d(z, y)$ for any $z \in X$ (triangle inequality).

Any function $d\colon X \times X \to \mathbb{R}$ with these three properties is called a *distance function* or *metric* on X.

Example 6.11 (a) $\mathbb{C}$, $\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{Z}$ are metric spaces with $d(x, y) := |\, y - x \,|$.
Any subsets of a metric space is again a metric space.
(b) The *real plane* $\mathbb{R}^2$ is a metric space with respect to

$$d_2((x_1, x_2), (y_1, y_2)) := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},$$
$$d_1((x_1, x_2), (y_1, y_2)) := |\, x_1 - x_2 \,| + |\, y_1 - y_2 \,|\,.$$

d_2 is called the *euclidean metric*.

(c) Let X be a set. Define

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

Then (X,d) becomes a metric space. It is called the *discrete metric space*.

Definition 6.9 Let E be a vector space over $\mathbb{C}$ (or $\mathbb{R}$). Suppose on E there is given a function $\|\cdot\| : E \to \mathbb{R}$ which associates to each $x \in E$ a real number $\|x\|$ such that the following three conditions are satisfied:

(i) $\|x\| \geq 0$ for every $x \in E$, and $\|x\| = 0$ if and only if $x = 0$,
(ii) $\|\lambda x\| = |\,\lambda\,| \, \|x\|$ for all $\lambda \in \mathbb{C}$ (in $\mathbb{R}$, resp.)
(iii) $\|x+y\| \leq \|x\| + \|y\|\,, \quad$ for all $\quad x, y \in E$.

Then E is called a *normed (vector) space* and $\|x\|$ is the *norm* of x.

$\|x\|$ generalizes the "length" of vector $x \in E$. Every normed vector space E is a metric space if we put $d(x,y) = \|x-y\|$. **Prove!**
However, there are metric spaces that are not normed spaces, for example $(\mathbb{N}, d(m,n) = |\,n-m\,|)$.

Example 6.12 (a) $E = \mathbb{R}^k$ or $E = \mathbb{C}^k$. Let $x = (x_1, \cdots, x_k) \in E$ and define

$$\|x\|_2 = \sqrt{\sum_{i=1}^{k} |\,x_i\,|^2}.$$

Then $\|\cdot\|$ is a norm on E. It is called the *Euclidean norm*.
There are other possibilities to define a norm on E. For example,

$$\begin{aligned} \|x\|_\infty &= \max_{1 \leq i \leq k} |\,x_i\,|\,, \\ \|x\|_1 &= \sum_{i=1}^{k} |\,x_k\,|\,, \\ \|x\|_{\mathrm{a}} &= \|x\|_2 + 3\,\|x\|_1\,, \quad \|x\|_{\mathrm{b}} = \max(\|x\|_1\,, \|x\|_2). \end{aligned}$$

(b) $E = \mathrm{C}([a,b])$. Let $p \geq 1$. Then

$$\begin{aligned} \|f\|_\infty &= \sup_{x \in [a,b]} |\,f(x)\,|\,, \\ \|f\|_p &= \left(\int_a^b |\,f(t)\,|^p \,\mathrm{d}t \right)^{\frac{1}{p}}. \end{aligned}$$

define norms on E. Note that $\|f\|_p \leq \sqrt[p]{b-a}\;\|f\|_\infty$.
(c) Hilbert's sequence space. $E = \ell_2 = \{(x_n) \mid \sum_{n=1}^{\infty} |\,x_n\,|^2 < \infty\}$. Then

$$\|x\|_2 = \left(\sum_{n=1}^{\infty} |\,x_n\,|^2 \right)^{\frac{1}{2}}$$

defines a norm on ℓ_2.
(d) The bounded sequences. $E = \ell_\infty = \{(x_n) \mid \sup_{n\in\mathbb{N}} |\, x_n \,| < \infty\}$. Then

$$\|x\|_\infty = \sup_{n\in\mathbb{N}} |\, x_n \,|$$

defines a norm on E.
(e) $E = \mathrm{C}([a,b])$. Then

$$\|f\|_1 = \int_a^b |\, f(t) \,| \, \mathrm{d}t$$

defines a norm on E.

6.4.3 Open and Closed Sets

Definition 6.10 Let X be a metric space with metric d. All points and subsets mentioned below are understood to be elements and subsets of X, in particular, let $E \subset X$ be a subset of X.

(a) The set $U_\varepsilon(x) = \{y \mid d(x,y) < \varepsilon\}$ with some $\varepsilon > 0$ is called the *ε-neighborhood* (or ε-ball with center x) of x. The number ε is called the radius of the neighborhood $U_\varepsilon(x)$.

(b) A point p is an *interior* or *inner* point of E if there is a neighborhood $U_\varepsilon(p)$ completely contained in E. E is *open* if every point of E is an interior point.

(c) A point p is called an *accumulation* or *limit* point of E if every neighborhood of p has a point $q \neq p$ such that $q \in E$.

(d) E is said to be *closed* if every accumulation point of E is a point of E. The *closure* of E (denoted by $\overline{E}$) is E together with all accumulation points of E. In other words $p \in \overline{E}$, if and only if every neighborhood of x has a non-empty intersection with E.

(e) The *complement* of E (denoted by E^{c}) is the set of all points $p \in X$ such that $p \notin E$.

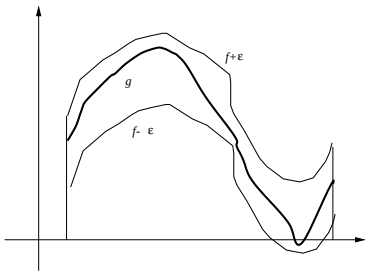
(f) E is *bounded* if there exists a real number $C > 0$ such that $d(x,y) \leq C$ for all $x, y \in E$.

(g) E is *dense* in X if $\overline{E} = X$.

Example 6.13 (a) $X = \mathbb{R}$ with the standard metric $d(x,y) = |\, x - y \,|$. $E = (a,b) \subset \mathbb{R}$ is an open set. Indeed, for every $x \in (a,b)$ we have $U_\varepsilon(x) \subset (a,b)$ if ε is small enough, say $\varepsilon \leq \min\{|\, x-a \,|, |\, x-b \,|\}$. Hence, x is an inner point of (a,b). Since x was arbitrary, (a,b) is open.
$F = [a,b)$ is not open since a is not an inner point of $[a,b)$. Indeed, $U_\varepsilon(a) \subsetneq [a,b)$ for every $\varepsilon > 0$.
We have

$$\overline{E} = \overline{F} = \text{set of accumulation points} = [a,b].$$

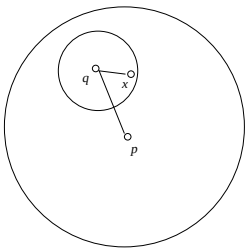
Indeed, a is an accumulation point of both (a, b) and $[a, b)$. This is true since every neighborhood $U_\varepsilon(a)$, $\varepsilon < b - a$, has $a + \varepsilon/2 \in (a, b)$ (resp. in $[a, b)$) which is different from a. For any point $x \notin [a, b]$ we find a neighborhood $U_\varepsilon(x)$ with $U_\varepsilon(x) \cap [a, b) = \varnothing$; hence $x \notin \overline{E}$.
The set of rational numbers $\mathbb{Q}$ is dense in $\mathbb{R}$. Indeed, every neighborhood $U_\varepsilon(r)$ of every real number r contains a rational number, see Proposition 1.11 (b).
For the real line one can prove: Every open set is the at most countable union of disjoint open (finite or infinite) intervals. A similar description for closed subsets of $\mathbb{R}$ is false. There is no similar description of open subsets of $\mathbb{R}^k$, $k \geq 2$.
(b) For every metric space X, both the whole space X and the empty set $\varnothing$ are open as well as closed.
(c) Let $B = \{x \in \mathbb{R}^k \mid \|x\|_2 < 1\}$ be the *open unit ball* in $\mathbb{R}^k$. B is open (see Lemma 6.16 below); B is not closed. For example, $x_0 = (1, 0, \ldots, 0)$ is an accumulation point of B since $x_n = (1 - 1/n, 0, \ldots, 0)$ is a sequence of elements of B converging to x_0, however, $x_0 \notin B$. The accumulation points of B are $\overline{B} = \{x \in \mathbb{R}^k \mid \|x\|_2 \leq 1\}$. This is also the closure of B in $\mathbb{R}^k$.



(d) Consider $E = \mathrm{C}([a, b])$ with the supremum norm. Then $g \in E$ is in the ε-neighborhood of a function $f \in E$ if and only if

$$|\, f(t) - g(t) \,| < \varepsilon, \quad \text{for all} \quad x \in [a, b].$$

Lemma 6.16 *Every neighborhood $U_r(p)$, $r > 0$, of a point p is an open set.*



Proof. Let $q \in U_r(p)$. Then there exists $\varepsilon > 0$ such that $d(q, p) = r - \varepsilon$. We will show that $U_\varepsilon(q) \subset U_r(p)$. For, let $x \in U_\varepsilon(q)$. Then by the triangle inequality we have

$$d(x, p) \leq d(x, q) + d(q, p) < \varepsilon + (r - \varepsilon) = r.$$

Hence $x \in U_r(p)$ and q is an interior point of $U_r(p)$. Since q was arbitrary, $U_r(p)$ is open. ∎

Remarks 6.4 (a) If p is an accumulation point of a set E, then every neighborhood of p contains infinitely many points of E.
(b) A finite set has no accumulation points; hence any finite set is closed.

Example 6.14 (a) The open complex unit disc, $\{z \in \mathbb{C} \mid |\, z \,| < 1\}$.
(b) The closed unit disc, $\{z \in \mathbb{C} \mid |\, z \,| \leq 1\}$.
(c) A finite set.
(d) The set $\mathbb{Z}$ of all integers.
(e) $\{1/n \mid n \in \mathbb{N}\}$.
(f) The set $\mathbb{C}$ of all complex numbers.
(g) The interval (a, b).

Here (d), (e), and (g) are regarded as subsets of $\mathbb{R}$. Some properties of these sets are tabulated below:

	Closed	Open	Bounded
(a)	No	Yes	Yes
(b)	Yes	No	Yes
(c)	Yes	No	Yes
(d)	Yes	No	No
(e)	No	No	Yes
(f)	Yes	Yes	No
(g)	No	Yes	Yes

Proposition 6.17 *A subset $E \subset X$ of a metric space X is open if and only if its complement E^c is closed.*

Proof. First, suppose E^c is closed. Choose $x \in E$. Then $x \notin E^c$, and x is not an accumulation point of E^c. Hence there exists a neighborhood U of x such that $U \cap E^c$ is empty, that is $U \subset E$. Thus x is an interior point of E and E is open.
Next, suppose that E is open. Let x be an accumulation point of E^c. Then every neighborhood of x contains a point of E^c, so that x is not an interior point of E. Since E is open, this means that $x \in E^c$. It follows that E^c is closed. ∎

6.4.4 Limits and Continuity

In this section we generalize the notions of convergent sequences and continuous functions to arbitrary metric spaces.

Definition 6.11 Let X be a metric space and (x_n) a sequence of elements of X. We say that (x_n) *converges to* $x \in X$ if $\lim_{n\to\infty} d(x_n, x) = 0$. We write $\lim_{n\to\infty} x_n = x$ or $x_n \underset{n\to\infty}{\longrightarrow} x$.

In other words, $\lim_{n\to\infty} x_n = x$ if for every neighborhood U_ε, $\varepsilon > 0$, of x there exists an $n_0 \in \mathbb{N}$ such that $n \geq n_0$ implies $x_n \in U_\varepsilon$.
Note that a subset F of a metric space X is closed if and only if F contains all limits of convergent sequences (x_n), $x_n \in F$.
Two metrics d_1 and d_2 on a space X are said to be *topologically equivalent* if $\lim_{n\to\infty} x_n = x$ w.r.t. d_1 if and only if $\lim_{n\to\infty} x_n = x$ w.r.t. d_2. In particular, two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on the same linear space E are said to be *equivalent* if the metric spaces are topologically equivalent.

Proposition 6.18 *Let $E_1 = (E, \|\cdot\|_1)$ and $E_2 = (E, \|\cdot\|_2)$ be normed vector spaces such that there exist positive numbers $c_1, c_2 > 0$ with*

$$c_1 \|x\|_1 \leq \|x\|_2 \leq c_2 \|x\|_1, \quad \textit{for all} \quad x \in E. \tag{6.42}$$

Then $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent.

Proof. Condition (6.42) is obviously symmetric with respect to E_1 and E_2 since $\|x\|_2 / c_2 \le \|x\|_1 \le \|x\|_2 / c_1$. Therefore, it is sufficient to show the following: If $x_n \underset{n\to\infty}{\longrightarrow} x$ w.r.t. $\|\cdot\|_1$ then $x_n \underset{n\to\infty}{\longrightarrow} x$ w.r.t. $\|\cdot\|_2$. Indeed, by Definition, $\lim_{n\to\infty} \|x_n - x\|_1 = 0$. By assumption,

$$c_1 \|x_n - x\|_1 \le \|x_n - x\|_2 \le c_2 \|x_n - x\|_1, \quad n \in \mathbb{N}.$$

Since the first and the last expressions tend to 0 as $n \to \infty$, the sandwich theorem shows that $\lim_{n\to\infty} \|x_n - x\|_2 = 0$, too. This proves $x_n \to x$ w.r.t. $\|\cdot\|_2$. ∎

Example 6.15 Let $E = \mathbb{R}^k$ or $E = \mathbb{C}^k$ with the norm $\|x\|_p = \sqrt[p]{|x_1|^p + \cdots |x_k|^p}$, $p \in [1, \infty]$. All these norms are equivalent. Indeed,

$$\begin{aligned} \|x\|_\infty^p \le \sum_{i=1}^k |x_i|^p \le \sum_{i=1}^k \|x\|_\infty^p = k \|x\|_\infty^p, \\ \implies \|x\|_\infty \le \|x\|_p \le \sqrt[p]{k}\, \|x\|_\infty. \end{aligned} \tag{6.43}$$

The following Proposition is quite analogous to Proposition 2.33 with $k = 2$. Recall that a complex sequence (z_n) converges if and only if both $\operatorname{Re} z_n$ and $\operatorname{Im} z_n$ converge.

Proposition 6.19 *Let (x_n) be a sequence of vectors of the euclidean space $(\mathbb{R}^k, \|\cdot\|_2)$,*

$$x_n = (x_{n1}, \ldots, x_{nk}).$$

Then (x_n) converges to $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$ if and only if

$$\lim_{n\to\infty} x_{ni} = a_i, \quad i = 1, \ldots, k.$$

Proof. Suppose that $\lim_{n\to\infty} x_n = a$. Given $\varepsilon > 0$ there is an $n_0 \in \mathbb{N}$ such that $n \ge n_0$ implies $\|x_n - a\|_2 < \varepsilon$. Thus, for $i = 1, \ldots, k$ we have

$$|x_{ni} - a_i| \le \|x_n - a\|_2 < \varepsilon;$$

hence $\lim_{n\to\infty} x_{ni} = a_i$.
Conversely, suppose that $\lim_{n\to\infty} x_{ni} = a_i$ for $i = 1, \ldots, k$. Given $\varepsilon > 0$ there are $n_{0i} \in \mathbb{N}$ such that $n \ge n_{0i}$ implies

$$|x_{ni} - a_i| < \frac{\varepsilon}{\sqrt{k}}.$$

For $n \ge \max\{n_{01}, \ldots, n_{0k}\}$ we have (see (6.43))

$$\|x_n - a\|_2 \le \sqrt{k}\, \|x_n - a\|_\infty < \varepsilon.$$

hence $\lim_{n\to\infty} x_n = a$. ∎

Corollary 6.20 *Let $B \subset \mathbb{R}^k$ be a bounded subset and (x_n) a sequence of elements of B. Then (x_n) has a converging subsequence.*

Proof. Since B is bounded all coordinates of B are bounded; hence there is a subsequence $(x_n^{(1)})$ of (x_n) such that the first coordinate converges. Further, there is a subsequence $(x_n^{(2)})$ of $(x_n^{(1)})$ such that the second coordinate converges. Finally there is a subsequence $(x_n^{(k)})$ of $(x_n^{(k-1)})$ such that all coordinates converge. By the above proposition the subsequence $(x_n^{(k)})$ converges in $\mathbb{R}^k$. ∎

The same statement is true for subsets $B \subset \mathbb{C}^k$.

Definition 6.12 A mapping $f\colon X \to Y$ from the metric space X into the metric space Y is said to be *continuous* at $a \in X$ if one of the following equivalent conditons is satisfied.
(a) For every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $x \in X$

$$d(x,a) < \delta \quad \text{implies} \quad d(f(x), f(a)) < \varepsilon. \tag{6.44}$$

(b) For any sequence (x_n), $x_n \in X$ with $\lim\limits_{n\to\infty} x_n = a$ it follows that $\lim\limits_{n\to\infty} f(x_n) = f(a)$.
The mapping f is said to be *continuous* on X if f is continuous at every point a of X.

Proposition 6.21 *The composition of two continuous mappings is continuous.*

The proof is completely the same as in the real case (see Proposition 3.4) and we omit it.
We give the topological description of continuous functions.

Proposition 6.22 *Let X and Y be metric spaces. A mapping $f\colon X \to Y$ is continuous if and only if the preimage of any open set in Y is open in X.*

Proof. Suppose that f is continuous and $G \subset Y$ is open. If $f^{-1}(G) = \varnothing$, there is nothing to prove; the empty set is open. Otherwise there exists $x_0 \in f^{-1}(G)$, and therefore $f(x_0) \in G$. Since G is open, there is $\varepsilon > 0$ such that $U_\varepsilon(f(x_0)) \subset G$. Since f is continuous at x_0, there exists $\delta > 0$ such that $x \in U_\delta(x_0)$ implies $f(x) \in U_\varepsilon(f(x_0)) \subset G$. That is, $U_\delta(x_0) \subset f^{-1}(G)$, and x_0 is an inner point of $f^{-1}(G)$; hence $f^{-1}(G)$ is open.
Suppose now that the condition of the proposition is fulfilled. We will show that f is continuous. Fix $x_0 \in X$ and $\varepsilon > 0$. Since $G = U_\varepsilon(f(x_0))$ is open by Lemma 6.16, $f^{-1}(G)$ is open by assumption. In particular, $x_0 \in f^{-1}(G)$ is an inner point. Hence, there exists $\delta > 0$ such that $U_\delta(x_0) \subset f^{-1}(G)$. It follows that $f(U_\delta(x_0)) \subset U_\varepsilon(x_0)$; this means that f is continuous at x_0. Since x_0 was arbitrary, f is continuous on X. ∎

Remark 6.5 Since the complement of an open set is a closed set, it is obvious that the proposition holds if we replace "open set" by "closed set."
In general, the image of an open set under a continuous function need not to be open; consider for example $f(x) = \sin x$ and $G = (0, 2\pi)$ which is open; however, $f((0,2\pi)) = [-1,1]$ is not open.

6.4.5 Comleteness and Compactness

(a) Completeness

Definition 6.13 Let (X, d) be a metric space. A sequence (x_n) of elements of X is said to be a *Cauchy sequence* if for every $\varepsilon > 0$ there exists a positive integer $n_0 \in \mathbb{N}$ such that

$$d(x_n, x_m) < \varepsilon \quad \text{for all} \quad m, n \geq n_0.$$

A metric space is said to be *complete* if every Cauchy sequence converges.
A complete normed vector space is called a *Banach* space.

Remark. The euclidean k-space $\mathbb{R}^k$ and $\mathbb{C}^k$ is complete.
The function space $\mathrm{C}([a, b]), \|\cdot\|_\infty)$ is complete, see homework 21.2. The Hilbert space ℓ_2 is complete

(b) Compactness

The notion of compactness is of great importance in analysis, especially in connection with continuity.
By an *open cover* of a set E in a metric space X we mean a collection $\{G_\alpha \mid \alpha \in I\}$ of open subsets of X such that $E \subset \bigcup_\alpha G_\alpha$. Here I is any index set and

$$\bigcup_{\alpha \in I} G_\alpha = \{x \in X \mid \ \exists\, \beta \in I \colon x \in G_\beta\}.$$

Definition 6.14 (Covering definition) A subset K of a metric space X is said to be *compact* if every open cover of K contains a finite subcover. More explicitly, if $\{G_\alpha\}$ is an open cover of K, then there are finitely many indices $\alpha_1, \ldots, \alpha_n$ such that

$$K \subset G_{\alpha_1} \cup \cdots \cup G_{\alpha_n}.$$

Note that the definition does not state that a set is compact if there exists a finite open cover—the whole space X is open and a cover consisting of only one member. Instead, *every* open cover has a finite subcover.

Example 6.16 (a) It is clear that every finite set is compact.
(b) Let (x_n) be a converging to x sequence in a metric space X. Then

$$A = \{x_n \mid n \in \mathbb{N}\} \cup \{x\}$$

is compact.
Proof. Let $\{G_\alpha\}$ be any open cover of A. In particular, the limit point x is covered by, say, G_0. Then there is an $n_0 \in \mathbb{N}$ such that $x_n \in G_0$ for every $n \geq n_0$. Finally, x_k is covered by some G_k, $k = 1, \ldots, n_0 - 1$. Hence the collection

$$\{G_k \mid k = 0, 1, \ldots, n_0 - 1\}$$

is a finite subcover of A; therefore A is compact. ∎

Proposition 6.23 (Sequence Definition) *A subset K of a metric space X is compact if and only if every sequence in K contains a convergent subsequence with limit in K.*

Proof. (a) Let K be compact and suppose to the contrary that (x_n) is a sequence in K without any convergent to some point of K subsequence. Then every $x \in K$ has a neighborhood U_x containing only finitely many elements of the sequence (x_n). (Otherwise x would be a limit point of (x_n) and there were a converging to x subsequence.) By construction,

$$K \subset \bigcup_{x \in X} U_x.$$

Since K is compact, there are finitely many points $y_1, \ldots, y_m \in K$ with

$$K \subset U_{y_1} \cup \cdots \cup U_{y_m}.$$

Since every U_{y_i} contains only finitely many elements of (x_n), there are only finitely many elements of (x_n) in K—a contradiction.
(b) The proof is an the appendix to this chapter. ∎

Remark 6.6 Further properties. (a) A compact subset of a metric space is closed and bounded.
(b) A closed subsets of a compact set is compact.
(c) A subset K of $\mathbb{R}^k$ or $\mathbb{C}^k$ is compact if and only if K is bounded and closed.
Proof. Suppose K is closed and bounded. Let (x_n) be a sequence in K. By Corollary 6.20 (x_n) has a convergent subsequence. Since K is closed, the limit is in K. By the above proposition K is compact. The other directions follows from (a) ∎

(c) Compactness and Continuity

As in the real case (see Theorem 3.6) we have the analogous results for metric spaces.

Proposition 6.24 *Let X be a compact metric space.*
(a) *Let $f\colon X \to Y$ be a continuous mapping into the metric space Y. Then $f(X)$ is compact.*
(b) *Let $f\colon X \to \mathbb{R}$ a continuous mapping. Then f is bounded and attains its maximum and minimum, that is there are points p and q in X such that*

$$f(p) = \sup_{x \in X} f(x), \quad f(q) = \inf_{x \in X} f(x).$$

Proof. (a) Let $\{G_\alpha\}$ be an open covering of $f(X)$. By Proposition 6.22 $f^{-1}(G_\alpha)$ is open for every α. Hence, $\{f^{-1}(G_\alpha)\}$ is an open cover of X. Since X is compact there is an open subcover of X, say $\{f^{-1}(G_{\alpha_1}), \ldots, f^{-1}(G_{\alpha_n})\}$. Then $\{G_{\alpha_1}, \ldots, G_{\alpha_n}\}$ is a finite subcover of $\{G_\alpha\}$ covering $f(X)$. Hence, $f(X)$ is compact. We skip (b). ∎

Similarly as for real function we have the following proposition about uniform continuity. The proof is in the appendix.

Proposition 6.25 *Let $f\colon K \to \mathbb{R}$ be a continuous function on a compact set $K \subset \mathbb{R}$. Then f is uniformly continuous on K.*

6.4.6 Continuous Functions in $\mathbb{R}^k$

Proposition 6.26 (a) *The projection mapping $\mathrm{p}_i\colon \mathbb{R}^k \to \mathbb{R}$, $i = 1, \ldots, k$, given by $\mathrm{p}_i(x_1, \ldots, x_k) = x_i$ is continuous.*
(b) *Let $U \subseteq \mathbb{R}^k$ be open and $f, g\colon U \to \mathbb{R}$ be continuous on U. Then $f + g$, fg, $|f|$, and, f/g ($g \neq 0$) are continuous functions on U.*
(c) *Let X be a metric space. A mapping*

$$f = (f_1, \ldots, f_k)\colon X \to \mathbb{R}^k$$

is continuous if and only if all components $f_i\colon X \to \mathbb{R}$, $i = 1, \ldots, k$, are continuous.

Proof. (a) Let (x_n) be a sequence converging to $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$. Then the sequence $(\mathrm{p}_i(x_n))$ converges to $a_i = \mathrm{p}_i(a)$ by Proposition 6.19. This shows continuity of p_i at a.
(b) The proofs are quite similar to the proofs in the real case, see Proposition 2.3. As a sample we carry out the proof in case fg. Let $a \in U$ and put $M = \max\{|f(a)|, |f(b)|\}$. Let $\varepsilon > 0$, $\varepsilon < 3M^2$, be given. Since f and g are continuous at a, there exists $\delta > 0$ such that

$$\begin{aligned} \|x - a\| < \delta \quad &\text{implies} \quad |f(x) - f(a)| < \frac{\varepsilon}{3M}, \\ \|x - a\| < \delta \quad &\text{implies} \quad |g(x) - g(a)| < \frac{\varepsilon}{3M}. \end{aligned} \tag{6.45}$$

Note that

$$fg(x) - fg(a) = (f(x) - f(a))(g(x) - g(a)) + f(a)(g(x) - g(a)) + g(a)(f(x) - f(a)).$$

Taking the absolute value of the above identity, using the triangle inequality as well as (6.45) we have that $\|x - a\| < \delta$ implies

$$\begin{aligned} |fg(x) - fg(a)| &\leq |f(x) - f(a)|\,|g(x) - g(a)| + |f(a)|\,|g(x) - g(a)| + |g(a)|\,|f(x) - f(a)| \\ &\leq \frac{\varepsilon^2}{9M^2} + M\frac{\varepsilon}{3M} + M\frac{\varepsilon}{3M} \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

This proves continuity of fg at a.
(c) Suppose first that f is continuous at $a \in X$. Since $f_i = \mathrm{p}_i \circ f$, f_i is continuous by the result of (a) and Proposition 6.21.
Suppose now that all the f_i, $i = 1, \ldots, k$, are continuous at a. Let (x_n), $x_n \neq a$, be a sequence in X with $\lim_{n\to\infty} x_n = a$ in X. Since f_i is continuous, the sequences $(f_i(x_n))$ of numbers converge to $f_i(a)$. By Proposition 6.19, the sequence of vectors $f(x_n)$ converges to $f(a)$; hence f is continuous at a. ∎

Example 6.17 Let $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$f(x,y,z) = \begin{pmatrix} \sin \frac{x^2+\mathrm{e}^z}{\sqrt{x^2+y^2+z^2+1}} \\ \log|\, x^2+y^2+z^2+1\,| \end{pmatrix}.$$

Then f is continuous on U. Indeed, since product, sum, and composition of continuous functions are continuous, $\sqrt{x^2+y^2+z^2+1}$ is a continuous function on $\mathbb{R}^3$. We also made use of Proposition 6.26 (a); the coordinate functions x, y, and z are continuous. Since the denominator is nonzero, $f_1(x,y,z) = \sin \frac{x^2+\mathrm{e}^z}{\sqrt{x^2+y^2+z^2+1}}$ is continuous. Since $|\, x^2+y^2+z^2+1\,| > 0$, $f_2(x,y,z) = \log|\, x^2+y^2+z^2+1\,|$ is continuous. By Proposition 6.26 (c) f is continuous.

6.5 Appendix E

(a) A compact subset is closed

Proof. Let K be a compact subset of a metric space X. We shall prove that the complement of K is an open subset of X.
Suppose that $p \in X$, $p \notin K$. If $q \in K$, let V^q and $U(q)$ be neighborhoods of p and q, respectively, of radius less than $d(p,q)/2$. Since K is compact, there are finitely many points $q_1, \ldots, q_n$ in K such that

$$K \subset U_{q_1} \cup \cdots \cup U_{q_n} =: U.$$

If $V = V^{q_1} \cap \cdots \cap V^{q_n}$, then V is a neighborhood of p which does not intersect U. Hence $U \subset K^{\mathsf{c}}$, so that p is an interior point of K^{c}, and K is closed. We show that K is bounded. Let $\varepsilon > 0$ be given. Since K is compact the open cover $\{U_\varepsilon(x) \mid x \in K\}$ of K has a finite subcover, say $\{U_\varepsilon(x_1), \ldots, U_\varepsilon(x_n)\}$. Let $U = \bigcup_{i=1}^n U_\varepsilon(x_i)$, then the maximal distance of two points x and y in U is bounded by

$$2\varepsilon + \sum_{1 \le i < j \le n} d(x_i, x_j).$$

■

A closed subset of a compact set is compact

Proof. Suppose $F \subset K \subset X$, F is closed in X, and K is compact. Let $\{U^\alpha\}$ be an open cover of F. Since F^{c} is open, $\{U^\alpha, F^{\mathsf{c}}\}$ is an open cover Ω of K. Since K is compact, there is a finite subcover Φ of Ω, which covers K. If F^{c} is a member of Φ, we may remove it from Φ and still retain an open cover of F. Thus we have shown that a finite subcollection of $\{U^\alpha\}$ covers F. ■

Equivalence of Compactness and Sequential Compactness

Proof of Proposition 6.23 (b). This direction is hard to proof. It does not work in arbitrary topological spaces and essentially uses that X is a metric space. The prove is roughly along the lines of Exercises 22 to 26 in [Rud76]. We give the proof of Bredon (see [Bre97, 9.4 Theorem]) Suppose that every sequence in K contains a converging in K subsequence.
1) K contains a countable dense set. For, we show that for every $\varepsilon > 0$, K can be covered by a finite number of ε-balls (ε is fixed). Suppose, this is not true, i. e. K can't be covered by any finite number of ε-balls. Then we construct a sequence (x_n) as follows. Take an arbitrary x_1. Suppose $x_1, \ldots, x_n$ are already found; since K is not covered by a finite number of ε-balls, we find x_{n+1} which distance to every preceding element of the sequence is greater than or equal to ε. Consider a limit point x of this sequence and an $\varepsilon/2$-neighborhood U of x. Almost all elements of a suitable subsequence of (x_n) belong to U, say x_r and x_s with $s > r$. Since both are in U their distance is less than ε. But this contradicts the construction of the sequence.
Now take the union of all those finite sets corresponding to $\varepsilon = 1/n$, $n \in \mathbb{N}$. This is a countable dense set of K.
2) Any open cover $\{U_\alpha\}$ of K has a countable subcover. Let $x \in K$ be given. Since $\{U_\alpha\}_{\alpha\in I}$ is an open cover of K we find $\beta \in I$ and $n \in \mathbb{N}$ such that $U_{2/n}(x) \subset U_\alpha$. Further, since $\{x_i\}_{i\in\mathbb{N}}$ is dense in K, we find $i, n \in \mathbb{N}$ such that $d(x, x_i) < 1/n$. By the triangle inequality

$$x \in U_{1/n}(x_i) \subset U_{2/n}(x) \subset U_\beta.$$

To each of the countably many $U_{1/n}(x_i)$ choose one $U_\beta \supset U_{1/n}(x_i)$. This is a countable subcover of $\{U_\alpha\}$.
3) Rename the countable open subcover by $\{V_n\}_{n\in\mathbb{N}}$ and consider the decreasing sequence C_n of closed sets

$$C_n = K \setminus \bigcup_{k=1}^{n} V_k, \quad C_1 \supset C_2 \supset \cdots .$$

If $C_k = \varnothing$ we have found a finite subcover, namely $V_1, V_2, \ldots, V_k$. Suppose that all the C_n are nonempty, say $x_n \in C_n$. Further, let x be the limit of the subsequence (x_{n_i}). Since $x_{n_i} \in C_m$ for all $n_i \geq m$ and C_m is closed, $x \in C_m$ for all m. Hence $x \in \bigcap_{m\in\mathbb{N}} C_m$. However,

$$\bigcap_{m\in\mathbb{N}} C_m = K \setminus \bigcup_{m\in\mathbb{N}} V_m = \varnothing.$$

This contradiction completes the proof. ∎

Proof of Proposition 6.25. Let $\varepsilon > 0$ be given. Since f is continuous, we can associate to each point $p \in K$ a positive number $\delta(p)$ such that $q \in K \cap U_{\delta(p)}(p)$ implies $|\, f(q) - f(p)\,| < \varepsilon/2$. Let $J(p) = \{q \in K \mid |\, p - q\,| < \delta(p)/2\}$.
Since $p \in J(p)$, the collection $\{J(p) \mid p \in K\}$ is an open cover of K; and since K is compact, there is a finite set of points $p_1, \ldots, p_n$ in K such that

$$K \subset J(p_1) \cup \cdots \cup J(p_n). \tag{6.46}$$

We put $\delta := \frac{1}{2}\min\{\delta(p_1), \ldots, \delta(p_n)\}$. Then $\delta > 0$. Now let p and q be points of K with $|\, x - y \,| < \delta$. By (6.46), there is an integer m, $1 \leq m \leq n$, such that $p \in J(p_m)$; hence

$$|\, p - p_m \,| < \frac{1}{2}\delta(p_m),$$

and we also have

$$|\, q - p_m \,| \leq |\, p - q \,| + |\, p - p_m \,| < \delta + \frac{1}{2}\delta(p_m) \leq \delta(p_m).$$

Finally, continuity at p_m gives

$$|\, f(p) - f(q) \,| \leq |\, f(p) - f(p_m) \,| + |\, f(p_m) - f(q) \,| < \varepsilon.$$

■

Proposition 6.27 *There exists a real continuous function on the real line which is nowhere differentiable.*

Proof. Define

$$\varphi(x) = |\, x \,|, \quad x \in [-1, 1]$$

and extend the definition of φ to all real x by requiring periodicity

$$\varphi(x + 2) = \varphi(x).$$

Then for all $s, t \in \mathbb{R}$,

$$|\, \varphi(s) - \varphi(t) \,| \leq |\, s - t \,|. \tag{6.47}$$

In particular, φ is continuous on $\mathbb{R}$. Define

$$f(x) = \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n \varphi(4^n x). \tag{6.48}$$

Since $0 \leq \varphi \leq 1$, Theorem 6.2 shows that the series (6.48) converges uniformly on $\mathbb{R}$. By Theorem 6.5, f is continuous on $\mathbb{R}$.
Now fix a real number x and a positive integer $m \in \mathbb{N}$. Put

$$\delta_m = \frac{\pm 1}{2 \cdot 4^m}$$

where the sign is chosen that no integer lies between $4^m x$ and $4^m(x + \delta_m)$. This can be done since $4^m\,|\, \delta_m \,| = \frac{1}{2}$. It follows that $|\, \varphi(4^m x) - \varphi(4^m x + 4^m \delta_m) \,| = \frac{1}{2}$. Define

$$\gamma_n = \frac{\varphi(4^n(x + \delta_m)) - \varphi(4^n x)}{\delta_m}.$$

When $n > m$, then $4^n\delta_m$ is an even integer, so that $\gamma_n = 0$ by peridicity of φ. When $0 \le n \le m$, (6.47) implies $|\,\gamma_n\,| \le 4^m$. Since $|\,\gamma_m\,| = 4^m$, we conclude that

$$\left|\,\frac{f(x+\delta_m)-f(x)}{\delta_m}\,\right| = \left|\,\sum_{n=0}^{m}\left(\frac{3}{4}\right)^n \gamma_n\,\right| \ge 3^m - \sum_{n=0}^{m-1} 3^n = \frac{1}{2}\,(3^m+1)\,.$$

As $m \to \infty$, $\delta_m \to 0$. It follows that f is not differentiable at x. ∎

Proof of Abel's Limit Theorem, Proposition 6.9. By Proposition 6.4, the series converges on $(-1,1)$ and the limit function is continuous there since the radius of convergence is at least 1, by assumption. Hence it suffices to proof continuity at $x = 1$, i. e. that $\lim_{x\to 1-0} f(x) = f(1)$. Put $r_n = \sum_{k=n}^{\infty} a_k$; then $r_0 = f(1)$ and $r_{n+1} - r_n = -c_n$ for all nonnegative integers $n \in \mathbb{Z}_+$ and $\lim_{n\to\infty} r_n = 0$. Hence there is a constant C with $|\,r_n\,| \le C$ and the series $\sum_{n=0}^{\infty} r_{n+1}x^n$ converges for $|\,x\,| < 1$ by the comparison test. We have

$$\begin{aligned}(1-x)\sum_{n=0}^{\infty} r_{n+1}x^n &= \sum_{n=0}^{\infty} r_{n+1}x^n + \sum_{n=0}^{\infty} r_{n+1}x^{n+1}\\ &= \sum_{n=0}^{\infty} r_{n+1}x^n - \sum_{n=0}^{\infty} r_n x^n + r_0 = -\sum_{n=0}^{\infty} a_n x^n + f(1),\end{aligned}$$

hence,

$$f(1) - f(x) = (1-x)\sum_{n=0}^{\infty} r_{n+1}x^n.$$

Let $\varepsilon > 0$ be given. Choose $N \in \mathbb{N}$ such that $n \ge N$ implies $|\,r_n\,| < \varepsilon$. Put $\delta = \varepsilon/(CN)$; then $x \in (1-\delta, 1)$ implies

$$\begin{aligned}|\,f(1)-f(x)\,| &\le (1-x)\sum_{n=0}^{N-1} |\,r_{n+1}\,|\,x^n + (1-x)\sum_{n=N}^{\infty} |\,r_{n+1}\,|\,x^n\\ &\le (1-x)CN + (1-x)\varepsilon\sum_{n=0}^{\infty} x^n = 2\varepsilon;\end{aligned}$$

hence f tends to $f(1)$ as $x \to 1-0$. ∎

Definition 6.15 If X is a metric space $\mathrm{C}(X)$ will denote the set of all continuous, bounded functions with domain X. We associate with each $f \in \mathrm{C}(X)$ its *supremum norm*

$$\|f\|_\infty = \|f\| = \sup_{x\in X} |\,f(x)\,|\,. \tag{6.49}$$

Since f is assumed to be bounded, $\|f\| < \infty$. Note that boundedness of X is redundant if X is a *compact* metric space (Proposition 6.24). Thus $\mathrm{C}(X)$ contains of all continuous functions in that case.

It is clear that $\mathrm{C}(X)$ is a vector space since the sum of bounded functions is again a bounded

function (see the triangle inequality below) and the sum of continuous functions is a continuous function (see Proposition 6.26). We show that $\|f\|_\infty$ is indeed a norm on $\mathrm{C}(X)$.
(i) Obviously, $\|f\|_\infty \geq 0$ since the absolute value $|\, f(x)\,|$ is nonnegative. Further $\|0\| = 0$. Suppose now $\|f\| = 0$. This implies $|\, f(x)\,| = 0$ for all x; hence $f = 0$.
(ii) Clearly, for every (real or complex) number λ we have

$$\|\lambda f\| = \sup_{x\in X} |\, \lambda f(x)\,| = |\,\lambda\,| \sup_{x\in X} |\, f(x)\,| = |\,\lambda\,|\,\|f\|\,.$$

(iii) If $h = f + g$ then

$$|\, h(x)\,| \leq |\, f(x)\,| + |\, g(x)\,| \leq \|f\| + \|g\|\,, \quad x \in X;$$

hence

$$\|f + g\| \leq \|f\| + \|g\|\,.$$

We have thus made $\mathrm{C}(X)$ into a normed vector space. Remark 6.1 can be rephrased as

> *A sequence (f_n) converges to f with respect to the norm in $\mathrm{C}(X)$ if and only if $f_n \to f$ uniformly on X.*

Accordingly, closed subsets of $\mathrm{C}(X)$ are sometimes called *uniformly closed*, the closure of a set $A \subset \mathrm{C}(X)$ is called the *uniform closure*, and so on.

Theorem 6.28 *The above norm makes $\mathrm{C}(X)$ into a Banach space (a complete normed space).*

Proof. Let (f_n) be a Cauchy sequence of $\mathrm{C}(X)$. This means to every $\varepsilon > 0$ corresponds an $n_0 \in \mathbb{N}$ such that $n, m \geq n_0$ implies $\|f_n - f_m\| < \varepsilon$. It follows by Proposition 6.1 that there is a function f with domain X to which (f_n) converges uniformly. By Theorem 6.5, f is continuous. Moreover, f is bounded, since there is an n such that $|\, f(x) - f_n(x)\,| < 1$ for all $x \in X$, and f_n is bounded.
Thus $f \in \mathrm{C}(X)$, and since $f_n \to f$ uniformly on X, we have $\|f - f_n\| \to 0$ as $n \to \infty$. ∎

Chapter 7

Calculus of Functions of Several Variables

In this chapter we consider functions $f\colon U \to \mathbb{R}$ or $f\colon U \to \mathbb{R}^m$ where $U \subset \mathbb{R}^n$ is an open set. In Subsection 6.4.6 we collected the main properties of *continuous* functions f. Now we will study differentiation and integration of such functions in more detail

The Norm of a linear Mapping

Proposition 7.1 *Let $T \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ be a linear mapping of the euclidean spaces $\mathbb{R}^n$ into $\mathbb{R}^m$.*
(a) *Then there exists some $C > 0$ such that*

$$\|T(x)\|_2 \le C\ \|x\|_2\,, \quad \textit{for all } x \in \mathbb{R}^n. \tag{7.1}$$

(b) *T is uniformly continuous on $\mathbb{R}^n$.*

Proof. (a) Using the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ we identify T with its matrix $T = (a_{ij})$, $T\mathrm{e}_j = \sum_{i=1}^m a_{ij}\mathrm{e}_i$. For $x = (x_1, \dots, x_n)$ we have

$$T(x) = \left(\sum_{j=1}^n a_{1j}x_j,\ \dots, \sum_{j=1}^n a_{mj}x_j\right);$$

hence by the Cauchy–Schwarz inequality we have

$$\begin{aligned}\|T(x)\|_2^2 = \sum_{i=1}^m \left|\sum_{j=1}^n a_{ij}x_j\right|^2 &\le \sum_{i=1}^n \left(\sum_{j=1}^m |\,a_{ij}x_j\,|\right)^2 \\ &\le \sum_{i=1}^m\sum_{j=1}^n |\,a_{ij}\,|^2 \sum_{j=1}^n |\,x_j\,|^2 = \left(\sum_{i,j} a_{ij}^2\right)\sum_{j=1}^n |\,x_j\,|^2 = C^2\ \|x\|^2\,,\end{aligned}$$

where $C = \sqrt{\sum_{i,j} a_{ij}^2}$. Consequently,

$$\|Tx\| \le C\,\|x\|\,.$$

(b) Let $\varepsilon > 0$. Put $\delta = \varepsilon/C$ with the above C. Then $\|x - y\| < \delta$ implies

$$\|Tx - Ty\| = \|T(x-y)\| \le C\,\|x-y\| < \varepsilon,$$

which proves (b). ∎

Definition 7.1 Let V and W normed vector spaces and $A \in \mathrm{L}(V, W)$. The smallest number C with (7.1) is called the *norm* of the linear map A and is denoted by $\|A\|$.

$$\|A\| = \inf\{C \mid \|Ax\| \leq C\,\|x\| \quad \text{for all } x \in V\}. \tag{7.2}$$

By definition,

$$\|Ax\| \leq \|A\|\ \|x\|\,. \tag{7.3}$$

Let $T \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ be a linear mapping. One can show that

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\| \leq 1} \|Tx\|\,.$$

7.1 Partial Derivatives

We consider functions $f\colon U \to \mathbb{R}$ where $U \subset \mathbb{R}^n$ is an open set. We want to find derivatives "one variable at a time."

Definition 7.2 Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ a real function. Then f is called *partial differentiable* at $a = (a_1, \ldots, a_n) \in U$ with respect to the ith coordinate if the limit

$$D_i f(a) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h} \tag{7.4}$$

exists where h is real and sufficiently small (such that $(a_1, \ldots, a_i + h, \ldots, a_n) \in U$).
$D_i f(x)$ is called the *ith partial derivative of f at a*. We also use the notations

$$D_i f(a) = \frac{\partial f}{\partial x_i}(a) = \frac{\partial f(a)}{\partial x_i} = f_{x_i}(a).$$

It is important that $D_i f(a)$ is the ordinary derivative of a certain function; in fact, if $g(x) = f(a_1, \ldots, x, \ldots, a_n)$, then $D_i f(a) = g'(a_i)$. That is, $D_i f(a)$ is the slope of the tangent line at $(a, f(a))$ to the curve obtained by intersecting the graph of f with the plane $x_j = a_j$, $j \neq i$. It also means that computation of $D_i f(a)$ is a problem we can already solve.

Example 7.1 (a) $f(x, y) = \sin(xy^2)$. Then $D_1 f(x, y) = y^2 \cos(xy^2)$ and $D_2 f(x, y) = 2xy \cos(xy^2)$.
(b) Consider the radius function $r\colon \mathbb{R}^n \to \mathbb{R}$

$$r(x) = \|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2},$$

$x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then r is partial differentiable on $\mathbb{R}^n \setminus 0$ with

$$\frac{\partial r}{\partial x_i}(x) = \frac{x_i}{r(x)}, \quad x \neq 0. \tag{7.5}$$

Indeed, the function

$$f(\xi) = \sqrt{x_1^2 + \cdots + \xi^2 + \cdots + x_n^2}$$

is differentiable, where $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ are considered to be constant. Using the chain rule one obtains (with $\xi = x_i$)

$$\frac{\partial r}{\partial x_i}(x) = f'(\xi) = \frac{1}{2}\frac{2\xi}{\sqrt{x_1^2 + \cdots + \xi^2 + \cdots + x_n^2}} = \frac{x_i}{r}.$$

(c) Let $f\colon (0, +\infty) \to \mathbb{R}$ be differentiable. The composition $x \mapsto f(r(x))$ (with the above radius function r) is denoted by $f(r)$, it is partial differentiable on $\mathbb{R}^n \setminus 0$. The chain rule gives

$$\frac{\partial}{\partial x_i} f(r) = f'(r)\frac{\partial r}{\partial x_i} = f'(r)\frac{x_i}{r}.$$

(d) Partial differentiability does not imply continuity. Define

$$f(x,y) = \begin{cases} \frac{xy}{(x^2+y^2)^2} = \frac{xy}{r^4}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

Obviously, f is partial differentiable on $\mathbb{R}^2 \setminus 0$. Indeed, by definition of the partial derivative

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0}\frac{f(h,0)}{h} = \lim_{h\to 0} 0 = 0.$$

Since f is symmetric in x and y, $\frac{\partial f}{\partial y}(0,0) = 0$, too. However, f is not continuous at 0 since $f(\varepsilon, \varepsilon) = 1/(4\varepsilon^2)$ becomes large as ε tends to 0.

Remark 7.1 In the next section we will become acquainted with stronger notion of differentiability which implies continuity. In particular, a *continuously* partial differentiable function is continuous.

Definition 7.3 Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ partial differentiable. The vector

$$\operatorname{grad} f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right) \tag{7.6}$$

is called the *gradient* of f at $x \in U$.

Example 7.2 (a) For the radius function $r(x)$ defined in Example 7.1 (b) we have

$$\operatorname{grad} r(x) = \frac{x}{r}.$$

Note that x/r is a unit vector (of the euclidean norm 1) in the direction x. With the notations of Example 7.1 (c),

$$\operatorname{grad} f(r) = f'(r)\frac{x}{r}.$$

(b) Let $f, g\colon U \to \mathbb{R}$ be partial differentiable functions. Then we have the following product rule

$$\operatorname{grad}(fg) = g\ \operatorname{grad} f + f\ \operatorname{grad} g. \tag{7.7}$$

This is immediate from the product rule for functions of one variable

$$\frac{\partial}{\partial x_i}(fg) = \frac{\partial f}{\partial x_i}g + f\frac{\partial g}{\partial x_i}.$$

(c) $f(x,y) = x^y$. Then $\operatorname{grad} f(x,y) = (yx^{y-1}, x^y \log x)$.

Notation. Instead of $\operatorname{grad} f$ one also writes ∇f ("Nabla f"). ∇ is a vector-valued differential operator:

$$\nabla = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right).$$

Definition 7.4 Let $U \subset \mathbb{R}^n$. A *vector field* on U is a mapping

$$v = (v_1, \ldots, v_n) \colon U \to \mathbb{R}^n. \tag{7.8}$$

To every point $x \in U$ there is associated a vector $v(x) \in \mathbb{R}^n$.
If the vector field v is partial differentiable (i. e. all components v_i are partial differentiable) then

$$\operatorname{div} v = \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i} \tag{7.9}$$

is called the *divergence* of the vector field v.

Formally the divergence of v can be written as a inner product of ∇ and v

$$\operatorname{div} v = \nabla \cdot v = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} v_i.$$

The product rule gives the following rule for the divergence. Let $f \colon U \to \mathbb{R}$ a partial differentiable function and

$$v = (v_1, \ldots, v_n) \colon U \to \mathbb{R}$$

a partial differentiable vector field, then

$$\frac{\partial}{\partial x_i}(f v_i) = \frac{\partial f}{\partial x_i} \cdot v_i + f \cdot \frac{\partial v_i}{\partial x_i}.$$

Summation over i gives

$$\operatorname{div}(fv) = \operatorname{grad} f \cdot v + f \operatorname{div} v. \tag{7.10}$$

Using the nabla operator this can be rewritten as

$$\nabla \cdot fv = \nabla f \cdot v + f \nabla \cdot v.$$

Example 7.3 Let $F \colon \mathbb{R}^n \setminus 0 \to \mathbb{R}^n$ be the vector field $F(x) = \frac{x}{r}$, $r = \|x\|$. Since

$$\operatorname{div} x = \sum_{i=1}^{n} \frac{\partial x_i}{\partial x_i} = n \quad \text{and} \quad x \cdot x = r^2,$$

Example 7.2 gives with $v = x$ and $f(r) = 1/r$

$$\operatorname{div} \frac{x}{r} = \operatorname{grad} \frac{1}{r} \cdot x + \frac{1}{r} \operatorname{div} x = -\frac{x}{r^3} \cdot x + \frac{n}{r} = \frac{n-1}{r}.$$

7.1.1 Higher Partial Derivatives

Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ a partial differentiable function. If all partial derivatives $D_i f\colon U \to \mathbb{R}$ are again partial differentiable, f is called *twice partial differentiable*. We can form the partial derivatives $D_j D_i f$ of the second order.
More general, $f\colon U \to \mathbb{R}$ is said to be $(k+1)$*-times partial differentiable* if it is k-times partial differentiable and all partial derivatives of order k

$$D_{i_k} D_{i_{k-1}} \cdots D_{i_1} f\colon U \to \mathbb{R}$$

are partial differentiable.
A function $f\colon U \to \mathbb{R}$ is said to be k*-times continuously partial differentiable* if it is k-times partial differentiable and all partial derivatives of order less than or equal to k are continuous. The set of all such functions on U is denoted by $\mathrm{C}^k(U)$.
We also use the notation

$$D_j D_i f = \frac{\partial^2 f}{\partial x_j \partial x_i} = f_{x_i x_j},\ D_i D_i f = \frac{\partial^2 f}{\partial x_i^2},\ D_{i_k} \cdots D_{i_1} f = \frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}.$$

Example. Let $f(x,y) = \sin(xy^2)$. One easily sees that

$$f_{yx} = f_{xy} = 2y\cos(xy^2) - y^2 \sin(xy^2) 2xy.$$

Proposition 7.2 (Schwarz's Lemma) *Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ be twice continuously partial differentiable.*
Then for every $a \in U$ and all $i, j = 1, \ldots, n$ we have

$$D_j D_i f(a) = D_i D_j f(a). \tag{7.11}$$

Proof. Without loss of generality we assume $n = 2$, $i = 1$, $j = 2$, and $a = 0$; we write (x,y) in place of (x_1, x_2). Since U is open, there is a small square of length $2\delta > 0$ completely contained in U:

$$\{(x,y) \in \mathbb{R}^2 \mid |\, x \,| < \delta, |\, y \,| < \delta\} \subset U.$$

For fixed $y \in U_\delta(0)$ define the function $F\colon (-\delta, \delta) \to \mathbb{R}$ via

$$F(x) = f(x,y) - f(x,0).$$

By the mean value theorem (Theorem 4.9) there is a ξ with $|\, \xi \,| \leq |\, x \,|$ such that

$$F(x) - F(0) = xF'(\xi).$$

But $F'(\xi) = f_x(\xi, y) - f_x(\xi, 0)$. Applying the mean value theorem to the function $h(y) = f_x(\xi, y)$, there is an η with $|\, \eta \,| \leq |\, y \,|$ and

$$f_x(\xi, y) - f_x(\xi, 0) = h'(y)y = \frac{\partial}{\partial y} f_x(\xi, \eta)\, y = f_{xy}(\xi, \eta)\, y.$$

Altogether we have

$$F(x) - F(0) = f(x,y) - f(x,0) - f(0,y) + f(0,0) = f_{xy}(\xi, \eta) xy. \tag{7.12}$$

The same arguments but starting with the function $G(y) = f(x,y) - f(0,y)$ show the existence of ξ' and η' with $|\,\xi'\,| \le |\,x\,|$, $|\,\eta'\,| \le |\,y\,|$ and

$$f(x,y) - f(x,0) - f(0,y) + f(0,0) = f_{xy}(\xi',\eta')\,xy. \tag{7.13}$$

From (7.12) and (7.13) for $xy \neq 0$ it follows that

$$f_{xy}(\xi,\eta) = f_{xy}(\xi',\eta').$$

If (x,y) approaches $(0,0)$ so do (ξ,η) and (ξ',η'). Since f_{xy} and f_{yx} are both continuous it follows from the above equation

$$D_2D_1f(0,0) = D_1D_2f(0,0).$$

■

Corollary 7.3 *Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}^n$ be k-times continuously partial differentiable. Then*

$$D_{i_k} \cdots D_{i_1} f = D_{i_{\pi(k)}} \cdots D_{i_{\pi(1)}} f$$

for every permutation π of $1,\ldots,k$.

Proof. The proof is by induction on k using the fact that any permutation can be written as a product of transpositions $(j \leftrightarrow j+1)$. ■

Example 7.4 Let $U \subset \mathbb{R}^3$ be open and let $v\colon U \to \mathbb{R}^3$ be a partial differentiable vector field. One defines a new vector field $\operatorname{curl} v\colon U \to \mathbb{R}^3$, the *curl* of v by

$$\operatorname{curl} v = \left(\frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3},\ \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1},\ \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2}\right). \tag{7.14}$$

Formally one can think of $\operatorname{curl} v$ as being the vector product of ∇ and v

$$\operatorname{curl} v = \nabla \times v = \begin{vmatrix} \mathrm{e}_1 & \mathrm{e}_2 & \mathrm{e}_3 \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ v_1 & v_2 & v_3 \end{vmatrix},$$

where $\mathrm{e}_1, \mathrm{e}_2$, and e_3 are the unit vectors in $\mathbb{R}^3$. If $f\colon U \to \mathbb{R}$ has continuous second partial derivatives then, by Proposition 7.2,

$$\operatorname{curl} \operatorname{grad} f = 0. \tag{7.15}$$

Indeed, the first coordinate of $\operatorname{curl} \operatorname{grad} f$ is by definition

$$\frac{\partial^2 f}{\partial x_2 \partial x_3} - \frac{\partial^2 f}{\partial x_3 \partial x_2} = 0.$$

The other two components are obtained by cyclic permutation of the indices.
We have found: $\operatorname{curl} v = 0$ is a necessary condition for a continuously partial differentiable vector field $v\colon U \to \mathbb{R}^3$ to be the gradient of a function $f\colon U \to \mathbb{R}$.

7.1.2 The Laplacian

Let $U \subset \mathbb{R}^n$ be open and $f \in \mathrm{C}(U)$. Put

$$\Delta f = \operatorname{div} \operatorname{grad} f = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}, \tag{7.16}$$

and call

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}$$

the *Laplacian* or *Laplace operator*. The equation $\Delta f = 0$ is called the *Laplace equation*; its solution are the *harmonic* functions. If f depends on an additional time variable t, $f : U \times I \to \mathbb{R}$, $(x,t) \mapsto f(x,t)$ one considers the so called *wave equation*

$$f_{tt} - a^2 \Delta f = 0, \tag{7.17}$$

and the so called *heat equation*

$$f_t - k\Delta f = 0. \tag{7.18}$$

Example 7.5 Let $f : (0, +\infty) \to \mathbb{R}$ be twice continuously differentiable. We want to compute the Laplacian $\Delta f(r)$, $r = \|x\|$, $x \in \mathbb{R}^n \setminus 0$. By Example 7.2 we have

$$\operatorname{grad} f(r) = f'(r)\frac{x}{r},$$

and by the product rule and Example 7.3 we obtain

$$\Delta f(r) = \operatorname{div} \operatorname{grad} f(r) = \operatorname{grad} f'(r) \cdot \frac{x}{r} + f'(r) \operatorname{div} \frac{x}{r} = f''(r)\frac{x}{r} \cdot \frac{x}{r} + f'(r)\frac{n-1}{r};$$

thus

$$\Delta f(r) = f''(r) + \frac{n-1}{r} f'(r).$$

In particular, $\Delta \frac{1}{r^{n-2}} = 0$ if $n \geq 3$ and $\Delta \log r = 0$ if $n = 2$. **Prove!**

7.2 Total Differentiation

In this section we define (total) differentiability of a function f from $\mathbb{R}^n$ to $\mathbb{R}^m$. Roughly speaking, f is differentiable (at some point) if it can be approximated by a linear mapping. In contrast to partial differentiability we need not to refer to single coordinates. Differentiable functions are continuous. In this section U always denotes an open subset of $\mathbb{R}^n$. The vector space of linear mappings f of a vector space V into a vector space W will be denoted by $\mathrm{L}(V, W)$.

Motivation: If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $x \in \mathbb{R}$ and $f'(x) = a$, then

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - a \cdot h}{h} = 0.$$

Note that the mapping $h \mapsto ah$ is linear from $\mathbb{R} \to \mathbb{R}$ and any linear mapping is of that form.

Definition 7.5 The mapping $f\colon U \to \mathbb{R}^m$ is said to be *differentiable* at a point $x \in U$ if there exist a linear map $A\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - A(h)\|}{\|h\|} = 0. \tag{7.19}$$

The linear map $A \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ is called the *derivative* of f at x and will be denoted by $Df(x)$.

In case $n = m = 1$ this notion coincides with the ordinary differentiability of a function.

Remark 7.2 We reformulate the definition of differentiability of f at $a \in U$: Define a function $\varphi_a\colon U_\varepsilon(0) \subset \mathbb{R}^n \to \mathbb{R}^m$ (depending on both a and h) by

$$f(a+h) = f(a) + A(h) + \varphi_a(h). \tag{7.20}$$

Then f is differentiable at a if and only if $\lim_{h\to 0} \frac{\|\varphi_a(h)\|}{\|h\|} = 0$. Replacing the r.h.s. of (7.20) by $f(a) + A(h)$ (forgetting about φ_a) and inserting x in place of $a+h$ and $Df(a)$ in place of A, we obtain the *linearization* $L\colon \mathbb{R}^n \to \mathbb{R}^m$ of f at a:

$$L(x) = f(a) + Df(a)(x-a). \tag{7.21}$$

Lemma 7.4 *If f is differentiable at $x \in U$ the linear mapping A is uniquely determined.*

Proof. Throughout we refer to the euclidean norms on $\mathbb{R}^n$ and $\mathbb{R}^m$. Suppose that $A' \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ is another linear mapping satisfying (7.19). Then for $h \in \mathbb{R}^n$, $h \neq 0$,

$$\begin{aligned}
\|A(h) - A'(h)\| &= \|f(x+h) - f(x) - A(h) - (f(x+h) - f(x) - A'(h))\| \\
&\leq \|f(x+h) - f(x) - A(h)\| + \|f(x+h) - f(x) - A'(h)\| \\
\frac{\|A(h) - A'(h)\|}{\|h\|} &\leq \frac{\|f(x+h) - f(x) - A(h)\|}{\|h\|} + \frac{\|f(x+h) - f(x) - A'(h)\|}{\|h\|}
\end{aligned}$$

Since the limit $h \to 0$ on the right exists and equals 0, the l.h.s also tends to 0 as $h \to 0$, that is

$$\lim_{h\to 0} \frac{\|A(h) - A'(h)\|}{\|h\|} = 0.$$

Now fix $h_0 \in \mathbb{R}^n$, $h_0 \neq 0$, and put $h = th_0$, $t \in \mathbb{R}$, $t \to 0$. Then $h \to 0$ and hence,

$$0 = \lim_{t\to 0} \frac{\|A(th_0) - A'(th_0)\|}{\|th_0\|} = \lim_{t\to 0} \frac{|\,t\,|\,\|A(h_0) - A'(h_0)\|}{|\,t\,|\,\|h_0\|} = \frac{\|A(h_0) - A'(h_0)\|}{\|h_0\|}.$$

Hence, $\|A(h_0) - A'(h_0)\| = 0$ which implies $A(h_0) = A'(h_0)$ such that $A = A'$. ∎

Definition 7.6 The matrix $(a_{ij}) \in \mathbb{R}^{m\times n}$ to the linear map $Df(x)$ with respect to the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$ is called the *Jacobi matrix* of f at x. It is denoted by $f'(x)$, that is

$$a_{ij} = (f'(x))_{ij} = (0,0,\ldots,\underbrace{1}_{i},0,\ldots,0)\cdot Df(x)\cdot \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \mathrm{e}_i \cdot Df(x)(\mathrm{e}_j).$$

Example Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be linear, $f(x) = B(x)$ with $B \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$. Then $Df(x) = B$ is the constant linear mapping. Indeed,

$$f(x+h) - f(x) - B(h) = B(x+h) - B(x) - B(h) = 0$$

since B is linear. Hence, $\lim\limits_{h\to 0} \|f(x+h) - f(x) - B(h)\| \, \|h\| = 0$ which proves the claim.

Remark 7.3 (a) Using a column vector $h = (h_1, \ldots, h_n)^\top$ the map $Df(x)(h)$ is then given by matrix multiplication

$$Df(x)(h) = f'(x) \cdot h = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n a_{1j}\, h_j \\ \vdots \\ \sum_{j=1}^n a_{mj}\, h_j \end{pmatrix}.$$

Once chosen the standard basis in $\mathbb{R}^m$, we can write $f(x) = (f_1(x), \ldots, f_m(x))$ as vector of m scalar functions $f_i\colon \mathbb{R}^n \to \mathbb{R}$. By Proposition 6.19 the limit of the vector function

$$\lim_{h\to 0} \frac{1}{\|h\|} \left(f(x+h) - f(x) - Df(x)(h) \right).$$

exists and is equal to 0 if and only if the limit exists for every coordinate $i = 1, \ldots, m$ and is 0

$$\lim_{h\to 0} \frac{1}{\|h\|} \left(f_i(x+h) - f_i(x) - \sum_{j=1}^n a_{ij} h_j \right) = 0, \quad i = 1, \ldots, m. \tag{7.22}$$

We see, f is differentiable at x if and only if all f_i, $i = 1, \ldots, m$, are. In this case the Jacobi matrix $f'(x)$ is just the collection of the row vectors $f_i'(x), i = 1, \ldots, m$:

$$f'(x) = \begin{pmatrix} f_1'(x) \\ \vdots \\ f_m'(x) \end{pmatrix},$$

where $f_i'(x) = (a_{i1}, a_{i2}, \ldots, a_{in})$.
(b) Case $m = 1$ ($f = f_1$), $f'(a) \in \mathbb{R}^{1\times n}$ is a linear functional (a row vector). Proposition 7.6 below will show that $f'(a) = \operatorname{grad} f(a)$. The linearization $L(x)$ of f at a is given by the linear functional $Df(a)$ from $\mathbb{R}^n \to \mathbb{R}$. The graph of L is an n-dimensional hyper plane in $\mathbb{R}^{n+1}$ touching the graph of f at the point $(a, f(a))$. In coordinates, the linearization (hyper plane equation) is

$$x_{n+1} = L(x) = f(a) + f'(a)\cdot(x - a).$$

Here $f'(a)$ is the row vector corresponding to the linear functional $Df(a)\colon \mathbb{R}^n \to \mathbb{R}$ w.r.t. the standard basis.

Example 7.6 Let $C = (c_{ij}) \in \mathbb{R}^{n\times n}$ be a symmetric $n \times n$ matrix, that is $c_{ij} = c_{ji}$ for all $i, j = 1, \ldots, n$ and define $f\colon \mathbb{R}^n \to \mathbb{R}$ by

$$f(x) = x\cdot C(x) = \sum_{i,j=1}^n c_{ij} x_i x_j, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

If $a, h \in \mathbb{R}^n$ we have

$$\begin{aligned} f(a+h) &= a + h{\cdot}C(a+h) = a{\cdot}C(a) + h{\cdot}C(a) + a{\cdot}C(h) + h{\cdot}C(h) \\ &= a{\cdot}Ca + 2C(a){\cdot}h + h{\cdot}C(h) \\ &= f(a) + v{\cdot}h + \varphi(h), \end{aligned}$$

where $v = 2C(a)$ and $\varphi(h) = h{\cdot}C(h)$. Since, by the Cauchy–Schwarz inequality,

$$|\,\varphi(h)\,| = |\,h{\cdot}C(h)\,| \le \|h\|\,\|C(h)\| \le \|h\|\,\|C\|\,\|h\| \le \|C\|\,\|h\|^2\,,$$

$\lim_{h\to 0} \frac{\varphi(h)}{\|h\|} = 0$. This proves f to be differentiable at $a \in \mathbb{R}^n$ with derivative $Df(x)(x) = 2C(a){\cdot}x$. The Jacobi matrix is a row vector $f'(a) = 2C(a)^\top$.

7.2.1 Basic Theorems

Lemma 7.5 *Let $f\colon U \to \mathbb{R}^m$ differentiable at x, then f is continuous at x.*

Proof. Define $\varphi_x(h)$ as in Remarks 7.2 with $Df(x) = A$, then

$$\lim_{h\to 0} \|\varphi_x(h)\| = 0$$

since f is differentiable at x. Since A is continuous by Prop. 7.1, $\lim_{h\to 0} A(h) = A(0) = 0$. This gives

$$\lim_{h\to 0} f(x+h) = f(x) + \lim_{h\to 0} A(h) + \lim_{h\to 0} \varphi_x(h) = f(x).$$

This shows continuity of f at x. ∎

Proposition 7.6 *Let $f\colon U \to \mathbb{R}^m$, $f(x) = (f_1(x), \ldots, f_m(x))$ be differentiable at $x \in U$. Then all partial derivatives $\frac{\partial f_i(x)}{\partial x_j}$, $i = 1, \ldots m$, $j = 1, \ldots, n$ exist and the Jacobi matrix $f'(x) \in \mathbb{R}^{m\times n}$ has the form*

$$(a_{ij}) = f'(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}(x) = \left(\frac{\partial f_i(x)}{\partial x_j}\right)_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}. \tag{7.23}$$

Notation. (a) For the Jacobi matrix we also use the notation

$$f'(x) = \left(\frac{\partial(f_1, \ldots, f_m)}{\partial(x_1, \ldots, x_n)}(x)\right).$$

(b) In case $n = m$ the determinant $\det(f'(x))$ of the Jacobi matrix is called the *Jacobian* or *functional determinant* of f at x. It is denoted by

$$\det(f'(x)) = \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)}(x)\,.$$

Proof. Inserting $h = te_j = (0, \dots, t, \dots, 0)$ into (7.22) (see Remark 7.3) we have, since $\|h\| = |t|$ and $h_k = t\delta_{kj}$ for all $i = 1, \dots, m$

$$\begin{aligned} 0 &= \lim_{t\to 0} \frac{\|f_i(x + te_j) - f_i(x) - \sum_{k=1}^n a_{ik}h_k\|}{\|te_j\|} \\ &= \lim_{t\to 0} \frac{|\, f_i(x_1, \dots, x_j + t, \dots, x_n) - f_i(x) - ta_{ij}\,|}{|\,t\,|} \\ &= \lim_{t\to 0} \left| \frac{f_i(x_1, \dots, x_j + t, \dots, x_n) - f_i(x)}{t} - a_{ij} \right| \\ &= \left| \frac{\partial f_i(x)}{\partial x_j} - a_{ij} \right|. \end{aligned}$$

Hence $a_{ij} = \dfrac{\partial f_i(x)}{\partial x_j}$. ∎

Hyper Planes

A plane in $\mathbb{R}^3$ is the set $H = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid a_1x_2 + a_2x_2 + a_3x_3 = a_4\}$ where, $a_i \in \mathbb{R}$, $i = 1, \dots, 4$. The vector $a = (a_1, a_2, a_3)$ is the *normal vector* to H; a is orthogonal to any vector $x - x'$, $x, x' \in H$. Indeed, $a \cdot (x - x') = a \cdot x - a \cdot x' = a_4 - a_4 = 0$.
The plane H is 2-dimensional since H can be written with two parameters $\alpha_1, \alpha_2 \in \mathbb{R}$ as $(x_1^0, x_2^0, x_3^0) + \alpha_1 v_1 + \alpha_2 v_2$, where (x_1^0, x_2^0, x_3^0) is some point in H and $v_1, v_2 \in \mathbb{R}^3$ are independent vectors spanning H.
This concept is can be generalized to $\mathbb{R}^n$. A *hyper plane* in $\mathbb{R}^n$ is the set of points

$$H = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid a_1x_1 + a_2x_2 + \cdots + a_nx_n = a_{n+1}\},$$

where $a_1, \dots, a_{n+1} \in \mathbb{R}$. The vector $(a_1, \dots, a_n) \in \mathbb{R}^n$ is called the *normal vector* to the hyper plane H. Note that a is unique only up to scalar multiples. A hyper plane in $\mathbb{R}^n$ is of dimension $n - 1$ since there are $n - 1$ linear independent vectors $v_1, \dots, v_{n-1} \in \mathbb{R}^n$ and a point $h \in H$ such that

$$H = \{h + \alpha_1 v_1 + \cdots + \alpha_n v_n \mid \alpha_1, \dots, \alpha_n \in \mathbb{R}\}.$$

Example 7.7 (a) Special case $m = 1$; let $f : U \to \mathbb{R}$ be differentiable. Then

$$f'(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) = \operatorname{grad} f(x).$$

It is a row vector and gives a linear functional on $\mathbb{R}^n$ which linearly associates to each vector $y = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$ a real number

$$Df(x) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \operatorname{grad} f(x) \cdot y = \sum_{j=1}^n f_{x_j}(x) y_j.$$

In particular by Remark 7.3 (b), the equation of the linearization of f at a (the touching hyper plane) is

$$\begin{aligned}
x_{n+1} &= L(x) = f(a) + \operatorname{grad} f(a)\cdot(x-a)\\
x_{n+1} &= f(a) + (f_{x_1}(a),\dots,f_{x_n}(a))\cdot(x_1-a_1,\cdots,x_n-a_n)\\
x_{n+1} &= f(a) + \sum_{j=1}^{n} f_{x_j}(a)(x_j-a_j)\\
0 &= \sum_{j=1}^{n} -f_{x_j}(a)(x_j-a_j) + 1\cdot(x_{n+1}-f(a))\\
0 &= \tilde{n}\cdot(\tilde{x}-\tilde{a}),
\end{aligned}$$

where $\tilde{x} = (x_1,\cdots,x_n,x_{n+1})$, $\tilde{a} = (a_1,\cdots a_n, f(a))$ and $\tilde{n} = (-\operatorname{grad} f(a), 1) \in \mathbb{R}^{n+1}$ is the normal vector to the hyper plane at $\tilde{a}$.
(b) Special case $n = 1$; let $f\colon (a,b) \to \mathbb{R}^m$, $f = (f_1,\dots,f_m)$ be differentiable. Then f is a *curve* in $\mathbb{R}^m$ with initial point $f(a)$ and end point $f(b)$. $f'(t) = (f_1'(t),\dots,f_m'(t)) \in \mathbb{R}^{m\times 1}$ is the Jacobi matrix of f at x (column vector). It is the *tangent vector* to the curve f at $t \in (a,b)$.
(c) Let $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$f(x,y,z) = (f_1,f_2) = \begin{pmatrix} x^3 - 3xy^2 + z \\ \sin(xyz^2) \end{pmatrix}.$$

Then

$$f'(x,y,z) = \left(\frac{\partial(f_1,f_2)}{\partial(x,y,z)}\right) = \begin{pmatrix} 3x^2-3y^2 & -6xy & 1 \\ yz^2\cos(xy^2z) & xz^2\cos(xy^2z) & 2xyz\cos(xy^2z) \end{pmatrix}.$$

The linearization of f at (a,b,c) is

$$L(x,y,z) = f(a,b,c) + f'(a,b,c)\cdot(x-a,y-b,z-c).$$

Remark 7.4 Note that the existence of all partial derivatives does not imply the existence of $f'(a)$. Recall Example 7.1 (d). There was given a function having partial derivatives at the origin not being continuous at $(0,0)$, and hence not being differentiable at $(0,0)$. However, the next proposition shows that the converse is true provided all partial derivatives are continuous.

Proposition 7.7 *Let $f\colon U \to \mathbb{R}^m$ be continuously partial differentiable at a point $a \in U$. Then f is differentiable at a and $f'\colon U \to \mathrm{L}(\mathbb{R}^n,\mathbb{R}^m)$ is continuous at a.*

The proof in case $n = 2$, $m = 1$ is in the appendix to this chapter.

Theorem 7.8 (Chain Rule) *If $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a point a and $g\colon \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $b = f(a)$, then the composition $k = g\circ f\colon \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at a and*

$$Dk(a) = Dg(b)\circ Df(a). \tag{7.24}$$

Using Jacobi matrices, this can be written as

$$k'(a) = g'(b) \cdot f'(a). \tag{7.25}$$

Proof. Let $A = Df(a)$, $B = Dg(b)$, and $y = f(x)$. Defining functions φ, ψ, and ρ by

$$\varphi(x) = f(x) - f(a) - A(x-a), \tag{7.26}$$
$$\psi(y) = g(y) - g(b) - B(y-b), \tag{7.27}$$
$$\rho(x) = g\circ f(x) - g\circ f(a) - B\circ A(x-a) \tag{7.28}$$

then

$$\lim_{x\to a} \frac{\|\varphi(x)\|}{\|x-a\|} = 0, \quad \lim_{y\to b} \frac{\|\psi(y)\|}{\|y-b\|} = 0 \tag{7.29}$$

and we have to show that

$$\lim_{x\to a} \frac{\|\rho(x)\|}{\|x-a\|} = 0.$$

Inserting (7.26) and (7.27) we find

$$\rho(x) = g(f(x)) - g(f(a)) - BA\,(x-a) = g(f(x)) - g(f(a)) - B(f(x) - f(a) - \varphi(x))$$
$$\rho(x) = [g(f(x)) - g(f(a)) - B(f(x) - f(a))] + B\circ\varphi(x)$$
$$\rho(x) = \psi(f(x)) + B(\varphi(x)).$$

Using $\|T(x)\| \le \|T\|\,\|x\|$ (see Proposition 7.1) this shows

$$\frac{\|\rho(x)\|}{\|x-a\|} \le \frac{\|\psi(f(x))\|}{\|x-a\|} + \frac{\|B\circ\varphi(x)\|}{\|x-a\|} \le \frac{\|\psi(y)\|}{\|y-b\|} \cdot \frac{\|f(x)-f(a)\|}{\|x-a\|} + \|B\|\,\frac{\|\varphi(x)\|}{\|x-a\|}.$$

Inserting (7.26) again into the above equation we continue

$$= \frac{\|\psi(y)\|}{\|y-b\|} \cdot \frac{\|\varphi(x) + A(x-a)\|}{\|a-x\|} + \|B\|\,\frac{\|\varphi(x)\|}{\|x-a\|}$$
$$\le \frac{\|\psi(y)\|}{\|y-b\|} \left(\frac{\|\varphi(x)\|}{\|a-x\|} + \|A\| \right) + \|B\|\,\frac{\|\varphi(x)\|}{\|x-a\|}.$$

All terms on the right side tend to 0 as x approaches a. This completes the proof. ∎

Remarks 7.5 (a) The chain rule in coordinates. If $A = f'(a)$, $B = g'(f(a))$, and $C = k'(a)$, then $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times m}$, and $C \in \mathbb{R}^{p\times n}$ and

$$\left(\frac{\partial(k_1,\dots,k_p)}{\partial(x_1,\dots,x_n)}\right) = \left(\frac{\partial(g_1,\dots,g_p)}{\partial(y_1,\dots,y_m)}\right) \circ \left(\frac{\partial(f_1,\dots,f_m)}{\partial(x_1,\dots,x_n)}\right) \tag{7.30}$$

$$\frac{\partial k_r}{\partial x_j}(a) = \sum_{i=1}^{m} \frac{\partial g_r}{\partial y_i}(f(a))\,\frac{\partial f_i}{\partial x_j}(a), \quad r = 1,\dots,p,\ j = 1,\dots,n. \tag{7.31}$$

(b) In particular, in case $p = 1$, $k(x) = g(f(x))$ we have,

$$\frac{\partial k}{\partial x_j} = \frac{\partial g}{\partial y_1}\,\frac{\partial f_1}{\partial x_j} + \cdots + \frac{\partial g}{\partial y_m}\,\frac{\partial f_m}{\partial x_j}.$$

Example 7.8 (a) Let $f(u,v) = uv$, $u = g(x,y) = x^2 + y^2$, $v = h(x,y) = xy$, and $z = f(g(x,y), h(x,y)) = (x^2+y^2)xy = x^3y + x^2y^3$.

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial u}\cdot\frac{\partial g}{\partial x} + \frac{\partial f}{\partial v}\cdot\frac{\partial h}{\partial x} = v\cdot 2x + u\cdot y = 2x^2y + y(x^2+y^2)$$
$$\frac{\partial z}{\partial x} = 3x^2y + y^3.$$

(b) Let $f(u,v) = u^v$, $u(t) = v(t) = t$. Then $F(t) = f(u(t), v(t)) = t^t$ and

$$\begin{aligned} F'(t) &= \frac{\partial f}{\partial u}u'(t) + \frac{\partial f}{\partial v}v'(t) = vu^{v-1}\cdot 1 + u^v \log u \cdot 1 \\ &= t\cdot t^{t-1} + t^t \log t = t^t(\log t + 1). \end{aligned}$$

7.3 Taylor's Formula

The gradient of f gives an approximation of a scalar function f by a linear functional. Taylor's formula generalizes this concept of approximation to higher order. We consider quadratic approximation of f to determine local extrema. Througout this section we refer to the euclidean norm $\|x\| = \|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$.

7.3.1 Directional Derivatives

Definition 7.7 Let $f\colon U \to \mathbb{R}$ be a function, $a \in U$, and $e \in \mathbb{R}^n$ a unit vector, $\|e\| = 1$. The *directional derivative* of f at a in the direction of the unit vector e is the limit

$$(D_e f)(a) = \lim_{t\to 0} \frac{f(a+te) - f(a)}{t}. \tag{7.32}$$

Note that for $e = \mathrm{e}_j$ we have $D_e f = D_j f = \frac{\partial f}{\partial x_j}$.

Proposition 7.9 *Let $f\colon U \to \mathbb{R}$ be continuously differentiable. Then for every $a \in U$ and every unit vector $e \in \mathbb{R}^n$, $\|e\| = 1$, we have*

$$D_e f(a) = e \cdot \operatorname{grad} f(a) \tag{7.33}$$

Proof. Define $g\colon \mathbb{R} \to \mathbb{R}^n$ by $g(t) = a + te = (a_1 + te_1, \ldots, a_n + te_n)$. For sufficiently small $t \in \mathbb{R}$, say $|\,t\,| \le \varepsilon$, the composition $k = f\circ g$

$$\mathbb{R} \xrightarrow{g} \mathbb{R}^n \xrightarrow{f} \mathbb{R}, \quad k(t) = f(g(t)) = f(a_1 + te_1, \ldots, a_n + te_n)$$

is defined. We compute $k'(t)$ using the chain rule:

$$k'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a+te)\, g_j'(t).$$

Since $g_j'(t) = (a_j + te_j)' = e_j$ and $g(0) = a$, it follows

$$k'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a + te)\, e_j, \tag{7.34}$$

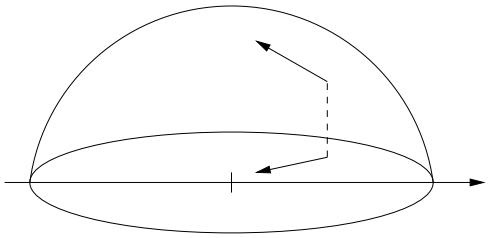
$$k'(0) = \sum_{j=1}^{n} f_{x_j}(a) e_j = \operatorname{grad} f(a) \cdot e.$$

On the other hand, by definition of the directional derivative

$$k'(0) = \lim_{t\to 0} \frac{k(t) - k(0)}{t} = \lim_{t\to 0} \frac{f(a+te) - f(a)}{t} = D_e f(a).$$

This completes the proof. ∎

Remark 7.6 (Geometric meaning of $\operatorname{grad} f$) Suppose that $\operatorname{grad} f(a) \neq 0$ and let e be a normed vector, $\|e\| = 1$. Varying e, $D_e f(x) = e \cdot \operatorname{grad} f(x)$ becomes maximal if and only if e and $\nabla f(a)$ have the same directions. Hence the vector $\operatorname{grad} f(a)$ points in the direction of maximal slope of f at a. Similarly, $-\operatorname{grad} f(a)$ is the direction of maximal decline.



For example $f(x,y) = \sqrt{1-x^2-y^2}$ has $\operatorname{grad} f(x,y) = \left(\frac{-x}{\sqrt{1-x^2-y^2}}, \frac{-y}{\sqrt{1-x^2-y^2}}\right)$. The maximal slope of f at (x,y) is in direction $e = (-x,-y)/\sqrt{x^2+y^2}$. In this case,the tangent line to the graph points to the z-axis and has maximal slope.

Corollary 7.10 *Let $f\colon U \to \mathbb{R}$ be k-times continuously differentiable, $a \in U$ and $x \in \mathbb{R}^n$ such that the whole segment $a + tx$, $t \in [0,1]$ is contained in U.*
Then the function $h\colon [0,1] \to \mathbb{R}$, $h(t) = f(a+tx)$ is k-times continuously differentiable, where

$$h^{(k)}(t) = \sum_{i_1,\dots,i_k=1}^{n} D_{i_k} \cdots D_{i_1} f(a+tx) x_{i_1} \cdots x_{i_k}. \tag{7.35}$$

In particular

$$h^{(k)}(0) = \sum_{i_1,\dots,i_k=1}^{n} D_{i_k} \cdots D_{i_1} f(a) x_{i_1} \cdots x_{i_k}. \tag{7.36}$$

Proof. The proof is by induction on k. For $k = 1$ it is exactly the statement of the Proposition. We demonstrate the step from $k = 1$ to $k = 2$. By (7.34)

$$h''(t) = \sum_{i_1=1}^{n} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial f(a+tx)}{\partial x_{i_1}} \right) x_{i_1} = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \frac{\partial f}{\partial x_{i_2} \partial x_{i_1}} (a+tx) x_{i_2} x_{i_1}.$$

In the second equality we applied the chain rule to $\tilde{h}(t) = f_{x_{i_1}}(a+tx)$. ∎

For brevity we use the following notation for the term on the right of (7.36):

$$(x\,\nabla)^k f(a) = \sum_{i_1,\dots,i_k=1}^{n} x_{i_1}\cdots x_{i_k} D_{i_k}\cdots D_{i_1} f(a).$$

In particular, $(x\nabla)f(a) = x_1 f_{x_1}(a) + x_2 f_{x_2}(a) + \cdots + x_n f_{x_n}(a)$ and $(\nabla\, x)^2 f(a) = \sum_{i,j=1}^{n} x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j}$.

7.3.2 Taylor's Formula

Theorem 7.11 *Let $f \in \mathrm{C}^{k+1}(U)$, $a \in U$, and $x \in \mathbb{R}^n$ such that $a + tx \in U$ for all $t \in [0,1]$. Then there exists $\theta \in [0,1]$ such that*

$$f(a+x) = \sum_{m=0}^{k} \frac{1}{m!}(x\,\nabla)^m f(a) + \frac{1}{(k+1)!}(x\,\nabla)^{k+1} f(a+\theta x) \tag{7.37}$$

$$f(a+x) = f(a) + \sum_{i=1}^{n} x_i f_{x_i}(a) + \frac{1}{2!}\sum_{i,j=1}^{n} x_i x_j f_{x_i x_j}(a) + \cdots + \\ + \frac{1}{(k+1)!}\sum_{i_1,\dots i_{k+1}} x_{i_1}\cdots x_{i_{k+1}} f_{x_{i_1}\cdots x_{i_{k+1}}}(a+\theta x).$$

The expression $R_{k+1}(a,x) = \frac{1}{(k+1)!}(x\,\nabla)^{k+1} f(a+\theta x)$ is called the Lagrange remainder term.

Proof. Consider the function $h\colon [0,1] \to \mathbb{R}$, $h(t) = f(a+tx)$. By Corollary 7.10, h is a $(k+1)$-times continuously differentiable. By Taylor's theorem for functions in one variable (Theorem 4.15 with $x = 1$ and $a = 0$ therein), we have

$$f(a+x) = h(1) = \sum_{m=0}^{k} \frac{h^{(m)}(0)}{m!} + \frac{h^{(k+1)}(\theta)}{(k+1)!}.$$

By Corollary 7.10 for $m = 1,\dots,k$ we have

$$\frac{h^{(m)}(0)}{m!} = \frac{1}{m!}(x\,\nabla)^m f(a).$$

and

$$\frac{h^{(k+1)}(\theta)}{(k+1)!} = \frac{1}{(k+1)!}(x\,\nabla)^{k+1} f(a+\theta x);$$

the assertion follows. ∎

It is often convenient to substitute $x := x + a$. Then the Taylor expansion reads

$$f(x) = \sum_{m=0}^{k} \frac{1}{m!}((x-a)\,\nabla)^m f(a) + \frac{1}{(k+1)!}((x-a)\,\nabla)^{(k+1)} f(a+\theta(x-a))$$

$$f(x) = f(a) + \sum_{i=1}^{n}(x_i - a_i) f_{x_i}(a) + \frac{1}{2!}\sum_{i,j=1}^{n}(x_i-a_i)(x_j-a_j) f_{x_i x_j}(a) + \cdots + \\ + \frac{1}{(k+1)!}\sum_{i_1,\dots,i_{k+1}} (x_{i_1}-a_{i_1})\cdots(x_{i_{k+1}}-a_{i_{k+1}}) f_{x_{i_1}\cdots x_{i_{k+1}}}(a+\theta(x-a)).$$

We write the Taylor formula for the case $n = 2$, $k = 3$:

$$\begin{aligned} f(a+x,b+y) = f(a,b) + (f_x(a,b)x + f_y(a,b)y) + \\ + \frac{1}{2!}\left(f_{xx}(a,b)x^2 + 2f_{xy}(a,b)\,xy + f_{yy}(a,b)y^2\right) + \\ + \frac{1}{3!}\left(f_{xxx}(a,b)x^3 + 3f_{xxy}(a,b)x^2y + 3f_{xyy}(a,b)xy^2 + f_{yyy}(a,b)y^3)\right) + R_4(a,x). \end{aligned}$$

If $f \in \bigcap_{k=0}^{\infty} \mathrm{C}^k(U) = \{f\colon f \in \mathrm{C}^k(U)\ \ \forall\, k \in \mathbb{N}\}$ and $\lim\limits_{k\to\infty} R_k(a,x) = 0$ for all $x \in U$, then

$$f(x) = \sum_{m=0}^{\infty} \frac{1}{m!}((x-a)\,\nabla)^m f(a).$$

The r.h.s. is called the *Taylor series* of f at a.

Example 7.9 (a) We compute the Taylor expansion of $f(x,y) = \cos x \sin y$ at $(0,0)$ to the third order. We have

$$\begin{aligned} f_x &= -\sin x \sin y, & f_y &= \cos x \cos y, \\ f_x(0,0) &= 0, & f_y(0,0) &= 1, \\ f_{xx} &= -\cos x \sin y, & f_{yy} &= -\cos x \sin y, & f_{xy} &= -\sin x \cos y \\ f_{xx}(0,0) &= 0, & f_{yy}(0,0) &= 0, & f_{xy}(0,0) &= 0. \\ f_{xxy} &= -\cos x \cos y, & f_{yyy} &= -\cos x \cos y, \\ f_{xxy}(0,0) &= -1, & f_{yyy}(0,0) &= -1, & f_{xyy}(0,0) &= f_{xxx}(0,0) = 0. \end{aligned}$$

Inserting this gives

$$f(x,y) = y + \frac{1}{3!}\left(-3x^2y - y^3\right) + R_4(x,y;0).$$

The same result can be obtained by multiplying the Taylor series for $\cos x$ and $\sin y$:

$$\left(1 - \frac{x^2}{2} + \frac{x^4}{4!} \mp \cdots\right)\left(y - \frac{y^3}{3!} \pm \cdots\right) = y - \frac{1}{2}x^2y - \frac{y^3}{6} + \cdots$$

(b) The Taylor series of $f(x,y) = \mathrm{e}^{xy^2}$ at $(0,0)$ is

$$\sum_{n=0}^{\infty} \frac{(xy^2)^n}{n!} = 1 + xy^2 + \frac{1}{2}x^2y^4 + \cdots;$$

it converges all over $\mathbb{R}^2$ to $f(x,y)$.

Corollary 7.12 (Mean Value Theorem) *Let $f\colon U \to \mathbb{R}$ be continuously differentiable, $a \in U$, $x \in \mathbb{R}^n$ such that $a + tx \in U$ for all $t \in [0,1]$. Then there exists $\theta \in [0,1]$ such that*

$$f(a+x) - f(a) = \nabla f(a+\theta x)\cdot x,\ f(y) - f(x) \quad = \nabla f((1-\theta)x + \theta y)\cdot(y-x). \qquad (7.38)$$

This is the special case of Taylor's formula with $k = 0$.

Corollary 7.13 *Let $f\colon U \to \mathbb{R}$ be k times continuously differentiable, $a \in U$, $x \in \mathbb{R}^n$ such that $a + tx \in U$ for all $t \in [0,1]$. Then there exists $\varphi\colon U \to \mathbb{R}$ such that*

$$f(a+x) = \sum_{m=0}^{k} \frac{1}{m!}(x\,\nabla)^m f(a) + \varphi(x), \tag{7.39}$$

where

$$\lim_{x\to 0} \frac{\varphi(x)}{\|x\|^k} = 0.$$

Proof. By Taylor's theorem for $f \in \mathrm{C}^k(U)$, there exists $\theta \in [0,1]$ such that

$$f(x+a) = \sum_{m=0}^{k-1} \frac{1}{m!}(x\,\nabla)^m f(a) + \frac{1}{k!}(x\,\nabla)^k f(a+\theta x) \overset{!}{=} \sum_{m=0}^{k} \frac{1}{m!}(x\,\nabla)^m f(a) + \varphi(x).$$

This implies

$$\varphi(x) = \frac{1}{k!}\left((x\,\nabla)^k f(a+\theta x) - (x\,\nabla)^k f(a)\right).$$

Since $|\, x_{i_1} \cdots x_{i_k} \,| \le \|x\| \ldots \|x\| = \|x\|^k$ for $x \neq 0$,

$$\frac{|\,\varphi(x)\,|}{\|x\|^k} \le \frac{1}{k!}\left|\,\nabla^k f(a+\theta x) - \nabla^k f(a)\,\right|.$$

Since all kth partial derivatives of f are continuous,

$$D_{i_1 i_2 \cdots i_k}(f(a+\theta x) - f(a)) \underset{x\to 0}{\longrightarrow} 0.$$

This proves the claim. ∎

Remarks 7.7 (a) With the above notations let

$$P_m(x) = \frac{((x-a)\,\nabla)^m}{m!} f(a).$$

Then P_m is a polynomial of degree m in the set of variables $x = (x_1, \ldots, x_n)$ and we have

$$f(x) = \sum_{m=0}^{k} P_m(x) + \varphi(x), \quad \lim_{x\to a} \frac{\|\varphi(x)\|}{\|x-a\|^k} = 0.$$

Let us consider in more detail the cases $m = 0, 1, 2$.
Case $m = 0$.

$$P_0(x) = \frac{D^0 f(a)}{0!} x^0 = f(a).$$

P_0 is the constant polynomial with value $f(a)$.
Case $m = 1$. We have

$$P_1(x) = \sum_{j=1}^{n} f_{x_j}(a)(x_j - a_j) = \operatorname{grad} f(a) \cdot (x-a)$$

Using Corollary 7.13 the first order approximation of a continuously differentiable function is

$$f(x) = f(a) + \operatorname{grad} f(a)\cdot(x-a) + \varphi(x), \quad \lim_{x\to a} \frac{\varphi(x)}{\|x-a\|} = 0. \tag{7.40}$$

The linearization of f at a is $L(x) = P_0(x) + P_1(x)$.
Case $m = 2$.

$$P_2(x) = \frac{1}{2}\sum_{i,j=1}^{n} f_{x_i x_j}(a)(x_i - a_i)(x_j - a_j).$$

Hence, $P_2(x)$ is quadratic with the corresponding matrix $\left(\frac{1}{2}f_{x_i x_j}(a)\right)$. As a special case of Corollary 7.13 ($m = 2$) we have for $f \in \mathrm{C}^2(U)$

$$f(a+x) = f(a) + \operatorname{grad} f(a)\cdot x + \frac{1}{2}x^\top \cdot \operatorname{Hess} f(a)\cdot x + \varphi(x), \quad \lim_{x\to 0} \frac{\varphi(x)}{\|x\|^2} = 0, \tag{7.41}$$

where

$$(\operatorname{Hess} f)(a) = \left(f_{x_i x_j}(a)\right)_{i,j=1}^{n} \tag{7.42}$$

is called the *Hessian matrix* of f at $a \in U$. The Hessian matrix is symmetric by Schwarz' lemma.

7.4 Extrema of Functions of Several Variables

Definition 7.8 Let $f\colon U \to \mathbb{R}$ be a function. The point $x \in U$ is called *local maximum (minimum)* of f if there exists a neighborhood $U_\varepsilon(x) \subset U$ of x such that

$$f(x) \geq f(y) \quad (f(x) \leq f(y)) \quad \text{for all } y \in U_\varepsilon(x).$$

A *local extremum* is either a local maximum or a local minimum.

Proposition 7.14 *Let $f\colon U \to \mathbb{R}$ be partial differentiable. If f has a local extremum at $x \in U$ then* $\operatorname{grad} f(x) = 0$.

Proof. For $i = 1, \ldots, n$ consider the function

$$g_i(t) = f(x + t\mathrm{e}_i).$$

This is a differentiable function of one variable, defined on a certain interval $(-\varepsilon, \varepsilon)$. If f has an extremum at x, then g_i has an extremum at $t = 0$. By Proposition 4.7

$$g_i'(0) = 0.$$

Since $g_i'(0) = \lim_{t\to 0} \frac{f(x+te_i)-f(x)}{t} = f_{x_i}(x)$ and i was arbitrary, it follows that

$$\operatorname{grad} f(x) = (D_i f(x), \ldots, D_n f(x)) = 0.$$

■

Example 7.10 Let $f(x,y) = \sqrt{1-x^2-y^2}$ be defined on the open unit disc $U = \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 < 1\}$. Then $\operatorname{grad} f(x,y) = (-x/r, -y/r) = 0$ if and only if $x = y = 0$. If f has an extremum in U then at the origin. Obviously, $f(x,y) = \sqrt{1-x^2-y^2} \leq 1 = f(0,0)$ for all points in U such that f attains its global (and local) maximum at $(0,0)$.

To obtain a sufficient criterion for the existence of local extrema we have to consider the Hessian matrix. Before, we need some facts from Linear Algebra.

Definition 7.9 Let $A \in \mathbb{R}^{n\times n}$ be a real, symmetric $n \times n$-matrix, that is $a_{ij} = a_{ji}$ for all $i,j = 1,\ldots,n$. The associated quadratic form

$$Q(x) = \sum_{i,j=1}^{n} a_{ij}x_ix_j = x^\top \cdot A \cdot x$$

is called

positive definite	if $Q(x) > 0$ for all $x \neq 0$,
negative definite	if $Q(x) < 0$ for all $x \neq 0$,
indefinite	if $Q(x) > 0$, $Q(y) < 0$ for some x, y,
positive semidefinite	if $Q(x) \geq 0$ for all x and Q is not positive definite,
negative semidefinite	if $Q(x) \leq 0$ for all x and Q is not negative definite.

Also, we say that the corresponding matrix A is *positive defininite* if $Q(x)$ is.

Example 7.11 Let $n = 2$, $Q(x) = Q(x_1,x_2)$. Then $Q_1(x) = 3x_1^2 + 7x_2^2$ is positive definite, $Q_2(x) = -x_1^2 - 2x_2^2$ is negative definite, $Q_3(x) = x_1^2 - 2x_2^2$ is indefinite, $Q_4(x) = x_1^2$ is positive semidefinite, and $Q_5(x) = -x_2^2$ is negative semidefinite.

Proposition 7.15 (Sylvester) *Let A be a real symmetric $n \times n$-matrix and $Q(x) = x\cdot Ax$ the corresponding quadratic form. For $k = 1,\cdots,n$ let*

$$A_k = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}, \quad D_k = \det A_k.$$

Let $\lambda_1,\ldots,\lambda_n$ be the eigenvalues of A. Then

(a) *Q is positive definite if and only if $\lambda_1 > 0$, $\lambda_2 > 0,\ldots,\lambda_n > 0$. This is the case if and only if $D_1 > 0$, $D_2 > 0$,..., $D_n > 0$.*
(b) *$Q(x)$ is negative definite if and only if $\lambda_1 < 0$, $\lambda_2 < 0$,...,$\lambda_n < 0$. This is the case if and only if $(-1)^k D_k > 0$ for all $k = 1,\ldots,n$.*
(c) *$Q(x)$ is indefinite if and only if, A has both positive and negative eigenvalues.*

Example 7.12 Case $n = 2$. Let $A \in \mathbb{R}^{2\times 2}$, $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, be a symmetric matrix. By Sylvester's criterion A is

(a) positive definite if and only if $\det A > 0$ and $a > 0$,

(b) negative definite if and only if $\det A > 0$ and $a < 0$,

(c) indefinite if and only if $\det A < 0$,

(d) semidefinite if and only if $\det A = 0$.

Proposition 7.16 *Let $f : U \to \mathbb{R}$ be twice continuously differentiable and let $\operatorname{grad} f(a) = 0$ at some point $a \in U$.*

(a) *If $\operatorname{Hess} f(a)$ is positive definite, then f has a local minimum at a.*
(b) *If $\operatorname{Hess} f(a)$ is negative definite, then f has a local maximum at a.*
(c) *If $\operatorname{Hess} f(a)$ is indefinite, then f has not a local extremum at a.*

Note that in general there is no information on a if $\operatorname{Hess} f(a)$ is semidefinit.
Proof. By (7.41) and since $\operatorname{grad} f(a) = 0$,

$$f(a+x) = f(a) + \frac{1}{2} x \cdot A(x) + \varphi(x), \quad \lim_{x \to 0} \frac{\varphi(x)}{\|x\|^2} = 0, \tag{7.43}$$

where $A = \operatorname{Hess} f(a)$.
(a) Let A be positive definite. Since the unit sphere $\mathrm{S} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ is compact (closed and bounded) and the map $Q(x) = x \cdot A(x)$ is continuous, the function attains its minimum, say m, on S, see Proposition 6.24,

$$m = \min\{x \cdot A(x) \mid x \in \mathrm{S}\}.$$

Since $Q(x)$ is positive definite and $0 \notin \mathrm{S}$, $m > 0$. If x is nonzero, $y = x/\|x\| \in \mathrm{S}$ and therefore

$$m \le y \cdot A(y) = \frac{1}{\|x\|} x \cdot A\left(\frac{x}{\|x\|}\right) = \frac{x}{\|x\|} \cdot \frac{A(x)}{\|x\|} = \frac{1}{\|x\|^2} x \cdot A(x),$$

This implies $Q(x) = x \cdot A(x) \ge m \|x\|^2$ for all $x \in U$.
Since $\varphi(x)/\|x\|^2 \longrightarrow 0$ as $x \to 0$, there exists $\delta > 0$ such that $\|x\| < \delta$ implies

$$-\frac{m}{4} \|x\|^2 \le \varphi(x) \le \frac{m}{4} \|x\|^2 .$$

From (7.43) it follows

$$f(a+x) = f(a) + \frac{1}{2} Q(x) + \varphi(x) \ge f(a) + \frac{1}{2} m \|x\|^2 - \frac{m}{4} \|x\|^2 \ge f(a) + \frac{m}{4} \|x\|^2 ,$$

hence

$$f(a+x) > f(a), \quad \text{if} \quad 0 < \|x\| < \delta,$$

and f has a strict (isolated) local minimum at a.
(b) If $A = \operatorname{Hess} f(a)$ is negative definite, consider $-f$ in place of f and apply (a).
(c) Let $A = \operatorname{Hess} f(a)$ indefinite. We have to show that in every neighborhood of a there exist x' and x'' such that $f(x'') < f(a) < f(x')$. Since A is indefinite, there is a vector $x \in \mathbb{R}^n \setminus 0$ such that $x \cdot A(x) = m > 0$. Then for small t we have

$$f(a+tx) = f(a) + \frac{1}{2} t x \cdot A(tx) + \varphi(tx) = f(a) + \frac{m}{2} t^2 + \varphi(tx).$$

If t is small enough, $-\frac{m}{4}t^2 \le \varphi(tx) \le \frac{m}{4}t^2$, hence

$$f(a+tx) > f(a), \quad \text{if} \quad 0 < |\,t\,| < \delta.$$

Similarly, if $y \in \mathbb{R}^n \setminus 0$ satisfies $y \cdot A(y) < 0$, for sufficiently small t we have $f(a+ty) < f(a)$. ∎

Example 7.13 (a) $f(x,y) = x^2 + y^2$. Here $\nabla f = (2x, 2y) \stackrel{!}{=} 0$ if and only if $x = y = 0$. Furthermore,

$$\operatorname{Hess} f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

is positive definite. f has a (strict) local minimum at $(0,0)$
(b)Find the local extrema of $z = f(x,y) = 4x^2 - y^2$ on $\mathbb{R}^2$. (the graph is a hyperbolic paraboloid). We find that the necesary condition $\nabla f = 0$ implies $f_x = 8x = 0$, $f_y = -2y = 0$; thus $x = y = 0$. Further,

$$\operatorname{Hess} f(x,y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 8 & 0 \\ 0 & -2 \end{pmatrix}.$$

The Hessian matrix at $(0,0)$ is indefinite; the function has not an extremum at the origin $(0,0)$.
(c) $f(x,y) = x^2 + y^3$. $\nabla f(x,y) = (2x, 3y^2)$ vanishes if and only if $x = y = 0$. Furthermore,

$$\operatorname{Hess} f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

is positive semidefinit. However, there is no local extremum at the origin since $f(\varepsilon, 0) = \varepsilon^2 > 0 = f(0,0) > -\varepsilon^3 = f(0, -\varepsilon)$.
(d) $f(x,y) = x^2 + y^4$. Again the Hessian matrix at $(0,0)$ is positive semidefinite. However, $(0,0)$ is a strict local minimum.

Local and Global Extrema

To compute the *global* extrema of a function $f\colon \overline{U} \to \mathbb{R}$ where $U \subset \mathbb{R}^n$ is open and $\overline{U}$ is the closure of U we have go along the following lines:

(a) Compute the local extrema on U;

(b) Compute the global extrema on the boundary $\partial U = \overline{U} \cap \overline{U^{\mathrm{c}}}$;

(c) If U is unbounded without boundary (as $U = \mathbb{R}$), consider the limits at infinity.

Note that

$$\sup_{x \in \overline{U}} f(x) = \max\{\text{maximum of all local maxima in } U,\ \sup_{x \in \partial U} f(x)\}.$$

To compute the global extremum of f on the boundary one has to find the local extrema on the interior point of the boundary and to compare them with the values on the boundary of the boundary.

Example 7.14 Find the global extrema of $f(x,y) = x^2 y$ on $\overline{U} = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \le 1\}$ (where U is the open unit disc.)
Since $\operatorname{grad} f = (f_x, f_y) = (2xy, x^2)$ local extrema can appear only on the y-axis $x = 0$, y is arbitrary. The Hessian matrix at $(0,y)$ is

$$\operatorname{Hess} f(0,y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}\Big|_{x=0} = \begin{pmatrix} 2y & 2x \\ 2x & 0 \end{pmatrix}\Big|_{x=0} = \begin{pmatrix} 2y & 0 \\ 0 & 0 \end{pmatrix}.$$

This matrix is positive semidefinite in case $y > 0$, negative semidefinite in case $y < 0$ and 0 at $(0,0)$. Hence, the above criterion gives *no answer*. We have to apply the definition directly. In case $y > 0$ we have $f(x,y) = x^2 y \ge 0$ for all x. In particular $f(x,y) \ge f(0,y) = 0$. Hence $(0,y)$ is a local minimum. Similarly, in case $y < 0$, $f(x,y) \le f(0,y) = 0$ for all x. Hence, f has a local maximum at $(0,y)$, $y < 0$. However f takes both positive and negative values in a neighborhood of $(0,0)$, for example $f(\varepsilon,\varepsilon) = \varepsilon^3$ and $f(\varepsilon,-\varepsilon) = -\varepsilon^3$. Thus $(0,0)$ is not a local extremum.
We have to consider the boundary $x^2 + y^2 = 1$. Inserting $x^2 = 1 - y^2$ we obtain

$$g(y) = f(x,y)|_{x^2+y^2=1} = x^2 y\big|_{x^2+y^2=1} = (1-y^2)y = y - y^3, \quad |\,y\,| \le 1.$$

We compute the local extrema of the boundary $x^2+y^2 = 1$ (note, that the circle has no boundary, such that the local extrema are actually the global extrema).

$$g'(y) = 1 - 3y^2 \stackrel{!}{=} 0, \quad |\,y\,| = \frac{1}{\sqrt{3}}.$$

Since $g''(1/\sqrt{3}) < 0$ and $g''(-1/\sqrt{3}) > 0$, g attains its maximum $\frac{2}{3\sqrt{3}}$ at $y = 1/\sqrt{3}$. Since this is greater than the local maximum of f at $(0,y)$, $y > 0$, f attains its global maximum at the two points

$$M_{1,2} = \left(\pm\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{3}}\right),$$

where $f(M_{1,2}) = x^2 y = \frac{2}{3\sqrt{3}}$. g attains its minimum $-\frac{2}{3\sqrt{3}}$ at $y = -1/\sqrt{3}$. Since this is less than the local minimum of f at $(0,y)$, $y < 0$, f attains its global minimum at the two points

$$m_{1,2} = \left(\pm\sqrt{\frac{2}{3}}, -\frac{1}{\sqrt{3}}\right),$$

where $f(m_{1,2}) = x^2 y = -\frac{2}{3\sqrt{3}}$.
The arithmetic-geometric mean inequality shows the same result for $x, y > 0$:

$$\frac{1}{3} \ge \frac{x^2+y^2}{3} = \frac{\frac{x^2}{2}+\frac{x^2}{2}+y^2}{3} \ge \left(\frac{x^2}{2}\frac{x^2}{2}y^2\right)^{\frac{1}{3}} \implies x^2 y \le \frac{2}{3\sqrt{3}}.$$

(b) Among all boxes with volume 1 find the one where the sum of the length of the 12 edges is minimal.
Let x, y and z denote the length of the three perpendicular edges of one vertex. By assumption $xyz = 1$; and $g(x,y,z) = 4(x+y+z)$ is the function to minimize.

Local Extrema. Inserting the constraint $z = 1/(xy)$ we have to minimize

$$f(x,y) = 4\left(x + y + \frac{1}{xy}\right) \quad \text{on} \quad U = \{(x,y) \mid x > 0,\, y > 0\}.$$

The necessary condition is

$$\left.\begin{aligned} f_x &= 4\left(1 - \frac{1}{x^2 y}\right) = 0, \\ f_y &= 4\left(1 - \frac{1}{xy^2}\right) = 0, \end{aligned}\right\} \Longrightarrow x^2 y = xy^2 = 1 \Longrightarrow x = y = 1.$$

Further,

$$f_{xx} = \frac{8}{x^3 y}, \quad f_{yy} = \frac{8}{xy^3}, \quad f_{xy} = \frac{4}{x^2 y^2}$$

such that

$$\det \operatorname{Hess} f(1,1) = \begin{vmatrix} 8 & 4 \\ 4 & 8 \end{vmatrix} = 64 - 16 > 0;$$

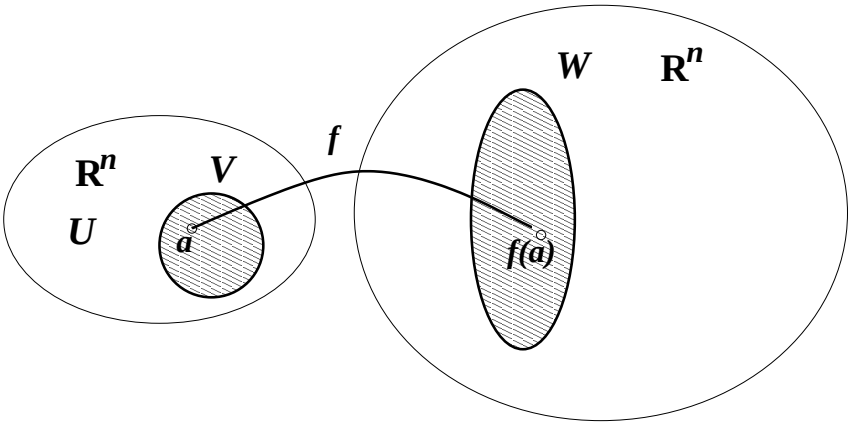
hence f has an extremum at $(1,1)$. Since $f_{xx}(1,1) = 8 > 0$, f has a local minimum at $(1,1)$.
Global Extrema. We show that $(1,1)$ is even the global minimum on the first quadrant U. Consider $N = \{(x,y) \mid \frac{1}{25} \le x, y \le 5\}$. If $(x,y) \notin N$,

$$f(x,y) \ge 4(5 + 0 + 0) = 20,$$

Since $f(x,y) \ge 12 = f(1,1)$, the global minimum of f on the right-upper quadrant is attained on the compact rectangle N. Inserting the four boundaries $x = 5$, $y = 5$, $x = 1/5$, and $y = 1/5$, in all cases, $f(x,y) \ge 20$ such that the local minimum $(1,1)$ is also the global minimum.

7.5 The Inverse Mapping Theorem

Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is differentiable on an open set $U \subset \mathbb{R}$, containing $a \in U$, and $f'(a) \neq 0$. If $f'(a) > 0$, then there is an open interval $V \subset U$ containing a such that $f'(x) > 0$ for all $x \in V$. Thus f is strictly increasing on V and therefore injective with an inverse function g defined on some open interval W containing $f(a)$. Moreover g is differentiable (see Proposition 4.5) and $g'(y) = 1/f'(x)$ if $f(x) = y$. An analogous result in higher dimensions is more involved but the result is very important.



Theorem 7.17 (Inverse Mapping Theorem) *Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable on an open set U containing a, and $\det f'(a) \neq 0$. Then there is an open set $V \subset U$ containing a and an open set W containing $f(a)$ such that $f\colon V \to W$ has a continuous inverse $g\colon W \to V$ which is differentiable and for all $y \in W$. For $y = f(x)$ we have*

$$g'(y) = (f'(x))^{-1}, \quad Dg(y) = (Df(x))^{-1}. \tag{7.44}$$

For the proof see [Rud76, 9.24 Theorem] or [Spi65, 2-11].

Corollary 7.18 *Let $U \subset \mathbb{R}^n$ be open, $f\colon U \to \mathbb{R}^n$ continuously differentiable and $\det f'(x) \neq 0$ for all $x \in U$. Then $f(U)$ is open in $\mathbb{R}^n$.*

Remarks 7.8 (a) One main part is to show that there is an open set $V \subset U$ which is mapped onto an *open* set W. In general, this is not true for *continuous* mappings. For example $\sin x$ maps the open interval $(0, 2\pi)$ onto the closed set $[-1, 1]$. Note that $\sin x$ does not satisfy the assumptions of the corollary since $\sin'(\pi/2) = \cos(\pi/2) = 0$.
(b) Note that continuity of $f'(x)$ in a neighborhood of a, continuity of the determinant mapping $\det\colon \mathbb{R}^{n\times n} \to \mathbb{R}$, and $\det f'(a) \neq 0$ implies that $\det f'(x) \neq 0$ in a neighborhood V_1 of a, see homework 10.4. This implies that the linear mapping $Df(x)$ is invertible for $x \in V_1$. Thus, $Df(x)^{-1}$ and $(f'(x))^{-1}$ exist for $x \in V_1$ — the linear mapping $Df(x)$ is regular.
(c) Let us reformulate the statement of the theorem. Suppose

$$\begin{aligned} y_1 &= f_1(x_1, \dots, x_n), \\ y_2 &= f_2(x_1, \dots, x_n), \\ &\vdots \\ y_n &= f_n(x_1, \dots, x_n) \end{aligned}$$

is a system of n equations in n variables $x_1, \dots, x_n$; $y_1, \dots, y_n$ are given in a neighborhood W of $f(a)$. Under the assumptions of the theorem, there exists a unique solution $x = g(y)$ of this system of equations

$$\begin{aligned} x_1 &= g_1(y_1, \dots, y_n), \\ x_2 &= g_2(y_1, \dots, y_n), \\ &\vdots \\ x_n &= g_n(y_1, \dots, y_n) \end{aligned}$$

in a certain neighborhood $(x_1, \dots, x_n) \in V$ of a. Note that the theorem states the existence of such a solution. It doesn't provide an explicit formula.
(d) Note that the inverse function g may exist even if $\det f'(x) = 0$. For example $f\colon \mathbb{R} \to \mathbb{R}$, defined by $f(x) = x^3$ has $f'(0) = 0$; however $g(y) = \sqrt[3]{x}$ is inverse to $f(x)$. One thing is certain if $\det f'(a) = 0$ then g cannot be differentiable at $f(a)$. If g were differentiable at $f(a)$, the chain rule applied to $g(f(x)) = x$ would give

$$g'(f(a)) \cdot f'(a) = \mathrm{id}$$

and consequently

$$\det g'(f(a)) \det f'(a) = \det \mathrm{id} = 1$$

contradicting $\det f'(a) = 0$.
(e) Note that the theorem states that under the given assumptions f is *locally* invertible. There is no information about the existence of an inverse function g to f on a fixed open set. See Example 7.15 (a) below.

Example 7.15 (a) Let $x = r\cos\varphi$ and $y = r\sin\varphi$ be the polar coordinates in $\mathbb{R}^2$. More precisely, let

$$f(r,\varphi) = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}, \quad f\colon \mathbb{R}^2 \to \mathbb{R}^2.$$

The Jacobian is

$$\frac{\partial(x,y)}{\partial(r,\varphi)} = \begin{vmatrix} x_r & x_\varphi \\ y_r & y_\varphi \end{vmatrix} = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r.$$

Let $f(r_0,\varphi_0) = (x_0,y_0) \neq (0,0)$, then $r_0 \neq 0$ and the Jacobian of f at (r_0,φ_0) is non-zero. Since all partial derivatives of f with respect to r and φ exist and they are continuous on $\mathbb{R}^2$, the assumptions of the theorem are satisfied. Hence, in a neighborhood U of (x_0,y_0) there exists a continuously differentiable inverse function $r = r(x,y)$, $\varphi = \varphi(x,y)$. In this case, the function can be given explicitly, $r = \sqrt{x^2+y^2}$, $\varphi = \arg(x,y)$. We want to compute the Jacobi matrix of the inverse function. Since the inverse matrix

$$\begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}^{-1} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix}$$

we obtain by the theorem

$$g'(x,y) = \begin{pmatrix} \frac{\partial(r,\varphi)}{\partial(x,y)} \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix};$$

in particular, the second row gives the partial derivatives of the argument function with respect to x and y

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

Note that we have not determined the explicit form of the argument function which is not unique since $f(r,\varphi+2k\pi) = f(r,\varphi)$, for all $k \in \mathbb{Z}$. However, the gradient takes always the above form. Note that $\det f'(r,\varphi) \neq 0$ for all $r \neq 0$ is not sufficient for f to be injective on $\mathbb{R}^2 \setminus \{(0,0)\}$.
(b) Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $(u,v) = f(x,y)$ where

$$u(x,y) = \sin x - \cos y, \quad v(x,y) = -\cos x + \sin y.$$

Since

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = \begin{vmatrix} \cos x & \sin y \\ \sin x & \cos y \end{vmatrix} = \cos x\cos y - \sin x\sin y = \cos(x+y)$$

f is locally invertible at $(x_0, y_0) = (\frac{\pi}{4}, -\frac{\pi}{4})$ since the Jacobian at (x_0, y_0) is $\cos 0 = 1 \neq 0$. Since $f(\frac{\pi}{4}, -\frac{\pi}{4}) = (0, -\sqrt{2})$, the inverse function $g(u, v) = (x, y)$ is defined in a neighborhood of $(0, -\sqrt{2})$ and the Jacobi matrix of g at $(0, -\sqrt{2})$ is

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Note that at point $(\frac{\pi}{4}, \frac{\pi}{4})$ the Jacobian of f vanishes. There is indeed no neighborhood of $(\frac{\pi}{4}, \frac{\pi}{4})$ where f is injective since for all $t \in \mathbb{R}$

$$f\left(\frac{\pi}{4} + t, \frac{\pi}{4} - t\right) = f\left(\frac{\pi}{4}, \frac{\pi}{4}\right) = (0, 0).$$

7.6 The Implicit Function Theorem

Motivation: Hyper Surfaces

Suppose that $F\colon U \to \mathbb{R}$ is a continuously differentiable function and $\operatorname{grad} F(x) \neq 0$ for all $x \in U$. Then

$$S = \{(x_1, \ldots, x_n) \mid F(x_1, \ldots, x_n) = 0\}$$

is called a *hyper surface* in $\mathbb{R}^n$. A hyper surface in $\mathbb{R}^n$ has dimension $n - 1$. Examples are hyper planes $a_1x_1 + \cdots + a_nx_n + c = 0$ ($(a_1, \ldots, a_n) \neq 0$), spheres $x_1^2 + \cdots + x_n^2 = r^2$. The graph of differentiable functions $f\colon U \to \mathbb{R}$ is also a hyper surface in $\mathbb{R}^{n+1}$

$$\Gamma_f = \{(x, f(x)) \in \mathbb{R}^{n+1} \mid x \in U\}.$$

Question: Is any hyper surface locally the graph of a differentiable function? More precisely, we may ask the following question: Suppose that $f\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is differentiable and $f(a_1, \ldots, a_n, b) = 0$. Can we find for each $(x_1, \ldots, x_n)$ near $(a_1, \ldots, a_n)$ a unique y near b such that $f(x_1, \ldots, x_n, y) = 0$? The answer to this question is provided by the Implicit Function Theorem (IFT).

Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = x^2 + y^2 - 1$. If we choose (a, b) with $a, b > 0$, there are open intervals A and B containig a and b with the following property: if $x \in A$, there is a unique $y \in B$ with $f(x, y) = 0$. We can therefore define a function $g\colon A \to B$ by the condition $g(x) \in B$ and $f(x, g(x)) = 0$. If $b > 0$ then $g(x) = \sqrt{1 - x^2}$; if $b < 0$ then $g(x) = -\sqrt{1 - x^2}$. Both functions g are differentiable. These functions are said to be defined *implicitly* by the equation $f(x, y) = 0$.

On the other hand, there exists no neighborhood of $(1, 0)$ such that $f(x, y) = 0$ can locally be solved for y. Note that $f_y(1, 0) = 0$. However it can be solved for $x = h(y) = \sqrt{1 - y^2}$.

Theorem 7.19 *Suppose that $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, $f = f(x, y)$, is continuously differentiable in an open set containing $(a, b) \in \mathbb{R}^{n+m}$ and $f(a, b) = 0$. Let $Df_y(x, y)$ be the linear mapping from $\mathbb{R}^m$ into $\mathbb{R}^m$ given by*

$$D_y f(x, y) = \big(D_{n+j} f^i(x, y)\big) = \left(\frac{\partial(f_1, \ldots, f_m)}{\partial(y_1, \ldots, y_m)}(x, y)\right) = \left(\frac{\partial f_i(x, y)}{\partial y_j}\right), \quad i, j = 1, \ldots, m. \tag{7.45}$$

If $\det D_y f(a,b) \neq 0$ *there is an open set* $A \subset \mathbb{R}^n$ *containing* a *and an open set* $B \subset \mathbb{R}^m$ *containing* b *with the following properties: There exists a unique continuously differentiable function* $g\colon A \to B$ *such that*

(a) $g(a) = b$,
(b) $f(x, g(x)) = 0$ *for all* $x \in A$.

For the derivative $Dg(x) \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ *we have*

$$Dg(x) = -(Df_y(x, g(x)))^{-1} \circ Df_x(x, g(x)),$$

$$g'(x) = -(f'_y(x, g(x)))^{-1} \cdot f'_x(x, g(x)).$$

The Jacobi matrix $g'(x)$ is given by

$$\left(\frac{\partial(g_1, \ldots, g_m)}{\partial(x_1, \ldots, x_n)}(x)\right) = -f'_y(x, g(x))^{-1} \cdot \left(\frac{\partial(f_1, \ldots, f_m)}{\partial(x_1, \ldots, x_n)}(x, g(x))\right) \tag{7.46}$$

$$\frac{\partial(g_k(x))}{\partial x_j} = -\sum_{l=1}^{n} (f'_y(x, g(x))^{-1})_{kl} \cdot \frac{\partial f_l(x, g(x))}{\partial x_j}, \quad k = 1, \ldots, m,\ j = 1, \ldots, n.$$

Idea of Proof. Define $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ by $F(x,y) = (x, f(x,y))$. Let $M = f'_y(a,b)$. Then

$$F'(a,b) = \begin{pmatrix} \mathbb{1}_n & 0_{n,m} \\ 0_{m,n} & M \end{pmatrix} \implies \det F'(a,b) = \det M \neq 0.$$

By the inverse mapping theorem Theorem 7.17 there exists an open set $W \subset \mathbb{R}^n \times \mathbb{R}^m$ containing $F(a,b) = (a,0)$ and an open set $V \subset \mathbb{R}^n \times \mathbb{R}^m$ containing (a,b) which may be of the form $A \times B$ such that $F\colon A \times B \to W$ has a differentiable inverse $h\colon W \to A \times B$.
Since g is differentiable, it is easy to find the Jacobi matrix. In fact, since $f_i(x, g(x)) = 0$, $i = 1, \ldots n$, taking the partial derivative $\frac{\partial f}{\partial x_j}$ on both sides gives by the chain rule

$$0 = \frac{\partial f_i(x, g(x))}{\partial x_j} + \sum_{k=1}^{m} \frac{\partial f_i(x, g(x))}{\partial y_k} \cdot \frac{\partial g_k(x)}{\partial x_j}$$

$$0 = \frac{\partial f_i(x, g(x))}{\partial x_j} + f'_y(x, g(x)) \cdot \left(\frac{\partial g_k(x)}{\partial x_j}\right) \quad \Big| - \frac{\partial f_i(x, g(x))}{\partial x_j}$$

$$-\left(\frac{\partial f_i(x, g(x))}{\partial x_j}\right) = f'_y(x, g(x)) \cdot \left(\frac{\partial g_k(x)}{\partial x_j}\right).$$

Since $\det f'_y(a,b) \neq 0$, $\det f'_y(x,y) \neq 0$ in a small neighborhood of (a,b). Hence $f'_y(x, g(x))$ is invertible and we can multiply the preceding equation from the left by $(f'_y(x, g(x)))^{-1}$ which gives (7.46).

Remarks 7.9 (a) The theorem gives a sufficient condition for "locally" solving the system of equations

$$0 = f_1(x_1, \ldots, x_n, y_1, \ldots y_m),$$
$$\vdots$$
$$0 = f_m(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

with given $x_1, \dots, x_n$ for $y_1, \dots, y_m$.
(b) We rewrite the statement in case $n = m = 1$: If $f(x, y)$ is continously differentiable on an open set $G \subset \mathbb{R}^2$ which contains (a, b) and $f(a, b) = 0$. If $f_y(a, b) \neq 0$ then there exist $\delta, \varepsilon > 0$ such that the following holds: for every $x \in U_\delta(a)$ there exists a unique $y = g(x) \in U_\varepsilon(b)$ with $f(x, y) = 0$. We have $g(a) = b$; the function $y = g(x)$ is continuously differentiable with

$$g'(x) = -\frac{f_x(x, g(x))}{f_y(x, g(x))}.$$

Be careful, note $f_x(x, g(x)) \neq \frac{\mathrm{d}}{\mathrm{d}x}(f(x, g(x)))$.

Example 7.16 (a) Let $f(x, y) = \sin(x + y) + \mathrm{e}^{xy} - 1$. Note that $f(0, 0) = 0$. Since

$$f_y(0, 0) = \cos(x + y) + x\mathrm{e}^{xy}|_{(0,0)} = \cos 0 + 0 = 1 \neq 0$$

$f(x, y) = 0$ can uniquely be solved for $y = g(x)$ in a neighborhood of $x = 0, y = 0$. Further

$$f_x(0, 0) = \cos(x + y) + y\mathrm{e}^{xy}|_{(0,0)} = 1.$$

By Remark 7.9 (b)

$$g'(x) = -\left.\frac{f_x(x, y)}{f_y(x, y)}\right|_{y=g(x)} = \frac{\cos(x + g(x)) + g(x)\,\mathrm{e}^{xg(x)}}{\cos(x + g(x)) + x\,\mathrm{e}^{xg(x)}}.$$

In particular $g'(0) = -1$.
Remark. Differentiating the equation $f_x + f_y g' = 0$ we obtain

$$\begin{aligned} 0 &= f_{xx} + f_{xy}g' + (f_{yx} + f_{yy}g')g' + f_y g'' \\ g'' &= -\frac{1}{f_y}\left(f_{xx} + 2f_{xy}g' + f_{yy}(g')^2\right) \\ g'' &\underset{g'=-\frac{f_x}{f_y}}{=} \frac{-f_{xx}f_y^2 + 2f_{xy}f_x f_y - f_{yy}f_x^2}{f_y^3}. \end{aligned}$$

Since

$$\begin{aligned} f_{xx}(0, 0) &= -\sin(x + y) + y^2\mathrm{e}^{xy}\big|_{(0,0)} = 0, \\ f_{yy}(0, 0) &= -\sin(x + y) + x^2\mathrm{e}^{xy}\big|_{(0,0)} = 0, \\ f_{xy}(0, 0) &= -\sin(x + y) + \mathrm{e}^{xy}(1 + xy)|_{(0,0)} = 1, \end{aligned}$$

we obtain $g''(0) = 2$. Therefore the Taylor expansion of $g(x)$ around 0 reads

$$g(x) = x + x^2 + r_3(x).$$

(b) Let $\gamma(t) = (x(t), y(t))$ be a differentiable curve $\gamma \in \mathrm{C}^2([0, 1])$ in $\mathbb{R}^2$. Suppose in a neighborhood of $t = 0$ the curve describes a function $y = g(x)$. Find the Taylor polynomial of degree 2 of g at $x_0 = x(0)$.

Inserting the curve into the equation $y = g(x)$ we have $y(t) = g(x(t))$. Differentiation gives

$$\dot{y} = g' \dot{x}, \qquad \ddot{y} = g''\dot{x}^2 + g'\ddot{x}$$

Thus

$$g'(x) = \frac{\dot{y}}{\dot{x}}, \qquad g''(x) = \frac{\ddot{y} - g'\ddot{x}}{\dot{x}^2} = \frac{\ddot{y}\dot{x} - \ddot{x}\dot{y}}{\dot{x}^3}$$

Now we have the Taylor ploynomial of g at x_0

$$T_2(g)(x) = x_0 + g'(x_0)(x - x_0) + \frac{g''(x_0)}{2}(x - x_0)^2.$$

(c) **The tangent hyper plane to a hyper surface.**

> Anyone who understands geometry can understand everything in this world
> (Galileo Galilei, 1564 – 1642)

Suppose that $F\colon U \to \mathbb{R}$ is continuously differentiable, $a \in U$, $F(a) = 0$, and $\operatorname{grad} F(a) \neq 0$. Then

$$\nabla F(a) \cdot (x - a) = \sum_{i=1}^{n} F_{x_i}(a)(x_i - a_i) = 0$$

is the equation of the tangent hyper plane to the surface $F(x) = 0$ at point a.

Proof. Indeed, since the gradient at a is nonzero we may assume without loss of generality that $F_{x_n}(a) \neq 0$. By the IFT, $F(x_1, \ldots, x_{n-1}, x_n) = 0$ is locally solvable for $x_n = g(x_1, \ldots, x_{n-1})$ in a neighborhood of $a = (a_1, \ldots, a_n)$ with $g(\tilde{a}) = a_n$, where $\tilde{a} = (a_1, \ldots, a_{n-1})$ and $\tilde{x} = (x_1, \ldots, x_{n-1})$. Define the tangent hyperplane to be the graph of the linearization of g at $(a_1, \ldots, a_{n-1}, a_n)$. By Example 7.7 (a) the hyperplane to the graph of g at $\tilde{a}$ is given by

$$x_n = g(\tilde{a}) + \operatorname{grad} g(\tilde{a}) \cdot \tilde{x}. \tag{7.47}$$

Since $F(\tilde{a}, g(\tilde{a})) = 0$, by the implicit function theorem

$$\frac{\partial g(\tilde{a})}{\partial x_j} = -\frac{F_{x_j}(a)}{F_{x_n}(a)}, \quad j = 1, \ldots, n-1.$$

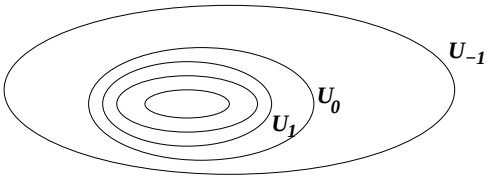
Inserting this into (7.47) we have

$$x_n - a_n = -\frac{1}{F_{x_n}(a)} \sum_{j=1}^{n-1} F_{x_j}(a)(x_j - a_j).$$

Multiplication by $-F_{x_n}(a)$ gives

$$-F_{x_n}(a)(x_n - a_n) = \sum_{j=1}^{n-1} F_{x_j}(a)(x_j - a_j) \Longrightarrow 0 = \operatorname{grad} F(a) \cdot (x - a).$$

■



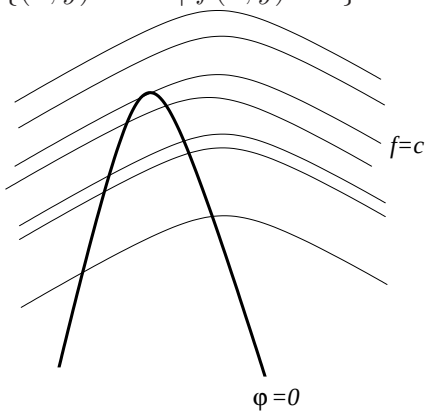
Let $f\colon U \to \mathbb{R}$ be differentiable. For $c \in \mathbb{R}$ define the *level set* $U_c = \{x \in U \mid f(x) = c\}$. The set U_c may be empty, may consist of a single point (in case of local extrema) or, in the "generic" case, that is if $\operatorname{grad} F(a) \neq 0$ and U_c is non-empty, U_c it is a $(n-1)$-dimensional hyper surface. $\{U_c \mid c \in \mathbb{R}\}$ is family of non-intersecting subsets of U which cover U.

7.7 Lagrange Multiplier Rule

This is a method to find local extrema of a function under certain constraints.
Consider the following problem: Find local extremma of a function $f(x, y)$ of two variables where x and y are not independent from each other but satisfy the constraint

$$\varphi(x, y) = 0.$$

Suppose further that f and φ are continuously differentiable. Note that the level sets $U_c = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}$ form a family of non-intersecting curves in the plane.



We have to find the curve $f(x, y) = c$ intersecting the constraint curve $\varphi(x, y) = 0$ where c is as large or as small as possible. Usually $f = c$ intersects $\varphi = 0$ if c monotonically changes. However if c is maximal, the curve $f = c$ *touches* the graph $\varphi = 0$. In other words, the tangent lines coincide. This means that the defining normal vectors to the tangent lines are scalar multiples of each other.

Theorem 7.20 (Lagrange Multiplier Rule) *Let $f, \varphi\colon U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ is open, be continuously differentiable and f has a local extrema at $a \in U$ under the constraint $\varphi(x) = 0$. Suppose that $\operatorname{grad} \varphi(a) \neq 0$.*
Then there exists a real number λ such that

$$\operatorname{grad} f(a) = \lambda \operatorname{grad} \varphi(a).$$

This number λ is called *Lagrange multiplier*.
Proof. The idea is to solve the constraint $\varphi(x) = 0$ for one variable and to consider the "free" extremum problem with one variable less. Suppose without loss of generality that $\varphi_{x_n}(a) \neq 0$. By the implicit function theorm we can solve $\varphi(x) = 0$ for $x_n = g(x_1, \ldots, x_{n-1})$ in a neighborhood of $x = a$. Differentiating $\varphi(\tilde{x}, g(\tilde{x})) = 0$ and inserting $a = (\tilde{a}, a_n)$ as before we have

$$\varphi_{x_j}(a) + \varphi_{x_n}(a) g_{x_j}(\tilde{a}) = 0, \quad j = 1, \ldots, n-1. \tag{7.48}$$

Since $h(\tilde{x}) = f(\tilde{x}, g(\tilde{x}))$ has a local extremum at $\tilde{a}$ all partial derivatives of h vanish at $\tilde{a}$:

$$f_{x_j}(a) + f_{x_n}(a) g_{x_j}(\tilde{a}) = 0, \quad j = 1, \dots, n-1. \tag{7.49}$$

Setting $\lambda = f_{x_n}(a)/\varphi_{x_n}(a)$ and comparing (7.48) and (7.49) we find

$$f_{x_j}(a) = \lambda \varphi_{x_j}(a), \quad j = 1, \dots, n-1.$$

Since by definition, $f_{x_n}(a) = \lambda \varphi_{x_n}(a)$ we finally obtain $\operatorname{grad} f(a) = \lambda \operatorname{grad} \varphi(a)$ which completes the proof. ∎

Example 7.17 (a) Let $A = (a_{ij})$ be a real symmetric $n \times n$-matrix, and define $f(x) = x \cdot Ax = \sum_{i,j} a_{ij} x_i x_j$. We aks for the local extrema of f on the unit sphere $\mathrm{S}^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$. This constraint can be written as $\varphi(x) = \|x\|^2 - 1 = \sum_{i=1}^n x_i^2 - 1 = 0$. Suppose that f attains a local minimum at $a \in \mathrm{S}^{n-1}$. By Example 7.6 (b)

$$\operatorname{grad} f(a) = 2A(a).$$

On the other hand

$$\operatorname{grad} \varphi(a) = (2x_1, \dots, 2x_n)|_{x=a} = 2a.$$

By Theorem 7.20 there exists a real number λ_1 such that

$$\operatorname{grad} f(a) = 2A(a) = \lambda_1 \operatorname{grad} \varphi(a) = 2a,$$

Hence $A(a) = \lambda_1 a$; that is, λ is an eigenvalue of A and a the corresponding eigenvector. In particular, A has a real eigenvalue. Since S^{n-1} has no boundary, the global minimum is also a local one. We find: if $f(a) = a \cdot A(a) = a \cdot \lambda a = \lambda$ is the global minimum, λ is the smallest eigenvalue.
(b) Let a be the point of a hypersurface $M = \{x \mid \varphi(x) = 0\}$ with minimal distance to a given point $b \notin M$. Then the line through a and b is orthogonal to M.
Indeed, the function $f(x) = \|x - b\|^2$ attains its minimum under the condition $\varphi(x) = 0$ at a. By the Theorem, there is a real number λ such that

$$\operatorname{grad} f(a) = 2(a - b) = \lambda \operatorname{grad} \varphi(a).$$

The assertion follows since by Example 7.16 (c), $\operatorname{grad} \varphi(a)$ is orthogonal to M at a and $b - a$ is a multiple of the normal vector $\nabla \varphi(a)$.

Theorem 7.21 (Lagrange Multiplier Rule — extended version) *Let $f, \varphi_i \colon U \to \mathbb{R}$, $i = 1, \dots, m$, $m < n$, be continuously differentiable functions. Let $M = \{x \in U \mid \varphi_1(x) = \cdots = \varphi_m(x) = 0\}$ and suppose that $f(x)$ has a local extrema at a under the constraints $x \in M$. Suppose further that the Jacobi matrix $\varphi'(a) \in \mathbb{R}^{m \times n}$ has maximal rank m.*
Then there exist real numbers $\lambda_1, \dots, \lambda_m$ such that

$$\operatorname{grad} f(a) = \operatorname{grad} (\lambda_1 \varphi_1 + \cdots + \lambda_m \varphi_m)(a) = 0.$$

Note that the rank condition ensures that there is a choice of m variables out of $x_1, \dots, x_n$ such that the Jacobian of $\varphi_1, \dots, \varphi_m$ with respect to this set of variable is nonzero at a.

7.8 Integrals depending on Parameters

Problem: Define $I(y) = \int_a^b f(x,y)\,\mathrm{d}x$; what are the relations between properties of $f(x,y)$ and of $I(y)$ for example with respect to continuity and differentiability.

7.8.1 Continuity of $I(y)$

Proposition 7.22 *Let $f(x,y)$ be continuous on the rectangle $R = [a,b] \times [c,d]$.*
Then $I(y) = \int_a^b f(x,y)\,\mathrm{d}x$ is continuous on $[c,d]$.

Proof. Let $\varepsilon > 0$. Since f is continuous on the compact set R, f is uniformly continuous on R (see Proposition 6.25). Hence, there is a $\delta > 0$ such that $|\,x - x'\,| < \delta$ and $|\,y - y'\,| < \delta$ and $(x,y),(x',y') \in R$ imply

$$|\,f(x,y) - f(x',y')\,| < \varepsilon.$$

Therefore, $|\,y - y_0\,| < \delta$ and $y, y_0 \in [c,d]$ imply

$$|\,I(y) - I(y_0)\,| = \left| \int_a^b (f(x,y) - f(x,y_0))\,\mathrm{d}x \right| \le \varepsilon(b-a).$$

This shows continuity of $I(y)$ at y_0. ∎

For example, $I(y) = \int_0^1 \arctan \frac{x}{y}\,\mathrm{d}x$ is continuous for $y > 0$.

Remark 7.10 (a) Note that continuity at y_0 means that we can interchange the limit and the integral, $\lim\limits_{y\to y_0} \int_a^b f(x,y)\,\mathrm{d}x = \int_a^b \lim\limits_{y\to y_0} f(x,y)\,\mathrm{d}x = \int_a^b f(x,y_0)\,\mathrm{d}x$.
(b) A similar statement holds for $y \to \infty$: Suppose that $f(x,y)$ is continuous on $[a,b]\times[c,+\infty)$ and $\lim_{y\to+\infty} f(x,y) = \varphi(x)$ exists *uniformly* for all $x \in [a,b]$ that is

$$\forall\,\varepsilon > 0\ \ \exists\,R > 0\ \ \forall\,x \in [a,b], y \ge R : |\,f(x,y) - \varphi(x)\,| < \varepsilon.$$

Then $\int_a^b \varphi(x)\,\mathrm{d}x$ exists and $\lim\limits_{y\to\infty} I(y) = \int_a^b \varphi(x)\,\mathrm{d}x$.

7.8.2 Differentiation of Integrals

Proposition 7.23 *Let $f(x,y)$ be defined on $R = [a,b]\times[c,d]$ and continuous as a function of x for every fixed y. Suppose that $f_y(x,y)$ exists for all $(x,y) \in R$ and is continuous as a function of the two variables x and y.*
Then $I(y)$ is differentiable and

$$I'(y) = \frac{\mathrm{d}}{\mathrm{d}y} \int_a^b f(x,y)\,\mathrm{d}x = \int_a^b f_y(x,y)\,\mathrm{d}x.$$

Proof. Let $\varepsilon > 0$. Since $f_y(x, y)$ is continuous, it is uniformly continuous on R. Hence there exists $\delta > 0$ such that $|\, x' - x'' \,| < \delta$ and $|\, y' - y'' \,| < \delta$ imply $|\, f_y(x', y') - f_y(x'', y'') \,| < \varepsilon$. We have for $|\, h \,| < \delta$

$$\left| \frac{I(y_0 + h) - I(y_0)}{h} - \int_a^b f_y(x, y_0)\,\mathrm{d}x \right| \le \left| \int_a^b \left(\frac{f(x, y_0 + h) - f(x, y_0)}{h} - f_y(x, y_0) \right) \mathrm{d}x \right|$$
$$\underset{\text{Mean value theorem}}{\le} \int_a^b |\, f_y(x, y_0 + \theta h) - f_y(x, y_0) \,|\, \mathrm{d}x < \varepsilon(b - a)$$

for some $\theta \in (0, 1)$. Since this inequality holds for all small h, it holds for the limit as $h \to 0$, too. Thus,

$$\left| I'(y_0) - \int_a^b f_y(x, y_0)\,\mathrm{d}x \right| \le \varepsilon(b - a).$$

Since ε was arbitrary, the claim follows. ∎

In case of variable integration limits we have the following theorem.

Proposition 7.24 *Let $f(x, y)$ be as in Proposition 7.23. Let $\alpha(y)$ and $\beta(y)$ be differentiable on $[c, d]$, and suppose that $\alpha([c, d])$ and $\beta([c, d])$ are contained in $[a, b]$.*
Let $I(y) = \int_{\alpha(y)}^{\beta(y)} f(x, y)\,\mathrm{d}x$. Then $I(y)$ is differentiable and

$$I'(y) = \int_{\alpha(y)}^{\beta(y)} f_y(x, y)\,\mathrm{d}x \; + \beta'(y) f(\beta(y), y) \; - \alpha'(y) f(\alpha(y), y). \tag{7.50}$$

Proof. Let $F(y, u, v) = \int_u^v f(x, y)\,\mathrm{d}x$; then $I(y) = F(y, \alpha(y), \beta(y))$. The fundamental theorem of calculus yields

$$\begin{aligned} \frac{\partial F}{\partial v}(y, u, v) &= \frac{\partial}{\partial v} \int_u^v f(x, y)\,\mathrm{d}x = f(v, y), \\ \frac{\partial F}{\partial u}(y, u, v) &= \frac{\partial}{\partial u} \left(- \int_v^u f(x, y)\,\mathrm{d}x \right) = -f(u, y). \end{aligned} \tag{7.51}$$

By the chain rule, the previous proposition and (7.51) we have

$$\begin{aligned} I'(y) &= \frac{\partial F}{\partial y} + \frac{\partial F}{\partial u}\,\alpha'(y) + \frac{\partial F}{\partial v}\,\beta'(y) \\ &= \frac{\partial F}{\partial y}(y, \alpha(y), \beta(y)) + \frac{\partial F}{\partial u}(y, \alpha(y), \beta(y))\,\alpha'(y) + \frac{\partial F}{\partial v}(y, \alpha(y), \beta(y))\,\beta'(y) \\ &= \int_{\alpha(y)}^{\beta(y)} f_y(x, y)\,\mathrm{d}x + \alpha'(y)(-f(\alpha(y), y)) + \beta'(y) f(\beta(y), y). \end{aligned}$$

∎

Example 7.18 (a) $I(y) = \int_3^4 \frac{\sin(xy)}{x}\,\mathrm{d}x$ is differentiable by Proposition 7.23 since $f_y(x,y) = \frac{\cos(xy)}{x}\,x = \cos(xy)$ is continuous. Hence

$$I'(y) = \int_3^4 \cos(xy)\,\mathrm{d}x = \frac{\sin(xy)}{y}\bigg|_3^4 = \frac{\sin 4y}{y} - \frac{\sin 3y}{y}.$$

(b) $I(y) = \int_{\log y}^{\sin y} \mathrm{e}^{x^2 y}\,\mathrm{d}x$ is differentiable with

$$I'(y) = \int_{\log y}^{\sin y} x^2 \mathrm{e}^{x^2 y}\,\mathrm{d}x + \cos y \mathrm{e}^{y \sin^2 y} - \frac{1}{y}\mathrm{e}^{y(\log y)^2}.$$

7.8.3 Improper Integrals with Parameters

Suppose that the improper integral $\int_a^\infty f(x,y)\,\mathrm{d}x$ exists for $y \in [c,d]$.

Definition 7.10 We say that the improper integral $\int_a^\infty f(x,y)\,\mathrm{d}x$ *converges uniformly* with respect to y on $[c,d]$ if for every $\varepsilon > 0$ there is an $A_0 > 0$ such that $A > A_0$ implies

$$\left| I(y) - \int_a^A f(x,y)\,\mathrm{d}x \right| \equiv \left| \int_A^\infty f(x,y)\,\mathrm{d}x \right| < \varepsilon$$

for all $y \in [c,d]$.

Note that the Cauchy and Weierstraß criteria (see Proposition 6.1 and Theorem 6.2) for uniform convergence of series of functions also hold for improper parametric integrals. For example the theorem of Weierstraß now reads as follows.

Proposition 7.25 *Suppose that $\int_a^A f(x,y)\,\mathrm{d}x$ exists for all $A \geq a$ and $y \in [c,d]$. Suppose further that $|\, f(x,y)\,| \leq \varphi(x)$ for all $x \geq a$ and $\int_a^\infty \varphi(x)\,\mathrm{d}x$ converges.*
Then $\int_a^\infty f(x,y)\,\mathrm{d}x$ converges uniformly with respect to $y \in [c,d]$.

Example 7.19 $I(y) = \int_1^\infty \mathrm{e}^{-xy} x^y y^2\,\mathrm{d}x$ converges uniformly on $[2,4]$ since

$$|\, f(x,y)\,| = \left|\, \mathrm{e}^{-xy} x^y y^2 \,\right| \leq \mathrm{e}^{-2x} x^4 4^2 = \varphi(x).$$

and $\int_1^\infty \mathrm{e}^{-2x} x^4 4^2\,\mathrm{d}x < \infty$ converges.

If we add the assumption of *uniform convergence* then the preceding theorems remain true for improper integrals.

Proposition 7.26 *Let $f(x,y)$ be continuous on $\{(x,y) \in \mathbb{R}^2 \mid a \leq x < \infty,\, c \leq y \leq d\}$. Suppose that $I(y) = \int_a^\infty f(x,y)\,\mathrm{d}x$ converges uniformly with respect to $y \in [c,d]$.*
Then $I(y)$ is continuous on $[c,d]$.

Proof. This proof was not carried out in the lecture. Let $\varepsilon > 0$. Since the improper integral converges uniformly, there exists $A_0 > 0$ such that for all $A \geq A_0$ we have

$$\left| \int_A^\infty f(x,y)\,\mathrm{d}x \right| < \varepsilon$$

for all $y \in [c,d]$. Let $A \geq A_0$ be fixed. On $\{(x,y) \in \mathbb{R}^2 \mid a \leq x \leq A,\ c \leq y \leq d\}$ $f(x,y)$ is uniformly continuous; hence there is a $\delta > 0$ such that $|\, x' - x'' \,| < \delta$ and $|\, y' - y'' \,| < \delta$ implies

$$|\, f(x',y') - f(x'',y'') \,| < \frac{\varepsilon}{A-a}.$$

Therefore,

$$\int_a^A |\, f(x,y) - f(x,y_0) \,|\, \mathrm{d}x < \frac{\varepsilon}{A-a}(A-a) = \varepsilon, \quad \text{for} \quad |\, y - y_0 \,| < \delta.$$

Finally,

$$|\, I(y) - I(y_0) \,| = \left(\int_a^A + \int_A^\infty \right) |\, f(x,y) - f(x,y_0) \,| \leq 2\varepsilon \quad \text{for} \quad |\, y - y_0 \,| < \delta.$$

■

We skip the proof of the following proposition.

Proposition 7.27 *Let $f_y(x,y)$ be continuous on $\{(x,y) \mid a \leq x < \infty,\ c \leq y \leq d\}$, $f(x,y)$ continuous with respect to x for all fixed $y \in [c,d]$.*
Suppose that for all $y \in [c,d]$ the integral $I(y) = \int_a^\infty f(x,y)\,\mathrm{d}x$ exists and the integral $\int_a^\infty f_y(x,y)\,\mathrm{d}x$ converges uniformly with respect to $y \in [c,d]$.
Then $I(y)$ is differentiable and $I'(y) = \int_a^\infty f_y(x,y)\,\mathrm{d}x$.

Combinig the results of the last Proposition and Proposition 7.25 we get the following corollary.

Corollary 7.28 *Let $f_y(x,y)$ be continuous on $\{(x,y) \mid a \leq x < \infty,\ c \leq y \leq d\}$, $f(x,y)$ continuous with respect to x for all fixed $y \in [c,d]$.*
Suppose that

(a) *for all $y \in [c,d]$ the integral $I(y) = \int_a^\infty f(x,y)\,\mathrm{d}x$ exists,*

(b) $|\, f_y(x,y) \,| \leq \varphi(x)$ *for all $x \geq a$ and all y*

(c) $\int_a^\infty \varphi(x)\,\mathrm{d}x$ *exists.*

Then $I(y)$ is differentiable and $I'(y) = \int_a^\infty f_y(x,y)\,\mathrm{d}x$.

Example 7.20 (a) $I(y) = \int_0^\infty \mathrm{e}^{-x^2}\cos(2yx)\,\mathrm{d}x$. $f(x,y) = \mathrm{e}^{-x^2}\cos(2yx)$, $f_y(x,y) = -2x\sin(2yx)\,\mathrm{e}^{-x^2}$ converges uniformly with respect to y since

$$|\, f_y(x,y) \,| \leq 2x\mathrm{e}^{-x^2} \leq K\mathrm{e}^{-x^2/2}.$$

Hence,

$$I'(y) = -\int_0^\infty 2x\sin(2yx)\,\mathrm{e}^{-x^2}\,\mathrm{d}x.$$

Integration by parts with $u = \sin(2yx)$, $v' = -\mathrm{e}^{-x^2}2x$ gives $u' = 2y\cos(2yx)$, $v = \mathrm{e}^{-x^2}$ and

$$\int_0^A -\mathrm{e}^{-x^2}2x\,\sin(2yx)\,\mathrm{d}x = \sin(2yA)\,\mathrm{e}^{-A^2} - \int_0^A 2y\cos(2yx)\,\mathrm{e}^{-x^2}\,\mathrm{d}x.$$

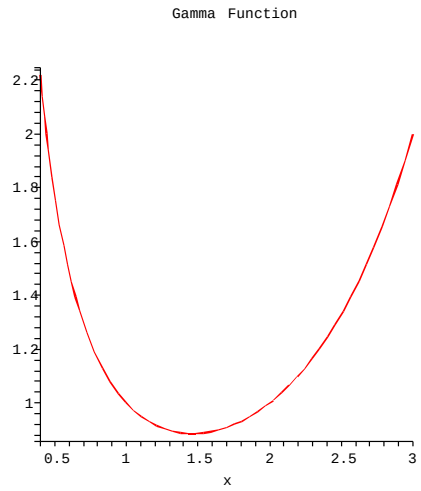
As $A \to \infty$ the first summand on the right tends to 0; thus $I(y)$ satisfies the ordinary differential equation

$$I'(y) = -2yI(y).$$

ODE: $y' = -2xy$; $\mathrm{d}y = -2xy\,\mathrm{d}x$; $\mathrm{d}y/y = -2x\,\mathrm{d}x$. Integration yields $\log y = -x^2 + c$; $y = c'\mathrm{e}^{-x^2}$.

The general solution is $I(y) = C\mathrm{e}^{-y^2}$. We determine the constant C. Insert $y = 0$. Since $I(0) = \int_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}/2$, we find

$$I(y) = \frac{\sqrt{\pi}}{2}\mathrm{e}^{-y^2}.$$



(b) The Gamma function $\Gamma(x) = \int_0^\infty t^{x-1}\mathrm{e}^{-t}\,\mathrm{d}t$ is in $\mathrm{C}^\infty(\mathbb{R}_+)$. Let $x > 0$, say $x \in [c, d]$. Recall from Subsection 5.3.3 the definition and the proof of the convergence of the improper integrals $\Gamma_1 = \int_0^1 f(x,t)\,\mathrm{d}t$ and $\Gamma_2 = \int_1^\infty f(x,t)\,\mathrm{d}t$, where $f(x,t) = t^{x-1}\mathrm{e}^{-t}$. Note that $\Gamma_1(x)$ is an improper integral at $t = 0 + 0$. By L'Hospital's rule $\lim\limits_{t\to 0+0} t^\alpha \log t = 0$ for all $\alpha > 0$. In particular, $|\log t| < t^{-c/2}$ if $0 < t < t_0 < 1$.

Since $\mathrm{e}^{-t} < 1$ and moreover $t^{x-1} < t^{c-1}$ for $t < t_0$ by Lemma 1.23 (b) we conclude that

$$\left|\frac{\partial}{\partial x} f(x,t)\right| = \left|\, t^{x-1}\log t\mathrm{e}^{-t}\,\right| \le |\log t|\; t^{c-1} \le t^{-\frac{c}{2}}\, t^{c-1} = \frac{1}{t^{1-\frac{c}{2}}},$$

for $0 < t < t_0$. Since $\varphi(t) = \frac{1}{t^{1-\frac{c}{2}}}$ is integrable over $[0,1]$, $\Gamma_1(x)$ is differentiable by the Corrollary with $\Gamma_1'(x) = \int_0^1 t^{x-1}\,\log t\mathrm{e}^{-t}\,\mathrm{d}t$. Similarly, $\Gamma_2(x)$ is an improper integral over an unbounded interval $[1, +\infty)$, for sufficiently large $t \ge t_0 > 1$, we have $\log t < t$ and $t^x < t^d$, such that

$$\frac{\partial}{\partial x} f(x,t) = t^{x-1}\log t\mathrm{e}^{-t} \le t^x\,\mathrm{e}^{-t} \le t^d\mathrm{e}^{-t} \le t^d\mathrm{e}^{-t/2}\,\mathrm{e}^{-t/2} \le M\mathrm{e}^{-t/2}.$$

Since $t^d\mathrm{e}^{-t/2}$ tends to 0 as $t \to \infty$, it is bounded by some constant M and $\mathrm{e}^{-t/2}$ is integrable on $[1, +\infty)$ such that $\Gamma_2(x)$ is differentiable with

$$\Gamma_2'(x) = \int_1^\infty t^{x-1}\log t\mathrm{e}^{-t}\,\mathrm{d}t.$$

Consequently, $\Gamma(x)$ is differentiable for all $x > 0$ with

$$\Gamma'(x) = \int_0^\infty t^{x-1}\,\log t\,\mathrm{e}^{-t}\,\mathrm{d}t.$$

Similarly one can show that $\Gamma \in \mathrm{C}^\infty(\mathbb{R}_{>0})$ with

$$\Gamma^{(k)}(x) = \int_0^\infty t^{x-1}\,(\log t)^k\,\mathrm{e}^{-t}\,\mathrm{d}t.$$

7.9 Appendix

Proof of Proposition 7.7. Let $A = (A_{ij}) = \left(\dfrac{\partial f_i}{\partial x_j}\right)$ be the matrix of partial derivatives considered as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Our aim is to show that

$$\lim_{h\to 0} \frac{\|f(a+h) - f(a) - A\,h\|}{\|h\|} = 0.$$

For, it suffices to prove the convergence to 0 for each coordinate $i = 1, \ldots, m$ by Proposition 6.26

$$\lim_{h\to 0} \frac{f_i(a+h) - f_i(a) - \sum_{j=1}^n A_{ij}h_j}{\|h\|} = 0. \tag{7.52}$$

Without loss of generality we assume $m = 1$ and $f = f_1$. For simplicity, let $n = 2$, $f = f(x, y)$, $a = (a, b)$, and $h = (h, k)$. Note first that by the mean value theorem we have

$$\begin{aligned} f(a+h, b+k) - f(a, b) &= f(a+h, b+k) - f(a, b+k) + f(a, b+k) - f(a, b) \\ &= \frac{\partial f}{\partial x}(\xi, b+k)h + \frac{\partial f}{\partial y}(a, \eta)k, \end{aligned}$$

where $\xi \in (a, a+h)$ and $\eta \in (b, b+k)$. Using this, the expression in(7.52) reads

$$\begin{aligned} &\frac{f(a+h, b+k) - f(a, b) - \frac{\partial f}{\partial x}(a, b)\,h - \frac{\partial f}{\partial y}(a, b)\,k}{\sqrt{h^2+k^2}} = \\ &= \frac{\left(\frac{\partial f}{\partial x}(\xi, b+k) - \frac{\partial f}{\partial x}(a, b)\right)\,h + \left(\frac{\partial f}{\partial y}(a, \eta) - \frac{\partial f}{\partial x}(a, b)\right)\,k}{\sqrt{h^2+k^2}}. \end{aligned}$$

Since both $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are continuous at (a, b), given $\varepsilon > 0$ we find $\delta > 0$ such that $(x, y) \in U_\delta((a, b))$ implies

$$\left|\frac{\partial f}{\partial x}(x, y) - \frac{\partial f}{\partial x}(a, b)\right| < \varepsilon, \quad \left|\frac{\partial f}{\partial y}(x, y) - \frac{\partial f}{\partial y}(a, b)\right| < \varepsilon.$$

This shows

$$\left\|\frac{\left(\frac{\partial f}{\partial x}(\xi, b+k) - \frac{\partial f}{\partial x}(a, b)\right)\,h + \left(\frac{\partial f}{\partial y}(a, \eta) - \frac{\partial f}{\partial x}(a, b)\right)\,k}{\sqrt{h^2+k^2}}\right\| \leq \frac{\varepsilon\,|\,h\,| + \varepsilon\,|\,k\,|}{\sqrt{h^2+k^2}} \leq 2\varepsilon,$$

hence f is differentiable at (a, b) with Jacobi matrix $A = (\frac{\partial f}{\partial x}(a, b)\ \frac{\partial f}{\partial y}(a, b))$.
Since both components of A—the partial derivatives—are continuous functions of (x, y), the assignment $x \mapsto f'(x)$ is continuous by Proposition 6.26. ∎

Chapter 8

Curves and Line Integrals

8.1 Rectifiable Curves

8.1.1 Curves in $\mathbb{R}^k$

We consider curves in $\mathbb{R}^k$. We define the tangent vector, regular points, angle of intersection.

Definition 8.1 A *curve* in $\mathbb{R}^k$ is a continuous mapping $\gamma\colon I \to \mathbb{R}^k$, where $I \subset \mathbb{R}$ is a closed interval consisting of more than one point.

The interval can be $I = [a, b]$, $I = [a, +\infty)$, or $I = \mathbb{R}$. In the first case $\gamma(a)$ and $\gamma(b)$ are called the *initial* and *end point* of γ. These two points derfine a natural *orientation* of the curve "from $\gamma(a)$ to $\gamma(b)$". Replacing $\gamma(t)$ by $\gamma(a+b-t)$ we obtain the curve from $\gamma(b)$ to $\gamma(a)$ with opposite orientation.
If $\gamma(a) = \gamma(b)$, γ is said to be a *closed curve*. The curve γ is given by a k-tupel $\gamma = (\gamma_1, \ldots, \gamma_k)$ of continuous real-valued functions. If γ is differentiable, the curve is said to be *differentiable*. Note that we have defined the curve to be *a mapping*, not a set of points in $\mathbb{R}^k$. Of course, with each curve γ in $\mathbb{R}^k$ there is associated a subset of $\mathbb{R}^k$, namely the image of γ,

$$C = \gamma(I) = \{\gamma(t) \in \mathbb{R}^k \mid t \in I\}.$$

but different curves γ may have the same image $C = \gamma(I)$. The curve is said to be *simple* if γ is injective on the inner points I° of I. A simple curve has no self-intersection.

Example 8.1 (a) A *circle* in $\mathbb{R}^2$ of radius $r > 0$ with center $(0, 0)$ is described by the curve

$$\gamma\colon [0, 2\pi] \to \mathbb{R}^2, \quad \gamma(t) = (r\cos t, r\sin t).$$

Note that $\tilde{\gamma}\colon [0, 4\pi] \to \mathbb{R}^2$ with $\tilde{\gamma}(t) = \gamma(t)$ has the same image but is different from γ. γ is a simple curve, $\tilde{\gamma}$ is not.
(b) Let $p, q \in \mathbb{R}^k$ be fixed points, $p \neq q$. Then

$$\gamma_1(t) = (1-t)p + tq, \quad t \in [0, 1],$$
$$\gamma_2(t) = (1-t)p + tq, \quad t \in \mathbb{R},$$

are the *segment* $\overline{pq}$ from p to q and the *line* pq *through* p *and* q, respectively. If $v \in \mathbb{R}^k$ is a vector, then $\gamma_3(t) = p + tv$, $t \in \mathbb{R}$, is the line through p with direction v.
(c) If $f\colon [a,b] \to \mathbb{R}$ is a continuous function, the graph of f is a curve in $\mathbb{R}^2$:

$$\gamma\colon [a,b] \to \mathbb{R}^2, \quad \gamma(t) = (t, f(t)).$$

(d) **Implicit Curves.** Let $F\colon U \subset \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable, $F(a,b) = 0$, and $\nabla F(a,b) \neq 0$ for some point $(a,b) \in U$. By the Implicit function theorem, $F(x,y) = 0$ can locally be solved for $y = g(x)$ or $x = f(y)$. In both cases $\gamma(t) = (t, g(t))$ and $\gamma(t) = (f(t), t)$ is a curve through (a,b). For example,

$$F(x,y) = y^2 - x^3 - x^2 = 0$$

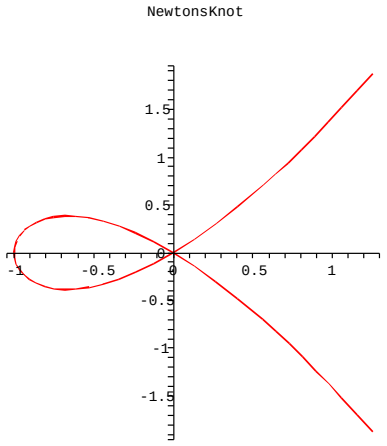
is locally solvable except for $(a,b) = (0,0)$. The corresponding curve is *Newton's knot*

Definition 8.2 (a) A simple curve $\gamma\colon I \to \mathbb{R}^k$ is said to be *regular* at t_0 if γ is continuously differentiable on I and $\gamma'(t_0) \neq 0$. γ is *regular* if it is regular at every point $t_0 \in I$.
(b) The vector $\gamma'(t_0)$ is called the *tangent vector*, $\alpha(t) = \gamma(t_0) + t\gamma'(t_0)$, $t \in \mathbb{R}$, is called the *tangent line* to the curve γ at point $\gamma(t_0)$.

Remark 8.1 The moving partice. Let t the time variable and $s(t)$ the coordinates of a point moving in $\mathbb{R}^k$. Then the tangent vector $v(t) = s'(t)$ is the *velocity vector* of the moving point. The *instantaneous velocity* is the euclidean norm of $v(t)$ $\|v(t)\| = \sqrt{s_1'(t)^2 + \cdots + s_k'(t)^2}$. The *acceleration vector* is the second derivative of $s(t)$, $a(t) = v'(t) = s''(t)$.

Let $\gamma_i\colon I_i \to \mathbb{R}^k$, $i = 1, 2$, be two regular curves with a common point $\gamma_1(t_1) = \gamma_2(t_2)$. The *angle of intersection* φ between the two curves γ_i at t_i is defined to be the angle between the two tangent lines $\gamma_1'(t_1)$ and $\gamma_2'(t_2)$. Hence,

$$\cos\varphi = \frac{\gamma_1'(t_1)\cdot\gamma_2'(t_2)}{\|\gamma_1'(t_1)\|\,\|\gamma_2'(t_2)\|}, \quad \varphi \in [0,\pi].$$



Example 8.2 (a) Newton's knot. The curve $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ given by $\gamma(t) = (t^2 - 1, t^3 - t)$ is not injective since $\gamma(-1) = \gamma(1) = (0,0) = x_0$. The point x_0 is a *double point* of the curve. In general γ has two different tangent lines at a double point. Since $\gamma'(t) = (2t, 3t^2 - 1)$ we have $\gamma'(-1) = (-2, 2)$ and $\gamma'(1) = (2, 2)$. The curve is regular since $\gamma'(t) \neq 0$ for all t.

Let us compute the angle of self-intersection. Since $\gamma(-1) = \gamma(1) = (0,0)$, the self-intersection angle φ satisfies

$$\cos\varphi = \frac{(-2,2)\cdot(2,2)}{8} = 0,$$

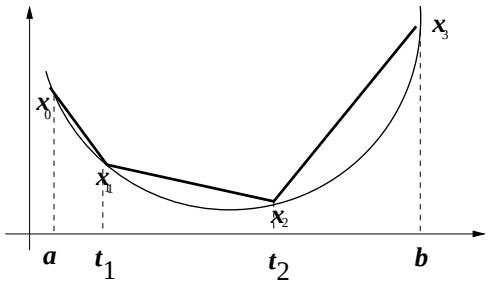
hence $\varphi = 90°$, the intersection is orthogonal.
(b) Neil's parabola. Let $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ be given by $\gamma(t) = (t^2, t^3)$. Since $\gamma'(t) = (2t, 3t^2)$, the origin is the only singular point.

8.1.2 Rectifiable Curves

The goal of this subsection is to define the *length of a curve*. For differentiable curves, there is a formula using the tangent vector. However, the "lenght of a curve" makes sense for some non-differentiable, continuous curves.
Let $\gamma\colon [a,b] \to \mathbb{R}^k$ be a curve. We associate to each partition $P = \{t_0, \dots, t_n\}$ of $[a,b]$ the points $x_i = \gamma(t_i)$, $i = 0, \dots, n$, and the number

$$\ell(P,\gamma) = \sum_{i=1}^{n} \|\gamma(t_i) - \gamma(t_{i-1})\|\,. \tag{8.1}$$

The ith term in this sum is the euclidean distance of the points $x_{i-1} = \gamma(t_{i-1})$ and $x_i = \gamma(t_i)$.



Hence $\ell(P,\gamma)$ is the length of the polygonal path with vertices $x_0, \dots, x_n$. As our partition becomes finer and finer, this polygon approaches the image of γ more and more closely.

Definition 8.3 A curve $\gamma\colon [a,b] \to \mathbb{R}^k$ is said to be *rectifiable* if the set of non-negative real numbers $\{\ell(P,\gamma) \mid P \quad \text{is a partition of } [a,b]\}$ is bounded. In this case

$$\ell(\gamma) = \sup \ell(P,\gamma),$$

where the supremum is taken over all partitions P of $[a,b]$, is called the *length* of γ.

In certain cases, $\ell(\gamma)$ is given by a Riemann integral. We shall prove this for *continuously differentiable* curves, i. e. for curves γ whose derivative γ' is continuous.

Proposition 8.1 *If γ' is continuous on $[a,b]$, then γ is rectifiable, and*

$$\ell(\gamma) = \int_a^b \|\gamma'(t)\|\ \mathrm{d}t.$$

Proof. If $a \le t_{i-1} < t_i \le b$, by Theorem 5.28, $\gamma(t_i) - \gamma(t_{i-1}) = \int_{t_{i-1}}^{t_i} \gamma'(t)\,\mathrm{d}t$. Applying Proposition 5.29 we have

$$\|\gamma(t_i) - \gamma(t_{i-1})\| = \left\| \int_{t_{i-1}}^{t_i} \gamma'(t)\,\mathrm{d}t \right\| \le \int_{t_{i-1}}^{t_i} \|\gamma'(t)\|\ \mathrm{d}t.$$

Hence

$$\ell(P,\gamma) \le \int_a^b \|\gamma'(t)\|\ \mathrm{d}t$$

for every partition P of $[a,b]$. Consequently,

$$\ell(\gamma) \le \int_a^b \|\gamma'(t)\| \, \mathrm{d}t.$$

To prove the opposite inequality, let $\varepsilon > 0$ be given. Since γ' is uniformly continuous on $[a,b]$, there exists $\delta > 0$ such that

$$\|\gamma'(s) - \gamma'(t)\| < \varepsilon \quad \text{if} \quad |\, s - t \,| < \delta.$$

Let P be a partition with $\Delta t_i \le \delta$ for all i. If $t_{i-1} \le t \le t_i$ it follows that

$$\|\gamma'(t)\| \le \|\gamma'(t_i)\| + \varepsilon.$$

Hence

$$\begin{aligned}
\int_{t_{i-1}}^{t_i} \|\gamma'(t)\| \, \mathrm{d}t &\le \|\gamma'(t_i)\| \, \Delta t_i + \varepsilon \Delta t_i \\
&= \left\| \int_{t_{i-1}}^{t_i} (\gamma'(t) - \gamma'(t_i) - \gamma'(t)) \, \mathrm{d}t \right\| + \varepsilon \Delta t_i \\
&\le \left\| \int_{t_{i-1}}^{t_i} \gamma'(t) \, \mathrm{d}t \right\| + \left\| \int_{t_{i-1}}^{t_i} (\gamma'(t_i) - \gamma'(t)) \, \mathrm{d}t \right\| + \varepsilon \Delta t_i \\
&\le \|\gamma'(t_i) - \gamma'(t_{i-1})\| + 2\varepsilon \Delta t_i.
\end{aligned}$$

If we add these inequalities, we obtain

$$\int_a^b \|\gamma'(t)\| \, \mathrm{d}t \le \ell(P,\gamma) + 2\varepsilon(b-a) \le \ell(\gamma) + 2\varepsilon(b-a).$$

Since ε was arbitrary,

$$\int_a^b \|\gamma'(t) \, \mathrm{d}t\| \le \ell(\gamma).$$

This completes the proof. ∎

Special Case $k = 2$

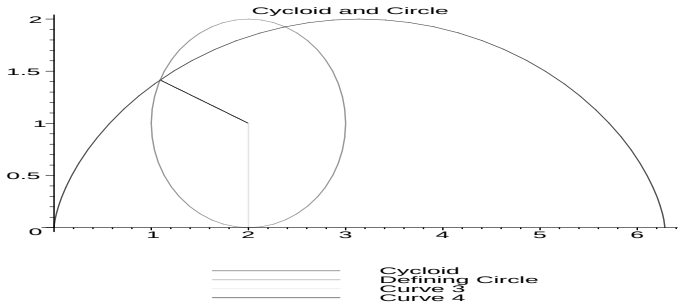
$k = 2$, $\gamma(t) = (x(t), y(t))$, $t \in [a,b]$. Then

$$\ell(\gamma) = \int_a^b \sqrt{x'(t)^2 + y'(t)^2} \, \mathrm{d}t.$$

In particular, let $\gamma(t) = (t, f(t))$ be the graph of a continuously differentiable function $f\colon [a,b] \to \mathbb{R}$. Then $\ell(\gamma) = \int_a^b \sqrt{1 + (f'(t))^2} \, \mathrm{d}t$.

Example 8.3 Catenary Curve. Let $f(t) = a \cosh \frac{t}{a}$, $t \in [0,b]$, $b > 0$. Then $f'(t) = \sinh \frac{t}{a}$ and moreover

$$\ell(\gamma) = \int_0^b \sqrt{1 + \left(\sinh \frac{t}{a}\right)^2} \, \mathrm{d}t = \int_0^b \cosh \frac{t}{a} \, \mathrm{d}t = a \sinh \frac{t}{b} \Big|_0^b = a \sinh \frac{b}{a}.$$



(b) The position of a bulge in a bicycle tire as it rolls down the street can be parametrized by an angle θ as shown in the figure.

Let the radius of the tire be a. It can be verified by plane trigonometry that

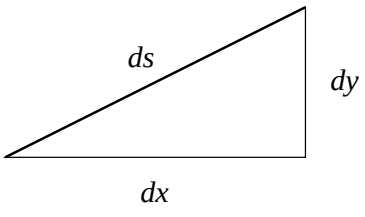
$$\gamma(\theta) = \begin{pmatrix} a(\theta - \sin\theta) \\ a(1 - \cos\theta) \end{pmatrix}.$$

This curve is called a *cycloid*.
Find the distance travelled by the bulge for $0 \leq \theta \leq 2\pi$.
Using $1 - \cos\theta = 2\sin^2\dfrac{\theta}{2}$ we have

$$\begin{aligned} \gamma'(\theta) &= a(1 - \cos\theta, \sin\theta) \\ \|\gamma'(\theta)\| &= a\sqrt{(1-\cos\theta)^2 + \sin^2\theta} = a\sqrt{2 - 2\cos\theta} \\ &= a\sqrt{2}\sqrt{1-\cos\theta} = 2a\sin\frac{\theta}{2}. \end{aligned}$$

Therefore,

$$\ell(\gamma) = 2a\int_0^{2\pi} \sin\frac{\theta}{2}\,\mathrm{d}\theta = -4a\left(\cos\frac{\theta}{2}\right)\Big|_0^{2\pi} = 4a(-\cos\pi + \cos 0) = 8a.$$



(c) *The arc element* $\mathrm{d}s$. Formally the arc element of a plane differentiable curve can be computed using the pythagorean theorem

$$\begin{aligned} (\,\mathrm{d}s)^2 = (\,\mathrm{d}x)^2 + (\,\mathrm{d}y)^2 &\Longrightarrow \mathrm{d}s = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} \\ \mathrm{d}s &= \mathrm{d}x\sqrt{1 + \frac{\mathrm{d}y^2}{\mathrm{d}x^2}} \\ \mathrm{d}s &= \sqrt{1 + (f'(x))^2}\,\mathrm{d}x. \end{aligned}$$

(d) **Arc of an Ellipse.** The ellipse with equation $x^2/a^2 + y^2/b^2 = 1$, $0 < b \leq a$, is parametrized by $\gamma(t) = (a\cos t, b\sin t)$, $t \in [0, t_0]$, such that $\gamma'(t) = (-a\sin t, b\cos t)$. Hence,

$$\begin{aligned} \ell(\gamma) &= \int_0^{t_0} \sqrt{a^2\sin^2 t + b^2\cos^2 t}\,\mathrm{d}t = \int_0^{t_0} \sqrt{a^2 - (a^2 - b^2)\cos^2 t}\,\mathrm{d}t \\ &= a\int_0^{t_0} \sqrt{1 - \varepsilon^2\cos^2 t}\,\mathrm{d}t, \end{aligned}$$

where $\varepsilon = \frac{\sqrt{a^2-b^2}}{a}$. This integral can be transformed into the function

$$E(\tau, \epsilon) = \int_0^{\tau} \sqrt{1 - \varepsilon^2\sin^2 t}\,\mathrm{d}t$$

is the elliptic integral of the second kind as defined in Chapter5.
(e) **A non-rectifiable curve.** Consider the graph $\gamma(t) = (t, f(t))$, $t \in [0,1]$, of the function f,

$$f(t) = \begin{cases} t\,\cos\frac{\pi}{2t}, & 0 < t \le 1, \\ 0, & t = 0. \end{cases}$$

Since $\lim_{t\to 0+0} f(t) = f(0) = 0$, f is continuous and $\gamma(t)$ is a curve. However, this curve is not rectifiable. Indeed, choose the partition $P_k = \{t_0 = 0, 1/(4k), 1/(4k-2), \ldots, 1/4, 1/2, t_{2k+1} = 1\}$ consisting of $2k+1$ points, $t_i = \frac{1}{4k-2i+2}$, $i = 1, \ldots, 2k$. Note that $t_0 = 0$ and $t_{2k+1} = 1$ play a special role and will be omitted in the calculations below. Then $\left(\cos\frac{\pi}{2t_i}\right) = (1, -1, 1, -1, \ldots, -1, 1)$, $i = 1, \ldots, 2k$. Thus

$$\begin{aligned} \ell(P_k, \gamma) &\ge \sum_{i=2}^{2k} \sqrt{(t_i - t_{i-1})^2 + (f(t_i) - f(t_{i-1}))^2} \ge \sum_{i=2}^{2k} |\, f(t_i) - f(t_{i-1})\,| \\ &\ge \left(\frac{1}{4k-2} + \frac{1}{4k}\right) + \left(\frac{1}{4k-4} + \frac{1}{4k-2}\right) + \cdots + \left(\frac{1}{2} + \frac{1}{4}\right) \\ &= \frac{1}{2} + 2\left(\frac{1}{4} + \cdots + \frac{1}{4k-2}\right) + \frac{1}{4k} \end{aligned}$$

which is unbounded for $k \to \infty$ since the harmonic series is unbounded. Hence γ is not rectifiable.

8.2 Line Integrals

A lot of physical applications are to be found in [MW85, Chapter 18]. Integration of vector fields along curves is of fundamental importance in both mathematics and physics. We use the concept of *work* to motivate the material in this section.
The motion of an object is described by a parametric curve $\vec{x} = \vec{x}(t) = (x(t), y(t), z(t))$. By differentiating this function, we obtain the velocity $\vec{v}(t) = \vec{x}'(t)$ and the acceleration $\vec{a}(t) = \vec{x}''(t)$. We use the physicist notation $\dot{\vec{x}}(t)$ and $\ddot{\vec{x}}(t)$ to denote derivatives with respect to the time t.
According to Newton's law, the total force $\tilde{F}$ acting on an object of mass m is

$$\tilde{F} = m\vec{a}.$$

Since the kinetic energy K is defined by $K = \frac{1}{2}m\vec{v}^2 = \frac{1}{2}m\vec{v}\cdot\vec{v}$ we have

$$\dot{K}(t) = \frac{1}{2}m(\dot{\vec{v}}\cdot\vec{v} + \vec{v}\cdot\dot{\vec{v}}) = m\vec{a}\cdot\vec{v} = \tilde{F}\cdot\vec{v}.$$

The total change of the kinetic energy from time t_1 to t_2, denoted W, is called the *work done by the force $\tilde{F}$ along the path $\vec{x}(t)$*:

$$W = \int_{t_1}^{t_2} \dot{K}(t)\,\mathrm{d}t = \int_{t_1}^{t_2} \tilde{F}\cdot\vec{v}\,\mathrm{d}t = \int_{t_1}^{t_2} \tilde{F}(t)\cdot\dot{\vec{x}}(t)\,\mathrm{d}t.$$

Let us now suppose that the force $\tilde{F}$ at time t depends only on the position $\vec{x}(t)$. That is, we assume that there is a vector field $\vec{F}(\vec{x})$ such that $\tilde{F}(t) = \vec{F}(\vec{x}(t))$ (gravitational and electrostatic attraction are position-dependent while magnetic forces are velocity-dependent). Then we may rewrite the above integral as

$$W = \int_{t_1}^{t_2} \vec{F}(\vec{x}(t)) \cdot \dot{\vec{x}}(t)\,\mathrm{d}t.$$

In the one-dimensional case, by a change of variables, this can be simplified to

$$W = \int_a^b F(x)\,\mathrm{d}x,$$

where a and b are the starting and ending positions.

Definition 8.4 Let $\Gamma = \{\vec{x}(t) \mid t \in [r,s]\}$, be a continuously differentiable curve $\vec{x}(t) \in \mathrm{C}^1([r,s])$ in $\mathbb{R}^n$ and $\vec{f}\colon \Gamma \to \mathbb{R}^n$ a continuous vector field on Γ. The integral

$$\int_\Gamma \vec{f}(\vec{x}) \cdot \mathrm{d}\vec{x} = \int_r^s \vec{f}(\vec{x}(t)) \cdot \dot{\vec{x}}(t)\,\mathrm{d}t$$

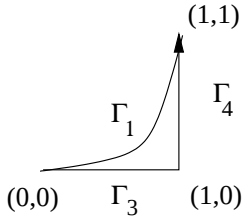
is called the *line integral* of the vector field $\vec{f}$ along the curve Γ.

Remark 8.2 (a) The definition of the line integral does not depend on the parametrization of Γ.
(b) If we take different curves between the same endpoints, the line integral may be different.
(c) If the vector field $\vec{f}$ is orthogonal to the tangent vector, then $\int_\Gamma \vec{f} \cdot \mathrm{d}\vec{x} = 0$.
(d) Other notations. If $\vec{f} = (P,Q)$ is a vector field in $\mathbb{R}^2$,

$$\int_\Gamma \vec{f} \cdot \mathrm{d}\vec{x} = \int_\Gamma P\,\mathrm{d}x + Q\,\mathrm{d}y,$$

where the right side is either a symbol or $\int_\Gamma P\,\mathrm{d}x = \int_\Gamma (P,0) \cdot \mathrm{d}\vec{x}$.

Example 8.4 (a) Find the line integral $\int_{\Gamma_i} y\,\mathrm{d}x + (x-y)\,\mathrm{d}y$, $i = 1,2$, where



$$\Gamma_1 = \{\vec{x}(t) = (t,t^2) \mid t \in [0,1]\} \quad \text{and} \quad \Gamma_2 = \Gamma_3 \cup \Gamma_4,$$

with $\Gamma_3 = \{(t,0) \mid t \in [0,1]\}$, $\Gamma_4 = \{(1,t) \mid t \in [0,1]\}$.
In the first case $\dot{\vec{x}}(t) = (1,2t)$; hence

$$\int_\Gamma y\,\mathrm{d}x + (x-y)\,\mathrm{d}y = \int_0^1 (t^2 \cdot 1 + (t-t^2)2t)\,\mathrm{d}t = \int_0^1 (3t^2 - 2t^3)\,\mathrm{d}t = \frac{1}{2}.$$

In the second case $\int_{\Gamma_2} f\,\mathrm{d}\vec{x} = \int_{\Gamma_3} f\,\mathrm{d}\vec{x} + \int_{\Gamma_4} f\,\mathrm{d}\vec{x}$. For the first part $(\mathrm{d}x,\,\mathrm{d}y) = (\mathrm{d}t, 0)$, for the second part $(\mathrm{d}x,\,\mathrm{d}y) = (0,\,\mathrm{d}t)$ such that

$$\int_\Gamma f\,\mathrm{d}\vec{x} = \int_\Gamma y\,\mathrm{d}x + (x-y)\,\mathrm{d}y = \int_0^1 0\,\mathrm{d}t + (t-0)\cdot 0 + \int_0^1 t\cdot 0 + (1-t)\,\mathrm{d}t = t - \frac{1}{2}t^2\Big|_0^1 = \frac{1}{2}.$$

(b) Find the work done by the force field $\vec{F}(x,y,z) = (y,-x,1)$ as a particle moves from $(1,0,0)$ to $(1,0,1)$ along the following paths $\varepsilon = \pm 1$:

$\vec{x}(t)_\varepsilon = (\cos t, \varepsilon \sin t, \frac{t}{2\pi})$, $t \in [0, 2\pi]$,

We find

$$\begin{aligned}\int_{\Gamma_\varepsilon} \vec{F} \cdot \mathrm{d}\vec{x} &= \int_0^{2\pi} (\varepsilon \sin t, -\cos t, 1) \cdot (-\sin t, \varepsilon \cos t, 1/(2\pi))\, \mathrm{d}t \\ &= \int_0^{2\pi} \left(-\varepsilon \sin^2 t - \varepsilon \cos^2 t + \frac{1}{2\pi} \right) \mathrm{d}t \\ &= -2\pi\varepsilon + 1.\end{aligned}$$

In case $\varepsilon = 1$, the motion is "with the force", so the work is positive; for the path $\varepsilon = -1$, the motion is against the force and the work is negative.

We can also define a *scalar line integral* in the following way. Let $\gamma\colon [a, b] \to \mathbb{R}^n$ be a continuously differentiable curve, $\Gamma = \gamma([a, b])$, and $f\colon \Gamma \to \mathbb{R}$ a continuous function. The integral

$$\int_\Gamma f(x)\, \mathrm{d}s := \int_a^b f(\gamma(t))\, \|\gamma'(t)\|\, \mathrm{d}t$$

is called the *scalar line integral* of f along Γ.

Properties of Line Integrals

Remark 8.3 (a) Linearity.

$$\int_\Gamma (\vec{f} + \vec{g})\, \mathrm{d}\vec{x} = \int_\Gamma \vec{f}\, \mathrm{d}\vec{x} + \int_\Gamma \vec{g}\, \mathrm{d}\vec{x}, \quad \int_\Gamma \lambda \vec{f}\, \mathrm{d}\vec{x} = \lambda \int_\Gamma \vec{f}\, \mathrm{d}\vec{x}.$$

(b) Change of orientation. If $\vec{x}(t)$, $t \in [r, s]$ defines a curve Γ which goes from $a = \vec{x}(r)$ to $b = \vec{x}(s)$, then $\vec{y}(t) = \vec{x}(r + s - t)$, $t \in [r, s]$, defines the curve $-\Gamma$ which goes in the opposite direction from b to a. It is easy to see that

$$\int_{-\Gamma} \vec{f}\, \mathrm{d}\vec{x} = -\int_\Gamma \vec{f}\, \mathrm{d}\vec{x}.$$

(c) Triangle inequality.

$$\left| \int_\Gamma \vec{f}\, \mathrm{d}\vec{x} \right| \le \ell(\Gamma) \sup_{x \in \Gamma} \left\| \vec{f}(x) \right\|.$$

Proof. Let $\vec{x}(t)$, $t \in [t_0, t_1]$ be a parametrization of Γ, then

$$\begin{aligned}\left| \int_\Gamma \vec{f}\, \mathrm{d}\vec{x} \right| &= \left| \int_{t_0}^{t_1} \vec{f}(\vec{x}(t)) \cdot \vec{x}'(t)\, \mathrm{d}t \right| \underset{\text{tri.in.,CSI}}{\le} \int_{t_0}^{t_1} \left\| \vec{f}(\vec{x}(t)) \right\| \, \|\vec{x}'(t)\|\, \mathrm{d}t \\ &\le \sup_{\vec{x} \in \Gamma} \left\| \vec{f}(\vec{x}) \right\| \int_{t_0}^{t_1} \|\vec{x}'(t)\|\, \mathrm{d}t = \sup_{\vec{x} \in \Gamma} \left\| \vec{f}(\vec{x}) \right\| \ell(\Gamma).\end{aligned}$$

■

(d) Splitting. If Γ_1 and Γ_2 are two curves such that the ending point of Γ_1 equals the starting point of Γ_2 then

$$\int_{\Gamma_1 \cup \Gamma_2} \vec{f}\, \mathrm{d}\vec{x} = \int_{\Gamma_1} \vec{f}\, \mathrm{d}\vec{x} + \int_{\Gamma_2} \vec{f}\, \mathrm{d}\vec{x}.$$

8.2.1 Path Independence

Problem: For which vector fields $\vec{f}$ the line integral from a to b does not depend upon the path (see Example 8.4 (a) Example 8.2)?

Definition 8.5 A vector field $\vec{f}\colon G \to \mathbb{R}^n$, $G \subset \mathbb{R}^n$, is called *conservative* if for any points a and b in G and any curves Γ_1 and Γ_2 from a to b we have

$$\int_{\Gamma_1} \vec{f}\,\mathrm{d}\vec{x} = \int_{\Gamma_2} \vec{f}\,\mathrm{d}\vec{x}.$$

In this case we say that the line integral $\int_\Gamma \vec{f}\,\mathrm{d}\vec{x}$ is *path independent* and we use the notation $\int_a^b \vec{f}\,\mathrm{d}\vec{x}$.

Definition 8.6 A vector field $\vec{f}\colon G \to \mathbb{R}^n$ is called *potential field* or *gradient vector field* if there exists a continuously differentiable function $U\colon G \to \mathbb{R}$ such that $\vec{f}(x) = \operatorname{grad} U(x)$ for $x \in G$. We call U the *potential* or *antiderivative* of $\vec{f}$.

Example 8.5 The gravitational force is given by

$$\vec{F}(x) = -\alpha \frac{x}{\|x\|^3},$$

where $\alpha = \gamma m M$. It is a potential field with potential

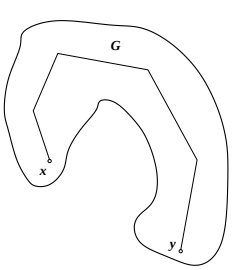
$$U(x) = \alpha \frac{1}{\|x\|}$$

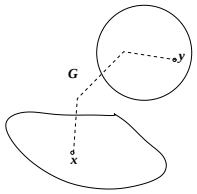
This follows from Example 7.2 (a), $\operatorname{grad} f(\|x\|) = f'(\|x\|)\frac{x}{\|x\|}$ with $f(y) = 1/y$ and $f'(y) = -1/y^2$.

Remark 8.4 (a) A vector field $\vec{f}$ is conservative if and only if the line integral over any *closed* curve in G is 0. Indeed, suppose that $\vec{f}$ is conservative and $\Gamma = \Gamma_1 \cup \Gamma_2$ is a closed curve, where Γ_1 is a curve from a to b and Γ_2 is a curve from b to a. By Remark 8.3 (b), changing the orientation of Γ_2, the sign of the line integral changes and $-\Gamma_2$ is again a curve from a to b:

$$\int_\Gamma \vec{f}\,\mathrm{d}\vec{x} = \left(\int_{\Gamma_1} + \int_{\Gamma_2}\right) \vec{f}\,\mathrm{d}\vec{x} = \left(\int_{\Gamma_1} - \int_{-\Gamma_2}\right) \vec{f}\,\mathrm{d}\vec{x} = 0.$$

The proof of the other direction is similar.





(b) *Uniqueness of a potential.* An open subset $G \subset \mathbb{R}^n$ is said to be *connected*, if any two points $x, y \in G$ can be connected by a polygonal path from x to y inside G. If it exists, $U(x)$ is uniquely determined up to a constant.

Indeed, if $\operatorname{grad} U_1(x) = \operatorname{grad} U_2(x) = \vec{f}$ put $U = U_1 - U_2$ then we have $\nabla U = \nabla U_1 - \nabla U_2 = \vec{f} - \vec{f} = 0$. Now suppose that $x, y \in G$ can be connected by a segment $\overline{xy} \subset G$ inside G. By the MVT (Corollary 7.12)

$$U(y) - U(x) = \nabla U((1-t)x + ty)\,(y - x) = 0,$$

since $\nabla U = 0$ on the segment $\overline{xy}$. This shows that $U = U_1 - U_2 = \operatorname{const.}$ on any polygonal path inside G. Since G is connected, $U_1 - U_2$ is constant on G.

Theorem 8.2 *Let $G \subset \mathbb{R}^n$ be a domain.*
(i) *If $U\colon G \to \mathbb{R}$ is continuously differentiable and $\vec{f} = \operatorname{grad} U$. Then $\vec{f}$ is conservative, and for every (piecewise continuously differentiable) curve Γ from a to b, $a, b \in G$, we have*

$$\int_\Gamma \vec{f}\,\mathrm{d}\vec{x} = U(b) - U(a).$$

(ii) *Let $\vec{f}\colon G \to \mathbb{R}^n$ be a continuous, conservative vector field and $a \in G$. Put*

$$U(x) = \int_a^x \vec{f}\,\mathrm{d}\vec{y}, \quad x \in G.$$

Then $U(x)$ is a potential for $\vec{f}$, that is $\operatorname{grad} U = \vec{f}$.
(iii) *A continuous vector field $\vec{f}$ is conservative in G if and only if it is a potential field.*

Proof. (i) Let $\Gamma = \{\vec{x}(t) \mid t \in [r, s]\}$, be a continuously differentiable curve from $a = \vec{x}(r)$ to $b = \vec{x}(s)$. We define $\varphi(t) = U(\vec{x}(t))$ and compute the derivative using the chain rule

$$\dot{\varphi}(t) = \operatorname{grad} U(\vec{x}(t)) \cdot \dot{\vec{x}}(t) = \vec{f}(\vec{x}(t)) \cdot \dot{\vec{x}}(t).$$

By definition of the line integral we have

$$\int_\Gamma \vec{f}\,\mathrm{d}\vec{x} = \int_r^s \vec{f}(\vec{x}(t))\,\dot{\vec{x}}(t)\,\mathrm{d}t.$$

Inserting the above expression and applying the fundamental theorem of calculus, we find

$$\int_\Gamma \vec{f}\,\mathrm{d}\vec{x} = \int_r^s \dot{\varphi}(t)\,\mathrm{d}t = \varphi(s) - \varphi(r) = U(\vec{x}(s)) - U(\vec{x}(r)) = U(b) - U(a).$$

(ii) Choose $h \in \mathbb{R}^n$ small such that $x + th \in G$ for all $t \in [0, 1]$. By the path independence of the line integral

$$U(x+h) - U(x) = \int_a^{x+h} \vec{f} \cdot\, \mathrm{d}\vec{y} - \int_a^x \vec{f} \cdot\, \mathrm{d}\vec{y} = \int_x^{x+h} \vec{f} \cdot\, \mathrm{d}\vec{y}$$

Consider the curve $\vec{x}(t) = x + th$, $t \in [0, 1]$ from x to $x + h$. Then $\dot{\vec{x}}(t) = h$. By the mean value theorem of integration (Theorem 5.18 with $\varphi = 1$, $a = 0$ and $b = 1$) we have

$$\int_x^{x+h} \vec{f} \cdot\, \mathrm{d}\vec{y} = \int_0^1 \vec{f}(\vec{x}(t)) \cdot h\,\mathrm{d}t = \vec{f}(x + \theta h) \cdot h,$$

where $\theta \in [0,1]$. We check $\operatorname{grad} U(x) = \vec{f}(x)$ using the definition of the derivative:

$$\frac{\left| U(x+h) - U(x) - \vec{f}(x)\cdot h \right|}{\|h\|} = \frac{\left| (\vec{f}(x+\theta h) - \vec{f}(x))\cdot h \right|}{\|h\|} \underset{\text{CSI}}{\leq} \frac{\left\| \vec{f}(x+\theta h) - \vec{f}(x) \right\| \|h\|}{\|h\|}$$
$$= \left\| \vec{f}(x+\theta h) - \vec{f}(x) \right\| \underset{h\to 0}{\longrightarrow} 0,$$

since $\vec{f}$ is continuous at x. This shows that $\nabla U = \vec{f}$.
(iii) follows immediately from (i) and (ii). ∎

Remark 8.5 (a) In case $n = 2$, a simple path to compute the line integral (and so the potential U) in (ii) consists of 2 segments: from $(0,0)$ via $(x,0)$ to (x,y). The line integral of $P\,\mathrm{d}x + Q\,\mathrm{d}y$ then reads as ordinary Riemann integrals

$$U(x,y) = \int_0^x P(t,0)\,\mathrm{d}t + \int_0^y Q(x,t)\,\mathrm{d}t.$$

(b) Case $n = 3$. You can also use just one single segment from the origin to the endpoint (x,y,z). This path is parametrized by the curve

$$\vec{x}(t) = (tx, ty, tz), \quad t \in [0,1], \quad \dot{\vec{x}}(t) = (x,y,z).$$

We obtain

$$U(x,y,z) = \int_{(0,0,0)}^{(x,y,z)} f_1\,\mathrm{d}x + f_2\,\mathrm{d}y + f_3\,\mathrm{d}z \tag{8.2}$$
$$= x\int_0^1 f_1(tx,ty,tz)\,\mathrm{d}t + y\int_0^1 f_2(tx,ty,tz)\,\mathrm{d}t + z\int_0^1 f_3(tx,ty,tz)\,\mathrm{d}t. \tag{8.3}$$

(c) Although Theorem 8.2 gives a necessary and sufficient condition for a vector field to be conservative, we are missing an easy criterion.

Recall from Example 7.4, that a necessary condition for $\vec{f} = (f_1, \ldots, f_n)$ to be a potential vector field is

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i}, \quad 1 \leq i < j \leq n.$$

which is a simple consequence from Schwarz's lemma since if $f_i = U_{x_i}$ then

$$U_{x_i x_j} = \frac{\partial U_{x_i}}{\partial x_j} = \frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i} = \frac{\partial U_{x_j}}{\partial x_i} = U_{x_j x_i}.$$

The condition $\frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i}, \quad 1 \leq i < j \leq n.$ is called *integrability condition* for $\vec{f}$. It is a necessary condition for $\vec{f}$ to be conservative. However, it is not sufficient.

Remark 8.6 Counter example. Let $G = \mathbb{R}^2 \setminus \{(0,0)\}$ and

$$\vec{f} = (P,Q) = \left(\frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2} \right).$$

The vector field satisfies the integrability condition $P_y = Q_x$. However, it is not conservative. For, consider the unit circle $\gamma(t) = (\cos t, \sin t)$, $t \in [0, 2\pi]$. Then $\gamma'(t) = (-\sin t, \cos t)$ and

$$\int_\gamma \vec{f}\,\mathrm{d}\vec{x} = \int_\gamma \frac{-y\,\mathrm{d}x}{x^2+y^2} + \frac{x\,\mathrm{d}y}{x^2+y^2} \quad = \int_0^{2\pi} \frac{\sin t \sin t\,\mathrm{d}t}{1} + \frac{\cos t \cos t\,\mathrm{d}t}{1} = \int_0^{2\pi} \mathrm{d}t = 2\pi.$$

This contradicts $\int_\Gamma \vec{f}\,\mathrm{d}\vec{x} = 0$ for conservative vector fields. Hence, $\vec{f}$ is not conservative.
$\vec{f}$ fails to be conservative since $G = \mathbb{R}^2 \setminus \{(0,0)\}$ has an hole.
For more details, see homework 30.1.

The next proposition shows that under one additional assumption this criterion is also sufficient. A connected open subset G (a region) of $\mathbb{R}^n$ is called *simply connected* if every closed polygonal path inside G can be shrunk inside G to a single point.
Roughly speaking, simply connected sets do not have holes.

convex subset of $\mathbb{R}^n$	simply connected
1-torus $\mathrm{S}^1 = \{z \in \mathbb{C} \mid \lvert z \rvert = 1\}$	not simply connected
annulus $\{(x,y) \in \mathbb{R}^2 \mid r^2 < x^2+y^2 \le R^2\}$, $0 \le r < R \le \infty$	not simply connected
$\mathbb{R}^2 \setminus \{(0,0)\}$	not simlpy connected
$\mathbb{R}^3 \setminus \{(0,0,0)\}$	simply connected

The precise mathematical term for a curve γ to be "shrinkable to a point" is to be null-homotopic.

Definition 8.7 (a) A closed curve $\gamma\colon [a,b] \to G$, $G \subset \mathbb{R}^n$ open, is said to be *null-homotopic* if there exists a continuous mapping $h\colon [a,b] \times [0,1] \to G$ and a point $x_0 \in G$ such that

(a) $h(t,0) = \gamma(t)$ for all t,
(b) $h(t,1) = x_0$ for all t,
(c) $h(a,s) = h(b,s) = x_0$ for all $s \in [0,1]$.

(b) G is *simply connected* if any curve in G is null homotopic.

Proposition 8.3 *Let $\vec{f} = (f_1, f_2, f_3)$ a continuously differentiable vector field on a region $G \subset \mathbb{R}^3$.*
(a) *If $\vec{f}$ is conservative then $\operatorname{curl} \vec{f} = 0$, i. e.*

$$\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3} = 0, \quad \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1} = 0, \quad \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} = 0.$$

(b) *If $\operatorname{curl} \vec{f} = 0$ and G is simply connected, then $\vec{f}$ is conservative.*

Proof. (a) Let $\vec{f}$ be conservative; by Theorem 8.2 there exists a potential U, $\operatorname{grad} U = \vec{f}$. However, $\operatorname{curl} \operatorname{grad} U = 0$ since

$$\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3} = \frac{\partial^2 U}{\partial x_2 \partial x_3} - \frac{\partial^2 U}{\partial x_3 \partial x_2} = 0$$

by Schwarz's Lemma.
(b) This will be an application of Stokes' theorem, see below. ∎

Example 8.6 Let on $\mathbb{R}^3$, $\vec{f} = (P, Q, R) = (6xy^2 + \mathrm{e}^x, 6x^2y, 1)$. Then

$$\operatorname{curl} \vec{f} = (R_y - Q_z, P_z - R_x, Q_x - P_y) = (0, 0, 12xy - 12xy) = 0;$$

hence, $\vec{f}$ is conservative with the potential $U(x, y, z)$.

First method to compute the potential U: ODE ansatz.

The ansatz $U_x = 6xy^2 + \mathrm{e}^x$ will be integrated with respect to x:

$$U(x, y, z) = \int U_x \,\mathrm{d}x + C(y, z) = \int (6xy^2 + \mathrm{e}^x)\,\mathrm{d}x + C(y, z) = 3x^2y^2 + \mathrm{e}^x + C(y, z).$$

Hence,

$$U_y = 6x^2y + C_y(y, z) \stackrel{!}{=} 6x^2y, \quad U_z = C_z \stackrel{!}{=} 1.$$

This implies $C_y = 0$ and $C_z = 1$. The solution here is $C(y, z) = z + c_1$ such that $U = 3x^2y^2 + \mathrm{e}^x + z + c_1$.

Second method: Line Integrals. See Remark 8.5 (b))

$$\begin{aligned} U(x, y, z) &= x \int_0^1 f_1(tx, ty, tz)\,\mathrm{d}t + y \int_0^1 f_2(tx, ty, tz)\,\mathrm{d}t + z \int_0^1 f_3(tx, ty, tz)\,\mathrm{d}t \\ &= x \int_0^1 (6t^3xy^2 + \mathrm{e}^{tx})\,\mathrm{d}t + y \int_0^1 6t^3x^2y\,\mathrm{d}t + z \int_0^1 \mathrm{d}t \\ &= 3x^2y^2 + \mathrm{e}^x + z. \end{aligned}$$

Chapter 9

Integration of Functions of Several Variables

References to this chapter are [O'N75, Section 4] which is quite elemantary and good accessible. Another elementary approach is [MW85, Chapter 17] (part III). A more advanced but still good accessible treatment is [Spi65, Chapter 3]. This will be our main reference here. Rudin's book [Rud76] is not recommendable for an introduction to integration.

9.1 Basic Definition

The definition of the Riemann integral of a function $f\colon A \to \mathbb{R}$, where $A \subset \mathbb{R}^n$ is a closed rectangle, is so similar to that of the ordinary integral that a rapid treatment will be given, see Section 5.1.
If nothing is specified otherwise, A denotes a rectangle. A *rectangle* A is the cartesian product of n intervals,

$$A = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_k \leq x_k \leq b_k,\ k = 1, \ldots, n\}.$$

Recall that a *partition* of a closed interval $[a, b]$ is a sequence $t_0, \ldots, t_k$ where $a = t_0 \leq t_1 \leq \cdots \leq t_k = b$. The partition divides the interval $[a, b]$ in to k subintervals $[t_{i-1}, t_i]$. A *partition* of a rectangle $[a_1, b_1] \times \cdots \times [a_n, b_n]$ is a collection $P = (P_1, \ldots, P_n)$ where each P_i is a partition of the interval $[a_i, b_i]$. Suppose for example that $P_1 = (t_0, \ldots, t_k)$ is a partition of $[a_1, b_1]$ and $P_2 = (s_0, \ldots, s_l)$ is a partition of $[a_2, b_2]$. Then the partition $P = (P_1, P_2)$ of $[a_1, b_1] \times [a_2, b_2]$ divides the closed rectangle $[a_1, b_1] \times [a_2, b_2]$ into kl subrectangles, a typical one being $[t_{i-1}, t_i] \times [s_{j-1}, s_j]$. In general, if P_i divides $[a_i, b_i]$ into N_i subintervals, then $P = (P_1, \ldots, P_n)$ divides $[a_1, b_1] \times \cdots \times [a_n, b_n]$ into $N_1 \cdots N_n$ subrectangles. These subrectangles will be called *subrectangles of the partition* P.
Suppose now A is a rectangle, $f\colon A \to \mathbb{R}$ is a bounded function, and P is a partition of A. For each subrectangle S of the partition let

$$m_S = \inf\{f(x) \mid x \in S\}, \quad M_S = \sup\{f(x) \mid x \in S\},$$

and let $v(S)$ be the volume of the rectangle S. Note that *volume* of the rectangle $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is

$$v(A) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n).$$

The *lower* and the *upper sums* of f for P are defined by

$$L(P, f) = \sum_S m_S \, v(S) \quad \text{and} \quad U(P, f) = \sum_S M_S \, v(S),$$

where the sum is taken over all subrectangles S of the partition P. Clearly, if f is bounded with $m \le f(x) \le M$ on the rectangle $x \in R$,

$$m \, v(R) \le L(P, f) \le U(P, f) \le M \, v(R),$$

so that the numbers $L(P, f)$ and $U(P, f)$ form bounded sets. Lemma 5.1 remains true; the proof is completely the same.

Lemma 9.1 *(a) Suppose the partition P^* is a refinement of P (that is, each subrectangle of P^* is contained in a subrectangle of P). Then*

$$L(P, f) \le L(P^*, f) \quad \textit{and} \quad U(P^*, f) \le U(P, f).$$

(b) If P and P' are any two partitions, then $L(P, f) \le U(P', f)$.

It follows from the above corollary that all lower sums are bounded above by any upper sum and vice versa.

Definition 9.1 Let $f \colon A \to \mathbb{R}$ be a bounded function. The function f is called *Riemann integrable* on the rectangle A if

$$\underline{\int_A} f \, \mathrm{d}x := \sup_P \{L(P, f)\} = \inf_P \{U(P, f)\} =: \overline{\int_A} f \, \mathrm{d}x,$$

where the supremum and the infimum are taken over all partitions P of A. This common number is the *Riemann integral* of f on A and is denoted by

$$\int_A f \, \mathrm{d}x \quad \text{or} \quad \int_A f(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

$\underline{\int}_A f \, \mathrm{d}x$ and $\overline{\int}_A f \, \mathrm{d}x$ are called the *lower* and the *upper* integral of f on A, respectively. They always exist. The set of integrable function on A is denoted by $\mathcal{R}(A)$.

As in the one dimensional case we have the following criterion.

Proposition 9.2 (Riemann Criterion) *A bounded function $f \colon A \to \mathbb{R}$ is integrable if and only if for every $\varepsilon > 0$ there exists a partition P of A such that $U(P, f) - L(P, f) < \varepsilon$.*

Example 9.1 (a) Let $f\colon A \to \mathbb{R}$ be a constant function $f(x) = c$. Then for any Partition P and any subrectangle S we have $m_S = M_S = c$, so that

$$L(P,f) = U(P,f) = \sum_S c\, v(S) = c \sum_S v(S) = cv(A).$$

Hence, $\int_A c\,\mathrm{d}x = cv(A)$.
(b) Let $f\colon [0,1] \times [0,1] \to \mathbb{R}$ be defined by

$$f(x,y) = \begin{cases} 0, & \text{if } x \text{ is rational,} \\ 1, & \text{if } x \text{ is irrational.} \end{cases}$$

If P is a partition, then every subrectangle S will contain points (x,y) with x rational, and also points (x,y) with x irrational. Hence $m_S = 0$ and $M_S = 1$, so

$$L(P,f) = \sum_S 0v(S) = 0,$$

and

$$U(P,f) = \sum_S 1v(S) = v(A) = v([0,1] \times [0,1]) = 1.$$

Therefore, $\overline{\int}_A f\,\mathrm{d}x = 1 \neq 0 = \underline{\int}_A f\,\mathrm{d}x$ and f is not integrable.

9.1.1 Properties of the Riemann Integral

We briefly write $\mathcal{R}$ for $\mathcal{R}(A)$.

Remark 9.1 (a) $\mathcal{R}$ is a linear space and $\int_A (\cdot)\ \mathrm{d}x$ is a linear functional, i. e. $f, g \in \mathcal{R}$ imply $\lambda f + \mu g \in \mathcal{R}$ for all $\lambda, \mu \in \mathbb{R}$ and

$$\int_A (\lambda f + \mu g)\,\mathrm{d}x = \lambda \int_A f\,\mathrm{d}x + \mu \int_A g\,\mathrm{d}x.$$

(b) $\mathcal{R}$ is a *lattice*, i. e., $f \in \mathcal{R}$ implies $|f| \in \mathcal{R}$. If $f, g \in \mathcal{R}$, then $\max\{f,g\} \in \mathcal{R}$ and $\min\{f,g\} \in \mathcal{R}$.

(c) $\mathcal{R}$ is an algebra, i. e., $f, g \in \mathcal{R}$ imply $fg \in \mathcal{R}$.

(d) The triangle inequality holds:

$$\left| \int_A f\,\mathrm{d}x \right| \leq \int_A |f|\,\mathrm{d}x.$$

(e) $\mathrm{C}(A) \subset \mathcal{R}(A)$.

(f) $f \in \mathcal{R}(A)$ and $f(A) \subset [a,b]$, $g \in \mathrm{C}[a,b]$. Then $g \circ f \in \mathcal{R}(A)$.

(g) If $f \in \mathcal{R}$ and $f = g$ except at finitely many points, then $g \in \mathcal{R}$ and $\int_A f\,\mathrm{d}x = \int_A g\,\mathrm{d}x$.

(h) Let $f\colon A \to \mathbb{R}$ and let P be a partition of A. Then $f \in \mathcal{R}(A)$ if and only if $f \restriction S$ is integrable for each subrectangle S. In this case

$$\int_A f\,\mathrm{d}x = \sum_S \int_S f \restriction S\,\mathrm{d}x.$$

9.2 Integrable Functions

We are going to characterize integrable functions. For, we need the notion of a set of *measure zero*.

Definition 9.2 Let A be a subset of $\mathbb{R}^n$. A has (n-dimensional) *measure zero* if for every $\varepsilon > 0$ there exists a sequence $(U_i)_{i\in\mathbb{N}}$ of closed rectangles U_i which cover A such that $\sum_{i=1}^{\infty} v(U_i) < \varepsilon$.

Open rectangles can also be used in the definition.

Remark 9.2 (a) Any finite set $\{a_1, \ldots, a_m\} \subset \mathbb{R}^n$ is of measure 0. Indeed, let $\varepsilon > 0$ and choose U_i be a rectangle with midpoint a_i and volume ε/m. Then $\{U_i \mid i = 1, \ldots, m\}$ covers A and $\sum_i v(U_i) \leq \varepsilon$.
(b) Any contable set is of measure 0.
(c) Any countable set has measure 0.
(d) If each $(A_i)_{i\in\mathbb{N}}$ has measure 0 then $A = A_1 \cup A_2 \cup \cdots$ has measure 0.
Proof. Let $\varepsilon > 0$. Since A_i has measure 0 there exist closed rectangles U_{ik}, $i \in \mathbb{N}$, $k \in \mathbb{N}$, such that for fixed i, the family $\{U_{ik} \mid k \in \mathbb{N}\}$ covers A_i, i.e. $\bigcup_{k\in\mathbb{N}} U_{ik} \supseteq A_i$ and $\sum_{k\in\mathbb{N}} v(U_{ik}) \leq \varepsilon/2^{i-1}$, $i \in \mathbb{N}$. In this way we have constructed an infinite array $\{U_{ik}\}$ which covers A. Arranging those sets in a sequence (cf. Cantor's first diagonal process), we obtain a sequence of rectangles which covers A and

$$\sum_{i,k=1}^{\infty} v(U_{ik}) \leq \sum_{i=1}^{\infty} \frac{\varepsilon}{2^{i-1}} = 2\varepsilon.$$

Hence, $\sum_{i,k=1}^{\infty} v(U_{ik}) \leq 2\varepsilon$ and A has measure 0. ∎

(e) Let $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be a non-singular rectangle, that is $a_i < b_i$ for all $i = 1, \ldots, n$. Then A is not of measure 0. Indeed, we use the following two facts about the volume of finite unions of rectangles:

(a) $v(U_1 \cup \cdots \cup U_n) \leq \sum_{i=1}^{n} v(U_i)$,
(b) $U \subseteq V$ implies $v(U) \leq v(V)$.

Now let $\varepsilon = v(A)/2 = (b_1 - a_1) \cdots (b_n - a_n)/2$ and suppose that the open rectangles $(U_i)_{i\in\mathbb{N}}$ cover the compact set A. Then there exists a finite subcover $U_1 \cup \cdots \cup U_m \supseteq A$. This and (a), (b) imply

$$\varepsilon < v(A) \leq v\left(\bigcup_{i=1}^{n} U_i\right) \leq \sum_{i=1}^{n} v(U_i) \leq \sum_{i=1}^{\infty} v(U_i).$$

This contradicts $\sum_i v(U_i) \leq \varepsilon$; thus, A has not measure 0.

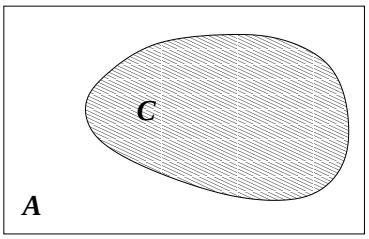
Theorem 9.3 *Let A be a closed rectangle and $f\colon A \to \mathbb{R}$ a bounded function. Let $B = \{x \in A \mid f$ is discontinuous at $x\}$.*
Then f is integrable if and only if B is a set of measure 0.

For the proof see [Spi65, 3-8 Theorem] or [Rud76, Theorem 11.33].

9.2.1 Integration over More General Sets

We have so far dealt only with integrals of functions over rectangles. Integrals over other sets are easily reduced to this type.
If $C \subset \mathbb{R}^n$, the *characteristic function* χ_C of C is defined by

$$\chi_C(x) = \begin{cases} 1, & x \in C, \\ 0, & x \notin C. \end{cases}$$



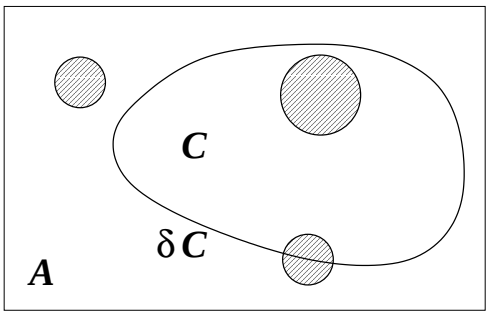
Definition 9.3 Let $f\colon C \to \mathbb{R}$ be bounded and A a rectangle, $C \subset A$. We call f *Riemann integrable* on C if the product function $f \cdot \chi_C\colon A \to \mathbb{R}$ is Riemann integrable on A. In this case we define

$$\int_C f \,\mathrm{d}x = \int_A f\chi_C \,\mathrm{d}x.$$

This certainly occurs if both f and χ_C are integrable on A. Note, that $\int_C 1 \,\mathrm{d}x = \int_A \chi_C \,\mathrm{d}x =: v(C)$ is defined to be the *volume or measure* of C.

Problem: Under which conditions on C, the volume $v(C) = \int_C \mathrm{d}x$ exists? By Theorem 9.3 χ_C is integrable if and only if the set B of discontinuities of χ_C in A has measure 0.

The boundary of a set C



Let $C \subset A$. For every $x \in A$ exactly one of the following three cases occurs:
(a) x has a neighborhood which is completely contained in C (x is an inner point of C),
(b) x has a neighborhood which is completely contained in C^{c} (x is an inner point of C^{c}),
(c) every neighborhood of x intersects both C and C^{c}. In this case we say, x belongs to the *boundary* ∂C of C. By definition $\partial C = \overline{C} \cap \overline{C^{\mathsf{c}}}$; also $\partial C = \overline{C} \setminus C^{\circ}$.

By the above discussion, A is the disjoint union of two open and a closed set:

$$A = C^{\circ} \cup \partial C \cup (C^{\mathsf{c}})^{\circ}.$$

Theorem 9.4 *The characteristic function $\chi_C\colon A \to \mathbb{R}$ is integrable if and only if the boundary of C has measure 0.*

Proof. Since the boundary ∂C is closed and inside the bounded set, ∂C is compact. Suppose first x is an inner point of C. Then there is an open set $U \subset C$ containing x. Thus $\chi_C(x) = 1$ on $x \in U$; clearly χ_C is continuous at x (since it is locally constant). Similarly, if x is an inner point of C^{c}, $\chi_C(x)$ is locally constant, namely $\chi_C = 0$ in a neighborhood of x. Hence χ_C is

continuous at x. Finally, if x is in the boundary of C for every open neighborhood U of x there is $y_1 \in U \cap C$ and $y_2 \in U \cap C^{\mathsf{c}}$, so that $\chi_C(y_1) = 1$ whereas $\chi_C(y_2) = 0$. Hence, χ_C is not continuous at x. Thus, the set of discontinuity of χ_C is exactly the boundary ∂C. The rest follows from Theorem 9.3. ∎

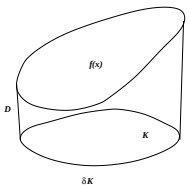
Definition 9.4 A bounded set C is called *Jordan measurable* or simply a *Jordan set* if its boundary has measure 0. The integral $v(C) = \int_C 1 \, \mathrm{d}x$ is called the n-dimensional *Jordan measure* of C or the n-dimensional *volume* of C; sometimes we write $\mu(C)$ in place of $v(C)$.

Naturally, the one-dimensional volume in the *length*, and the two-dimensional volume is the *area*.

Typical Examples of Jordan Measurable Sets

Hyper planes $\sum_{i=1}^{n} a_i x_i = c$, and, more general, hyper surfaces $f(x_1, \ldots, x_n) = c$, $f \in \mathrm{C}^1(G)$ are sets with measure 0 in $\mathbb{R}^n$. Curves in $\mathbb{R}^n$ have measure 0. Graphs of functions $\Gamma_f = \{(x, f(x)) \in \mathbb{R}^{n+1} \mid x \in G\}$, f continuous, are of measure 0 in $\mathbb{R}^{n+1}$. If G is a bounded region in $\mathbb{R}^n$, the boundary ∂G has measure 0. If $G \subset \mathbb{R}^n$ is a region, the *cylinder* $C = \partial G \times \mathbb{R} = \{(x, x_{n+1}) \mid x \in \partial G\} \subset \mathbb{R}^{n+1}$ is a measure 0 set.
Let $D \subset \mathbb{R}^{n+1}$ be given by

$$D = \{(x, x_{n+1}) \mid x \in K,\, 0 \le x_{n+1} \le f(x)\},$$



where $K \subset \mathbb{R}^n$ is a compact set and $f\colon K \to \mathbb{R}$ is continuous. Then D is Jordan measurable. Indeed, D is bounded by the graph Γ_f, the hyper plane $x_{n+1} = 0$ and the cylinder $\partial D \times \mathbb{R} = \{(x, x_{n+1}) \mid x \in \partial K\}$ and all have measure 0 in $\mathbb{R}^{n+1}$.

9.2.2 Fubini's Theorem and Iterated Integrals

Our goal is to evaluate Riemann integrals; however, so far there was no method to compute multiple integrals. The following theorem fills this gap.

Theorem 9.5 (Fubini's Theorem) *Let $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ be closed rectangles, and let $f\colon A \times B \to \mathbb{R}$ be integrable. For $x \in A$ let $g_x\colon B \to \mathbb{R}$ be defined by $g_x(y) = f(x, y)$ and let*

$$\mathcal{L}(x) = \underline{\int_B} g_x \, \mathrm{d}y = \underline{\int_B} f(x, y) \, \mathrm{d}y,$$
$$\mathcal{U}(x) = \overline{\int_B} g_x \, \mathrm{d}y = \overline{\int_B} f(x, y) \, \mathrm{d}y.$$

Then $\mathcal{L}(x)$ and $\mathcal{U}(x)$ are integrable on A and

$$\int_{A\times B} f\ \mathrm{d}x\mathrm{d}y = \int_A \mathcal{L}(x)\,\mathrm{d}x = \int_A \left(\underline{\int}_B f(x,y)\,\mathrm{d}y\right)\,\mathrm{d}x,$$
$$\int_{A\times B} f\ \mathrm{d}x\mathrm{d}y = \int_A \mathcal{U}(x)\,\mathrm{d}x = \int_A \left(\overline{\int}_B f(x,y)\,\mathrm{d}y\right)\,\mathrm{d}x.$$

The integrals on the right are called iterated integrals.

The proof is in the appendix to this chapter.

Remarks 9.3 (a) A similar proof shows that we can exchange the order of integration

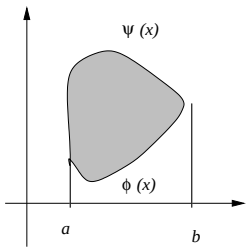
$$\int_{A\times B} f\ \mathrm{d}x\mathrm{d}y = \int_B \left(\underline{\int}_A f(x,y)\,\mathrm{d}x\right)\,\mathrm{d}y = \int_B \left(\overline{\int}_A f(x,y)\,\mathrm{d}x\right)\,\mathrm{d}y.$$

These integrals are called *iterated integrals* for f.
(b) In practice it is often the case that each g_x is integrable so that $\int_{A\times B} f\,\mathrm{d}x\mathrm{d}y = \int_A \left(\int_B f(x,y)\,\mathrm{d}y\right)\,\mathrm{d}x$. This certainly occurs if f is continuous.
(c) If $A = [a_1,b_1]\times\cdots\times[a_n,b_n]$ and $f\colon A\to\mathbb{R}$ is continuous, we can apply Fubini's theorem repeatedly to obtain

$$\int_A f\,\mathrm{d}x = \int_{a_n}^{b_n}\left(\cdots\left(\int_{a_1}^{b_1} f(x_1,\ldots,x_n)\,\mathrm{d}x_1\right)\cdots\right)\,\mathrm{d}x_n.$$

(d) If $C\subset A\times B$, Fubini's theorem can be used to compute $\int_C f\,\mathrm{d}x$ since this is by definition $\int_{A\times B} f\chi_C\,\mathrm{d}x$. Here are two examples in case $n=2$ and $n=3$.
Let $a<b$ and $\varphi(x)$ and $\psi(x)$ continuous real valued functions on $[a,b]$ with $\varphi(x)<\psi(x)$ on $[a,b]$. Put

$$C = \{(x,y)\in\mathbb{R}^2 \mid a\le x\le b,\quad \varphi(x)\le y\le\psi(x)\}.$$



Let $f(x,y)$ be continuous on C. Then f is integrable on C and

$$\iint_C f\,\mathrm{d}x\mathrm{d}y = \int_a^b\left(\int_{\varphi(x)}^{\psi(x)} f(x,y)\,\mathrm{d}y\right)\,\mathrm{d}x.$$

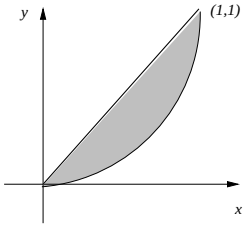
Let

$$G = \{(x,y,z)\in\mathbb{R}^3 \mid a\le x\le b,\ \varphi(x)\le y\le\psi(x),\ \alpha(x,y)\le z\le\beta(x,y)\},$$

where all functions are sufficiently nice. Then

$$\iiint_G f(x,y,z)\ \mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_a^b\left(\int_{\varphi(x)}^{\psi(y)}\left(\int_{\alpha(x,y)}^{\beta(x,y)} f(x,y,z)\,\mathrm{d}z\right)\,\mathrm{d}y\right)\,\mathrm{d}x.$$

(e) **Cavalieri's Principle.** Let A and B be Jordan sets in $\mathbb{R}^3$ and let $A_c = \{(x,y)\mid(x,y,c)\in A\}$ be the section of A with the plane $z=c$; B_c is defined similar. Suppose each A_c and B_c is Jordan measurable (in $\mathbb{R}^2$) and they have the same area $v(A_c)=v(B_c)$ for all $c\in\mathbb{R}$.
Then A and B have the same volume $v(A)=v(B)$.



Example 9.2 (a) Let $f(x,y) = xy$ and

$$C = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ x^2 \le y \le x\}$$
$$= \{(x,y) \in \mathbb{R}^2 \mid 0 \le y \le 1,\ y \le x \le \sqrt{y}\}.$$

Then

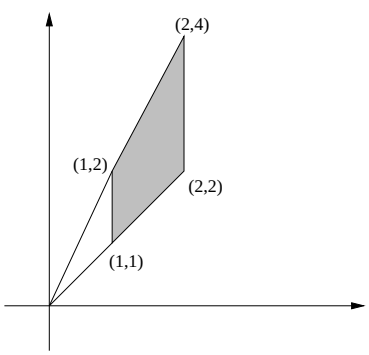
$$\iint\limits_C xy \,\mathrm{d}x\mathrm{d}y = \int_0^1 \int_{x^2}^x xy \,\mathrm{d}y \,\mathrm{d}x = \int_0^1 \frac{xy^2}{2}\Big|_{y=x^2}^x \mathrm{d}x = \frac{1}{2}\int_0^1 (x^3 - x^5)\,\mathrm{d}x$$
$$= \frac{x^4}{8} - \frac{x^6}{12}\Big|_0^1 = \frac{1}{8} - \frac{1}{12} = \frac{1}{24}.$$

Interchanging the order of integration we obtain

$$\iint\limits_C xy \,\mathrm{d}x\mathrm{d}y = \int_0^1 \int_y^{\sqrt{y}} xy \,\mathrm{d}x \,\mathrm{d}y = \int_0^1 \frac{x^2 y}{2}\Big|_y^{\sqrt{y}} = \frac{1}{2}\int_0^1 (y^2 - y^3)\,\mathrm{d}y$$
$$= \frac{y^3}{6} - \frac{y^4}{8}\Big|_0^1 = \frac{1}{6} - \frac{1}{8} = \frac{1}{24}.$$

(b) Let $G = \{(x,y,z) \in \mathbb{R}^3 \mid x,y,z \ge 0,\ x+y+z \le 1\}$ and $f(x,y,z) = 1/(x+y+z+1)^3$. The set G can be parametrized as follows

$$\iiint\limits_G f \,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_0^1 \left(\int_0^{1-x} \left(\int_0^{1-x-y} \frac{\mathrm{d}z}{(1+x+y+z)^3}\right) \mathrm{d}y\right) \mathrm{d}x$$
$$= \int_0^1 \left(\int_0^{1-x} \frac{1}{2}\frac{-1}{(1+x+y+z)^2}\Big|_0^{1-x-y} \mathrm{d}y\right) \mathrm{d}x$$
$$= \int_0^1 \left(\int_0^{1-x} \left(\frac{1}{2}\frac{1}{(1+x+y)^2} - \frac{1}{8}\right) \mathrm{d}y\right) \mathrm{d}x$$
$$= \frac{1}{2}\int_0^1 \left(\frac{1}{x+1} + \frac{x-3}{4}\right) \mathrm{d}x = \frac{1}{2}\left(\log 2 - \frac{5}{8}\right).$$



(c) Let $f(x,y) = \mathrm{e}^{y/x}$ and D the above region. Compute the integral of f on D.
D can be parametrized as follows $D = \{(x,y) \mid 1 \le x \le 2,\ x \le y \le 2x\}$ Hence,

$$\iint\limits_D f \,\mathrm{d}x\mathrm{d}y = \int_1^2 \mathrm{d}x \int_x^{2x} \mathrm{e}^{\frac{y}{x}} \,\mathrm{d}y$$
$$= \int_1^2 \mathrm{d}x \ x\mathrm{e}^{\frac{y}{x}}\Big|_x^{2x} = \int_1^2 (\mathrm{e}^2 x - \mathrm{e}x)\,\mathrm{d}x = \frac{3}{2}(\mathrm{e}^2 - \mathrm{e}).$$

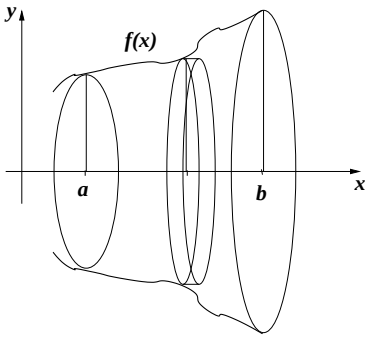
But trying to reverse the order of integration we encounter two problems. First, we must break D in several regions:

$$\iint\limits_D f\,\mathrm{d}x\mathrm{d}y = \int_1^2 \mathrm{d}y \int_1^y \mathrm{e}^{y/x}\,\mathrm{d}x + \int_2^4 \mathrm{d}y \int_{y/2}^2 \mathrm{e}^{y/x}\,\mathrm{d}x.$$

This is not a serious problem. A greater problem is that $\mathrm{e}^{1/x}$ has no elementary antiderivative, so $\int_1^y \mathrm{e}^{y/x}\,\mathrm{d}x$ and $\int_{y/2}^2 \mathrm{e}^{y/x}\,\mathrm{d}x$ are very difficult to evaluate. In this example, there is a considerable advantage in one order of integration over the other.

The Cosmopolitan Integral

Let $f \geq 0$ be a continuous real-valued function on $[a, b]$.



There are some very special solids whose volumes can be expressed by integrals. The simplest such solid G is a "volume of revolution" obtained by revolving the region under the graph of $f \geq 0$ on $[a, b]$ around the horizontal axis. We apply Fubini's theorem to the set

$$G = \{(x, y, z) \in \mathbb{R}^3 \mid a \leq x \leq b, \quad y^2 + z^2 \leq f(x)^2\}.$$

Consequently, the volume of $v(G)$ is given by

$$v(G) = \iiint_G \mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_a^b \mathrm{d}x \left(\iint_{G_x} \mathrm{d}y\mathrm{d}z \right), \tag{9.1}$$

where $G_x = \{(y, z) \in \mathbb{R}^2 \mid y^2 + z^2 \leq f(x)^2\}$ is the closed disc of radius $f(x)$ around $(0, 0)$. For any fixed $x \in [a, b]$ its area is $v(G_x) = \iint_{G_x} \mathrm{d}y\mathrm{d}z = \pi f(x)^2$. Hence

$$v(G) = \pi \int_a^b f(x)^2\,\mathrm{d}x. \tag{9.2}$$

Example 9.3 We compute the volume of the ellipsoid obtained by revolving the graph of the ellipse

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

around the x-axis. We have $y^2 = f(x)^2 = b^2 \left(1 - \frac{x^2}{a^2}\right)$; hence

$$v(G) = \pi b^2 \int_{-a}^a \left(1 - \frac{x^2}{a^2}\right) \mathrm{d}x = \pi b^2 \left(x - \frac{x^3}{3a^2}\right)\Big|_{-a}^a = \pi b^2 \left(2a - \frac{2a^3}{3a^2}\right) = \frac{4\pi}{3} b^2 a.$$

9.3 Change of Variable

We want to generalize change of variables formula $\int_{g(a)}^{g(b)} f(x)\,\mathrm{d}x = \int_a^b f(g(y))g'(y)\,\mathrm{d}y$.

If B_R is the ball in $\mathbb{R}^3$ with radius R around the origin we have in cartesian coordinates

$$\iiint_{B_R} f \,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_{-R}^{R} \mathrm{d}x \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \mathrm{d}y \int_{-\sqrt{R^2-x^2-y^2}}^{\sqrt{R^2-x^2-y^2}} \mathrm{d}z f(x,y,z).$$

Usually, the complicated limits yield hard computations. Here spherical coordinates are appropriate.

To motivate the formula consider the area of a parallelogram D in the x-y-plane spanned by the two vectors $a = (a_1, a_2)$ and $b = (b_1, b_2)$.

$$D = \{\lambda a + \mu b \mid \lambda, \mu \in [0,1]\} = \left\{ \begin{pmatrix} g_1(\lambda,\mu) \\ g_2(\lambda,\mu) \end{pmatrix} \middle| \, (\lambda, \mu \in [0,1] \right\},$$

where $g_1(\lambda,\mu) = \lambda a_1 + \mu b_1$ and $g_2(\lambda,\mu) = \lambda a_2 + \mu b_2$. As known from linear algebra the area of D equals the norm of the vector product

$$v(D) = \|a \times b\| = \left\| \det \begin{pmatrix} \mathrm{e}_1 & \mathrm{e}_2 & \mathrm{e}_3 \\ a_1 & a_2 & 0 \\ b_1 & b_2 & 0 \end{pmatrix} \right\| = \|(0,0,a_1b_2 - a_2b_1)\| = |\, a_1b_2 - a_2b_1 \,| =: d$$

Introducing new variables λ and μ with

$$x = \lambda a_1 + \mu b_1, \quad y = \lambda a_2 + \mu b_2,$$

the parallelogram D in the x-y-plane is now the unit square $C = [0,1] \times [0,1]$ in the λ-μ-plane and $D = g(C)$. We want to compare the area d of D with the area 1 of C. Note that d is exactly the absolute value of the Jacobian $\frac{\partial(g_1,g_2)}{\partial(\lambda,\mu)}$; indeed

$$\frac{\partial(g_1,g_2)}{\partial(\lambda,\mu)} = \det \begin{pmatrix} \frac{\partial g_1}{\partial \lambda} & \frac{\partial g_1}{\partial \mu} \\ \frac{\partial g_2}{\partial \lambda} & \frac{\partial g_2}{\partial \mu} \end{pmatrix} = \det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} = a_1b_2 - a_2b_1.$$

Hence,

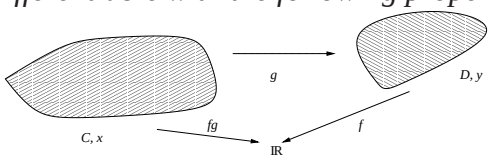
$$\iint_D \mathrm{d}x\mathrm{d}y = \iint_C \left| \frac{\partial(g_1,g_2)}{\partial(\lambda,\mu)} \right| \mathrm{d}\lambda\mathrm{d}\mu.$$

This is true for any $\mathbb{R}^n$ and any regular map, $g\colon C \to D$.

Theorem 9.6 (Change of variable) *Let C and D be compact Jordan set in $\mathbb{R}^n$; let $M \subset C$ a set of measure 0. Let $g\colon C \to D$ be continuously differentiable with the following properties*

(i) *g is injective on $C \setminus M$.*

(ii) *$g'(x)$ is regular on $C \setminus M$.*



Let $f\colon D \to \mathbb{R}$ be continuous.

Then

$$\int_D f(y)\,\mathrm{d}y = \int_C f(g(x)) \left| \frac{\partial(g_1,\ldots,g_n)}{\partial(x_1,\ldots,x_n)}(x) \right| \mathrm{d}x. \tag{9.3}$$

Remark 9.4 Why the *absolute value* of the Jacobian? In $\mathbb{R}^1$ we don't have the absolute value. But in contrast to $\mathbb{R}^n$, $n \geq 1$, we have an orientation of the integration set $\int_a^b f \,\mathrm{d}x = -\int_b^a f \,\mathrm{d}x$.

For the proof see [Rud76, 10.9 Theorem]. The main steps of the proof are: 1) In a small open set g can be written as the composition of n "flips" and n "primitive mappings". A flip changes two variables x_i and x_k, wheras a primitive mapping H is equal to the identity except for one variable, $H(x) = x + (h(x) - x)\mathrm{e}_m$ where $h\colon U \to \mathbb{R}$.
2) If the statement is true for transformations S and T, then it is true for the composition $S \circ T$ which follows from $\det(AB) = \det A \det B$.
3) Use a partition of unity.

Example 9.4 (a) Polar coordinates. Let $A = \{(r, \varphi) \mid 0 \leq r \leq R,\ 0 \leq \varphi < 2\pi\}$ be a rectangle in polar coordinates. The mapping $g(r, \varphi) = (x, y)$, $x = r\cos\varphi$, $y = r\sin\varphi$ maps this rectangle continuously differentiable onto the disc D with radius R. Let $M = \{(r, \varphi) \mid r = 0\}$. Since $\frac{\partial(x,y)}{\partial(r,\varphi)} = r$, the map g is bijective and regular on $A \setminus M$. The assumptions of the theorem are satisfied and we have

$$\begin{aligned}\iint_D f(x, y)\,\mathrm{d}x\mathrm{d}y &= \iint_A f(r\cos\varphi, r\sin\varphi) r \mathrm{d}r\mathrm{d}\varphi \\ &\underset{\text{Fubini}}{=} \int_0^R \int_0^{2\pi} f(r\cos\varphi, r\sin\varphi) r\,\mathrm{d}r\mathrm{d}\varphi.\end{aligned}$$

(b) Spherical coordinates. Recall from the exercise class the spherical coordinates $r \in [0, \infty)$, $\varphi \in [0, 2\pi]$, and $\vartheta \in [0, \pi]$

$$\begin{aligned}x &= r\sin\vartheta\cos\varphi, \\ y &= r\sin\vartheta\sin\varphi, \\ z &= r\cos\vartheta.\end{aligned}$$

The Jacobian reads

$$\frac{\partial(x, y, z)}{\partial(r, \vartheta, \varphi)} = \begin{vmatrix} x_r & x_\vartheta & x_\varphi \\ y_r & y_\vartheta & y_\varphi \\ z_r & z_\vartheta & z_\varphi \end{vmatrix} = \begin{vmatrix} \sin\vartheta\cos\varphi & r\cos\vartheta\cos\varphi & -r\sin\vartheta\sin\varphi \\ \sin\vartheta\sin\varphi & r\cos\vartheta\sin\varphi & r\sin\vartheta\cos\varphi \\ \cos\vartheta & -r\sin\vartheta & 0 \end{vmatrix} = r^2\sin\vartheta$$

Sometimes one uses

$$\frac{\partial(x, y, z)}{\partial(r, \varphi, \vartheta)} = -r^2\sin\vartheta.$$

Hence

$$\iiint_{B_1} f(x, y, z)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_0^1 \int_0^{2\pi} \int_0^\pi f(x, y, z)\sin^2\vartheta\,\mathrm{d}r\,\mathrm{d}\varphi\,\mathrm{d}\vartheta.$$

This example was not covered in the lecture. Compute the volume of the ellipsoid E given by $u^2/a^2 + v^2/b^2 + w^2/c^2 = 1$. We use scaled spherical coordinates:

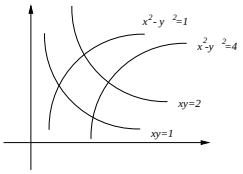
$$\begin{aligned}u &= ar\sin\vartheta\cos\varphi, \\ v &= br\sin\vartheta\sin\varphi, \\ w &= cr\cos\vartheta,\end{aligned}$$

where $r \in [0,1]$, $\vartheta \in [0,\pi]$, $\varphi \in 0, 2\pi]$. Since the rows of the spherical Jacobian matrix $\frac{\partial(x,y,z)}{\partial(r,\vartheta,\varphi)}$ are simply multiplied by a, b, and c, respectively, we have

$$\frac{\partial(u,v,w)}{\partial(r,\vartheta,\varphi)} = abcr^2 \sin\vartheta.$$

Hence, if B_1 is the unit ball around 0 we have using iterated integrals

$$\begin{aligned} v(E) &= \iiint_E \mathrm{d}u\mathrm{d}v\mathrm{d}w = abc \iiint_{B_1} r^2 \sin\vartheta \,\mathrm{d}r\mathrm{d}\vartheta\mathrm{d}\varphi \\ &= abc \int_0^1 \mathrm{d}r r^2 \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \sin\vartheta \mathrm{d}\vartheta \\ &= \frac{1}{3} abc\, 2\pi \;(-\cos\vartheta)|_0^{\pi} = \frac{4\pi}{3} abc. \end{aligned}$$



(c) $\iint_C (x^2+y^2)\,\mathrm{d}x\mathrm{d}y$ where C is bounded by the four hyperbolas $xy = 1$, $xy = 2$, $x^2 - y^2 = 1$, $x^2 - y^2 = 4$.
We change coordinates $g(x,y) = (u,v)$

$$u = xy, \quad v = x^2 - y^2.$$

The Jacobian is

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} y & x \\ 2x & -2y \end{vmatrix} = -2(x^2+y^2).$$

The Jacobian of the inverse transform is

$$\frac{\partial(x,y)}{\partial(u,v)} = -\frac{1}{2(x^2+y^2)}.$$

In the (u,v)-plane, the region is a rectangle $D = \{(u,v) \in \mathbb{R}^2 \mid 1 \le u \le 2,\ 1 \le v \le 4\}$. Hence,

$$\iint_C (x^2+y^2)\,\mathrm{d}x\mathrm{d}y = \iint_D (x^2+y^2) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| \mathrm{d}u\mathrm{d}v = \iint_D \frac{x^2+y^2}{2(x^2+y^2)}\,\mathrm{d}u\mathrm{d}v = \frac{1}{2}v(D) = \frac{3}{2}.$$

Physical Applications

If $\varrho(x) = \rho(x_1,x_2,x_3)$ is a mass density of a solid $C \subset \mathbb{R}^3$, then

$m = \int_C \rho\,\mathrm{d}x$ is the mass of C and

$\overline{x_i} = \frac{1}{m}\int_C x_i\,\rho(x)\,\mathrm{d}x,\ i = 1,\dots,3$ are the coordinates of the mass center $\overline{x}$ of C.

The *moments of inertia* of C are defined as follows

$$I_{xx}=\iiint\limits_C (y^2+z^2)\rho\,\mathrm{d}x\mathrm{d}y\mathrm{d}z, I_{yy}=\iiint\limits_C (x^2+z^2)\rho\,\mathrm{d}x\mathrm{d}y\mathrm{d}z, I_{zz}=\iiint\limits_C (x^2+y^2)\rho\,\mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

$$I_{xy}=\iiint\limits_C xy\rho\,\mathrm{d}x\mathrm{d}y\mathrm{d}z, \qquad I_{xz}=\iiint\limits_C xz\rho\,\mathrm{d}x\mathrm{d}y\mathrm{d}z, \qquad I_{yz}=\iiint\limits_C yz\rho\,\mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

Here I_{xx}, I_{yy}, and I_{zz} are the moments of inertia of the solid with respect to the x-axis, y-axis, and z-axis, respectively.

Example 9.5 Compute the mass center of a homogeneous half-plate of radius R, $C=\{(x,y)\mid x^2+y^2\le R^2,\ y\ge 0\}$.
Solution. By the symmetry of C with respect to the y-axis, $\overline{x}=0$. Using polar coordinates we find

$$\overline{y}=\frac{1}{m}\iint\limits_C y\,\mathrm{d}x\mathrm{d}y=\frac{1}{m}\int_0^R\int_0^\pi r\sin\varphi r\mathrm{d}\varphi\,\mathrm{d}r=\frac{1}{m}\int_0^R r^2\,\mathrm{d}r\,(-\cos\varphi)\,|_0^\pi=\frac{1}{m}\frac{2R^3}{3}.$$

Since the mass is proportional to the area, $m=\pi\frac{R^2}{2}$ and we find $(0,\frac{4R}{3\pi})$ is the mass center of the half-plate.

9.4 Appendix

Proof of Fubini's Theorem. Let P_A be a partition of A and P_B a partition of B. Together they give a partition P of $A\times B$ for which any subrectangle S is of the form $S_A\times S_B$, where S_A is a subrectangle of the partition P_A, and S_B is a subrectangle of the partition P_B. Thus

$$\begin{aligned}L(P,f)=\sum_S m_S\,v(S)&=\sum_{S_A,S_B} m_{S_A\times S_B}\,v(S_A\times S_B)\\&=\sum_{S_A}\left(\sum_{S_B} m_{S_A\times S_B}\,v(S_B)\right)v(S_A).\end{aligned}$$

Now, if $x\in S_A$, then clearly $m_{S_A\times S_B}(f)\le m_{S_B}(g_x)$ since the reference set $S_A\times S_B$ on the left is bigger than the reference set $\{x\}\times S_B$ on the right. Consequently, for $x\in S_A$ we have

$$\begin{aligned}&\sum_{S_B} m_{S_A\times S_B}\,v(S_B)\le\sum_{S_B} m_{S_B}(g_x)\,v(S_B)\le\underline{\int_B} g_x\,\mathrm{d}y=\mathcal{L}(x)\\&\sum_{S_B} m_{S_A\times S_B}\,v(S_B)\le m_{S_A}(\mathcal{L}(x)).\end{aligned}$$

Therefore,

$$\sum_{S_A}\left(\sum_{S_B} m_{S_A\times S_B}\,v(S_B)\right)v(S_A)\le\sum_{S_A} m_{S_A}(\mathcal{L}(x))v(S_A)=L(P_A,\mathcal{L}).$$

We thus obtain

$$L(P,f) \le L(P_A,\mathfrak{L}) \le U(P_A,\mathfrak{L}) \le U(P_A,\mathfrak{U}) \le U(P,f),$$

where the proof of the last inequality is entirely analogous to the proof of the first. Since f is integrable, $\sup\{L(P,f)\} = \inf\{U(P,f)\} = \int_{A\times B} f\,\mathrm{d}x\mathrm{d}y$. Hence,

$$\sup\{L(P_A,\mathfrak{L})\} = \inf\{U(P_A,\mathfrak{L})\} = \int_{A\times B} f\,\mathrm{d}x\mathrm{d}y.$$

In other words, $\mathfrak{L}(x)$ is integrable on A and $\int_{A\times B} f\,\mathrm{d}x\mathrm{d}y = \int_A \mathfrak{L}(x)\,\mathrm{d}x$.
The assertion for $\mathfrak{U}(x)$ follows similarly from the inequalities

$$L(P,f) \le L(P_A,\mathfrak{L}) \le L(P_A,\mathfrak{U}) \le U(P_A,\mathfrak{U}) \le U(P,f).$$

■

Chapter 10

Surface Integrals

10.1 Surfaces in $\mathbb{R}^3$

Recall that a *domain* G is an open and *connected* subset in $\mathbb{R}^n$; connected means that for any two points x and y in G, there exist points $x_0, x_1, \ldots, x_k$ with $x_0 = x$ and $x_k = y$ such that every segment $\overline{x_{i-1}x_i}$, $i = 1, \ldots, k$, is completely contained in G.

Definition 10.1 Let $G \subset \mathbb{R}^2$ be a domain and $F\colon G \to \mathbb{R}^3$ continuously differentiable. The mapping F as well as the set $\mathcal{F} = F(G) = \{F(s,t) \mid (s,t) \in G\}$ is called an *open regular surface* if the Jacobian matrix $F'(s,t)$ has rank 2 for all $(s,t) \in G$.

If

$$F(s,t) = \begin{pmatrix} x(s,t) \\ y(s,t) \\ z(s,t) \end{pmatrix},$$

the Jacobian matrix of F is

$$F'(s,t) = \begin{pmatrix} x_s & x_t \\ y_s & y_t \\ z_s & z_t \end{pmatrix}.$$

The two column vectors of $F'(s,t)$ span the tangent plane to F at (s,t):

$$D_1F(s,t) = \left(\frac{\partial x}{\partial s}(s,t), \frac{\partial y}{\partial s}(s,t), \frac{\partial z}{\partial s}(s,t)\right),$$
$$D_2F(s,t) = \left(\frac{\partial x}{\partial t}(s,t), \frac{\partial y}{\partial t}(s,t), \frac{\partial z}{\partial t}(s,t)\right).$$

Justification: Suppose $(s,t_0) \in G$ where t_0 is fixed. Then $\gamma(s) = F(s,t_0)$ defines a curve in $\mathcal{F}$ with tangent vector $\gamma'(s) = D_1F(s,t_0)$. Similarly, for fixed s_0 we obtain another curve $\tilde{\gamma}(t) = F(s_0,t)$ with tangent vector $\tilde{\gamma}'(t) = D_2F(s_0,t)$. Since $F'(s,t)$ has rank 2 at every point of G, the vectors D_1F and D_2F are linearly independent; hence they span a plane.

Definition 10.2 Let $F\colon G \to \mathbb{R}^3$ be an open regular surface, and $(s_0, t_0) \in G$. Then

$$\vec{x} = F(s_0, t_0) + \alpha D_1 F(s_0, t_0) + \beta D_2 F(s_0, t_0), \quad \alpha, \beta \in \mathbb{R}$$

is called the *tangent plane* E to F at $F(s_0, t_0)$. The line through $F(s_0, t_0)$ which is orthogonal to E is called the *normal line* to F at $F(s_0, t_0)$.

Recall that the vector product $\vec{x} \times \vec{y}$ of vectors $\vec{x} = (x_1, x_2, x_3)$ and $\vec{y} = (y_1, y_2, y_3)$ from $\mathbb{R}^3$ is the vector

$$\vec{x} \times \vec{y} = \begin{vmatrix} \mathrm{e}_1 & \mathrm{e}_2 & \mathrm{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = (x_2 y_3 - y_2 x_3, x_3 y_1 - y_3 x_1, x_1 y_2 - y_1 x_2).$$

It is orthogonal to the plane spanned by the parallelogram P with edges $\vec{x}$ and $\vec{y}$. Its length is the area of the parallelogram P.
A vector which points in the direction of the normal line is

$$D_1 F(s_0, t_0) \times D_2 F(s_0, t_0) = \begin{vmatrix} \mathrm{e}_1 & \mathrm{e}_2 & \mathrm{e}_3 \\ x_s & y_s & z_s \\ x_t & y_t & z_t \end{vmatrix} \tag{10.1}$$

$$\vec{n} = \pm \frac{D_1 F \times D_2 F}{\| D_1 F \times D_2 F \|}, \tag{10.2}$$

where $\vec{n}$ is the unit normal vector at (s_0, t_0).

Example 10.1 (Graph of a function) Let F be given by the graph of a function $f\colon G \to \mathbb{R}$, namely $F(x, y) = (x, y, f(x, y))$. By definition

$$D_1 F = (1, 0, f_x), \quad D_2 F = (0, 1, f_y),$$

hence

$$D_1 f \times D_2 f = \begin{vmatrix} \mathrm{e}_1 & \mathrm{e}_2 & \mathrm{e}_3 \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix} = (-f_x, -f_y, 1).$$

Therefore, the tangent plane has the equation

$$-f_x(x - x_0) - f_y(y - y_0) + 1(z - z_0) = 0.$$

Further, the unit normal vector to the tangent plane is

$$\vec{n} = \pm \frac{(f_x, f_y, -1)}{\sqrt{f_x^2 + f_y^2 + 1}}.$$

10.1.1 The Area of a Surface

Let F and $\mathcal{F}$ be as above. We assume that the continuous vector fields D_1F and D_2F on G can be extended to continuous functions on the closure $\overline{G}$.

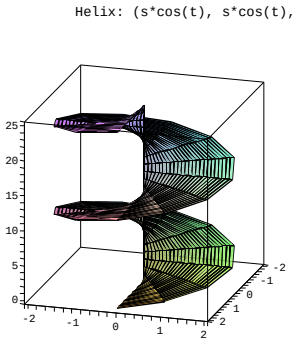
Definition 10.3 The number

$$|\,\mathcal{F}\,| = \left|\,\overline{\mathcal{F}}\,\right| := \iint\limits_G \|D_1F \times D_2F\|\,\mathrm{d}s\mathrm{d}t \tag{10.3}$$

is called the *area* of $\mathcal{F}$ and of $\overline{\mathcal{F}}$.
We call

$$\mathrm{d}S = \|D_1F \times D_2F\|\,\mathrm{d}s\mathrm{d}t$$

the scalar surface element of F. In this notation, $|\,\mathcal{F}\,| = \iint\limits_G \mathrm{d}S$.



Example 10.2 Let $\mathcal{F} = \{(s\cos t, s\sin t, 2t) \mid s \in [0,2],\ t \in [0,4\pi]\}$ be the surfaced spanned by a helix. We shall compute its area. The normal vector is

$$D_1F = (\cos t, \sin t, 0), \quad D_2F = (-s\sin t, s\cos t, 2)$$

such that

$$D_1F \times D_2F = \begin{vmatrix} \mathrm{e}_1 & \mathrm{e}_2 & \mathrm{e}_3 \\ \cos t & \sin t & 0 \\ -s\sin t & s\cos t & 2 \end{vmatrix} = (2\sin t, -2\cos t, s).$$

Therefore,

$$|\,\mathcal{F}\,| = \int_0^{4\pi}\int_0^2 \sqrt{4\cos^2 t + 4\sin^2 t + s^2}\,\mathrm{d}s\mathrm{d}t = 4\pi\int_0^2 \sqrt{4+s^2}\,\mathrm{d}s = 8\pi(\sqrt{2} - \log(\sqrt{2}-1)).$$

Example 10.3 (Guldin's Rule (Paul Guldin, 1577–1643, Swiss Mathematician)) Let f be a continuously differentiable function on $[a,b]$ with $f(x) \geq 0$ for all $x \in [a,b]$. Let the graph of f revolve around the x-axis and let $\mathcal{F}$ be the corresponding surface. We have

$$|\,\mathcal{F}\,| = 2\pi\int_a^b f(x)\sqrt{1+f'(x)^2}\,\mathrm{d}x.$$

Proof. Using polar coordinates in the y-z-plane, we obtain a parametrization of $\mathcal{F}$

$$\mathcal{F} = \{(x, f(x)\cos\varphi, f(x)\sin\varphi) \mid x \in [a,b], \varphi \in [0,2\pi]\}.$$

We have

$$D_1F = (1, f'(x)\cos\varphi, f'(x)\sin\varphi), \quad D_1F = (0, -f\sin\varphi, f\cos\varphi),$$
$$D_1F \times D_2F = (ff', -f\cos\varphi, -f\sin\varphi);$$

so that $\mathrm{d}S = f(x)\sqrt{1+f'(x)^2}\,\mathrm{d}x\mathrm{d}\varphi$. Hence

$$|\,\mathcal{F}\,| = \int_a^b \int_0^{2\pi} f(x)\sqrt{1+f'(x)^2}\mathrm{d}\varphi\,\mathrm{d}x = 2\pi\int_a^b f(x)\sqrt{1+f'(x)^2}\,\mathrm{d}x.$$

■

10.2 Scalar Surface Integrals

Let $\mathcal{F}$ and $\overline{\mathcal{F}}$ be as above, and $f\colon \overline{\mathcal{F}} \to \mathbb{R}$ a continuous function on the compact subset $\overline{\mathcal{F}} \subset \mathbb{R}^3$.

Definition 10.4 The number

$$\iint\limits_{\overline{\mathcal{F}}} f(\vec{x})\,\mathrm{d}S := \iint\limits_{\overline{G}} f(F(s,t))\,\|D_1F(s,t)\times D_2(s,t)\|\,\mathrm{d}s\mathrm{d}t$$

is called the *scalar surface integral of f on $\overline{\mathcal{F}}$*.

10.2.1 Other Forms for $\mathrm{d}S$

(a) Let the surface $\mathcal{F}$ be given as the graph of a function $F(x,y) = (x,y,f(x,y))$, $(x,y)\in G$. Then, see Example 10.1,

$$\mathrm{d}S = \sqrt{1+f_x^2+f_y^2}\,\mathrm{d}x\mathrm{d}y.$$

(b) Let the surface be given implicitly as $G(x,y,z) = 0$. Suppose G is locally solvable for z in a neighborhood of some point (x_0,y_0,z_0). Then the surface element (up to the sign) is given by

$$\mathrm{d}S = \frac{\sqrt{F_x^2+F_y^2+F_z^2}}{|\,F_z\,|}\,\mathrm{d}x\mathrm{d}y = \frac{\|\operatorname{grad}F\|}{|\,F_z\,|}\,\mathrm{d}x\mathrm{d}y.$$

One checks that $DF_1 \times DF_2 = (F_x, F_y, F_z)/F_z$.
(c) If $\mathcal{F}$ is given by $F(s,t) = (x(s,t), y(s,t), z(s,t))$ we have

$$\mathrm{d}S = \sqrt{EG-H^2}\,\mathrm{d}s\mathrm{d}t,$$

where

$$E = x_s^2+y_s^2+z_s^2, \quad G = x_t^2+y_t^2+z_t^2, \quad H = x_sx_t+y_sy_t+z_sz_t.$$

Indeed, using $\left\|\vec{a}\times\vec{b}\right\| = \|\vec{a}\|\ \left\|\vec{b}\right\|\sin\varphi$ and $\sin^2\varphi = 1-\cos^2\varphi$, where φ is the angle spanned by $\vec{a}$ and $\vec{b}$ we get

$$\begin{aligned} EG-H^2 &= \|D_1F\|^2\ \|D_2F\|^2 - (D_1F\cdot D_2F)^2 = \|D_1F\|^2\ \|D_2F\|^2\,(1-\cos^2\varphi)\\ &= \|D_1F\|^2\ \|D_2F\|^2\sin^2\varphi = \|D_1F\times D_2F\|^2 \end{aligned}$$

which proves the claim.

Example 10.4 (a) We give two different forms for the scalar surface element of a sphere. By (b), the sphere $x^2+y^2+z^2=R^2$ has surface element

$$\mathrm{d}S = \frac{\|(2x,2y,2z)\|}{2z}\,\mathrm{d}x\mathrm{d}y = \frac{R}{z}\,\mathrm{d}x\mathrm{d}y.$$

If

$$x = R\cos\varphi\sin\vartheta,\quad y = R\sin\varphi\sin\vartheta,\quad z = R\cos\vartheta,$$

we obtain

$$\begin{aligned} D_1 = F_\vartheta &= R(\cos\varphi\cos\vartheta, \sin\varphi\cos\vartheta, -\sin\vartheta),\\ D_2 = F_\varphi &= R(-\sin\varphi\sin\vartheta, \cos\varphi\sin\vartheta, 0),\\ D_1\times D_2 &= R^2(\cos\varphi\sin^2\vartheta, \sin\varphi\sin^2\vartheta, \sin\vartheta\cos\vartheta). \end{aligned}$$

Hence,

$$\mathrm{d}S = \|D_1\times D_2\|\,\mathrm{d}\vartheta\mathrm{d}\varphi = R^2\sin\vartheta\mathrm{d}\vartheta\mathrm{d}\varphi.$$

(b) **Riemann integral in $\mathbb{R}^3$ and surface integral over spheres.** Let $M = \{(x,y,z)\in\mathbb{R}^3 \mid \rho\le\|(x,y,z)\|\le R\}$ where $R>\rho\ge 0$. Let $f\colon M\to\mathbb{R}$ be continuous. Let us denote the sphere of radius r by $\mathrm{S}_r = \{(x,y,z)\in\mathbb{R}^3 \mid x^2+y^2+z^2=r^2\}$. Then

$$\iiint\limits_M f\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_\rho^R \mathrm{d}r\left(\iint\limits_{\mathrm{S}_r} f(\vec{x})\,\mathrm{d}S\right) = \int_\rho^R r^2\left(\iint\limits_{\mathrm{S}_1} f(r\vec{x})\,\mathrm{d}S(\vec{x})\right)\mathrm{d}r.$$

Indeed, by the previous example, and by our knowledge of spherical coordinates (r,ϑ,φ).

$$\mathrm{d}x\mathrm{d}y\mathrm{d}z = r^2\sin\vartheta\ \mathrm{d}r\,\mathrm{d}\vartheta\,\mathrm{d}\varphi = \mathrm{d}r\,\mathrm{d}S_r.$$

On the other hand, on the unit sphere S_1, $\mathrm{d}S = \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi$ such that

$$\mathrm{d}x\mathrm{d}y\mathrm{d}z = r^2\,\mathrm{d}r\,\mathrm{d}S$$

which establishes the second formula.

10.2.2 Physical Application

(a) If $\rho(x,y,z)$ is the mass density on a surface $\overline{\mathcal{F}}$, $\iint\limits_{\mathcal{F}} \rho \,\mathrm{d}S$ is the total mass of $\mathcal{F}$. The mass center of $\mathcal{F}$ has coordinates (x_c, y_c, z_c) with

$$x_c = \frac{1}{\iint\limits_{\mathcal{F}} \rho \,\mathrm{d}S} \iint\limits_{\mathcal{F}} x\,\rho(x,y,z)\,\mathrm{d}S,$$

and similarly for y_c and z_c.
(b) If $\sigma(\vec{x})$ is a charge density on a surface $\mathcal{F}$. Then

$$U(\vec{y}) = \iint\limits_{\mathcal{F}} \frac{\sigma(\vec{x})}{\|\vec{y}-\vec{x}\|}\,\mathrm{d}S(\vec{x}), \quad \vec{y} \notin \mathcal{F}$$

is the potential generated by $\mathcal{F}$.

10.3 Surface Integrals

10.3.1 Orientation

We want to define the notion of *orientation* for a regular surface. Let $\mathcal{F}$ be a regular (injective) surface with or without boundary. Then for every point $x_0 \in \mathcal{F}$ there exists the tangent plane E_{x_0}; the normal line to $\mathcal{F}$ at x_0 is uniquely defined.
However, a unit vector on the normal line can have two different directions.

Definition 10.5 (a) Let $\mathcal{F}$ be a surface as above. A *unit normal field* to $\mathcal{F}$ is a continuous function $\vec{n}\colon \mathcal{F} \to \mathbb{R}^3$ with the following two properties for every $x_0 \in \mathcal{F}$

(i) $\vec{n}(x_0)$ is orthogonal to the tangent plane to $\mathcal{F}$ at x_0.
(ii) $\|\vec{n}(x_0)\| = 1$.

(b) A regular surface $\mathcal{F}$ is called *orientable*, if there exists a unit normal field on $\mathcal{F}$.

Suppose $\mathcal{F}$ is an oriented, open, regular surface with piecewise smooth boundary $\partial\mathcal{F}$. Let $F(s,t)$ be a parametrization of $\mathcal{F}$. We assume that the vector functions F, DF_1, and DF_2 can be extended to continuous functions on $\overline{\mathcal{F}}$. The unit normal vector is given by

$$\vec{n} = \varepsilon \frac{D_1F \times D_2F}{\|D_1F \times D_2F\|},$$

where $\varepsilon = +1$ or $\varepsilon = -1$ fixes the *orientation* of $\mathcal{F}$. It turns out that for a regular surface $\mathcal{F}$ there either exists exactly two unit normal fields or there is no such field. If $\mathcal{F}$ is provided with an orientation we write $\mathcal{F}_+$ for the pair $(\mathcal{F}, \vec{n})$. For $\mathcal{F}$ with the opposite orientation, we write $\mathcal{F}_-$.

Examples of non-orientable surfaces are the *Möbius band* and the *real projective plane*. Analytically the Möbius band is given by

$$F(s,t) = \begin{pmatrix} \left(1 + t\cos\frac{s}{2}\right)\sin s \\ \left(1 + t\cos\frac{s}{2}\right)\cos s \\ t\sin\frac{s}{2} \end{pmatrix}, \quad (s,t) \in [0,2\pi] \times \left(-\frac{1}{2}, \frac{1}{2}\right).$$

Definition 10.6 Let $\vec{f}\colon \overline{\mathcal{F}} \to \mathbb{R}^3$ be a continuous vector field on $\overline{\mathcal{F}}$. The number

$$\iint\limits_{\mathcal{F}_+} \vec{f}(\vec{x}) \cdot \vec{n}\, \mathrm{d}S \tag{10.4}$$

is called the *surface integral of the vector field* $\vec{f}$ *on* $\mathcal{F}_+$. We call

$$\mathrm{d}\vec{S} = \vec{n}\, \mathrm{d}S = \varepsilon\, D_1F \times D_2F\, \mathrm{d}s\mathrm{d}t$$

the surface element of $\mathcal{F}$.

Remark 10.1 (a) The surface integral is independent of the parametrization of $\mathcal{F}$ but depends on the orientation; $\iint_{\mathcal{F}_+} \vec{f}\cdot \mathrm{d}\vec{S} = -\iint_{\mathcal{F}_-} \vec{f}\cdot \mathrm{d}\vec{S}$.
For, let $(s,t) = (s(\xi,\eta), t(\xi,\eta))$ be a new parametrization with $F(s(\xi,\eta), t(\xi,\eta)) = G(\xi,\eta)$. Then the Jacobian is

$$\mathrm{d}s\mathrm{d}t = \frac{\partial(s,t)}{\partial(\xi,\eta)}\, \mathrm{d}\xi\mathrm{d}\eta = (s_\xi t_\eta - s_\eta t_\xi)\, \mathrm{d}\xi\mathrm{d}\eta.$$

Further

$$D_1G = D_1F\, s_\xi + D_2F\, t_\xi, \quad D_2G = D_1F\, s_\eta + D_2F t_\eta,$$

so that using $\vec{x}\times\vec{x} = 0$, $\vec{x}\times\vec{y} = -\vec{y}\times\vec{x}$

$$\begin{aligned} D_1G \times D_2G \mathrm{d}\xi\mathrm{d}\eta &= (D_1F\, s_\xi + D_2F\, t_\xi) \times (D_1F\, s_\eta + D_2F t_\eta)\mathrm{d}\xi\mathrm{d}\eta, \\ &= (s_\xi t_\eta - s_\eta t_\xi) D_1F \times D_2F \mathrm{d}\xi\mathrm{d}\eta \\ &= D_1F \times D_2F\, \mathrm{d}s\mathrm{d}t. \end{aligned}$$

(b) The scalar surface integral is a special case of the surface integral, namely $\iint f\, \mathrm{d}S = \iint f\vec{n}\cdot\vec{n}\, \mathrm{d}S$.
(c) Special cases. Let $\mathcal{F}$ be the graph of a function f, $\mathcal{F} = \{(x,y,f(x,y)) \mid (x,y) \in C\}$, then

$$\mathrm{d}\vec{S} = \pm(-f_x, -f_y, 1)\, \mathrm{d}x\mathrm{d}y.$$

If the surface is given implicitly by $F(x,y,z) = 0$ and it is locally solvable for z, then

$$\mathrm{d}\vec{S} = \pm\frac{\operatorname{grad} F}{F_z}\, \mathrm{d}x\mathrm{d}y.$$

(d) Still another form of $\vec{\mathrm{d}S}$.

$$\iint\limits_{\overline{\mathcal{F}}} \vec{f}\,\vec{\mathrm{d}S} = \varepsilon \iint\limits_{\overline{G}} f(F(s,t)) \cdot (D_1F \times D_2F)\,\mathrm{d}s\mathrm{d}t$$

$$\vec{f}(F(s,t)) \cdot (D_1F \times D_2F) = \begin{vmatrix} f_1(F(s,t)) & f_2(F(s,t)) & f_3(F(s,t)) \\ x_s(s,t) & y_s(s,t) & z_s(s,t) \\ x_t(s,t) & y_t(s,t) & z_t(s,t) \end{vmatrix}. \tag{10.5}$$

(e) Again another notation. Computing the previous determinant or the determinant (10.1) explicitely we have

$$\vec{f}\cdot(D_1F\times D_2F) = f_1 \begin{vmatrix} y_s & z_s \\ y_t & z_t \end{vmatrix} + f_2 \begin{vmatrix} z_s & x_s \\ z_t & x_t \end{vmatrix} + f_3 \begin{vmatrix} x_s & y_s \\ x_t & y_t \end{vmatrix} = f_1 \frac{\partial(y,z)}{\partial(s,t)} + f_2 \frac{\partial(z,x)}{\partial(s,t)} + f_3 \frac{\partial(x,y)}{\partial(s,t)}.$$

Hence,

$$\vec{\mathrm{d}S} = \varepsilon D_1F \times D_2F\,\mathrm{d}s\mathrm{d}t = \varepsilon \left(\frac{\partial(y,z)}{\partial(s,t)}\,\mathrm{d}s\mathrm{d}t,\ \frac{\partial(z,x)}{\partial(s,t)}\,\mathrm{d}s\mathrm{d}t,\ \frac{\partial(x,y)}{\partial(s,t)}\,\mathrm{d}s\mathrm{d}t \right)$$

$$\vec{\mathrm{d}S} = \varepsilon\,(\,\mathrm{d}y\mathrm{d}z,\ \mathrm{d}z\mathrm{d}x,\ \mathrm{d}x\mathrm{d}y)\,.$$

Therefore we can write

$$\iint\limits_{\overline{\mathcal{F}}} \vec{f}\cdot\ \vec{\mathrm{d}S} = \iint\limits_{\overline{\mathcal{F}}} (f_1\,\mathrm{d}y\mathrm{d}z + f_2\,\mathrm{d}z\mathrm{d}x + f_3\,\mathrm{d}x\mathrm{d}y)\,.$$

In this setting

$$\iint\limits_{\overline{\mathcal{F}}} f_1\,\mathrm{d}y\mathrm{d}z = \iint\limits_{\overline{\mathcal{F}}} (f_1,0,0)\cdot\ \vec{\mathrm{d}S} = \pm \iint\limits_{\overline{G}} f_1(F(s,t)) \frac{\partial(y,z)}{\partial(s,t)}\,\mathrm{d}s\mathrm{d}t.$$

Sometimes one uses

$$\vec{\mathrm{d}S} = (\cos(\vec{n},\mathrm{e}_1), \cos(\vec{n},\mathrm{e}_2), \cos(\vec{n},\mathrm{e}_3))\,\mathrm{d}S,$$

since $\cos(\vec{n},\mathrm{e}_i) = \vec{n}\cdot\mathrm{e}_i = n_i$ and $\vec{\mathrm{d}S} = \vec{n}\,\mathrm{d}S$.
Note that we have surface integrals in the last two lines, not ordinary double integrals since $\overline{\mathcal{F}}$ is a surface in $\mathbb{R}^3$ and $f_1 = f_1(x,y,z)$ can also depend on x.
The physical meaning of $\iint_{\overline{\mathcal{F}}} \vec{f}\cdot\ \vec{\mathrm{d}S}$ is the flow of the vector field $\vec{f}$ through the surface $\mathcal{F}$. The flow is (locally) positive if $\vec{n}$ and $\vec{f}$ are on the same side of the tangent plane to $\mathcal{F}$ and negative in the other case.

Example 10.5 (a) Compute the surface integral

$$\iint\limits_{\mathcal{F}_+} f\,\mathrm{d}z\mathrm{d}x$$

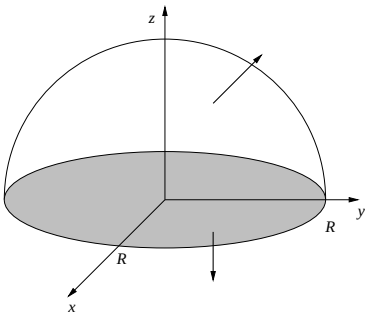
of $f(x,y,z) = x^2yz$ where $\mathcal{F}$ is the graph of $g(x,y) = x^2 + y$ over the unit square $G = [0,1]\times[0,1]$ with the downward directed unit normal field.

By Remark 10.1 (c)

$$\vec{\mathrm{d}S} = (g_x, g_y, -1)\,\mathrm{d}x\mathrm{d}y = (2x, 1, -1)\,\mathrm{d}x\mathrm{d}y.$$

Hence

$$\begin{aligned}\iint\limits_{\mathcal{F}_+} f\,\mathrm{d}z\mathrm{d}x &= \iint\limits_{\mathcal{F}_+} (0, f, 0)\cdot\,\vec{\mathrm{d}S}\\ &= (R)\iint\limits_G x^2y(x^2+y)\,\mathrm{d}x\mathrm{d}y = \int_0^1 \mathrm{d}x\int_0^1 (x^4y + x^2y^2)\,\mathrm{d}y = \frac{19}{90}.\end{aligned}$$



(b) Let G denote the upper half ball of radius R in $\mathbb{R}^3$:

$$G = \{(x, y, z) \mid x^2 + y^2 + z^2 \le R^2, \quad z \ge 0\},$$

and let $\mathcal{F}$ be the boundary of G with the orientation of the outer normal. Then $\mathcal{F}$ consists of the upper half sphere $\mathcal{F}_1$

$$\mathcal{F}_1 = \{(x, y, \sqrt{R^2 - x^2 - y^2}) \mid x^2 + y^2 \le R^2\}, \quad z = g(x, y) = \sqrt{R^2 - x^2 - y^2},$$

with the upper orientation of the unit normal field and of the disc $\mathcal{F}_2$ in the x-y-plane

$$\mathcal{F}_2 = \{(x, y, 0) \mid x^2 + y^2 \le R^2\}, \quad z = g(x, y) = 0,$$

with the downward directed normal. Let $\vec{f}(x, y, z) = (ax, by, cz)$. We want to compute

$$\iint\limits_{\mathcal{F}_+} \vec{f}\cdot\,\vec{\mathrm{d}S}.$$

By Remark 10.1 (c), the surface element of the half-sphere $\mathcal{F}_1$ is $\vec{\mathrm{d}S} = \frac{1}{z}(x, y, z)\,\mathrm{d}x\mathrm{d}y$. Hence

$$I_1 = \iint\limits_{\mathcal{F}_{1+}} \vec{f}\cdot\,\vec{\mathrm{d}S} = \iint\limits_{B_R} (ax, by, cz)\cdot\frac{1}{z}(x, y, z)\,\mathrm{d}x\mathrm{d}y = \iint\limits_{B_R} \frac{1}{z}(ax^2 + by^2 + cz^2)\Big|_{z=g(x,y)}\,\mathrm{d}x\mathrm{d}y.$$

Using polar coordinates $x = r\cos\varphi$, $y = r\sin\varphi$, $r \in [0, R]$, and $z = \sqrt{R^2 - x^2 - y^2} = \sqrt{R^2 - r^2}$ we get

$$I_1 = \int_0^{2\pi} \mathrm{d}\varphi \int_0^R \frac{ar^2\cos^2\varphi + br^2\sin^2\varphi + c(R^2 - r^2)}{\sqrt{R^2 - r^2}} r\mathrm{d}r.$$

Noting $\int_0^{2\pi}\sin^2\varphi\mathrm{d}\varphi = \int_0^{2\pi}\cos^2\varphi\mathrm{d}\varphi = \pi$ we continue

$$I_1 = \pi\int_0^R \left(\frac{ar^3}{\sqrt{R^2 - r^2}} + \frac{br^3}{\sqrt{R^2 - r^2}} + 2cr\sqrt{R^2 - r^2}\right)\,\mathrm{d}r.$$

Using $r = R\sin t,\ \mathrm{d}r = R\cos t\,\mathrm{d}t$ we have

$$\int_0^R \frac{r^3}{\sqrt{R^2-r^2}}\,\mathrm{d}r = \int_0^{\frac{\pi}{2}} \frac{R^3\sin^3 t R\cos t\,\mathrm{d}t}{R\sqrt{1-\sin^2 t}} = R^3\int_0^{\frac{\pi}{2}} \sin^3 t\,\mathrm{d}t = \frac{2}{3}R^3.$$

Hence,

$$\begin{aligned} I_1 &= \frac{2\pi}{3}R^3(a+b) + \pi c\int_0^R (R^2-r^2)^{\frac{1}{2}}\mathrm{d}(r^2) \\ &= \frac{2\pi}{3}R^3(a+b) + \pi c\left[-\frac{2}{3}(R^2-r^2)^{\frac{3}{2}}\right|_0^R \\ &= \frac{2\pi}{3}R^3(a+b+c). \end{aligned}$$

In case of the disc $\mathcal{F}_2$ we have $z = f(x,y) = 0$, such that $f_x = f_y = 0$ and

$$\vec{\mathrm{d}S} = (0,0,-1)\,\mathrm{d}x\mathrm{d}y$$

by Remark 10.1 (c). Hence

$$\iint\limits_{\mathcal{F}_{2+}} \vec{f}\cdot\,\vec{\mathrm{d}S} = \iint\limits_{B_R} (ax,by,cz)\cdot(0,0,-1)\,\mathrm{d}x\mathrm{d}y = -c\iint\limits_{B_R} z\,\mathrm{d}x\mathrm{d}y \underset{z=0}{=} 0.$$

Hence,

$$\iint\limits_{\mathcal{F}} (ax,by,cz)\cdot\,\vec{\mathrm{d}S} = \frac{2\pi}{3}R^3(a+b+c).$$

10.4 Gauß' Divergence Theorem

The aim is to generalize the fundamental theorem of calculus to higher dimensions:

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

Note that a and b form the boundary of the segment $[a,b]$. There are three possibilities to do this

$$\begin{array}{lcll} \iiint\limits_G g\,\mathrm{d}x\mathrm{d}y\mathrm{d}z & \Longrightarrow & \iint\limits_{(\partial G)_+} \vec{f}\cdot\,\vec{\mathrm{d}S} & \text{Gauß' theorem in } \mathbb{R}^3, \\ \iint\limits_G g\,\mathrm{d}x\mathrm{d}y & \Longrightarrow & \int\limits_{(\partial G)_+} \vec{f}\cdot\,\mathrm{d}\vec{x} & \text{Green's theorem in } \mathbb{R}^2, \\ \iint\limits_{\mathcal{F}_+} \vec{g}\cdot\,\vec{\mathrm{d}S} & \Longrightarrow & \int\limits_{(\partial G)_+} \vec{f}\cdot\,\mathrm{d}\vec{x} & \text{Stokes' theorem .} \end{array}$$

Let $G \subset \mathbb{R}^3$ be a bounded domain (open, connected) such that its boundary $\mathcal{F} = \partial G$ satisfies the following assumptions:

1. $\mathcal{F}$ is a union of regular, orientable surfaces $\mathcal{F}_i$. The parametrization $F_i(s,t)$, $(s,t) \in \overline{C_i}$, of $\mathcal{F}_i$ as well as D_1F_i and D_2F_i are continuous vector functions on $\overline{C_i}$; C_i is a domain in $\mathbb{R}^2$.

2. Let $\mathcal{F}_i$ be oriented by the **outer normal** (with respect to G).

3. There is given a continuously differentiable vector field $\vec{f}\colon \overline{G} \to \mathbb{R}^3$ on G (More precisely, there exist an open set $U \supset G$ and a continuously differentiable function $\tilde{f}\colon U \to \mathbb{R}^3$ such that $\tilde{f}\restriction\overline{G} = \vec{f}$.)

Theorem 10.1 (Gauß' Divergence Theorem) *Under the above assumptions we have*

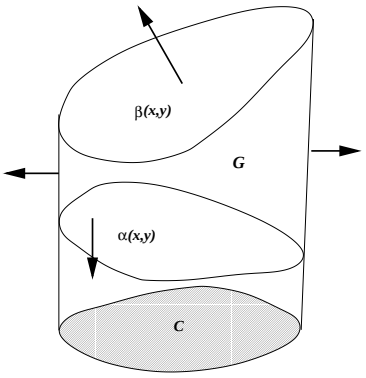
$$\iiint\limits_{\overline{G}} \operatorname{div} \vec{f}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint\limits_{\partial G_+} \vec{f}\cdot \mathrm{d}\vec{S} \tag{10.6}$$

Sometimes the theorem is called Gauß–Ostrogadski theorem or simply Ostrogadski theorem. Other writings:

$$\iiint\limits_{\overline{G}} \left(\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z}\right) \mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint\limits_{\partial G} (f_1\,\mathrm{d}y\mathrm{d}z + f_2\,\mathrm{d}z\mathrm{d}x + f_3\,\mathrm{d}x\mathrm{d}y) \tag{10.7}$$

The theorem holds for more general regions $G \subset \mathbb{R}^3$.

Proof.



We give a proof for

$$G = \{(x,y,z) \mid (x,y) \in C,\ \alpha(x,y) \le z \le \beta(x,y)\},$$

where $C \subset \mathbb{R}^2$ is a domain and $\alpha, \beta \in \mathrm{C}^1(C)$ define regular top and bottom surfaces $\mathcal{F}_1$ and $\mathcal{F}_2$ of $\mathcal{F}$, respectively. We prove only one part of (10.7) namely $\vec{f} = (0,0,f_3)$.

$$\iiint\limits_{\overline{G}} \frac{\partial f_3}{\partial z}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint\limits_{\partial G} f_3\,\mathrm{d}x\mathrm{d}y. \tag{10.8}$$

By Fubini's theorem, the left side reads

$$\begin{aligned}\iiint\limits_{\overline{G}} \frac{\partial f_3}{\partial x}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z &= \iint\limits_{C} \left(\int_{\alpha(x,y)}^{\beta(x,y)} \frac{\partial f_3}{\partial z}\,\mathrm{d}z\right) \mathrm{d}x\mathrm{d}y \\ &= \iint\limits_{C} (f_3(x,y,\beta(x,y)) - f_3(x,y,\alpha(x,y)))\,\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{10.9}$$

where the last equality is by the fundamental theorem of calculus.
Now we are going to compute the surface integral. The outer normal for the top surface is $(-\beta_x(x,y), -\beta_y(x,y), 1)$ such that

$$\begin{aligned} I_1 &= \iint\limits_{\mathcal{F}_{1+}} f_3\,\mathrm{d}x\mathrm{d}y = \iint\limits_{C} (0,0,f_3)\cdot(-\beta_x(x,y), -\beta_y(x,y), 1)\,\mathrm{d}x\mathrm{d}y \\ &= \iint\limits_{C} f_3(x,y,\beta(x,y))\,\mathrm{d}x\mathrm{d}y. \end{aligned}$$

Since the bottom surface $\mathcal{F}_2$ is oriented downward, the outer normal is $(\alpha_x(x,y), \alpha_y(x,y), -1)$ such that

$$I_2 = \iint\limits_{\mathcal{F}_{2+}} f_3 \,\mathrm{d}x\mathrm{d}y = -\iint\limits_C f_3(x,y,\alpha(x,y)) \,\mathrm{d}x\mathrm{d}y.$$

Finally, the shell $\mathcal{F}_3$ is parametrized by an angle φ and z:

$$\mathcal{F}_3 = \{(r(\varphi)\cos\varphi, r(\varphi)\sin\varphi, z) \mid \alpha(x,y) \le z \le \beta(x,y), \varphi \in [0, 2\pi]\}.$$

Since $D_2F = (0,0,1)$, the normal vector is orthogonal to the z-axis, $\vec{n} = (n_1, n_2, 0)$. Therefore,

$$I_3 = \iint\limits_{\mathcal{F}_{3+}} f_3 \,\mathrm{d}x\mathrm{d}y = \iint\limits_{\mathcal{F}_{3+}} (0,0,f_3) \cdot (n_1, n_2, 0) \,\mathrm{d}S = 0.$$

Comparing $I_1 + I_2 + I_3$ with (10.9) proves the theorem in this special case. ∎

Remarks 10.2 (a) Gauß' divergence theorem can be used to compute the volume of the domain $G \subset \mathbb{R}^3$. Suppose the boundary ∂G of G has the orientation of the outer normal. Then

$$v(G) = \iint\limits_{\partial G} x \,\mathrm{d}y\mathrm{d}z = \iint\limits_{\partial G} y \,\mathrm{d}z\mathrm{d}x = \iint\limits_{\partial G} z \,\mathrm{d}x\mathrm{d}y.$$

(b) Applying the mean value theorem to the left-hand side of Gauß' formula we have for any bounded region G containing x_0

$$\operatorname{div} \vec{f}(x_0 + h) \iiint\limits_G \mathrm{d}x\mathrm{d}y\mathrm{d}z = \operatorname{div} \vec{f}(x_0 + h) v(G) = \iint\limits_{\partial G_+} \vec{f} \cdot \mathrm{d}\vec{S},$$

where h is a small vector. The integral on the left is the volume $v(G)$. Hence

$$\operatorname{div} \vec{f}(x_0) = \lim_{G \to x_0} \frac{1}{v(G)} \iint\limits_{\partial G_+} \vec{f} \cdot \mathrm{d}\vec{S} = \lim_{\varepsilon \to 0} \frac{1}{\frac{4\pi\varepsilon^3}{3}} \iint\limits_{\mathrm{S}_\varepsilon(x_0)_+} \vec{f} \cdot \mathrm{d}\vec{S},$$

where the region G tends to x_0. In the second formula, we have chosen $G = B_\varepsilon(x_0)$ the open ball of radius ε with center x_0. The right hand side can be thought as to be the *source density* of the field $\vec{f}$. In particular, the right side gives a basis independent description of $\operatorname{div} \vec{f}$.

Example 10.6 We want to compute the surface integral from Example 10.5 (b) using Gauß' theorem:

$$\iint\limits_{\mathcal{F}_+} \vec{f} \cdot \mathrm{d}\vec{S} = \iiint\limits_C \operatorname{div} f \,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \iiint\limits_{x^2+y^2+z^2 \le R^2,\, z \ge 0} (a+b+c) \,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \frac{2\pi R^3}{3}(a+b+c).$$

Gauß' divergence theorem which play an important role in partial differential equations.
Recall (Proposition 7.9 (Prop. 8.9)) that the *directional derivative* of a function $v\colon U \to \mathbb{R}$, $U \subset \mathbb{R}^n$, at x_0 in the direction of the unit vector $\vec{n}$ is given by $D_{\vec{n}} f(x_0) = \operatorname{grad} f(x_0) \cdot \vec{n}$.
Notation. Let $U \subset \mathbb{R}^3$ be open and $\mathcal{F}_+ \subset U$ be an oriented, regular open surface with the unit normal vector $\vec{n}(x_0)$ at $x_0 \in \mathcal{F}$. Let $g\colon U \to \mathbb{R}$ be differentiable.
Then

$$\frac{\partial g}{\partial \vec{n}}(x_0) = \operatorname{grad} g(x_0) \cdot \vec{n}(x_0) \tag{10.10}$$

is called the *normal derivative* of g on $\mathcal{F}_+$ at x_0.

Proposition 10.2 *Let G be a region as in Gauß' theorem, the boundary ∂G is oriented with the outer normal, u, v are twice continuously differentiable on an open set U with $\overline{G} \subset U$. Then we have Green's identities:*

$$\iiint_G \nabla(u) \cdot \nabla(v)\, \mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint_{\partial G} u \frac{\partial v}{\partial \vec{n}}\, \mathrm{d}S - \iiint_G u\, \Delta(v)\, \mathrm{d}x\mathrm{d}y\mathrm{d}z, \tag{10.11}$$

$$\iiint_G (u\Delta(v) - v\Delta(u))\, \mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint_{\partial G} \left(u \frac{\partial v}{\partial \vec{n}} - v \frac{\partial u}{\partial \vec{n}} \right) \mathrm{d}S, \tag{10.12}$$

$$\iiint_G \Delta(u)\, \mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint_{\partial G} \frac{\partial u}{\partial \vec{n}}\, \mathrm{d}S. \tag{10.13}$$

Proof. Put $f = u\nabla(v)$. Then by nabla calculus

$$\begin{aligned} \operatorname{div} f = \nabla(u\nabla v) &= \nabla(u) \cdot \nabla(v) + u\nabla \cdot (\nabla v) \\ &= \operatorname{grad} u \cdot \operatorname{grad} v + u\Delta(v). \end{aligned}$$

Applying Gauß' theorem, we obtain

$$\begin{aligned} \iiint_G \operatorname{div} f\, \mathrm{d}x\mathrm{d}y\mathrm{d}z &= \iiint_G (\operatorname{grad} u \cdot \operatorname{grad} v)\, \mathrm{d}x\mathrm{d}y\mathrm{d}z + \iiint_G u\Delta(v)\, \mathrm{d}x\mathrm{d}y\mathrm{d}z \\ &= \iint_{\partial G} u \operatorname{grad} v \cdot \vec{n}\, \mathrm{d}S = \iint_{\partial G} u \frac{\partial v}{\partial \vec{n}}\, \mathrm{d}S. \end{aligned}$$

This proves Green's first identity. Changing the role of u and v and taking the difference, we obtain the second formula.
Inserting $v = -1$ into (10.12) we get (10.13). ∎

Application to Laplace equation

Let u_1 and u_2 be functions on G with $\Delta u_1 = \Delta u_2$, which coincide on the boundary ∂G, $u_1(x) = u_2(x)$ for all $x \in \partial G$. Then $u_1 \equiv u_2$ in G.

Proof. Put $u = u_1 - u_2$ and apply Green's first formula (10.11) to $u = v$. Note that $\Delta(u) = \Delta(u_1) - \Delta(u_2) = 0$ (U is harmonic ini G) and $u(x) = u_1(x) - u_2(x) = 0$ on the boundary $x \in \partial G$. In other words, a harmonic function is uniquely determined by its boundary values.

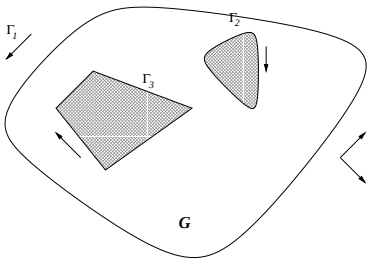
$$\iiint_G \nabla(u)\cdot\nabla(u)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint_{\partial G} \underbrace{u}_{0,x\in\partial G} \frac{\partial u}{\partial \vec{n}}\,\mathrm{d}S - \iiint u\,\underbrace{\Delta(u)}_{0}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = 0.$$

Since $\nabla(u)\cdot\nabla(u) = \|\nabla(u)\|^2 \geq 0$ and $\|\nabla(u)\|^2$ is a continuous function on $\overline{G}$, by homework 14.3, $\|\nabla(u)\|^2 = 0$ on $\overline{G}$; hence $\nabla(u) = 0$ on $\overline{G}$. By the Mean Value Theorem, Corollary 7.12, u is constant on $\overline{G}$. Since $u = 0$ on ∂G, $u(x) = 0$ for all $x \in G$. ∎

10.5 Stokes' Theorem

Roughly speaking, Stokes' theorem relates a surface integral over a surface $\mathcal{F}$ with a line integral over the boundary $\partial\mathcal{F}$. In case of a plane surface in $\mathbb{R}^2$, it is called Green's theorem.

10.5.1 Green's Theorem



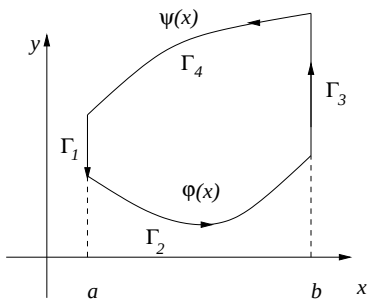
Let G be a domain in $\mathbb{R}^2$ with picewise smooth (differentiable) boundaries $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$. We give an orientation to the boundary: the outer curve is oriented counter clockwise (mathematical positive), the inner boundaries are oriented in the opposite direction.

Theorem 10.3 (Green's Theorem) *Let (P, Q) be a continuously differentiable vector field on $\overline{G}$ and let the boundary $\Gamma = \partial G$ be oriented as above. Then*

$$\iint_G \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathrm{d}x\mathrm{d}y = \int_\Gamma P\,\mathrm{d}x + Q\,\mathrm{d}y. \tag{10.14}$$

Proof. (a) First, we consider a region G of type 1 in the plane, as shown in the figure and we will prove that

$$-\iint_G \frac{\partial P}{\partial y}\,\mathrm{d}x\mathrm{d}y = \int_\Gamma P\,\mathrm{d}x. \tag{10.15}$$



The double integral on the left may be evaluated as an iterated integral (Fubini's theorem), we have

$$\begin{aligned}\iint_G \frac{\partial P}{\partial y}\,\mathrm{d}x\mathrm{d}y &= \int_a^b \left(\int_{\varphi(x)}^{\psi(x)} P_y(x,y)\,\mathrm{d}y\right)\mathrm{d}x \\ &= \int_a^b (P(x,\psi(x)) - P(x,\varphi(x)))\,\mathrm{d}x.\end{aligned}$$

The latter equality is due to the fundamental theorem of calculus. To compute the line integral, we parametrize the four parts of Γ in a natural way:

$$\begin{aligned}
-\Gamma_1, &\qquad \vec{x}_1(t) = (a,t), &&\qquad t \in [\varphi(a), \psi(a)], &&\qquad \mathrm{d}x = 0, \quad \mathrm{d}y = \mathrm{d}t,\\
\Gamma_2, &\qquad \vec{x}_2(t) = (t, \varphi(t)), &&\qquad t \in [a,b], &&\qquad \mathrm{d}x = \mathrm{d}t, \quad \mathrm{d}y = \varphi'(t)\,\mathrm{d}t,\\
\Gamma_3, &\qquad \vec{x}_3(t) = (b,t), &&\qquad t \in [\varphi(b), \psi(b)], &&\qquad \mathrm{d}x = 0, \quad \mathrm{d}y = \mathrm{d}t,\\
-\Gamma_4, &\qquad \vec{x}_4(t) = (t, \psi(t)), &&\qquad t \in [a,b], &&\qquad \mathrm{d}x = \mathrm{d}t, \quad \mathrm{d}y = \psi'(t)\,\mathrm{d}t.
\end{aligned}$$

Since $\mathrm{d}x = 0$ on Γ_1 and Γ_3 we are left with the line integrals over Γ_2 and Γ_4:

$$\int_\Gamma P\,\mathrm{d}x = \int_a^b P(t, \varphi(t))\,\mathrm{d}t - \int_a^b P(t, \psi(t))\,\mathrm{d}t$$

Let us prove the second part, $-\iint\limits_G \frac{\partial Q}{\partial x}\,\mathrm{d}x\mathrm{d}y = \int_\Gamma Q\,\mathrm{d}y$. Using Proposition 7.24 we have

$$\int_{\varphi(x)}^{\psi(x)} \frac{\partial}{\partial x} Q(x,y)\,\mathrm{d}y = \frac{\mathrm{d}}{\mathrm{d}x}\left(\int_{\varphi(x)}^{\psi(x)} Q(x,y)\,\mathrm{d}y\right) - \psi'(x)Q(x,\psi(x)) + \varphi'(x)Q(x,\varphi(x)).$$

Inserting this into $\iint\limits_G \frac{\partial Q}{\partial x}\,\mathrm{d}x\mathrm{d}y = \int_a^b \left(\int_{\varphi(x)}^{\psi(x)} \frac{\partial Q}{\partial x}\,\mathrm{d}y\right)\,\mathrm{d}x$, we get

$$\begin{aligned}
\iint\limits_G \frac{\partial Q}{\partial x}\,\mathrm{d}x\mathrm{d}y &= \int_a^b \left(\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_{\varphi(x)}^{\psi(x)} Q(x,y)\,\mathrm{d}y\right) - \psi'(x)Q(x,\psi(x)) + \varphi'(x)Q(x,\varphi(x))\right)\,\mathrm{d}x\\
&= \int_{\varphi(b)}^{\psi(b)} Q(b,y)\,\mathrm{d}y - \int_{\varphi(a)}^{\psi(a)} Q(a,y)\,\mathrm{d}y - \int_a^b Q(x,\psi(x))\,\psi'(x)\,\mathrm{d}x + \qquad (10.16)\\
&\qquad + \int_a^b Q(x,\varphi(x))\,\varphi'(x)\,\mathrm{d}x. \qquad (10.17)
\end{aligned}$$

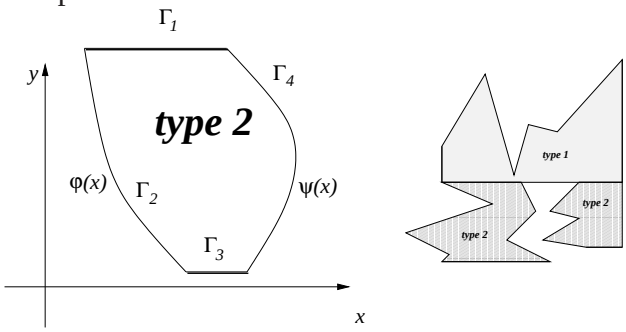
We compute the line integrals:

$$-\int_{-\Gamma_1} Q\ \mathrm{d}y = -\int_{\varphi(a)}^{\psi(a)} Q(a,y)\,\mathrm{d}y, \quad \int_{\Gamma_3} Q\ \mathrm{d}y = \int_{\varphi(b)}^{\psi(b)} Q(b,y)\,\mathrm{d}y.$$

Further,

$$\int_{\Gamma_2} Q\ \mathrm{d}y = \int_a^b Q(t,\varphi(t))\,\varphi'(t)\,\mathrm{d}t, \quad -\int_{-\Gamma_4} Q\ \mathrm{d}y = -\int_a^b Q(t,\psi(t))\,\psi'(t)\,\mathrm{d}t.$$

Adding up these integrals and comparing the result with (10.17), the proof for type 1 regions is complete.



Exactly in the same way, we can prove that if G is a type 2 region then (10.14) holds.

(b) Breaking a region G up into smaller regions, each of which is both of type 1 and 2, Green's theorem is valid for G. The line integrals along the inner boundary cancel leaving the line integral around the boundary of G.

(c) If the region has a hole, one can split it into two simply connected regions, for which Green's theorem is valid by the arguments of (b). ∎

Application: Area of a Region

If Γ is a curve which bounds a region G, then the area of G is $A = \int_\Gamma (1-\alpha)x\,\mathrm{d}y - \alpha y\,\mathrm{d}x$ where $\alpha \in \mathbb{R}$ is arbitrary, in particular,

$$A = \frac{1}{2}\int_\Gamma x\,\mathrm{d}y - y\,\mathrm{d}x = \int_\Gamma x\,\mathrm{d}y = -\int_\Gamma y\,\mathrm{d}x. \tag{10.18}$$

Proof. Choosing $Q = (1-\alpha)x$, $P = -\alpha y$ one has

$$\begin{aligned} A &= \iint\limits_G \mathrm{d}x\mathrm{d}y = \iint\limits_G ((1-\alpha)-(-\alpha))\,\mathrm{d}x\mathrm{d}y = \iint\limits_G (Q_x - P_y)\,\mathrm{d}x\mathrm{d}y = \int_\Gamma P\,\mathrm{d}x + Q\,\mathrm{d}y \\ &= -\alpha\int_\Gamma y\,\mathrm{d}x + (1-\alpha)\int_\Gamma x\,\mathrm{d}y. \end{aligned}$$

Inserting $\alpha = 0$, $\alpha = 1$, and $\alpha = \frac{1}{2}$ yields the assertion. ∎

Example 10.7 Find the area bounded by the ellipse Γ: $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$. We parametrize Γ by $\vec{x}(t) = (a\cos t, b\sin t)$, $t \in [0, 2\pi]$, $\dot{\vec{x}}(t) = (-a\sin t, b\cos t)$. Then (10.18) gives

$$A = \frac{1}{2}\int_0^{2\pi} a\cos t\, b\sin t\,\mathrm{d}t - b\sin t(-a\sin t)\,\mathrm{d}t = \frac{1}{2}\int_0^{2\pi} ab\,\mathrm{d}t = \pi ab.$$

10.5.2 Stokes' Theorem

Conventions: Let $\mathcal{F}_+$ be a regular, oriented surface. Let $\Gamma = \partial\mathcal{F}_+$ be the boundary of $\mathcal{F}$ with the *induced orientation:* the orientation of the surface (normal vector) together with the orientation of the boundary form a right-oriented screw. A second way to get the induced orientation: sitting in the arrowhead of the unit normal vector to the surface, the boundary curve has counter clockwise orientation.

Theorem 10.4 (Stokes' theorem) *Let $\mathcal{F}_+$ be a smooth regular oriented surface with a parametrization $F \in \mathrm{C}^2(G)$ and G is a plane region to which Green's theorem applies. Let*

$\Gamma = \partial\mathcal{F}$ be the boundary with the above orientation. Further, let $\vec{f}$ be a continuously differentiable vector field on $\overline{\mathcal{F}}$.
Then we have

$$\iint\limits_{\mathcal{F}_+} \operatorname{curl} \vec{f} \cdot \mathrm{d}\vec{S} = \int_{\partial\mathcal{F}_+} \vec{f} \cdot \mathrm{d}\vec{x}. \tag{10.19}$$

This can also be written as

$$\iint\limits_{\mathcal{F}_+} \left(\frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}\right) \mathrm{d}y\mathrm{d}z + \left(\frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x}\right) \mathrm{d}z\mathrm{d}x + \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right) \mathrm{d}x\mathrm{d}y = \int_\Gamma f_1 \,\mathrm{d}x + f_2 \,\mathrm{d}y + f_3 \,\mathrm{d}z.$$

Proof. Main idea: Reduction to Green's theorem. Since both sides of the equation are additive with respect to the vector field $\vec{f}$, it suffices to proof the statement for the vector fields $(f_1, 0, 0)$, $(0, f_2, 0)$, and $(0, 0, f_3)$. We show the theorem for $\vec{f} = (f, 0, 0)$, the other cases are quite analogous:

$$\iint\limits_{\mathcal{F}} \left(\frac{\partial f}{\partial z}\,\mathrm{d}z\mathrm{d}x - \frac{\partial f}{\partial y}\,\mathrm{d}x\mathrm{d}y\right) = \int_{\partial\mathcal{F}} f \,\mathrm{d}x.$$

Let $F(u, v)$, $u, v \in G$ be the parametrization of the surface $\mathcal{F}$. Then

$$\mathrm{d}x = \frac{\partial x}{\partial u}\,\mathrm{d}u + \frac{\partial x}{\partial v}\,\mathrm{d}v,$$

such that the line integral on the right reads with $P(u, v) = f(x(u, v), y(u, v), z(u, v)) \frac{\partial x}{\partial u}(u, v)$ and $Q(u, v) = f \frac{\partial x}{\partial v}$.

$$\begin{aligned}
\int_{\partial\mathcal{F}} f \,\mathrm{d}x &= \int_{\partial G} f\, x_u \,\mathrm{d}u + f\, x_v \,\mathrm{d}v = \int_{\partial G} P \,\mathrm{d}u + Q \,\mathrm{d}v \underset{\text{Green's th.}}{=} \iint\limits_G \left(-\frac{\partial P}{\partial v} + \frac{\partial Q}{\partial u}\right) \mathrm{d}u\,\mathrm{d}v \\
&= \iint\limits_G -(f_u x_u + f\, x_{vu}) + (f_u x_v + f\, x_{uv})\,\mathrm{d}u\,\mathrm{d}v = \iint\limits_G -f_v x_u + f_u x_v \,\mathrm{d}u\,\mathrm{d}v \\
&= \iint\limits_G -(f_x\, x_v + f_y\, y_v + f_z\, z_v) x_u + (f_x\, x_u + f_y\, y_u + f_z\, z_u) x_v \,\mathrm{d}u\,\mathrm{d}v \\
&= \iint\limits_G (-f_y(x_u y_v - x_v y_u) + f_z(z_u x_v - z_v x_u))\,\mathrm{d}u\,\mathrm{d}v \\
&= \iint\limits_G \left(-f_y \frac{\partial(x, y)}{\partial(u, v)} + f_z \frac{\partial(z, x)}{\partial(u, v)}\right) \mathrm{d}u\,\mathrm{d}v = \iint\limits_{\mathcal{F}} -f_y \,\mathrm{d}x\mathrm{d}y + f_z \,\mathrm{d}z\mathrm{d}x.
\end{aligned}$$

This completes the proof. ∎

Remark 10.3 (a) Green's theorem is a special case with $\mathcal{F} = G \times \{0\}$, $\vec{n} = (0, 0, 1)$ (orientation) and $\vec{f} = (P, Q, 0)$.

(b) The right side of (10.19) is called the *circulation* of the vector field $\vec{f}$ over the closed curve Γ. Now let $\vec{x}_0 \in \mathcal{F}$ be fixed and consider smaller and smaller neighborhoods $\mathcal{F}_{0+}$ of $\vec{x}_0$ with boundaries Γ_0. By Stokes' theorem and by the Mean Value Theorem of integration,

$$\int_{\Gamma_0} \vec{f}\,\mathrm{d}\vec{x} = \iint_{\mathcal{F}_0} \operatorname{curl}\vec{f}\cdot\vec{n}\,\mathrm{d}S = \operatorname{curl}\vec{f}(x_0)\vec{n}(x_0)\operatorname{area}(\mathcal{F}_0).$$

Hence,

$$\operatorname{curl}\vec{f}(x_0)\cdot\vec{n}(x_0) = \lim_{\mathcal{F}_0\to x_0} \frac{\int_{\partial\mathcal{F}_0}\vec{f}\,\mathrm{d}\vec{x}}{|\,\mathcal{F}_0\,|}.$$

We call $\operatorname{curl}\vec{f}(x_0)\cdot\vec{n}(x_0)$ the *infinitesimal circulation* of the vector field $\vec{f}$ at x_0 corresponding to the unit normal vector $\vec{n}$.
(c) Stokes' theorem then says that the integral over the infinitesimal circulation of a vector field $\vec{f}$ corresponding to the unit normal vector $\vec{n}$ over $\mathcal{F}$ equals the circulation of the vector field along the boundary of $\mathcal{F}$.

Path Independence of Line Integrals

We are going complete the proof of Proposition 8.3 and show that for a

> *simply* connected region $G \subset \mathbb{R}^3$ and
> a twice continuously differentiable vector field $\vec{f}$ with
> $\operatorname{curl}\vec{f} = 0$ for all $x \in G$

the vector field $\vec{f}$ is conservative.
Proof. Indeed, let Γ be a closed, regular, piecewise differentiable curve $\Gamma \subset G$ and let Γ be the the boundary of a smooth regular oriented surface $\mathcal{F}_+$, $\Gamma = \partial\mathcal{F}_+$ such that Γ has the induced orientation. Inserting $\operatorname{curl}\vec{f} = 0$ into Stokes' theorem gives

$$\iint_{\mathcal{F}_+} \operatorname{curl} f\cdot\mathrm{d}\vec{S} = 0 = \int_{\Gamma}\vec{f}\cdot\mathrm{d}\vec{x};$$

the line integral is path independent and hence, $\vec{f}$ is conservative. Note that the region must be simply connected; otherwise its in general impossible to find $\mathcal{F}$ with boundary Γ. ∎

10.5.3 Vector Potential and the Inverse Problem of Vector Analysis

Let $\vec{f}$ be a continuously differentiable vector field on the *simply connected* region $G \subset \mathbb{R}^3$.

Definition 10.7 The vector field $\vec{f}$ on G is called a *source-free* field (solenoidal field) if there exists a vector field $\vec{g}$ on G with $\vec{f} = \operatorname{curl}\vec{g}$. Then $\vec{g}$ is called the *vector potential* to $\vec{f}$.

Theorem 10.5 *$\vec{f}$ is source-free if and only if* $\operatorname{div}\vec{f} = 0$.

Proof. (a) If $\vec{f} = \operatorname{curl} \vec{g}$ then $\operatorname{div} \vec{f} = \operatorname{div}(\operatorname{curl} \vec{g}) = 0$.
(b) To simplify notations, we skip the arrows. We explicitly construct a vector potential g to f with $g = (g_1, g_2, 0)$ and $\operatorname{curl} g = f$. This means

$$\begin{aligned} f_1 &= -\frac{\partial g_2}{\partial z}, \\ f_2 &= \frac{\partial g_1}{\partial z}, \\ f_3 &= \frac{\partial g_2}{\partial x} - \frac{\partial g_1}{\partial y}. \end{aligned}$$

Integrating the first two equations, we obtain

$$\begin{aligned} g_2 &= -\int_{z_0}^{z} f_1(x, y, t)\,\mathrm{d}t + h(x, y), \\ g_1 &= \int_{z_0}^{z} f_2(x, y, t)\,\mathrm{d}t, \end{aligned}$$

where $h(x, y)$ is the integration constant, not depending on z. Inserting this into the third equation, we obtain

$$\begin{aligned} \frac{\partial g_2}{\partial x} - \frac{\partial g_1}{\partial y} &= -\int_{z_0}^{z} \frac{\partial f_1}{\partial x}(x, y, t)\,\mathrm{d}t + h_x(x, y) - \int_{z_0}^{z} \frac{\partial f_2}{\partial y}(x, y, t)\,\mathrm{d}t \\ &= -\int_{z_0}^{z} \left(\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} \right) \mathrm{d}t + h_x \\ &\underset{\operatorname{div} f = 0}{=} \int_{z_0}^{z} \frac{\partial f_3}{\partial z}(x, y, t)\,\mathrm{d}t + h_x \\ f_3(x, y, z) &= f_3(x, y, z) - f_3(x, y, z_0) + h_x(x, y). \end{aligned}$$

This yields, $h_x(x, y) = f_3(x, y, z_0)$. Integration with respect to x finally gives $h(x, y) = \int_{x_0}^{x} f_3(t, y, z_0)\,\mathrm{d}t$; the third equation is satisfied and $\operatorname{curl} g = f$. ∎

Remarks 10.4 (a) The proof of second direction is a constructive one; you can use this method to calculate a vector potential explicitly. You can also try another ansatz, say $g = (0, g_2, g_3)$ or $g = (g_1, 0, g_3)$.
(b) If g is a vector potential for f and $U \in \mathrm{C}^2(G)$, then $\tilde{g} = g + \operatorname{grad} U$ is also a vector potential for f. Indeed

$$\operatorname{curl} \tilde{g} = \operatorname{curl} g + \operatorname{curl} \operatorname{grad} U = f.$$

The Inverse Problem of Vector Analysis

Let h be a function and $\vec{a}$ be a vector field on G; both continuously differentiable.
Problem: Does there exist a vector field $\vec{f}$ such that

$$\operatorname{div} \vec{f} = h \quad \text{and} \quad \operatorname{curl} \vec{f} = \vec{a}.$$

Proposition 10.6 *The above problem has a solution if and only if* $\operatorname{div} \vec{a} = 0$.

Proof. The condition is necessary since $\operatorname{div} \vec{a} = \operatorname{div} \operatorname{curl} \vec{f} = 0$. We skip the vector arrows. For the other direction we use the ansatz $f = r + s$ with

$$\operatorname{curl} r = 0, \qquad \operatorname{div} r = h, \tag{10.20}$$

$$\operatorname{curl} s = a, \qquad \operatorname{div} s = 0. \tag{10.21}$$

Since $\operatorname{curl} r = 0$, by Proposition 8.3 there exists a potential U with $r = \operatorname{grad} U$. Then $\operatorname{curl} r = 0$ and $\operatorname{div} r = \operatorname{div} \operatorname{grad} U = \Delta(U)$. Hence (10.20) is satisfied if and only if $r = \operatorname{grad} U$ and $\Delta(U) = h$.
Since $\operatorname{div} a = 0$ by asssumption, there exists a vector potential g such that $\operatorname{curl} g = a$. Let φ be twice continuously differentiable on G and set $s = g + \operatorname{grad} \varphi$. Then $\operatorname{curl} s = \operatorname{curl} g = a$ and $\operatorname{div} s = \operatorname{div} g + \operatorname{div} \operatorname{grad} \varphi = \operatorname{div} g + \Delta(\varphi)$. Hence, $\operatorname{div} s = 0$ if and only if $\Delta(\varphi) = -\operatorname{div} g$. Both equations $\Delta(U) = h$ and $\Delta(\varphi) = -\operatorname{div} g$ are so called Poisson equations which can be solved within the theory of partial differential equations (PDE). ∎

The inverse problem has *not* a unique solution. Choose a harmonic function ψ, $\Delta(\psi) = 0$ and put $f_1 = f + \operatorname{grad} \psi$. Then

$$\begin{aligned} \operatorname{div} f_1 &= \operatorname{div} f + \operatorname{div} \operatorname{grad} \psi = \operatorname{div} f + \Delta(\psi) = \operatorname{div} f = h, \\ \operatorname{curl} f_1 &= \operatorname{curl} f + \operatorname{curl} \operatorname{grad} \psi = \operatorname{curl} f = a. \end{aligned}$$

Chapter 11

Differential Forms on $\mathbb{R}^n$

We show that Gauß', Green's and Stokes' theorems are three cases of a "general" theorem which is also named after Stokes. The simple formula now reads $\int_c \mathrm{d}\omega = \int_{\partial c} \omega$. The appearance of the Jacobian in the change of variable theorem will become clear. We formulate the Poincaré lemma.
Good references are [Spi65], [AF01], and [vW81].

11.1 The Exterior Algebra $\Lambda(\mathbb{R}^n)$

Although we are working with the ground field $\mathbb{R}$ all constructions make sense for arbitrary fields $\mathbb{K}$, in particular, $\mathbb{K} = \mathbb{C}$. Let $\{\mathrm{e}_1, \ldots, \mathrm{e}_n\}$ be the standard basis of $\mathbb{R}^n$; for $h \in \mathbb{R}^n$ we write $h = (h_1, \ldots, h_n)$ with respect to the standard basis, $h = \sum_i h_i \mathrm{e}_i$.

11.1.1 The Dual Vector Space V^*

The interplay between a normed space E and its dual space E' forms the basis of *functional analysis*. We start with the definition of the (algebraic) dual.

Definition 11.1 Let V be a linear space. The *dual* vector space V^* to V is the set of all linear functionals $f\colon V \to \mathbb{R}$,

$$V^* = \{f\colon V \to \mathbb{R} \mid \ f \text{ is linear}\}.$$

It turns out that V^* is again a linear space if we introduce addition and scalar multiples in the natural way. For $f, g \in V^*$, $\alpha \in \mathbb{R}$ put

$$(f+g)(v) := f(v) + g(v), \quad (\alpha f)(v) := \alpha f(v).$$

The *evaluation* of $f \in V^*$ on $v \in V$ is sometimes denoted by

$$f(v) = \langle f\,,\, v\rangle \in \mathbb{K}.$$

In this case, the brackets denote the *dual pairing* between V^* and V. By definition, the pairing is linear in both components. That is, for all $v, w \in V$ and for all $\lambda,\, \mu \in \mathbb{R}$

$$\begin{aligned}\langle \lambda f + \mu g\,,\, v\rangle &= \lambda\,\langle f\,,\, v\rangle + \mu\,\langle g\,,\, v\rangle\,,\\ \langle f\,,\, \lambda v + \mu w\rangle &= \lambda\,\langle f\,,\, v\rangle + \mu\,\langle f\,,\, w\rangle\,.\end{aligned}$$

Example 11.1 (a) Let $V = \mathbb{R}^n$ with the above standard basis. For $i = 1, \ldots, n$ define the *ith coordinate functional* $\mathrm{d}x_i \colon \mathbb{R}^n \to \mathbb{R}$ by

$$\mathrm{d}x_i(h) = \mathrm{d}x_i(h_1, \ldots, h_n) = h_i, \quad h \in \mathbb{R}^n.$$

The functional $\mathrm{d}x_i$ associates to each vector $h \in \mathbb{R}^n$ its ith coordinate h_i. The functional $\mathrm{d}x_i$ is indeed linear since for all $v, w \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, $\mathrm{d}x_i(\alpha v + \beta w) = (\alpha v + \beta w)_i = \alpha v_i + \beta w_i = \alpha\,\mathrm{d}x_i(v) + \beta\,\mathrm{d}x_i(w)$.
The linear space $(\mathbb{R}^n)^*$ has also dimension n. We will show that $\{\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_n\}$ is a basis of $(\mathbb{R}^n)^*$. We call it the *dual basis* of to $\{\mathrm{e}_1, \ldots, \mathrm{e}_n\}$. Using the Kronecker symbol the evaluation of $\mathrm{d}x_i$ on e_j reads as follows

$$\mathrm{d}x_i(\mathrm{e}_j) = \delta_{ij}, \quad i, j = 1, \ldots, n.$$

$\{\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_n\}$ *generates* V^*. Indeed, let $f \in V^*$. Then $f = \sum_{i=1}^n f(\mathrm{e}_i)\,\mathrm{d}x_i$ since both coincide for all $h \in V$:

$$\sum_{i=1}^n f(\mathrm{e}_i)\;\mathrm{d}x_i(h) = \sum_{i=1}^n f(\mathrm{e}_i)\,h_i \underset{f \text{ homog.}}{=} \sum_i f(h_i\mathrm{e}_i) = f\left(\sum_i h_i\mathrm{e}_i\right) = f(h).$$

In Propostition 11.1 below, we will see that $\{\mathrm{d}x_1, \ldots, \mathrm{d}x_n\}$ is not only generating but linearly independent.
(b) If $V = \mathrm{C}([0,1])$, the continuous functions on $[0,1]$ and α is an increasing on $[0,1]$ function, then the Riemann-Stieltjes integral

$$\varphi_\alpha(f) = \int_0^1 f\,\mathrm{d}\alpha, \quad f \in V$$

defines a linear functional φ_α on V.
If $a \in [0,1]$,

$$\delta_a(f) = f(a), \quad f \in V$$

defines the *evaluation functional* of f at a. In case $a = 0$ this is Dirac's δ-functional playing an important role in the theory of distributions (generalized functions).
(c) Let $a \in \mathbb{R}^n$. Then $\langle a\,,\, x\rangle = \sum_{i=1}^n a_i x_i$, $x \in \mathbb{R}^n$ defines a linear functional on $\mathbb{R}^n$. By (a) this is already the most general form of a linear functional on $\mathbb{R}^n$.

Definition 11.2 Let $k \in \mathbb{N}$. An *alternating (or skew-symmetric) multilinear form of degree k on $\mathbb{R}^n$*, a *k-form* for short, is a mapping $\omega\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$, k factors $\mathbb{R}^n$, which is *multilinear* and *skew-symmetric*, i. e.

$$\text{MULT} \quad \omega(\cdots, \alpha v_i + \beta w_i, \cdots) = \alpha\omega(\cdots, v_i, \cdots) + \beta\omega(\cdots, w_i, \cdots), \tag{11.1}$$

$$\text{SKEW} \quad \omega(\cdots, v_i, \cdots, v_j, \cdots) = -\omega(\cdots, v_j, \cdots, v_i, \cdots), \quad i, j = 1, \ldots, k, i \neq j, \tag{11.2}$$

for all vectors $v_1, v_2, \ldots, v_k, w_i \in \mathbb{R}^n$.

We denote the linear space of all k-forms on $\mathbb{R}^n$ by $\Lambda^k(\mathbb{R}^n)$ with the convention $\Lambda^0(\mathbb{R}^n) = \mathbb{R}$. In case $k = 1$ property (11.2) is an empty condition such that $\Lambda^1(\mathbb{R}^n) = (\mathbb{R}^n)^*$ is just the dual space.
Let $f_1, \ldots, f_k \in (\mathbb{R}^n)^*$ be linear functionals on $\mathbb{R}^n$. Then we define the k-form $f_1 \wedge \cdots \wedge f_k \in \Lambda^k(\mathbb{R}^n)$ (read: "f_1 wedge f_2 … wedge f_k") as follows

$$f_1 \wedge \cdots \wedge f_k(h_1, \ldots, h_k) = \begin{vmatrix} f_1(h_1) & \cdots & f_1(h_k) \\ \vdots & & \vdots \\ f_k(h_1) & \cdots & f_k(h_k) \end{vmatrix} \tag{11.3}$$

In particular, let $i_1, \ldots, i_k \in \{1, \ldots, n\}$ be fixed and choose $f_j = \mathrm{d}x_{i_j}$, $j = 1, \ldots, k$. Then

$$\mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}(h_1, \ldots, h_k) = \begin{vmatrix} h_{1i_1} & \cdots & h_{ki_1} \\ \vdots & & \vdots \\ h_{1i_k} & \cdots & h_{ki_k} \end{vmatrix}$$

$f_1 \wedge \cdots \wedge f_k$ is indeed a k-form since the f_i are linear, the determinant is multilinear

$$\begin{vmatrix} \lambda a + \mu a' & b & c \\ \lambda d + \mu d' & e & f \\ \lambda g + \mu g' & h & i \end{vmatrix} = \lambda \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} + \mu \begin{vmatrix} a' & b & c \\ d' & e & f \\ g' & h & i \end{vmatrix},$$

and skew-symmetric

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = - \begin{vmatrix} b & a & c \\ e & d & f \\ h & g & i \end{vmatrix}.$$

For example, let $y = (y_1, \ldots, y_n)$, $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$,

$$\mathrm{d}x_3 \wedge \mathrm{d}x_1(y, z) = \begin{vmatrix} y_3 & z_3 \\ y_1 & z_1 \end{vmatrix} = y_3 z_1 - y_1 z_3.$$

If $f_r = f_s = f$ for some $r \neq s$, we have $f_1 \wedge \cdots \wedge f \wedge \cdots \wedge f \wedge \cdots \wedge f_k = 0$ since determinants with identical rows vanish. Also, for any $\omega \in \Lambda^k(\mathbb{R}^n)$,

$$\omega(h_1, \ldots, h, \ldots, h, \ldots, h_k) = 0, \quad h_1, \ldots, h_k, h \in \mathbb{R}^n$$

since the defining determinant has two identical columns.

Proposition 11.1 *For $k \le n$ the k-forms $\{\, \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k} \mid 1 \le i_1 < i_2 < \cdots < i_k \le n\}$ form a basis of the vector space $\Lambda^k(\mathbb{R}^n)$. A k-form with $k > n$ is identically zero. We have*

$$\dim \Lambda^k(\mathbb{R}^n) = \binom{n}{k}.$$

Proof. Any k-form ω is uniquely determined by its values on the k-tuple of vectors $(\mathrm{e}_{i_1}, \ldots, \mathrm{e}_{i_k})$ with $1 \le i_1 < i_2 < \cdots < i_k \le n$. Indeed, using skew-symmetry of ω, we know ω on all k-tuples of basis vectors; using linearity in each component, we get ω on all k-tuples of vectors. This shows that the $\mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}$ with $1 \le i_1 < i_2 < \cdots < i_k \le n$ generate the linear space $\Lambda^k(\mathbb{R}^n)$. We make this precise in case $k = 2$. With $y = \sum_i y_i \mathrm{e}_i$, $z = \sum_j z_j \mathrm{e}_j$ we have by linearity and skew-symmetry of ω

$$\begin{aligned}\omega(y,z) &= \sum_{i,j=1}^{n} y_i z_j \omega(\mathrm{e}_i, \mathrm{e}_j) \underset{\text{SKEW}}{=} \sum_{1 \le i < j \le n} (y_i z_j - y_j z_i)\omega(\mathrm{e}_i, \mathrm{e}_j) \\ &= \sum_{i<j} \omega(\mathrm{e}_i, \mathrm{e}_j) \begin{vmatrix} y_i & z_i \\ y_j & z_j \end{vmatrix} = \sum_{i<j} \omega(\mathrm{e}_i, \mathrm{e}_j)\, \mathrm{d}x_i \wedge \mathrm{d}x_j(y,z).\end{aligned}$$

Hence,

$$\omega = \sum_{i<j} \omega(\mathrm{e}_i, \mathrm{e}_j)\, \mathrm{d}x_i \wedge \mathrm{d}x_j.$$

This shows that the $\binom{n}{2}$ 2-forms $\{\, \mathrm{d}x_i \wedge \mathrm{d}x_j \mid i < j\}$ generate $\Lambda^2(\mathbb{R}^n)$.
We show its linear independence. Suppose that $\sum_{i<j} \alpha_{ij}\, \mathrm{d}x_i \wedge \mathrm{d}x_j = 0$ for some $\alpha_{ij} \in \mathbb{R}$. Evaluating this on $(\mathrm{e}_r, \mathrm{e}_s)$, $r < s$, gives

$$0 = \sum_{i<j} \alpha_{ij}\, \mathrm{d}x_i \wedge \mathrm{d}x_j(\mathrm{e}_r, \mathrm{e}_s) = \sum_{i<j} \alpha_{ij} \begin{vmatrix} \delta_{ri} & \delta_{si} \\ \delta_{rj} & \delta_{sj} \end{vmatrix} = \sum_{i<j} \alpha_{ij}(\delta_{ri}\delta_{sj} - \delta_{rj}\delta_{si}) = \alpha_{rs};$$

hence, the above 2-forms are linearly independent. The arguments for general k are similar. ∎

In general, let $\omega \in \Lambda^k(\mathbb{R}^n)$ then there exist unique numbers $a_{i_1 \cdots i_k} = \omega(\mathrm{e}_{i_1}, \ldots, \mathrm{e}_{i_k}) \in \mathbb{R}$, $i_1 < i_2 < \cdots < i_k$ such that

$$\omega = \sum_{1 \le i_1 < \cdots < i_k \le n} a_{i_1 \cdots i_k}\, \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}.$$

Example 11.2 Let $n = 3$.
$k = 1$ $\{\, \mathrm{d}x_1,\ \mathrm{d}x_2,\ \mathrm{d}x_3\}$ is a basis of $\Lambda^1(\mathbb{R}^3)$.
$k = 2$ $\{\, \mathrm{d}x_1 \wedge \mathrm{d}x_2,\ \mathrm{d}x_1 \wedge \mathrm{d}x_3,\ \mathrm{d}x_2 \wedge \mathrm{d}x_3\}$ is a basis of $\Lambda^2(\mathbb{R}^3)$.
$k = 3$ $\{\, \mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}x_3\}$ is a basis of $\Lambda^3(\mathbb{R}^3)$.
$\Lambda^k(\mathbb{R}^3) = \{0\}$ for $k \ge 4$.

Definition 11.3 An *algebra* A over $\mathbb{R}$ is a linear space together with a product map $(a, b) \to ab$, $A \times A \to A$, such that the following holds for all $a, b, c \in A$ an d $\alpha \in \mathbb{R}$

(i) $a(bc) = (ab)c$ (associative),
(ii) $(a+b)c = ac + bc, \quad a(b+c) = ab + ac,$
(iii) $\alpha(ab) = (\alpha a)b = a(\alpha b).$

Standard examples are $\mathrm{C}(X)$, the continuous functions on a metric space X or $\mathbb{R}^{n\times n}$, the full $n \times n$-matrix algebra over $\mathbb{R}$ or the algebra of polynomials $\mathbb{R}[X]$.

Let $\Lambda(\mathbb{R}^n) = \bigoplus_{k=0}^{n} \Lambda^k(\mathbb{R}^n)$ be the direct sum of linear spaces.

Proposition 11.2 (i) *$\Lambda(\mathbb{R}^n)$ is an $\mathbb{R}$-algebra with unity 1 and product $\wedge$ defined by*

$$(\,\mathrm{d}x_{i_1}\wedge\cdots\wedge\mathrm{d}x_{i_k})\wedge(\,\mathrm{d}x_{j_1}\wedge\cdots\wedge\mathrm{d}x_{j_l}) = \mathrm{d}x_{i_1}\wedge\cdots\wedge\mathrm{d}x_{i_k}\wedge\mathrm{d}x_{j_1}\wedge\cdots\wedge\mathrm{d}x_{j_l}$$

(ii) *If $\omega_k \in \Lambda^k(\mathbb{R}^n)$ and $\omega_l \in \Lambda^l(\mathbb{R}^n)$ then $\omega_k\wedge\omega_l \in \Lambda^{k+l}(\mathbb{R}^n)$ and*

$$\omega_k\wedge\omega_l = (-1)^{kl}\omega_l\wedge\omega_k.$$

Proof. (i) Associativity is clear since concatanation of strings is associative. The distributive laws are used to extend multiplication from the basis to the entire space $\Lambda(\mathbb{R}^n)$.
We show (ii) for $\omega_k = \mathrm{d}x_{i_1}\wedge\cdots\wedge\mathrm{d}x_{i_k}$ and $\omega_l = \mathrm{d}x_{j_1}\wedge\cdots\wedge\mathrm{d}x_{j_l}$. We already know $\mathrm{d}x_i\wedge\mathrm{d}x_j = -\,\mathrm{d}x_j\wedge\mathrm{d}x_i$. There are kl transpositions $\mathrm{d}x_{i_r} \leftrightarrow \mathrm{d}x_{j_s}$ necessary to transport all $\mathrm{d}x_{j_s}$ from the right to the left of ω_k. Hence the sign is $(-1)^{kl}$. ∎

In particular, $\mathrm{d}x_i\wedge\mathrm{d}x_i = 0$. The formula $\mathrm{d}x_i\wedge\mathrm{d}x_j = -\,\mathrm{d}x_j\wedge\mathrm{d}x_i$ determines the product in $\Lambda(\mathbb{R}^n)$ uniquely.
We call $\Lambda(\mathbb{R}^n)$ is the *exterior algebra* of the vector space $\mathbb{R}^n$.
The following formula will be used in the next subsection. Let $\omega \in \Lambda^k(\mathbb{R}^n)$ and $\eta \in \Lambda^l(\mathbb{R}^n)$ then for all $v_1, \ldots, v_{k+l} \in \mathbb{R}^n$

$$(\omega\wedge\eta)(v_1,\ldots,v_{k+l}) = \frac{1}{k!l!}\sum_{\sigma\in \mathrm{S}_{k+l}} (-1)^\sigma \omega(v_{\sigma(1)},\ldots,v_{\sigma(k)})\,\eta(v_{\sigma(k+1)},\ldots,v_{\sigma(k+l)}). \qquad (11.4)$$

Indeed, let $\omega = f_1\wedge\cdots\wedge f_k$, $\eta = f_{k+1}\wedge\cdots\wedge f_{k+l}$, $f_i \in (\mathbb{R}^n)^*$. The above formula can be obtained from

$$(f_1\wedge\cdots\wedge f_k)\wedge(f_{k+1}\wedge\cdots\wedge f_{k+l})(v_1,\ldots,v_{k+l}) =$$

$$= \begin{vmatrix} f_1(v_1) & \cdots f_1(v_k) & f_1(v_{k+1}) & \cdots f_1(v_{k+l}) \\ f_2(v_1) & \cdots f_2(v_k) & f_2(v_{k+1}) & \cdots f_2(v_{k+l}) \\ \vdots & \vdots & \vdots & \\ f_k(v_1) & \cdots f_k(v_k) & f_k(v_{k+1}) & \cdots f_k(v_{k+l}) \\ f_{k+1}(v_1) & \cdots f_{k+1}(v_k) & f_{k+1}(v_{k+1}) & \cdots f_{k+1}(v_{k+l}) \\ \vdots & \vdots & \vdots & \\ f_{k+l}(v_1) & \cdots f_{k+l}(v_k) & f_{k+l}(v_{k+1}) & \cdots f_{k+l}(v_{k+l}) \end{vmatrix}$$

when expanding this determinant with respect to the last l rows. This can be done using Laplace expansion:

$$|A| = \sum_{1\le j_1<\cdots<j_k\le n} (-1)^{\sum_{m=1}^k (i_m+j_m)} \begin{vmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_k} \\ \vdots & & \vdots \\ a_{i_k j_1} & \cdots & a_{i_k j_k} \end{vmatrix} \begin{vmatrix} a_{i_{k+1} j_{k+1}} & \cdots & a_{i_{k+1} j_{k+l}} \\ \vdots & & \vdots \\ a_{i_{k+l} j_{k+1}} & \cdots & a_{i_{k+l} j_{k+l}} \end{vmatrix},$$

where $(i_1,\dots,i_k)$ is any fixed orderd multi-index and $(j_{k+1},\dots,j_{k+l})$ is the complementary orderd multi-index to $(j_1,\dots,j_k)$ such that all inegers $1,2,\dots,k+l$ appear.

11.1.2 The Pull-Back of k-forms

Definition 11.4 Let $A\in \mathrm{L}(\mathbb{R}^n,\mathbb{R}^m)$ a linear mapping and $k\in\mathbb{N}$. For $\omega\in \Lambda^k(\mathbb{R}^m)$ we define a k-form $A^*(\omega)\in\Lambda^k(\mathbb{R}^n)$ by

$$(A^*\omega)(h_1,\dots,h_k) = \omega(A(h_1),A(h_2),\dots,A(h_k)),\quad h_1,\dots,h_k\in\mathbb{R}^n$$

We call $A^*(\omega)$ the *pull-back* of ω under A.

Note that $A^*\in \mathrm{L}(\Lambda^k(\mathbb{R}^m),\Lambda^k(\mathbb{R}^n))$ is a linear mapping. In case $k=1$ we call A^* the *dual mapping* to A. In case $k=0$, $\omega\in\mathbb{R}$ we simply set $A^*(\omega)=\omega$. We have $A^*(\omega\wedge\eta) = A^*(\omega)\wedge A^*(\eta)$. Indeed, let $\omega\in\Lambda^k(\mathbb{R}^n)$, $\eta\in\Lambda^l(\mathbb{R}^n)$, and $h_i\in\mathbb{R}^n$, $i=1,\dots,k+l$, then by (11.4)

$$\begin{aligned} A^*(\omega\wedge\eta)(h_1,\dots,h_{k+l}) &= (\omega\wedge\eta)(A(h_1),\dots,A(h_{k+l})) \\ &= \frac{1}{k!l!}\sum_{\sigma\in \mathrm{S}_{k+l}} (-1)^\sigma \omega(A(v_{\sigma(1)}),\dots,A(v_{\sigma(k)}))\,\eta(A(v_{\sigma(k+1)}),\dots,A(v_{\sigma(k+l)})) \\ &= \frac{1}{k!l!}\sum_{\sigma\in \mathrm{S}_{k+l}} (-1)^\sigma A^*(\omega)(v_{\sigma(1)},\dots,v_{\sigma(k)})\,A^*(\eta)(v_{\sigma(k+1)},\dots,v_{\sigma(k+l)}) \\ &= (A^*(\omega)\wedge A^*(\eta))(h_1,\dots,h_{k+l}). \end{aligned}$$

Example 11.3 (a) Let $A=\begin{pmatrix}1&0&3\\2&1&0\end{pmatrix}\in\mathbb{R}^{2\times 3}$ be a linear map $A\colon \mathbb{R}^3\to\mathbb{R}^2$, defined by matrix multiplication, $A(v)=A\cdot v$, $v\in\mathbb{R}^3$. Let $\{\mathrm{e}_1,\mathrm{e}_2,\mathrm{e}_3\}$ and $\{f_1,f_2\}$ be the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$ resp. and let $\{\mathrm{d}x_1,\mathrm{d}x_2,\mathrm{d}x_3\}$ and $\{\mathrm{d}y_1,\mathrm{d}y_2\}$ their dual bases, respectvely. First we compute $A^*(\mathrm{d}y_1)$ and $A^*(\mathrm{d}y_2)$.

$$A^*(\mathrm{d}y_1)(\mathrm{e}_i) = \mathrm{d}y_1(A(\mathrm{e}_i)) = a_{i1},\quad A^*(\mathrm{d}y_2)(\mathrm{e}_i) = a_{i2}.$$

In particular,

$$A^*(\mathrm{d}y_1) = 1\,\mathrm{d}x_1 + 0\,\mathrm{d}x_2 + 3\,\mathrm{d}x_3,\quad A^*(\mathrm{d}y_2) = 2\,\mathrm{d}x_1 + \mathrm{d}x_2.$$

Compute $A^*(\mathrm{d}y_2\wedge\mathrm{d}y_1)$. By definition, for $1\le i<j\le 3$,

$$A^*(\mathrm{d}y_2\wedge\mathrm{d}y_1)(\mathrm{e}_i,\mathrm{e}_j) = \mathrm{d}y_2\wedge\mathrm{d}y_1(A(\mathrm{e}_i),A(\mathrm{e}_j)) = \mathrm{d}y_2\wedge\mathrm{d}y_1(A_i,A_j) = \begin{vmatrix} a_{i2} & a_{j2} \\ a_{j2} & a_{j1}\end{vmatrix}.$$

In particular

$$A^*(\,\mathrm{d}y_2\wedge\mathrm{d}y_1)(\mathrm{e}_1,\mathrm{e}_2)=\begin{vmatrix}2&1\\1&0\end{vmatrix}=-1,\quad A^*(\,\mathrm{d}y_2\wedge\mathrm{d}y_1)(\mathrm{e}_1,\mathrm{e}_3)=\begin{vmatrix}2&0\\1&3\end{vmatrix}=6,$$
$$A^*(\,\mathrm{d}y_2\wedge\mathrm{d}y_1)(\mathrm{e}_2,\mathrm{e}_3)=\begin{vmatrix}1&0\\3&0\end{vmatrix}=3.$$

Hence,

$$A^*(\,\mathrm{d}y_2\wedge\mathrm{d}y_1)=-\,\mathrm{d}x_1\wedge\mathrm{d}x_2+6\,\mathrm{d}x_1\wedge\mathrm{d}x_3+3\,\mathrm{d}x_2\wedge\mathrm{d}x_3.$$

On the other hand

$$A^*(\,\mathrm{d}y_2)\wedge A^*(\,\mathrm{d}y_1)=(2\,\mathrm{d}x_1+\mathrm{d}x_2)\wedge(\,\mathrm{d}x_1+3\,\mathrm{d}x_3)=-\,\mathrm{d}x_1\wedge\mathrm{d}x_2+3\,\mathrm{d}x_1\wedge\mathrm{d}x_3+6\,\mathrm{d}x_1\wedge\mathrm{d}x_3.$$

(b) Let $A\in\mathbb{R}^{n\times n}$, $A\colon\mathbb{R}^n\to\mathbb{R}^n$ and $\omega=\mathrm{d}x_1\wedge\cdots\wedge\mathrm{d}x_n\in\Lambda^n(\mathbb{R}^n)$. Then $A^*(\omega)=\det(A)\,\omega$.

11.1.3 Orientation of $\mathbb{R}^n$

If $\{e_1,\ldots,e_n\}$ and $\{f_1,\ldots,f_n\}$ are two bases of $\mathbb{R}^n$ there exists a unique regular matrix $A=(a_{ij})$ ($\det A\neq 0$) such that $e_i=\sum_j a_{ij}f_j$. We say that $\{e_1,\ldots,e_n\}$ and $\{f_1,\ldots,f_n\}$ are *equivalent* if and only if $\det A>0$. Since $\det A\neq 0$, there are exactly two equivalence classes. We say that the two bases $\{e_i\mid i=1,\ldots,n\}$ and $\{f_i\mid i=1,\ldots,n\}$ define *the same orientation* if and only if $\det A>0$.

Definition 11.5 An *orientation* of $\mathbb{R}^n$ is given by fixing one of the two equivalence classes.

Example 11.4 (a) In $\mathbb{R}^2$ the bases $\{\mathrm{e}_1,\mathrm{e}_2\}$ and $\{\mathrm{e}_2,\mathrm{e}_1\}$ have different orientations since $A=\begin{pmatrix}0&1\\1&0\end{pmatrix}$ and $\det A=-1$.
(b) In $\mathbb{R}^3$ the bases $\{\mathrm{e}_1,\mathrm{e}_2,\mathrm{e}_3\}$, $\{\mathrm{e}_3,\mathrm{e}_1,\mathrm{e}_2\}$ and $\{\mathrm{e}_2,\mathrm{e}_3,\mathrm{e}_1\}$ have the same orientation whereas $\{\mathrm{e}_1,\mathrm{e}_3,\mathrm{e}_2\}$, $\{\mathrm{e}_2,\mathrm{e}_1,\mathrm{e}_3\}$, and $\{\mathrm{e}_3,\mathrm{e}_2,\mathrm{e}_1\}$ have opposite orientation.
(c) The standard basis $\{\mathrm{e}_1,\ldots,\mathrm{e}_n\}$ and $\{\mathrm{e}_2,\mathrm{e}_1,\mathrm{e}_3,\ldots,\mathrm{e}_n\}$ define different orientations.

11.2 Differential Forms

11.2.1 Definition

Throughout this section let $U\subset\mathbb{R}^n$ be an open and connected set.

Definition 11.6 (a) A *differential k-form on U* is a mapping $\omega\colon U\to\Lambda^k(\mathbb{R}^n)$, i. e. to every point $p\in U$ we associate a k-form $\omega(p)\in\Lambda^k(\mathbb{R}^n)$. The linear space of differential k-forms on U is denoted by $\Omega^k(U)$.

(b) Let ω be a differential k-form on U. Since $\{\, \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k} \mid 1 \le i_1 < i_2 < \cdots < i_k \le n\}$ forms a basis of $\Lambda^k(\mathbb{R}^n)$ there exist uniquely determined functions $a_{i_1 \dots i_k}$ on U such that

$$\omega(p) = \sum_{1 \le i_1 < \cdots < i_k \le n} a_{i_1 \dots i_k}(p)\ \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}. \tag{11.5}$$

If all functions $a_{i_1 \dots i_k}$ are in $\mathrm{C}^r(U)$, $r \in \mathbb{N} \cup \{\infty\}$ we say ω is an *r times continuosly differentiable differential k-form* on U. The set of those differential k-forms is denoted by $\Omega_r^k(U)$

We define $\Omega_r^0(U) = \mathrm{C}^r(U)$ and $\Omega(U) = \bigoplus_{k=0}^{n} \Omega^k(U)$. The product in $\Lambda(\mathbb{R}^n)$ defines a product in $\Omega(U)$:

$$(\omega \wedge \eta)(x) = \omega(x) \wedge \eta(x), \quad x \in U,$$

hence $\Omega(U)$ is an algebra. For example, if

$$\omega_1 = x^2\, \mathrm{d}y \wedge \mathrm{d}z + xyz\, \mathrm{d}x \wedge \mathrm{d}y \quad \omega_2 = (xy^2 + 3z^2)\, \mathrm{d}x$$

define a differential 2-form and a 1-form on $\mathbb{R}^3$, $\omega_1 \in \Omega^2(\mathbb{R}^3), \omega_2 \in \Omega^1(\mathbb{R}^3)$ then $\omega_1 \wedge \omega_2 = (x^3y^2 + 3x^2z^2)\, \mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z$.

11.2.2 Differentiation

Definition 11.7 Let $f \in \Omega^0(U) = \mathrm{C}^r(U)$ and $p \in U$. We define

$$\mathrm{d}f(p) = Df(p);$$

then $\mathrm{d}f$ is a differential 1-form on U.
If $\omega(p) = \sum_{1 \le i_1 < \cdots < i_k \le n} a_{i_1 \dots i_k}(p)\ \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}$ is a differential k-form, we define

$$\mathrm{d}\omega(p) = \sum_{1 \le i_1 < \cdots < i_k \le n} \mathrm{d}a_{i_1 \dots i_k}(p) \wedge \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}. \tag{11.6}$$

Then $\mathrm{d}\omega$ is a differential $(k+1)$-form. The linear operator $\mathrm{d}\colon \Omega^k(U) \to \Omega^{k+1}(U)$ is called the *exterior differential*.

Remarks 11.1 (a) Note, that for a function $f\colon U \to \mathbb{R}$, $Df \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}) = \Lambda^1(\mathbb{R}^n)$. By Example 7.7 (a)

$$Df(x)(h) = \operatorname{grad} f(x) \cdot h = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) h_i = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\, \mathrm{d}x_i(h),$$

hence

$$\mathrm{d}f(x) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)\ \mathrm{d}x_i. \tag{11.7}$$

Viewing $x_i\colon U \to \mathbb{R}$ as a C^∞-function, by the above formula

$$\mathrm{d}x_i(x) = \mathrm{d}x_i.$$

This justifies the notation $\mathrm{d}x_i$. If $f \in \mathrm{C}^\infty(\mathbb{R})$ we have $\mathrm{d}f(x) = f'(x)\,\mathrm{d}x$.
(b) One can show that the definition of $\mathrm{d}\omega$ does not depend on the choice of the basis $\{\mathrm{d}x_1, \ldots, \mathrm{d}x_n\}$ of $\Lambda^1(\mathbb{R}^n)$.

Example 11.5 (a) $G = \mathbb{R}^2$, $\omega = \mathrm{e}^{xy}\,\mathrm{d}x + xy^3\,\mathrm{d}y$. Then

$$\begin{aligned}\mathrm{d}\omega &= \mathrm{d}(\mathrm{e}^{xy})\wedge \mathrm{d}x + \mathrm{d}(xy^3)\wedge \mathrm{d}y\\ &= (y\mathrm{e}^{xy}\,\mathrm{d}x + x\mathrm{e}^{xy}\,\mathrm{d}y)\wedge \mathrm{d}x + (y^3\,\mathrm{d}x + 3xy^2\,\mathrm{d}y)\wedge \mathrm{d}y\\ &= (-x\mathrm{e}^{xy} + y^3)\,\mathrm{d}x\wedge \mathrm{d}y.\end{aligned}$$

(b) Let f be continuously differentiable. Then

$$\mathrm{d}f = f_x\,\mathrm{d}x + f_y\,\mathrm{d}y + f_z\,\mathrm{d}z = \operatorname{grad} f \cdot (\mathrm{d}x,\, \mathrm{d}y,\, \mathrm{d}z) = \operatorname{grad} f \cdot \mathrm{d}\vec{x}.$$

(c) Let $v = (v_1, v_2, v_3)$ be a C^1-vector field. Put $\omega = v_1\,\mathrm{d}x + v_2\,\mathrm{d}y + v_3\,\mathrm{d}z$. Then we have

$$\begin{aligned}\mathrm{d}\omega &= \left(\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}\right)\mathrm{d}y\wedge \mathrm{d}z + \left(\frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x}\right)\mathrm{d}z\wedge \mathrm{d}x + \left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\right)\mathrm{d}x\wedge \mathrm{d}y\\ &= \operatorname{curl}(v) \cdot (\mathrm{d}y\wedge \mathrm{d}z,\, \mathrm{d}z\wedge \mathrm{d}x,\, \mathrm{d}x\wedge \mathrm{d}y) = \operatorname{curl}(v) \cdot \mathrm{d}S.\end{aligned}$$

(d) Let v be as above. Put $\omega = v_1\,\mathrm{d}y\wedge \mathrm{d}z + v_2\,\mathrm{d}z\wedge \mathrm{d}x + v_3\,\mathrm{d}x\wedge \mathrm{d}y$. Then we have

$$\mathrm{d}\omega = \operatorname{div}(v)\,\mathrm{d}x\wedge \mathrm{d}y\wedge \mathrm{d}z.$$

Proposition 11.3 *The exterior differential* d *is a linear mapping which satisfies*

(i) $\mathrm{d}(\omega\wedge\eta) = \mathrm{d}\omega\wedge\eta + (-1)^k\omega\wedge\mathrm{d}\eta, \quad \omega \in \Omega_1^k(U),\ \eta \in \Omega_1(U).$
(ii) $\mathrm{d}(\mathrm{d}\omega) = 0, \quad \omega \in \Omega_2(U).$

Proof. (i) We first prove Leibniz' rule for functions $f, g \in \Omega_1^0(U)$. By Remarks 11.1 (a),

$$\begin{aligned}\mathrm{d}(fg) &= \sum_i \frac{\partial}{\partial x_i}(fg)\,\mathrm{d}x_i = \sum_i \left(\frac{\partial f}{\partial x_i}g + f\,\frac{\partial g}{\partial x_i}\right)\mathrm{d}x_i\\ &= \sum_i \frac{\partial f}{\partial x_i}\,\mathrm{d}x_i\, g + \sum_i \frac{\partial g}{\partial x_i}\,\mathrm{d}x_i\, f = \mathrm{d}f\, g + f\mathrm{d}g.\end{aligned}$$

For $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_l)$ we abbreviate $\mathrm{d}x_I = \mathrm{d}x_{i_1}\wedge\cdots\wedge\mathrm{d}x_{i_k}$ and $\mathrm{d}x_J =$

$\mathrm{d}x_{j_1}\wedge\cdots\wedge\mathrm{d}x_{j_l}$. Let $\omega=\sum_I a_I\,\mathrm{d}x_I$ and $\eta=\sum_J b_J\,\mathrm{d}x_J$. By definition

$$\begin{aligned}\mathrm{d}(\omega\wedge\eta)&=\mathrm{d}\left(\sum_{I,J}a_Ib_J\,\mathrm{d}x_I\wedge\mathrm{d}x_J\right)=\sum_{I,J}\mathrm{d}(a_Ib_j)\wedge\mathrm{d}x_I\wedge\mathrm{d}x_J\\&=\sum_{I,J}(\mathrm{d}a_I\,b_J+a_I\,\mathrm{d}b_J)\wedge\mathrm{d}x_I\wedge\mathrm{d}x_J\\&=\sum_{I,J}\mathrm{d}a_I\wedge\mathrm{d}x_I\wedge b_J\,\mathrm{d}x_J+\sum_{I,J}a_I\,\mathrm{d}x_I\wedge\mathrm{d}b_J\wedge\mathrm{d}x_J(-1)^k\\&=\mathrm{d}\omega\wedge\eta+(-1)^k\omega\wedge\mathrm{d}\eta,\end{aligned}$$

where in the third line we used $\mathrm{d}b_J\wedge\mathrm{d}x_I=(-1)^k\,\mathrm{d}x_I\wedge\mathrm{d}b_J$.
(ii) Again by the definition of d:

$$\begin{aligned}\mathrm{d}(\mathrm{d}\omega)&=\sum_I\mathrm{d}\left(\mathrm{d}a_I\wedge\mathrm{d}x_I\right)=\sum_{I,j}\mathrm{d}\left(\frac{\partial a_I}{\partial x_j}\,\mathrm{d}x_j\wedge\mathrm{d}x_I\right)=\sum_{I,i,j}\frac{\partial^2a_I}{\partial x_i\partial x_j}\,\mathrm{d}x_i\wedge\mathrm{d}x_j\wedge\mathrm{d}x_I\\&\underset{\text{Schwarz' lemma}}{=}\sum_{I,i,j}\frac{\partial^2a_I}{\partial x_j\partial x_i}\,(-\mathrm{d}x_j\wedge\mathrm{d}x_i\wedge\mathrm{d}x_I)=-\mathrm{d}(\mathrm{d}\omega).\end{aligned}$$

It follows that $\mathrm{d}(\mathrm{d}\omega)=\mathrm{d}^2\,\omega=0$. ∎

11.2.3 Pull-Back

Definition 11.8 Let $f\colon U\to V$ be a differentiable function with open sets $U\subset\mathbb{R}^n$ and $V\subset\mathbb{R}^m$. Let $\omega\in\Omega^k(V)$ be a differential k-form. We define a differential k-form $f^*(\omega)\in\Omega^k(U)$ by

$$\begin{aligned}(f^*\omega)(p)&=(Df(p)^*)\omega(f(p)),\\(f^*\omega)(p;h_1,\ldots,h_k)&=\omega(f(p);Df(p)(h_1),\ldots,Df(p)(h_k)),\quad p\in U,\ h_1,\ldots,h_k\in\mathbb{R}^n.\end{aligned}$$

In case $k=0$ and $\omega\in\Omega^0(V)=\mathrm{C}^\infty(V)$ we simply set

$$(f^*\omega)(p)=\omega(f(p)),\quad f^*\omega=\omega\circ f.$$

We call $f^*(\omega)$ the *pull-back* of the differential k-form ω with respect to f.

Note that by definition the pull-back f^* is a linear mapping from the space of differential k-forms on V to the space of differential k-forms on U, $f^*\colon\Omega^k(V)\to\Omega^k(U)$.

Proposition 11.4 *Let f be as above and $\omega, \eta \in \Omega(V)$. Let $\{\mathrm{d}y_1, \ldots, \mathrm{d}y_m\}$ be the dual basis to the standard basis in $(\mathbb{R}^m)^*$. Then we have with $f = (f_1, \ldots, f_m)$*

$$\textit{(a)} \quad f^*(\mathrm{d}y_i) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}\, \mathrm{d}x_j = \mathrm{d}f_i, \quad i = 1, \ldots, m. \tag{11.8}$$

$$\textit{(b)} \quad f^*(\mathrm{d}\omega) = \mathrm{d}(f^*\omega), \tag{11.9}$$

$$\textit{(c)} \quad f^*(a\omega) = (a \circ f) f^*(\omega), \quad a \in \mathrm{C}^\infty(V), \tag{11.10}$$

$$\textit{(d)} \quad f^*(\omega \wedge \eta) = f^*(\omega) \wedge f^*(\eta). \tag{11.11}$$

If $n = m$, then

$$\textit{(e)} \quad f^*(\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_n) = \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)}\, \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n. \tag{11.12}$$

Proof. We show (a). Let $h \in \mathbb{R}^n$; by Definition 11.7 and the definition of the derivative we have

$$\begin{aligned} f^*(\mathrm{d}y_i)(h) = \mathrm{d}y_i(Df(p)(h)) &= \left\langle \mathrm{d}y_i\,,\, \left(\sum_{j=1}^{n} \left(\frac{\partial f_k(p)}{\partial x_j} \right) h_j \right)_{k=1,\ldots,m} \right\rangle \\ &= \sum_{j=1}^{n} \left(\frac{\partial f_i(p)}{\partial x_j} \right) h_j = \sum_{j=1}^{n} \left(\frac{\partial f_i(p)}{\partial x_j} \right) \mathrm{d}x_j(h). \end{aligned}$$

This shows (a). Equation (11.10) is a special case of (11.11); we prove (d). Let $p \in U$. Using the pull-back formula for k forms we obtain

$$\begin{aligned} f^*(\omega \wedge \eta)(p) &= (Df(p))^*(\omega \wedge \eta(f(p))) = (Df(p))^*(\omega(f(p)) \wedge \eta(f(p))) \\ &= (Df(p))^*(\omega(f(p))) \wedge (Df(p))^*(\eta(f(p))) \\ &= f^*(\omega)(p) \wedge f^*(\eta)(p) = (f^*(\omega) \wedge f^*(\eta))(p) \end{aligned}$$

To show (11.9) we start with a 0-form g and prove that $f^*(\mathrm{d}g) = \mathrm{d}(f^* g)$ for *functions* $g\colon U \to \mathbb{R}$. By (11.7) and (11.10) we have

$$\begin{aligned} f^*(\mathrm{d}g)(p) = f^* \left(\sum_{i=1}^{m} \frac{\partial g}{\partial y_i}\, \mathrm{d}y_i \right)(p) &= \sum_{i=1}^{m} \frac{\partial g(f(p))}{\partial y_i} f^*(\mathrm{d}y_i) \\ &= \sum_{i=1}^{m} \frac{\partial g(f(p))}{\partial y_i} \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(p)\, \mathrm{d}x_j \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{m} \frac{\partial g(f(p))}{\partial y_i} \frac{\partial f_i(p)}{\partial x_j} \right) \mathrm{d}x_j \\ &\underset{\text{chain rule}}{=} \sum_{j=1}^{n} \frac{\partial}{\partial x_j}(g \circ f)(p)\, \mathrm{d}x_j \\ &= \mathrm{d}(g \circ f)(p) = \mathrm{d}(f^* g)(p). \end{aligned}$$

Now let $\omega = \sum_I a_I\, \mathrm{d}x_I$ be an arbitrary form. Since by Leibniz rule

$$\mathrm{d}f^*(\mathrm{d}x_I) = \mathrm{d}(\mathrm{d}(f_{i_1}) \wedge \cdots \wedge \mathrm{d}(f_{i_k})) = 0,$$

we get by Leibniz rule

$$\begin{aligned}\mathrm{d}\,(f^*\omega) &= \mathrm{d}\left(\sum_I f^*(a_I) f^*(\,\mathrm{d}x_I)\right)\\ &= \sum_I \left(\mathrm{d}(f^*(a_I)) \wedge f^*(\,\mathrm{d}x_I) + f^*(a_I)\,\mathrm{d}(f^*(\,\mathrm{d}x_I))\right) = \sum_I \mathrm{d}(f^*(a_I))i \wedge f^*(\,\mathrm{d}x_I).\end{aligned}$$

On the other hand, by (d) we have

$$f^*\left(\mathrm{d}\sum_I a_i\,\mathrm{d}x_I\right) = f^*\left(\sum_I \mathrm{d}a_I \wedge\,\mathrm{d}x_I\right) = \sum_I f^*(\mathrm{d}a_I) \wedge f^*(\,\mathrm{d}x_I).$$

By the first part of (b), both expressions coincide. This completes the proof of (b).
We finally prove (e). By (b) and (d) we have

$$\begin{aligned}f^*(\,\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_n) &= f^*(\,\mathrm{d}y_1) \wedge \cdots \wedge f^*(\,\mathrm{d}y_n)\\ &= \sum_{i_1=1}^n \frac{\partial f_1}{\partial x_{i_1}}\,\mathrm{d}x_{i_1} \wedge \cdots \wedge \sum_{i_n=1}^n \frac{\partial f_n}{\partial x_{i_n}}\,\mathrm{d}x_{i_n}\\ &= \sum_{i_1,\ldots,i_n=1}^n \frac{\partial f_1}{\partial x_{i_1}} \cdots \frac{\partial f_n}{\partial x_{i_n}}\,\mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_n}.\end{aligned}$$

Since the square of a 1-form vanishes, the only non-vanishing terms in the above sum are the permutations $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$. Using skew-symmetry to write $\mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_n}$ as a multiple of $\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n$, we obtain the sign of the permutation $(i_1, \ldots, i_n)$:

$$\begin{aligned}f^*(\,\mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_n) &= \sum_{I=(i_1,\ldots,i_n)\in \mathrm{S}_n} \mathrm{sign}\,(I) \frac{\partial f_1}{\partial x_{i_1}} \cdots \frac{\partial f_n}{\partial x_{i_n}}\,\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n\\ &= \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)}\,\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n.\end{aligned}$$

■

Example 11.6 (a) Let $f(r, \varphi) = (r\cos\varphi, r\sin\varphi)$ be given on $\mathbb{R}^2 \setminus (0 \times \mathbb{R})$ and let $\{\,\mathrm{d}r, \mathrm{d}\varphi\}$ and $\{\,\mathrm{d}x,\,\mathrm{d}y\}$ be the dual bases to $\{\mathrm{e}_r, \mathrm{e}_\varphi\}$ and $\{\mathrm{e}_1, \mathrm{e}_2\}$. We have

$$\begin{aligned} f^*(x) &= r\cos\varphi, & f^*(y) &= r\sin\varphi,\\ f^*(\,\mathrm{d}x) &= \cos\varphi\,\mathrm{d}r - r\sin\varphi\mathrm{d}\varphi, & f^*(\,\mathrm{d}y) &= \sin\varphi\,\mathrm{d}r + r\cos\varphi\mathrm{d}\varphi,\\ f^*(\,\mathrm{d}x \wedge \mathrm{d}y) &= r\,\mathrm{d}r \wedge \mathrm{d}\varphi, \end{aligned}$$

$$f^*\left(\frac{-y}{x^2+y^2}\,\mathrm{d}x + \frac{x}{x^2+y^2}\,\mathrm{d}y\right) = \mathrm{d}\varphi.$$

(b) Let $k \in \mathbb{N}$, $r \in \{1, \ldots, k\}$, and $\alpha \in \mathbb{R}$. Define a mapping a mapping from $I\colon \mathbb{R}^k \to \mathbb{R}^{k+1}$ and $\omega \in \Omega^k(\mathbb{R}^{k+1})$ by

$$\begin{aligned} I(x_1, \ldots, x_k) &= (x_1, \ldots, x_{r-1}, \alpha, x_r, \ldots x_k),\\ \omega(y_1, \ldots, y_{k+1}) &= \sum_{i=1}^{k+1} f_i(y)\,\mathrm{d}y_1 \wedge \cdots \widehat{\mathrm{d}y_i} \cdots \wedge \mathrm{d}y_{k+1}, \end{aligned}$$

where $f_i \in \mathrm{C}^\infty(\mathbb{R}^{k+1})$ for all i; the hat means omission of the factor $\mathrm{d}y_i$. Then

$$I^*(\omega)(x) = f_r(x_1, \ldots, x_{r-1}, \alpha, x_r, \ldots, x_k)\, \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_k.$$

This follows from

$$\begin{aligned} I^*(\,\mathrm{d}y_i) &= \mathrm{d}x_i, \quad i = 1, \ldots, r-1, \\ I^*(\,\mathrm{d}y_r) &= 0, \\ I^*(\,\mathrm{d}y_{i+1}) &= \mathrm{d}x_i, \quad i = r, \ldots, k. \end{aligned}$$

Roughly speaking: $f^*(\omega)$ is obtained by substituting the new variables at all places.

11.2.4 Closed and Exact Forms

Motivation: Let $f(x)$ be a continuous function on $\mathbb{R}$. Then $\omega = f(x)\,\mathrm{d}x$ is a 1-form. By the fundamental theorem of calculus, there exists an antiderivative $F(x)$ to $f(x)$ such that $\mathrm{d}\,F(x) = f(x)\,\mathrm{d}x = \omega$.
Problem: Given $\omega \in \Omega^k(U)$. Does there exist $\eta \in \Omega^{k-1}(U)$ with $\mathrm{d}\eta = \omega$?

Definition 11.9 $\omega \in \Omega^k(U)$ is called *closed* if $\mathrm{d}\,\omega = 0$.
$\omega \in \Omega^k(U)$ is called *exact* if there exists $\eta \in \Omega^{k-1}(U)$ such that $\mathrm{d}\,\eta = \omega$.

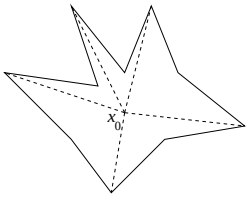
Remarks 11.2 (a) An exact form ω is closed; indeed, $\mathrm{d}\omega = \mathrm{d}(\mathrm{d}\eta) = 0$.
(b) A 1-form $\omega = \sum_i f_i\,\mathrm{d}x_i$ is closed if and only if $\operatorname{curl} \vec{f} = 0$ for the corresponding vector field $\vec{f} = (f_1, \ldots, f_n)$. Here the general curl can be defined as a vector with $n(n-1)/2$ components

$$(\operatorname{curl} \vec{f})_{ij} = \frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j}.$$

The form ω is exact if and only if $\vec{f}$ is conservative, that is, $\vec{f}$ is a gradient vector field with $\vec{f} = \operatorname{grad}(U)$. Then $\omega = \mathrm{d}\,U$.
(c) There are closed forms that are not exact; for example, the winding form

$$\omega = \frac{-y}{x^2+y^2}\,\mathrm{d}x + \frac{x}{x^2+y^2}\,\mathrm{d}y$$

on $\mathbb{R}^2 \setminus \{(0,0)\}$ is not exact, cf. homework 30.1.
(d) If $\mathrm{d}\,\eta = \omega$ then $\mathrm{d}(\eta + \mathrm{d}\xi) = \omega$, too, for all $\xi \in \Omega^{k-2}(U)$.



Definition 11.10 An open set U is called *star-shaped* if there exists an $x_0 \in U$ such that for all $x \in U$ the segment from x_0 to x is in U, i. e. $(1-t)x_0 + tx \in U$ for all $t \in [0,1]$.

Convex sets U are star-shaped (take any $x_0 \in U$); any star-shaped set is connected and simply connected.

Lemma 11.5 *Let $U \subset \mathbb{R}^n$ be star-shaped with respect to the origin. Let $\omega = \sum_{i_1<\cdots<i_k} a_{i_1\cdots i_k}\,\mathrm{d}x_{i_1}\wedge\cdots\wedge\mathrm{d}x_{i_k} \in \Omega^k(U)$. Define*

$$I(\omega)(x) = \sum_{i_1<\cdots<i_k}\sum_{r=1}^{k}(-1)^{r-1}\left(\int_0^1 t^{k-1}a_{i_1\cdots i_k}(tx)\,\mathrm{d}t\right)x_{i_r}\,\mathrm{d}x_{i_1}\wedge\cdots\wedge\widehat{\mathrm{d}x_{i_r}}\wedge\cdots\wedge\mathrm{d}x_{i_k}, \tag{11.13}$$

where the hat means omission of the factor $\mathrm{d}x_{i_r}$. Then we have

$$I(\mathrm{d}\,\omega) + \mathrm{d}(I\,\omega) = \omega. \tag{11.14}$$

(Without proof.)

Example 11.7 (a) Let $k = 1$, $n = 3$, and $\omega = a_1\,\mathrm{d}x_1 + a_2\,\mathrm{d}x_2 + a_3\,\mathrm{d}x_3$. Then

$$I(\omega) = x_1\int_0^1 a_1(tx)\,\mathrm{d}t + x_2\int_0^1 a_2(tx)\,\mathrm{d}t + x_3\int_0^1 a_3(tx)\,\mathrm{d}t.$$

Note that this is exactly the formula for the potential $U(x_1, x_2, x_3)$ from Remark 8.5 (b). Let (a_1, a_2, a_3) be a vector field on U with $\mathrm{d}\omega = 0$. This is equivalent to $\operatorname{curl} a = 0$ by Example 11.5 (c). The above lemma shows $\mathrm{d}U = \omega$ for $U = I(\omega)$; this means $\operatorname{grad} U = (a_1, a_2, a_3)$, U is the potential to the vector field (a_1, a_2, a_3).
(b) Let $k = 2$, $n = 3$, and $\omega = a_1\,\mathrm{d}x_2\wedge\mathrm{d}x_3 + a_2\,\mathrm{d}x_3\wedge\mathrm{d}x_1 + a_3\,\mathrm{d}x_1\wedge\mathrm{d}x_2$ where a is a C^1-vector field on U. Then

$$\begin{aligned} I(\omega) = {} & \left(x_3\int_0^1 ta_2(tx)\,\mathrm{d}t - x_2\int_0^1 ta_3(tx)\,\mathrm{d}t\right)\mathrm{d}x_1 + \\ & + \left(x_1\int_0^1 ta_3(tx)\,\mathrm{d}t - x_3\int_0^1 ta_1(tx)\,\mathrm{d}t\right)\mathrm{d}x_2 + \\ & + \left(x_2\int_0^1 ta_1(tx)\,\mathrm{d}t - x_1\int_0^1 ta_2(tx)\,\mathrm{d}t\right)\mathrm{d}x_3. \end{aligned}$$

By Example 11.5 (d), ω is closed if and only if $\operatorname{div}(a) = 0$ on U. Let $\eta = b_1\,\mathrm{d}x_1 + b_2\,\mathrm{d}x_2 + b_3\,\mathrm{d}x_3$ such that $\mathrm{d}\eta = \omega$. This means $\operatorname{curl} b = a$. The Poincaré lemma shows that b with $\operatorname{curl} b = a$ exists if and only if $\operatorname{div}(a) = 0$. Then b is the vector potential to a. In case $\mathrm{d}\,\omega = 0$ we can choose $\vec{b}\,\mathrm{d}\vec{x} = I(\omega)$.

Theorem 11.6 (Poincaré Lemma) *Let U be star-shaped. Then every closed differential form is exact.*

Proof. Without loss of generality let U be star-shaped with respect to the origin and $\mathrm{d}\omega = 0$. By Lemma 11.5, $\mathrm{d}(I\omega) = \omega$. ∎

Remarks 11.3 (a) Let U be star-shaped, $\omega \in \Omega^k(U)$. Suppose $\mathrm{d}\,\eta_0 = \omega$ for some $\eta \in \Omega^{k-1}(U)$. Then the general solution of $\mathrm{d}\,\eta = \omega$ is given by $\eta_0 + \mathrm{d}\xi$ with $\xi \in \Omega^{k-2}(U)$. Indeed, let η be a second solution of $\mathrm{d}\,\eta = \omega$. Then $\mathrm{d}(\eta - \eta_0) = 0$. By the Poincaré lemma, there exists $\xi \in \Omega^{k-2}(U)$ with $\eta - \eta_0 = \mathrm{d}\xi$, hence $\eta = \eta_0 + \mathrm{d}\xi$.
(b) Let V be a linear space and W a linear subspace of V. We define an equivalence relation on V by $v_1 \sim v_2$ if $v_1 - v_2 \in W$. The equivalence class of v is denoted by $v + W$. One easily sees that the set of equivalence classes, denoted by V/W, is again a linear space: $\alpha(v + W) + \beta(u + W) := \alpha v + \beta u + W$.
Let U be an arbitrary open subset of $\mathbb{R}^n$. We define

$$C^k(U) = \{\omega \in \Omega^k(U) \mid \mathrm{d}\,\omega = 0\}, \text{ the } \textit{cocycles} \text{ on } U,$$
$$B^k(U) = \{\omega \in \Omega^k(U) \mid \omega \quad \text{is exact}\}, \text{ the } \textit{coboundaries} \text{ on } U.$$

Since exact forms are closed, $B^k(U)$ is a linear subspace of $C^k(U)$. The factor space

$$H^k_{\mathrm{deR}}(U) = C^k(U)/B^k(U)$$

is called the *de Rham cohomology* of U. If U is star-shaped, $H^k_{\mathrm{deR}}(U) = 0$ for $k \geq 1$, by Poincaré's lemma. The first de Rham cohomology H^1_{deR} of $\mathbb{R}^2 \setminus \{(0,0)\}$ is non-zero. The winding form is a non-zero element. We have

$$H^0_{\mathrm{deR}}(U) \cong \mathbb{R}^p,$$

if and only if U has exactly p components which are not connected $U = U_1 \cup \cdots \cup U_p$ (disjoint union). Then, the characteristic functions χ_{U_i}, $i = 1, \ldots, p$, form a basis of the 0-cycles $C^0(U)$ ($B^0(U) = 0$).

11.3 Stokes' Theorem

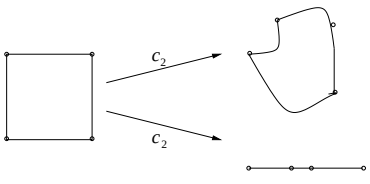
11.3.1 Singular Cubes, Singular Chains, and the Boundary Operator

A very nice treatment of the topics to this section is [Spi65, Chapter 4]. The set $[0,1]^k = [0,1] \times \cdots \times [0,1] = \{x \in \mathbb{R}^k \mid 0 \leq x_i \leq 1, i = 1, \ldots, k\}$ is called the *k-dimensional unit cube*. Let $U \subset \mathbb{R}^n$ be open.

Definition 11.11 (a) A *singular k-cube* in $U \subset \mathbb{R}^n$ is a continuously differentiable mapping $c_k \colon [0,1]^k \to U$.
(b) A *singular k-chain* in U is a formal sum

$$s_k = n_1 c_{k,1} + \cdots + n_r c_{k,r}$$

with singular k-cubes $c_{k,i}$ and integers $n_i \in \mathbb{Z}$.



A singular 0-cube is a point, a singular 1-cube is a curve, in general, a singular 2-cube (in $\mathbb{R}^3$) is a surface with a boundary of 4 pieces which are differentiable curves. Note that a singular 2-cube can also be a single point—that is where the name "singular" comes from.

Let $I_k\colon [0,1]^k \to \mathbb{R}^k$ be the identity map, i. e. $I_k(x) = x$, $x \in [0,1]^k$. It is called the *standard k-cube in $\mathbb{R}^k$*. We are going to define the *boundary* ∂s_k of a singular k-chain s_k. For $i = 1, \ldots, k$ define

$$I^k_{(i,0)}(x_1, \ldots, x_{k-1}) = (x_1, \ldots, x_{i-1}, 0, x_i, \ldots, x_{k-1}),$$
$$I^k_{(i,1)}(x_1, \ldots, x_{k-1}) = (x_1, \ldots, x_{i-1}, 1, x_i, \ldots, x_{k-1}).$$

Insert a 0 and a 1 at the ith component, respectively.
The boundary of the standard k-cube I_k is now defined by $\partial I_k\colon [0,1]^{k-1} \to [0,1]^k$

$$\partial I_k = \sum_{i=1}^{k} (-1)^i \left(I^k_{(i,0)} - I^k_{(i,1)} \right). \tag{11.15}$$

It is the formal sum of $2k$ singular $(k-1)$-cubes, the faces of the k-cube.
The boundary of an arbitrary singular k-cube $c_k\colon [0,1]^k \to U \subset \mathbb{R}^n$ is defined by the composition of the above mapping $\partial\, I_k\colon [0,1]^{k-1} \to [0,1]^k$ and the k-cube c_k:

$$\partial c_k = c_k \circ \partial I_k = \sum_{i=1}^{k} (-1)^i \left(c_k \circ I^k_{(i,0)} - c_k \circ I^k_{(i,1)} \right), \tag{11.16}$$

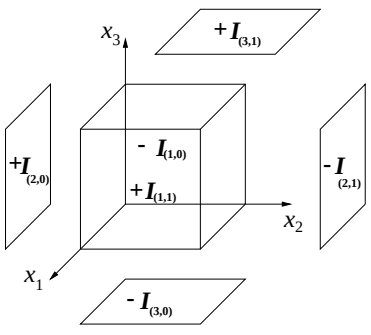
and for a singular k-chain $s_k = n_1 c_{k,1} + \cdots + n_r c_{k,r}$ we set

$$\partial s_k = n_1 \partial c_{k,1} + \cdots + n_r \partial c_{k,r}.$$

The boundary operator ∂c_k associates to each singular k-chain a singular $(k-1)$-chain (since both $I^k_{(i,0)}$ and $I^k_{(i,1)}$ depend on $k-1$ variables, all from the segment $[0,1]$).
One can show that

$$\partial(\partial s_k) = 0$$

for any singular k-chain s_k.



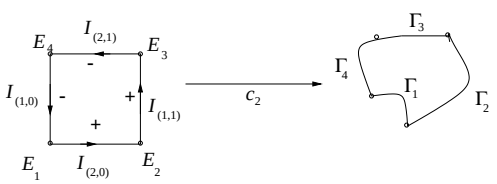
Example 11.8 (a) In case $n = k = 3$ have

$$\partial I_3 = -I^3_{(1,0)} + I^3_{(1,1)} + I^3_{(2,0)} - I^3_{(2,1)} - I^3_{(3,0)} + I^3_{(3,1)},$$

where

$$-I^3_{(1,0)}(x_1, x_2) = -(0, x_1, x_2), \quad +I^3_{(1,1)}(x_1, x_2) = +(1, x_1, x_2),$$
$$+I^3_{(2,0)}(x_1, x_2) = +(x_1, 0, x_2), \quad -I^3_{(2,1)}(x_1, x_2) = -(x_1, 1, x_2),$$
$$-I^3_{(3,0)}(x_1, x_2) = -(x_1, x_2, 0), \quad +I^3_{(3,1)}(x_1, x_2) = +(x_1, x_2, 1).$$

Note, if we take care of the signs in (11.15) all 6 unit normal vectors $D_1 I^k_{(i,j)} \times D_2 I^k_{(i,j)}$ to the faces have the orientation of the outer normal with respect to the unit 3-cube $[0,1]^3$. The above sum ∂I_3 is a *formal* sum of singular 2-cubes. You are not allowed to add componentwise: $-(0, x_1, x_2) + (1, x_1, x_2) \neq (1, 0, 0)$.



(b) In case $k = 2$ we have

$$\begin{aligned}\partial I_2(x) &= I^2_{(1,1)} - I^2_{(1,0)} + I^2_{(2,0)} - I^2_{(2,1)}\\ &= (1,x) - (0,x) + (x,0) - (x,1).\\ \partial\partial I_2 &= (E_3 - E_2) - (E_4 - E_1) +\\ &\quad + (E_2 - E_1) - (E_3 - E_4) = 0.\end{aligned}$$

Here we have $\partial c_2 = \Gamma_1 + \Gamma_2 - \Gamma_3 - \Gamma_4$.

(c) Let $c_2 \colon [0, 2\pi] \times [0, \pi] \to \mathbb{R}^3 \setminus \{(0,0,0)\}$ be the singular 2-cube

$$c(s,t) = (\cos s \sin t, \sin s \sin t, \cos t).$$

By (b)

$$\begin{aligned}\partial c_2(x) &= c_2 \circ \partial I_2 =\\ &= c_2(2\pi, x) - c_2(0,x) + c_2(x,0) - c_2(x,\pi)\\ &= (\cos 2\pi \sin x, \sin 2\pi \sin x, \cos 2\pi) - (\cos 0 \sin x, \sin 0 \sin x, \cos x) +\\ &+ (\cos x \sin 0, \sin x \sin 0, \cos 0) - (\cos x \sin \pi, \sin x \sin \pi, \cos \pi)\\ &= (\sin x, 0, \cos x) - (\sin x, 0, \cos x) + (0,0,1) - (0,0,-1)\\ &= (0,0,1) - (0,0,-1).\end{aligned}$$

Hence, the boundary ∂c_2 of the singular 2-cube c_2 is a degenerate singular 1-chain. We come back to this example.

11.3.2 Integration

Definition 11.12 Let $c_k \colon [0,1]^k \to U \subset \mathbb{R}^n$, $\vec{x} = c_k(t_1, \ldots, t_k)$, be a singular k-cube and ω a k-form on U. Then $(c_k)^*(\omega)$ is a k-form on the unit cube $[0,1]^k$. Thus there exists a unique function $f(t)$, $t \in [0,1]^k$, such that

$$(c_k)^*(\omega) = f(t)\, \mathrm{d}t_1 \wedge \cdots \wedge \mathrm{d}t_k.$$

Then

$$\int_{c_k} \omega := \int_{I_k} (c_k^*)\,(\omega) := \int_{[0,1]^k} f(t)\, \mathrm{d}t_1 \cdots \mathrm{d}t_k$$

is called the *integral of ω over the singular cube c_k*; on the right there is the k-dimensional Riemann integral.

If $s_k = \sum_{i=1}^{r} n_i c_{k,i}$ is a k-chain, set

$$\int_{s_k} \omega = \sum_{i=1}^{r} n_i \int_{c_{k,i}} \omega.$$

If $k = 0$, a 0-cube is a single point $c_0(0) = x_0$ and a 0-form is a function $\omega \in \mathrm{C}^\infty(G)$. We set $\int_{c_0} \omega = c_0^*(\omega)|_{t=0} = \omega(c_0(0)) = \omega(x_0)$. We discuss two special cases $k = 1$ and $k = n$.

Example 11.9 (a) $k = 1$. Let $c\colon [0,1] \to \mathbb{R}^n$ be an oriented, smooth curve $\Gamma = c([0,1])$. Let $\omega = f_1(x)\,\mathrm{d}x_1 + \cdots + f_n(x)\,\mathrm{d}x_n$ be a 1-form on $\mathbb{R}^n$, then

$$c^*(\omega) = (f_1(c(t))c_1'(t) + \cdots + f_n(c(t))c_n'(t))\ \mathrm{d}t$$

is a 1-form on $[0,1]$ such that

$$\int_c \omega = \int_{[0,1]} c^*\,\omega = \int_0^1 f_1(c(t))c_1'(t) + \cdots + f_n(c(t))c_n'(t)\,\mathrm{d}t = \int_\Gamma \vec{f}\cdot\,\mathrm{d}\vec{x}.$$

Obviously, $\int_c \omega$ is the line integral of $\vec{f}$ over Γ.
(b) $k = n$. Let $c\colon [0,1]^k \to \mathbb{R}^k$ be continuously differentiable and let $x = c(t)$. Let $\omega = f(x)\,\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_k$ be a differential k-form on $\mathbb{R}^k$. By Proposition 11.4 (e),

$$c^*(\omega) = f(c(t))\,\frac{\partial(c_1,\ldots,c_k)}{\partial(t_1,\ldots,t_k)}\,\mathrm{d}t_1 \wedge \ldots \wedge \mathrm{d}t_k.$$

Therefore,

$$\int_c \omega = \int_{[0,1]^k} f(c(t))\,\frac{\partial(c_1,\ldots,c_k)}{\partial(t_1,\ldots,t_k)}\,\mathrm{d}t_1 \ldots \mathrm{d}t_k \tag{11.17}$$

Let $c = I_k$ be the standard k-cube in $[0,1]^k$. Then

$$\int_{I_k} \omega = \int_{[0,1]^k} f(x)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$$

is the k-dimensional Riemann integral of f over $[0,1]^k$.
Let $\tilde{I}_k(x_1,\ldots,x_k) = (x_2, x_1, x_3, \ldots, x_k)$. Then $I_k([0,1]^k) = \tilde{I}_k([0,1]^k) = [0,1]^k$, however

$$\begin{aligned}\int_{\tilde{I}_k} \omega &= \int_{[0,1]^k} f(x_2, x_1, x_3, \ldots, x_k)(-1)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k \\ &= -\int_{[0,1]^k} f(x_1, x_2, x_3, \ldots, x_k)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k = -\int_{I_k} \omega.\end{aligned}$$

We see that $\int_{I_k} \omega$ is an *oriented* Riemann integral. Note that in the above formula (11.17) we do *not* have the absolute value of the Jacobian.

11.3.3 Stokes' Theorem

Theorem 11.7 *Let U be an open subset of $\mathbb{R}^n$, $k \geq 0$ a non-negative integer and let s_{k+1} be a singular $(k+1)$-chain on U. Let ω be a differential k-form on U, $\omega \in \Omega_1^k(U)$. Then we have*

$$\int_{\partial s_{k+1}} \omega = \int_{s_{k+1}} \mathrm{d}\,\omega.$$

Proof. (a) Let $s_{k+1} = I_{k+1}$ be the standard $(k+1)$-cube, in particular $n = k+1$.
Let $\omega = \sum_{i=1}^{k+1} f_i(x)\ \mathrm{d}x_1 \wedge \cdots \wedge \widehat{\mathrm{d}x_i} \wedge \cdots \wedge \mathrm{d}x_{k+1}$. Then

$$\mathrm{d}\,\omega = \sum_{i=1}^{k+1} (-1)^{i+1} \frac{\partial f_i(x)}{\partial x_i}\ \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_{k+1},$$

hence by Example 11.6 (b), Fubini's theorem and the fundamental theorem of calculus

$$\begin{aligned}
\int\limits_{I_{k+1}} \mathrm{d}\,\omega &= \sum_{i=1}^{k+1} (-1)^{i+1} \int\limits_{[0,1]^{k+1}} \frac{\partial f_i}{\partial x_i}\, \mathrm{d}x_1 \cdots \mathrm{d}x_{k+1} \\
&= \sum_{i=1}^{k+1} (-1)^{i+1} \int\limits_{[0,1]^k} \left(\int_0^1 \frac{\partial f_i}{\partial x_i}(x_1, \ldots, \underset{\widehat{i}}{t}, \ldots x_{k+1})\, \mathrm{d}t \right) \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \cdots \mathrm{d}x_{k+1} \\
&= \sum_{i=1}^{k+1} (-1)^{i+1} \int\limits_{[0,1]^k} \left(f_i(x_1, \ldots, \underset{\widehat{i}}{1}, \ldots, x_{k+1}) - f_i(x_1, \ldots, \underset{\widehat{i}}{0}, \ldots, x_{k+1}) \right) \mathrm{d}x_1 \cdot \widehat{\cdot} \cdot\, \mathrm{d}x_{k+1} \\
&\underset{\text{Example 11.6 (b)}}{=} \sum_{i=1}^{k+1} (-1)^{i+1} \left(\int\limits_{[0,1]^k} \left(I_{(i,1)}^{k+1} \right)^* \omega - \int\limits_{[0,1]^k} \left(I_{(i,0)}^{k+1} \right)^* \omega \right) \\
&= \int\limits_{\partial I_{k+1}} \omega,
\end{aligned}$$

by definition of ∂I_{k+1}. The assertion is shown in case of identity map.
(b) The general case. Let I_{k+1} be the standard $(k+1)$-cube. Since the pull-back and the differential commute (Proposition 11.4) we have

$$\begin{aligned}
\int\limits_{c_{k+1}} \mathrm{d}\,\omega = \int\limits_{I_{k+1}} (c_{k+1})^* (\mathrm{d}\,\omega) = \int\limits_{I_{k+1}} \mathrm{d}\,((c_{k+1})^* \omega) = \int\limits_{\partial I_{k+1}} (c_{k+1})^* \omega \\
= \sum_{i=1}^{k+1} (-1)^i \left(\int\limits_{I_{(i,0)}^{k+1}} (c_{k+1})^* \omega - \int\limits_{I_{(i,0)}^{k+1}} (c_{k+1})^* \omega \right) \\
= \sum_{i=1}^{k+1} (-1)^i \int\limits_{c_{k+1} \circ I_{(i,0)}^{k+1} - c_{k+1} \circ I_{(i,1)}^{k+1}} \omega = \int\limits_{\partial c_{k+1}} \omega.
\end{aligned}$$

(c) Finally, let $s_{k+1} = \sum_i n_i c_{ki}$ with $n_i \in \mathbb{Z}$ and singular $(k+1)$-cubes c_{ki}. By definition and by (b),

$$\int_{s_{k+1}} \mathrm{d}\omega = \sum_i n_i \int_{c_{k+1}} \mathrm{d}\omega = \sum_i n_i \int_{\partial c_{k+1}} \omega = \int_{\partial s_{k+1}} \omega.$$

■

Remark 11.4 Stokes' theorem is valid for arbitrary oriented compact differentiable k-dimensional manifolds $\mathcal{F}$ and continuously differentiable $(k-1)$-forms ω on $\mathcal{F}$.

Example 11.10 We come back to Example 11.8 (c). Let $\omega = (x\,\mathrm{d}y\wedge\mathrm{d}z + y\,\mathrm{d}z\wedge\mathrm{d}x + z\,\mathrm{d}x\wedge\mathrm{d}y)/r^3$ be a 2-form on $\mathbb{R}^3\setminus\{(0,0,0)\}$. It is easy to show that ω is closed, $\mathrm{d}\,\omega = 0$. We compute $\int_{c_2}\omega$. First note that $c_2^*(r^3) = 1$, $c_2^*(x) = \cos s\sin t$, $c_2^*(y) = \sin s\sin t$, $c_2^*(z) = \cos t$ such that

$$\begin{aligned} c_2^*(\,\mathrm{d}x) &= \mathrm{d}(\cos s\sin t) = -\sin s\sin t\,\mathrm{d}s + \cos s\cos t\,\mathrm{d}t,\\ c_2^*(\,\mathrm{d}y) &= \mathrm{d}(\sin s\sin t) = \cos s\sin t\,\mathrm{d}s + \sin s\cos t\,\mathrm{d}t,\\ c_2^*(\,\mathrm{d}z) &= -\sin t\,\mathrm{d}t. \end{aligned}$$

and

$$\begin{aligned} c_2^*(\omega) &= c_2^*(x\,\mathrm{d}y\wedge\mathrm{d}z + y\,\mathrm{d}z\wedge\mathrm{d}x + z\,\mathrm{d}x\wedge\mathrm{d}y)\\ &= c_2^*(x)\,c_2^*(\,\mathrm{d}y)\wedge c_2^*(\,\mathrm{d}z) + c_2^*(y\,\mathrm{d}z\wedge\mathrm{d}x) + c_2^*(z\,\mathrm{d}x\wedge\mathrm{d}y)\\ &= (-\cos^2 s\sin^3 t - \sin^2 s\sin^3 t - \cos t(\sin^2 s\sin t\cos t + \cos^2 s\sin t\cos t)\,\mathrm{d}s\wedge\mathrm{d}t\\ &= -\sin t\;\mathrm{d}s\wedge\mathrm{d}t, \end{aligned}$$

such that

$$\int_{c_2}\omega = \int\limits_{[0,2\pi]\times[0,\pi]} c_2^*(\omega) = \int\limits_{[0,2\pi]\times[0,\pi]} -\sin t\,\mathrm{d}s\wedge\mathrm{d}t = \int_0^{2\pi}\int_0^{\pi}(-\sin t)\,\mathrm{d}s\mathrm{d}t = -4\pi.$$

Stokes' theorem shows that ω is not exact on $\mathbb{R}^3\setminus\{(0,0,0)\}$. Suppose to the contrary that $\omega = \mathrm{d}\,\eta$ for some $\eta\in\Omega^1(\mathbb{R}^3\setminus\{(0,0,0)\})$. Since by Example 11.8 (c), ∂c_2 is a degenerate 1-chain (it consists of two points), the pull-back $(\partial c_2)^*(\eta)$ is 0 and so is the integral

$$0 = \int_{I_1}(\partial c_2)^*(\eta) = \int_{\partial c_2}\eta = \int_{c_2}\mathrm{d}\eta = \int_{c_2}\omega = -4\pi,$$

a contradiction; hence, ω is not exact.

We come back to the two special cases $k=1$, $n=2$ and $k=1$, $n=3$.

11.3.4 Special Cases

$k=1, n=3$. Let $c\colon[0,1]^2\to U\subseteq\mathbb{R}^3$ be a singular 2-cube, $\mathcal{F} = c([0,1]^2)$ is a regular smooth surface in $\mathbb{R}^3$. Then $\partial\mathcal{F}$ is a closed path consisting of 4 parts with the counter-clockwise orientation. Let $\omega = f_1\,\mathrm{d}x_1 + f_2\,\mathrm{d}x_2 + f_3\,\mathrm{d}x_3$ be a differential 1-form on U. By Example 11.9 (a)

$$\int_{\partial c_2}\omega = \int_{\partial\mathcal{F}} f_1\,\mathrm{d}x_1 + f_2\,\mathrm{d}x_2 + f_3\,\mathrm{d}x_3$$

On the other hand by Example 11.5 (c)

$$\mathrm{d}\,\omega = \operatorname{curl} f\cdot(\,\mathrm{d}x_2\wedge\mathrm{d}x_2,\,\mathrm{d}x_3\wedge\mathrm{d}x_2,\,\mathrm{d}x_1\wedge\mathrm{d}x_2).$$

In this case Stokes' theorem gives

$$\int_{\partial c_2} \omega = \int_{c_2} \mathrm{d}\omega$$

$$\int_{\partial \mathcal{F}} f_1\,\mathrm{d}x_1 + f_2\,\mathrm{d}x_2 + f_3\,\mathrm{d}x_3 = \int_{\mathcal{F}} \operatorname{curl} f \cdot (\,\mathrm{d}x_2 \wedge \mathrm{d}x_3,\ \mathrm{d}x_3 \wedge \mathrm{d}x_1,\ \mathrm{d}x_1 \wedge \mathrm{d}x_2)$$

If $\mathcal{F}$ is in the x_1–x_2 plane, we get Green's theorem.
$k = 2$, $n = 3$. Let c_3 be a singular 3-cube in $\mathbb{R}^3$ and $G = c_3([0,1])$. Further let

$$\omega = v_1\,\mathrm{d}x_2 \wedge \mathrm{d}x_3 + v_2\,\mathrm{d}x_3 \wedge \mathrm{d}x_1 + v_3\,\mathrm{d}x_1 \wedge \mathrm{d}x_2,$$

with a continuously differentiable vector field $v \in \mathrm{C}^1(G)$. By Example 11.5 (d), $\mathrm{d}\,\omega = \operatorname{div}(v)\,\mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}x_3$. The boundary of G consists of the 6 faces $\partial c_3([0,1]^3)$. They are oriented with the outer unit normal vector. Stokes' theorem then gives

$$\int_{c_3} \mathrm{d}\omega = \int_{\partial c_3} v_1\,\mathrm{d}x_2 \wedge \mathrm{d}x_3 + v_2\,\mathrm{d}x_3 \wedge \mathrm{d}x_1 + v_3\,\mathrm{d}x_1 \wedge \mathrm{d}x_2,$$

$$\int_G \operatorname{div} v\ \mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_{\partial G} \vec{v} \cdot \mathrm{d}\vec{S}.$$

This is Gauß' divergence theorem.

11.3.5 Applications

The following two applications were not covered by the lecture.

(a) Brower's Fixed Point Theorem

Proposition 11.8 (Retraction Theorem) *Let $G \subset \mathbb{R}^n$ be a compact, connected, simply connected set with smooth boundary ∂G.*
There exist no vector field $f\colon G \to \mathbb{R}^n$, $f_i \in \mathrm{C}^2(G)$, $i = 1, \ldots, n$ such that $f(G) \subset \partial G$ and $f(x) = x$ for all $x \in \partial G$.

Proof. Suppose to the contrary that such f exists; consider $\omega \in \Omega^{n-1}(U)$, $\omega = x_1\ \mathrm{d}x_2 \wedge \mathrm{d}x_3 \wedge \cdots \wedge \mathrm{d}x_n$. First we show that $f^*(\mathrm{d}\omega) = 0$. By definition, for $v_1, \ldots, v_n \in \mathbb{R}^n$ we have

$$f^*(\mathrm{d}\omega)(p)(v_1, \ldots, v_n) = \mathrm{d}\omega(f(p))(Df(p)v_1, Df(p)v_2, \ldots, Df(p)v_n)).$$

Since $\dim f(G) = \dim \partial G = n - 1$, the n vectors $Df(p)v_1, Df(p)v_2, \ldots, Df(p)v_n$ can be thought as beeing n vectors in an $n-1$ dimensional linear space; hence, they are linearly dependent. Consequently, any alternating n-form on those vectors is 0. Thus $f^*(\mathrm{d}\omega) = 0$. By Stokes' theorem

$$\int_{\partial G} f^*(\omega) = 0 = \int_G f^*(\mathrm{d}\omega).$$

On the other hand, $f = \mathrm{id}$ on ∂G such that

$$f^*(\omega) = \omega \mid_{\partial G} = x_1 \, \mathrm{d}x_2 \wedge \cdots \wedge \mathrm{d}x_n \mid_{\partial G} .$$

Again, by Stokes' theorem,

$$0 = \int_{\partial G} f^*(\omega) = \int_G \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n = |\, G \,| \,;$$

a contradiction. ∎

Theorem 11.9 (Brower's Fixed Point Theorem) *Let $g\colon B_1 \to B_1$ a continuous mapping of the closed unit ball $B_1 \subset \mathbb{R}^n$ into itself.*
Then f has a fixed point p, $f(p) = p$.

Proof. (a) We first prove that the theorem holds true for a twice continuously differentiable mapping g. Suppose to the contrary that g has no fixed point. For any $p \in B_1$ the line through p and $f(p)$ is then well defined. Let $h(p)$ be those intersection point of the above line with the the unit sphere S^{n-1} such that $h(p) - p$ is a positive multiple of $f(p) - p$. In particular, h is a C^2-mapping from B_1 into S^{n-1} which is the identity on S^{n-1}. By the previous proposition, such a mapping does not exist. Hence, f has a fixed point.

(b) For a continuous mapping one needs Stone–Weierstraß to approximate the continuous functions by polynomials. ∎

In case $n = 1$ Brower's theorem is just the intermediate value theorem.

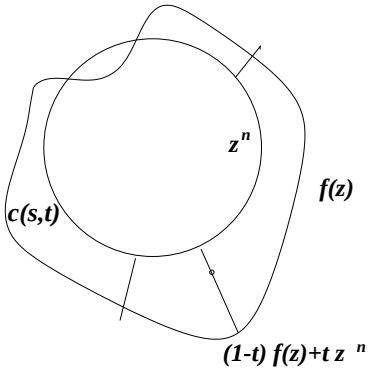
(b) The Fundamental Theorem of Algebra

We give a first proof of the fundamental theorem of algebra, Theorem 5.19:

> Every polynomial $f(z) = z^n + a_1 z^{n-1} + \cdots + a_n$ with complex coefficients $a_i \in \mathbb{C}$ has a root in $\mathbb{C}$.

We use two facts, the winding form ω on $\mathbb{R}^2 \setminus \{(0,0)\}$ is closed but not exact and z^n and $f(z)$ are "close together" for sufficiently large $|\, z \,|$.

We view $\mathbb{C}$ as $\mathbb{R}^2$ with $(a, b) = a + b\mathrm{i}$. Define the following singular 1-cubes on $\mathbb{R}^2$



$$c_{R,n}(s) = (R^n \cos(2\pi n s), R^n \sin(2\pi n s)) = z^n,$$
$$c_{R,f}(s) = f \circ c_{R,1}(s) = f(R\cos(2\pi s), R\sin(2\pi s)) = f(z),$$

where $z = z(s) = R(\cos 2\pi s + \mathrm{i} \sin 2\pi s)$, $s \in [0,1]$. Note that $|\, z \,| = R$. Further, let

$$\begin{aligned} c(s,t) &= (1-t)c_{R,f}(s) + t c_{R,n} \\ &= (1-t)f(z) + tz^n, \quad (s,t) \in [0,1]^2, \\ b(s,t) &= f((1-t)\, R(\cos 2\pi s, \sin 2\pi s)) \\ &= f((1-t)z), \quad (s,t) \in [0,1]^2 \end{aligned}$$

be singular 2-cubes in $\mathbb{R}^2$.

Lemma 11.10 *If $|\, z \,| = R$ is sufficiently large, then*

$$|\, c(s,t) \,| \geq \frac{R^n}{2}, \quad (s,t) \in [0,1]^2.$$

Proof. Since $f(z) - z^n$ is a polynomial of degree less than n,

$$\left| \frac{f(z) - z^n}{z^n} \right| \underset{z \to \infty}{\longrightarrow} 0,$$

in particular $|\, f(z) - z^n \,| \leq R^n/2$ if R is sufficiently large. Then we have

$$\begin{aligned} |\, c(s,t) \,| &= |\, (1-t)f(z) + tz^n \,| = |\, z^n + (1-t)(f(z) - z^n) \,| \\ &\geq |\, z^n \,| - (1-t)\, |\, f(z) - z^n \,| \geq R^n - \frac{R^n}{2} = \frac{R^n}{2}. \end{aligned}$$

■

The only fact we need is $c(s,t) \neq 0$ for sufficiently large R; hence, c maps the unit square into $\mathbb{R}^2 \setminus \{(0,0)\}$.

Lemma 11.11 *Let $\omega = \omega(x,y) = (-y\,\mathrm{d}x + x\,\mathrm{d}y)/(x^2 + y^2)$ be the winding form on $\mathbb{R}^2 \setminus \{(0,0)\}$. Then we have*
(a)

$$\partial c = c_{R,f} - c_{R,n},$$
$$\partial b = f(z) - f(0).$$

(b) *For sufficiently large R, c, $c_{R,n}$, and $c_{R,f}$ are chains in $\mathbb{R}^2 \setminus \{(0,0)\}$ and*

$$\int_{c_{R,n}} \omega = \int_{c_{R,f}} \omega = 2\pi n.$$

Proof. (a) Note that $z(0) = z(1) = R$. Since $\partial I_2(x) = (x,0) - (x,1) + (1,x) - (0,x)$ we have

$$\begin{aligned}\partial c(s) &= c(s,0) - c(s,1) + c(1,s) - c(0,s)\\ &= f(z) - z^n - ((1-s)f(R) + sR^n) + ((1-s)f(R) + sR^n) = f(z) - z^n.\end{aligned}$$

This proves (a). Similarly, we have

$$\begin{aligned}\partial b(s) &= b(s,0) - b(s,1) + b(1,s) - b(0,s)\\ &= f(z) - f(0) + f((1-s)R) - f((1-s)R) = f(z) - f(0).\end{aligned}$$

(b) By the Lemma 11.10, c is a singular 2-chain in $\mathbb{R}^2 \setminus \{(0,0,0)\}$ for sufficiently large R. Hence ∂c is a 1-chain in $\mathbb{R}^2 \setminus \{(0,0)\}$. In particular, both $c_{R,n}$ and $c_{R,f}$ take values in $\mathbb{R}^2 \setminus \{(0,0)\}$. Hence $(\partial c)^*(\omega)$ is well-defined. We compute $c^*_{R,n}(\omega)$ using the pull-backs of $\mathrm{d}x$ and $\mathrm{d}y$

$$\begin{aligned}{c_{R,n}}^*(x^2+y^2) &= R^{2n},\\ {c_{R,n}}^*(\,\mathrm{d}x) &= -2\pi n R^n \sin(2\pi n s)\,\mathrm{d}s,\\ {c_{R,n}}^*(\,\mathrm{d}y) &= 2\pi n R^n \cos(2\pi n s)\,\mathrm{d}s,\\ {c_{R,n}}^*(\omega) &= 2\pi n\,\mathrm{d}s.\end{aligned}$$

Hence

$$\int_{c_{R,n}} \omega = \int_0^1 2\pi n\,\mathrm{d}s = 2\pi n.$$

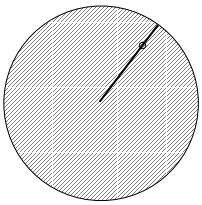
By Stokes' theorem and since ω is closed,

$$\int_{\partial c} \omega = \int_c \mathrm{d}\,\omega = 0,$$

such that by (a), and the above calculation

$$0 = \int_{\partial c} \omega = \int_{c_{R,n}} \omega - \int_{c_{R,f}} \omega, \quad \text{hence} \quad \int_{c_{R,n}} \omega = \int_{c_{R,f}} \omega = 2\pi n.$$

■



We complete the proof of the fundamental theorem of algebra. Suppose to the contrary that the polynomial $f(z)$ is non-zero in $\mathbb{C}$, then b as well as ∂b are singular chains in $\mathbb{R}^2 \setminus \{(0,0)\}$.

By Lemma 11.11 (b) and again by Stokes' theorem we have

$$\int_{c_{R,f}} \omega = \int_{c_{R,f}-f(0)} \omega = \int_{\partial b} \omega = \int_b \mathrm{d}\omega = 0.$$

But this is a contradiction to Lemma 11.11 (b). Hence, b is not a 2-chain in $\mathbb{R}^2 \setminus \{(0,0)\}$, that is there exist $s,t \in [0,1]$ such that $b(s,t) = f((1-t)z) = 0$. We have found that $(1-$

$t)R(\cos(2\pi s) + \mathrm{i}\sin(2\pi s))$ is a zero of f. Actually, we have shown a little more. There is a zero of f in the disc $\{z \in \mathbb{C} \mid |\, z \,| \leq R\}$ where $R \geq \max\{1, 2\sum_i |\, a_i \,|\}$. Indeed, in this case

$$|\, f(z) - z^n \,| \leq \sum_{k=1}^{n-1} |\, a_k \,| \left|\, z^{n-k} \,\right| \leq \sum_{k=1}^{n-1} |\, a_k \,|\, R^{n-1} \leq \frac{R^n}{2}$$

and this condition ensures $|\, c(s,t) \,| \neq 0$ as in the proof of Lemma 11.10.

Chapter 12

Measure Theory and Integration

12.1 Measure Theory

Citation from Rudins book, [Rud66, Chapter 1]: Towards the end of the 19th century it became clear to many mathematicians that the Riemann integral should be replaced by some other type of integral, more general and more flexible, better suited for dealing with limit processes. Among the attempts made in this direction, the most notable ones were due to Jordan, Borel, W.H. Young, and Lebesgue. It was Lebesgue's construction which turned out to be the most successful.
In a brief outline, here is the main idea: The Riemann integral of a function f over an interval $[a, b]$ can be approximated by sums of the form

$$\sum_{i=1}^{n} f(t_i)\, m(E_i),$$

where $E_1, \ldots, E_n$ are disjoint intervals whose union is $[a, b]$, $m(E_i)$ denotes the length of E_i and $t_i \in E_i$ for $i = 1, \ldots, n$. Lebesgue discovered that a completely satisfactory theory of integration results if the sets E_i in the above sum are allowed to belong to a larger class of subsets of the line, the so-called "measurable sets," and if the class of functions under consideration is enlarged to what we call "measurable functions." The crucial set-theoretic properties involved are the following: The union and the intersection of any countable family of measurable sets are measurable;... the notion of "length" (now called "measure") can be extended to them in such a way that

$$m(E_1 \cup E_2 \cup \cdots) = m(E_1) + m(E_2) + \cdots$$

for any countable collection $\{E_i\}$ of pairwise disjoint measurable sets. This property of m is called *countable additivity*.

The passage from Riemann's theory of integration to that of Lebesgue is a process of completion. It is of the same fundamental importance in analysis as the construction of the real number system from rationals.

12.1.1 Algebras, σ-algebras, and Borel Sets

(a) The Measure Problem. Definitions

Lebesgue (1904) states the following problem: We want to associate to each bounded subset E of the real line a positive real number $m(E)$, called measure of E, such that the following properties are satisfied:

(1) Any two congruent sets (by shift or reflexion) have the same measure.
(2) The measure is countably additive.
(3) The measure of the unit interval $[0, 1]$ is 1.

He emphasized that he was not able to solve this problem in full detail, but for a certain class of sets which he called measurable. We will see that this restriction to a large family of bounded sets is unavoidable—the measure problem has no solution.

Definition 12.1 Let X be a set. A family (non-empty) family $\mathcal{A}$ of subsets of X is called an *algebra* if $A, B \in \mathcal{A}$ implies $A^{\mathsf{c}} \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$.
An algebra $\mathcal{A}$ is called a σ*-algebra* if for all countable families $\{A_n \mid n \in \mathbb{N}\}$ with $A_n \in \mathcal{A}$ we have

$$\bigcup_{n\in\mathbb{N}} A_n = A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots \in \mathcal{A}.$$

Remarks 12.1 (a) Since $A \in \mathcal{A}$ implies $A \cup A^{\mathsf{c}} \in \mathcal{A}$; $X \in \mathcal{A}$ and $\varnothing = X^{\mathsf{c}} \in \mathcal{A}$.
(b) If $\mathcal{A}$ is an algebra, then $A \cap B \in \mathcal{A}$ for all $A, B \in \mathcal{A}$. Indeed, by de Morgan's rule, $(\bigcup_\alpha A_\alpha)^{\mathsf{c}} = \bigcap_\alpha A_\alpha^{\mathsf{c}}$ and $(\bigcap_\alpha A_\alpha)^{\mathsf{c}} = \bigcup_\alpha A_\alpha^{\mathsf{c}}$, we have $A \cap B = (A^{\mathsf{c}} \cup B^{\mathsf{c}})^{\mathsf{c}}$, and all the members on the right are in $\mathcal{A}$ by the definition of an algebra.
(c) Let $\mathcal{A}$ be a σ-algebra. Then $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$ if $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$. Again by de Morgan's rule

$$\bigcap_{n\in\mathbb{N}} A_n = \left(\bigcup_{n\in\mathbb{N}} A_n^{\mathsf{c}}\right)^{\mathsf{c}}.$$

(d) The family $\mathcal{P}(X)$ of all subsets of X is both an algebra as well as a σ-algebra.
(e) Any σ-algebra is an algebra but there are algebras not being σ-algebras.
(f) The family of finite and cofinite subsets (these are complements of finite sets) of an infinite set form an algebra. Do they form a σ-algebra?

(b) Elementary Sets and Borel Sets in $\mathbb{R}^n$

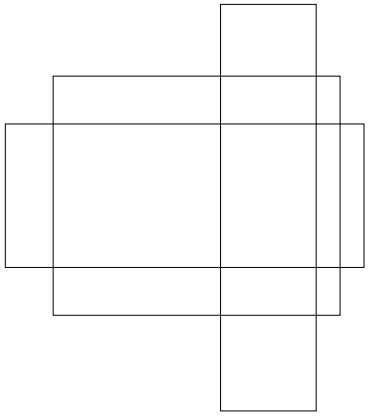
Let $\overline{\mathbb{R}}$ be the extended real axis together with $\pm\infty$, $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. We use the old rules as introduced in Section 3.1.1 at page 79. The new rule which is used in measure theory only is

$$0 \cdot \pm\infty = \pm\infty \cdot 0 = 0.$$

The set

$$I = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i \preceq x_i \preceq b_i, \quad i = 1, \ldots, n\}$$

is called a *rectangle* or a *box* in $\mathbb{R}^n$, where $\preceq$ either stands for $<$ or for $\leq$ where $a_i, b_i \in \overline{\mathbb{R}}$. For example $a_i = -\infty$ and $b_i = +\infty$ yields $I = \mathbb{R}^n$, whereas $a_1 = 2$, $b_1 = 1$ yields $I = \varnothing$. A subset of $\mathbb{R}^n$ is called an *elementary set* if it is the union of a finite number of rectangles in $\mathbb{R}^n$. Let $\mathcal{E}_n$ denote the set of elementary subsets of $\mathbb{R}^n$. $\mathcal{E}_n = \{I_1 \cup I_2 \cup \cdots \cup I_r \mid r \in \mathbb{N}, \quad I_j \text{ is a box in } \mathbb{R}^n\}$.



Lemma 12.1 *$\mathcal{E}_n$ is an algebra but not a σ-algebra.*

Proof. The complement of a finite interval is the union of two intervals, the complement of an infinite interval is an infinite interval. Hence, the complement of a rectangle in $\mathbb{R}^n$ is the finite union of rectangles.
The countable (disjoint) union $M = \bigcup_{n\in\mathbb{N}}[n, n + \frac{1}{2}]$ is not an elementary set. ∎

Note that any elementary set is the *disjoint* union of a finite number of rectangles.

Let $\mathcal{B}$ be any (nonempty) family of subsets of X. Let $\sigma(\mathcal{B})$ denote the intersection of all σ-algebras containing $\mathcal{B}$, i. e. $\sigma(\mathcal{B}) = \bigcap_{i\in I} \mathcal{A}_i$, where $\{\mathcal{A}_i \mid i \in I\}$ is the family of all σ-algebras $\mathcal{A}_i$ which contain $\mathcal{B}$, $\mathcal{B} \subseteq \mathcal{A}_i$ for all $i \in I$.
Note that the σ-algebra $\mathcal{P}(X)$ of all subsets is always a member of that family $\{\mathcal{A}_i\}$ such that $\sigma(\mathcal{B})$ exists. We call $\sigma(\mathcal{B})$ the σ-algebra *generated by* $\mathcal{B}$. By definition, $\sigma(\mathcal{B})$ is the smallest σ-algebra which contains the sets of $\mathcal{B}$. It is indeed a σ-algebra. Moreover, $\sigma(\sigma(\mathcal{B})) = \sigma(\mathcal{B})$ and if $\mathcal{B}_1 \subset \mathcal{B}_2$ then $\sigma(\mathcal{B}_1) \subset \sigma(\mathcal{B}_2)$.

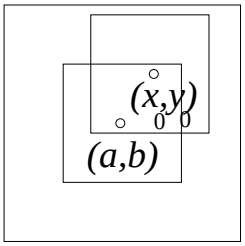
Definition 12.2 The *Borel algebra* $\mathcal{B}_n$ in $\mathbb{R}^n$ is the σ-algebra generated by the elementary sets $\mathcal{E}_n$. Its elements are called *Borel sets*.

The Borel algebra $\mathcal{B}_n = \sigma(\mathcal{E}_n)$ is the smallest σ-algebra which contains all boxes in $\mathbb{R}^n$ We will see that the Borel algebra is a huge family of subsets of $\mathbb{R}^n$ which contains "all sets we are ever interested in." Later, we will construct a non-Borel set.

Proposition 12.2 *Open and closed subsets of $\mathbb{R}^n$ are Borel sets.*

Proof. We give the proof in case of $\mathbb{R}^2$. Let $I_\varepsilon(x_0, y_0) = (x_0 - \varepsilon, x_0 + \varepsilon) \times (y_0 - \varepsilon, y_0 + \varepsilon)$ denote the open square of size 2ε by 2ε with midpoint (x_0, y_0). Then $I_{\frac{1}{n+1}} \subseteq I_{\frac{1}{n}}$ for $n \in \mathbb{N}$. Let $M \subset \mathbb{R}^2$ be open. To every point (x_0, y_0) with rational coordinates x_0, y_0 we choose the largest square $I_{1/n}(x_0, y_0) \subseteq M$ in M and denote it by $J(x_0, y_0)$. We show that

$$M = \bigcup_{(x_0,y_0)\in M,\, x_0,y_0\in\mathbb{Q}} J(x_0, y_0).$$



Since the number of rational points in M is at least countable, the right side is in $\sigma(\mathcal{E}_2)$. Now let $(a,b) \in M$ arbitrary. Since M is open, there exists $n \in \mathbb{N}$ such that $I_{2/n}(a,b) \subseteq M$. Since the rational points are dense in $\mathbb{R}^2$, there is rational point (x_0, y_0) which is contained in $I_{1/n}(a,b)$. Then we have

$$I_{\frac{1}{n}}(x_0,y_0) \subseteq I_{\frac{2}{n}}(a,b) \subseteq M.$$

Since $(a,b) \in I_{\frac{1}{n}}(x_0,y_0) \subseteq J(x_0,y_0)$, we have shown that M is the union of the countable family of sets J. Since closed sets are the complements of open sets and complements are again in the σ-algebra, the assertion follows for closed sets. ∎

Remarks 12.2 (a) We have proved that any open subset M of $\mathbb{R}^n$ is the countable union of rectangles $I \subseteq M$.
(b) The Borel algebra $\mathcal{B}_n$ is also the σ-algebra generated by the family of open or closed sets in $\mathbb{R}^n$, $\mathcal{B}_n = \sigma(\mathcal{G}_n) = \sigma(\mathcal{F}_n)$. Countable unions and intersections of open or closed sets are in $\mathcal{B}_n$.
Let us look in more detail at some of the sets in $\sigma(\mathcal{E}_n)$. Let $\mathcal{G}$ and $\mathcal{F}$ be the families of all open and closed subsets of $\mathbb{R}^n$, respectively. Let $\mathcal{G}_\delta$ be the collection of all intersection of sequences of open sets (from $\mathcal{G}$), and let $\mathcal{F}_\sigma$ be the collection of all unions of sequences of sets of $\mathcal{F}$. One can prove that $\mathcal{F} \subset \mathcal{G}_\delta$ and $\mathcal{G} \subset \mathcal{F}_\sigma$. These inclusions are strict. Since countable intersection and unions of countable intersections and union are still countable operations, $\mathcal{G}_\delta, \mathcal{F}_\sigma \subset \sigma(\mathcal{E}_n)$
For an arbitrary family $\mathcal{S}$ of sets let $\mathcal{S}_\sigma$ be the collection of all unions of sequences of sets in $\mathcal{S}$, and let $\mathcal{S}_\delta$ be the collection of all unions of sequences of sets in $\mathcal{S}$. We can iterate the operations represented by σ and δ, obtaining from the class $\mathcal{G}$ the classes $\mathcal{G}_\delta$, $\mathcal{G}_{\delta\sigma}$, $\mathcal{G}_{\delta\sigma\delta}$,... and from $\mathcal{F}$ the classes $\mathcal{F}_\sigma$, $\mathcal{F}_{\sigma\delta}$,.... It turns out that we have inclusions

$$\mathcal{G} \subset \mathcal{G}_\delta \subset \mathcal{G}_{\delta\sigma} \subset \cdots \subset \sigma(\mathcal{E}_n)$$
$$\mathcal{F} \subset \mathcal{F}_\sigma \subset \mathcal{F}_{\sigma\delta} \subset \cdots \subset \sigma(\mathcal{E}_n).$$

No two of these classes are equal. There are Borel sets that belong to none of them.

12.1.2 Additive Functions and Measures

Definition 12.3 (a) Let $\mathcal{A}$ be an algebra over X. An *additive function* or *content* μ on $\mathcal{A}$ is a function $\mu\colon \mathcal{A} \to [0, +\infty]$ such that

(i) $\mu(\varnothing) = 0$,
(ii) $\mu(A \cup B) = \mu(A) + \mu(B)$ for all $A, B \in \mathcal{A}$ with $A \cap B = \varnothing$.

(b) An additive function μ is called *countably additive* (or σ*-additive* in the German literature) on $\mathcal{A}$ if for any disjoint family $\{A_n \mid A_n \in \mathcal{A},\ n \in \mathbb{N}\}$, that is $A_i \cap A_j = \varnothing$ for all $i \neq j$, with

$\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ we have

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

(c) A *measure* is a countably additive function on a σ-algebra $\mathcal{A}$.
If X is a set, $\mathcal{A}$ a σ-algebra on X and μ a measure on $\mathcal{A}$, then the tripel $(X, \mathcal{A}, \mu)$ is called a *measure space*. Likewise, if X is a set and $\mathcal{A}$ a σ-algebra on X, the pair $(X, \mathcal{A})$ is called a *measurable space*.

Notation.
We write $\sum_{n\in\mathbb{N}} A_n$ in place of $\bigcup_{n\in\mathbb{N}} A_n$ if $\{A_n\}$ is a disjoint family of subsets. The countable additivity reads as follows

$$\mu(A_1 \cup A_2 \cup \cdots) = \mu(A_1) + \mu(A_2) + \cdots,$$
$$\mu\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

We say μ is *finite* if $\mu(X) < \infty$. If $\mu(X) = 1$, we call $(X, \mathcal{A}, \mu)$ a *probability space*. We call μ *σ-finite* if there exist sets $A_n \in \mathcal{A}$, with $\mu(A_n) < \infty$ and $X = \bigcup_{n=1}^{\infty} A_n$.

Example 12.1 (a) Let X be a set, $x_0 \in X$ and $\mathcal{A} = \mathcal{P}(X)$. Then

$$\mu(A) = \begin{cases} 1, & x_0 \in A, \\ 0, & x_0 \notin A \end{cases}$$

defines a finite measure on $\mathcal{A}$. μ is called the *point mass concentrated at* x_0.
(b1) Let X be a set and $\mathcal{A} = \mathcal{P}(X)$. Put

$$\mu(A) = \begin{cases} n, & \text{if } A \text{ has } n \text{ elements} \\ +\infty, & \text{if } A \text{ has infinitely many elements.} \end{cases}$$

μ is a measure on $\mathcal{A}$, the so called *counting measure*.
(b2) Let X be a set and $\mathcal{A} = \mathcal{P}(X)$. Put

$$\mu(A) = \begin{cases} 0, & \text{if } A \text{ has finitely many or countably many elements} \\ +\infty, & \text{if } A \text{ has uncountably many elements.} \end{cases}$$

μ is countably additive, not σ-finite.
(b3) Let X be a set and $\mathcal{A} = \mathcal{P}(X)$. Put

$$\mu(A) = \begin{cases} 0, & \text{if } A \text{ is finite} \\ +\infty, & \text{if } A \text{ is infinite.} \end{cases}$$

μ is additive, not σ-additive.

(c) $X = \mathbb{R}^n$, $\mathcal{A} = \mathcal{E}_n$ is the algebra of elementary sets of $\mathbb{R}^n$. Every $A \in \mathcal{E}_n$ is the finite disjoint union of rectangles $A = \sum_{k=1}^m I_k$. We set $\mu(A) = \sum_{k=1}^m \mu(I_k)$ where

$$\mu(I) = (b_1 - a_1) \cdots (b_n - a_n),$$

if $I = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i \preceq x_i \preceq b_i,\ i = 1, \ldots, n\}$ and $a_i \preceq b_i$; $\mu(\varnothing) = 0$. Then μ is an additive function on $\mathcal{E}_n$. It is called the *Lebesgue content* on $\mathbb{R}^n$. Note that μ is not a measure (since $\mathcal{A}$ is not a σ-algebra and μ is not yet shown to be countably additive). However, we will see in Proposition 12.5 below that μ is even *countably additive*. By definition, $\mu(\text{line in } \mathbb{R}^2) = 0$ and $\mu(\text{plane in } \mathbb{R}^3) = 0$.
(d) Let $X = \mathbb{R}$, $\mathcal{A} = \mathcal{E}_1$, and α an increasing function on $\mathbb{R}$. For a, b in $\overline{\mathbb{R}}$ with $a < b$ set

$$\begin{aligned}
\mu_\alpha([a, b]) &= \alpha(b + 0) - \alpha(a - 0), \\
\mu_\alpha([a, b)) &= \alpha(b - 0) - \alpha(a - 0), \\
\mu_\alpha((a, b]) &= \alpha(b + 0) - \alpha(a + 0), \\
\mu_\alpha((a, b)) &= \alpha(b - 0) - \alpha(a + 0).
\end{aligned}$$

Then μ_α is an additive function on $\mathcal{E}_1$ if we set

$$\mu_\alpha(A) = \sum_{i=1}^n \mu_\alpha(I_i), \quad \text{if} \quad A = \sum_{i=1}^n I_i.$$

We call μ_α the *Lebesgue–Stieltjes content*.
On the other hand, if $\mu\colon \mathcal{E}_1 \to \overline{\mathbb{R}}$ is an additive function, then $\alpha_\mu\colon \mathbb{R} \to \mathbb{R}$ defined by

$$\alpha_\mu(x) = \begin{cases} \mu((0, x]), & x \geq 0, \\ -\mu((x, 0]), & x < 0, \end{cases}$$

defines an increasing, right-continuous function α_μ on $\mathbb{R}$ such that $\mu = \mu_{\alpha_\mu}$. In general $\alpha \neq \alpha_{\mu_\alpha}$ since the function on the right hand side is continuous from the right whereas α is, in general, not .

Properties of Additive Functions

Proposition 12.3 *Let $\mathcal{A}$ be an algebra over X and μ an additive function on $\mathcal{A}$. Then*

(a) $\mu\left(\sum_{k=1}^n A_k\right) = \sum_{k=1}^n \mu(A_k)$ *if $A_k \in \mathcal{A}$, $k = 1, \ldots, n$ form a disjoint family of n subsets.*

(b) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$, $A, B \in \mathcal{A}$.

(c) $A \subseteq B$ *implies* $\mu(A) \leq \mu(B)$ (μ is monotone).

(d) *If* $A \subseteq B$, $A, B \in \mathcal{A}$, *and* $\mu(A) < +\infty$, *then* $\mu(B \setminus A) = \mu(B) - \mu(A)$, ($\mu$ is subtractive).

(e)
$$\mu\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} \mu(A_k),$$
if $A_k \in \mathcal{A}$, $k = 1, \ldots, n$, (μ is finitely subadditive).

(f) *If* $\{A_k \mid k \in \mathbb{N}\}$ *is a disjoint family in* $\mathcal{A}$ *and* $\sum_{k=1}^{\infty} A_k \in \mathcal{A}$. *Then*
$$\mu\left(\sum_{k=1}^{\infty} A_k\right) \geq \sum_{k=1}^{\infty} \mu(A_k).$$

Proof. (a) is by induction. (d), (c), and (b) are easy (cf. Homework 34.4).
(e) We can write $\bigcup_{k=1}^{n} A_k$ as the finite *disjoint* union of n sets of $\mathcal{A}$:
$$\bigcup_{k=1}^{n} A_k = A_1 \cup \big(A_2 \setminus A_1\big) \cup \big(A_3 \setminus (A_1 \cup A_2)\big) \cup \cdots \cup \big(A_n \setminus (A_1 \cup \cdots \cup A_{n-1})\big).$$
Since μ is additive,
$$\mu\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mu\left(A_k \setminus (A_1 \cup \cdots A_{k-1})\right) \leq \sum_{k=1}^{n} \mu(A_k),$$
where we used $\mu(B \setminus A) \leq \mu(B)$ (from (d)).
(f) Since μ is additive, and monotone
$$\sum_{k=1}^{n} \mu(A_k) = \mu\left(\sum_{k=1}^{n} A_k\right) \leq \mu\left(\sum_{k=1}^{\infty} A_k\right).$$
Taking the supremum on the left gives the assertion. ∎

Proposition 12.4 *Let* μ *be an additive function on the algebra* $\mathcal{A}$. *Consider the following statements*

(a) μ *is countably additive.*

(b) *For any increasing sequence* $A_n \subseteq A_{n+1}$, $A_n \in \mathcal{A}$, *with* $\bigcup_{n=1}^{\infty} A_n = A \in \mathcal{A}$ *we have* $\lim_{n\to\infty} \mu(A_n) = \mu(A)$.

(c) *For any decreasing sequence* $A_n \supseteq A_{n+1}$, $A_n \in \mathcal{A}$, *with* $\bigcap_{n=1}^{\infty} A_n = A \in \mathcal{A}$ *and* $\mu(A_n) < \infty$ *we have* $\lim_{n\to\infty} \mu(A_n) = \mu(A)$.

(d) *Statement* (c) *with* $A = \varnothing$ *only.*

We have (a) $\leftrightarrow$ (b) $\rightarrow$ (c) $\rightarrow$ (d). *In case* $\mu(X) < \infty$ (μ *is finite), all statements are equivalent.*

Proof. (a) $\to$ (b). Without loss of generality $A_1 = \varnothing$. Put $B_n = A_n \setminus A_{n-1}$ for $n = 2, 3, \ldots$. Then $\{B_n\}$ is a disjoint family with $A_n = B_2 \cup B_3 \cup \cdots \cup B_n$ and $A = \bigcup_{n=1}^{\infty} B_n$. Hence, by countable additivity of μ,

$$\mu(A) = \sum_{k=2}^{\infty} \mu(B_k) = \lim_{n\to\infty} \sum_{k=2}^{n} \mu(B_k) = \lim_{n\to\infty} \mu\left(\sum_{k=2}^{n} B_k\right) = \lim_{n\to\infty} \mu(A_n).$$

(b) $\to$ (a). Let $\{A_n\}$ be a family of disjoint sets in $\mathcal{A}$ with $\bigcup A_n = A \in \mathcal{A}$; put $B_k = A_1 \cup \cdots \cup A_k$. Then B_k is an increasing to A sequence. By (b)

$$\mu(B_n) = \mu\left(\sum_{k=1}^{n} A_k\right) \underset{\mu \text{ is additive}}{=} \sum_{k=1}^{n} \mu(A_k) \underset{n\to\infty}{\longrightarrow} \mu(A) = \mu\left(\sum_{k=1}^{\infty} A_k\right);$$

Thus, $\displaystyle\sum_{k=1}^{\infty} \mu(A_k) = \mu\left(\sum_{k=1}^{\infty} A_k\right)$.

(b) $\to$ (c). Since A_n is decreasing to A, $A_1 \setminus A_n$ is increasing to $A_1 \setminus A$. By (b)

$$\mu(A_1 \setminus A_n) \underset{n\to\infty}{\longrightarrow} \mu(A_1 \setminus A),$$

hence $\mu(A_1) - \mu(A_n) \underset{n\to\infty}{\longrightarrow} \mu(A_1) - \mu(A)$ which implies the assertion.

(c) $\to$ (d) is trivial.

Now let μ be finite, in particular, $\mu(B) < \infty$ for all $B \in \mathcal{A}$. We show (d) $\to$ (b). Let (A_n) be an increasing to A sequence of subsets $A, A_n \in \mathcal{A}$. Then $(A \setminus A_n)$ is a decreasing to $\varnothing$ sequence. By (d), $\mu(A \setminus A_n) \underset{n\to\infty}{\longrightarrow} 0$. Since μ is subtractive (Proposition 12.3 (d)) and all values are finite, $\mu(A_n) \underset{n\to\infty}{\longrightarrow} \mu(A)$. ∎

Proposition 12.5 *Let α be a right-continuous increasing function $\alpha\colon \mathbb{R} \to \mathbb{R}$, and μ_α the corresponding Lebesgue–Stieltjes content on $\mathcal{E}_1$. Then μ_α is countably additive if.*

Proof. For simplicity we write μ for μ_α. Recal that $\mu((a, b]) = \alpha(b) - \alpha(a)$. We will perform the proof in case of

$$(a, b] = \bigcup_{k=1}^{\infty} (a_k, b_k]$$

with a disjoint family $[a_k, b_k)$ of intervals. By Proposition 12.3 (f) we already know

$$\mu\left((a, b]\right) \geq \sum_{k=1}^{\infty} \mu((a_k, b_k]). \tag{12.1}$$

We prove the opposite direction. Let $\varepsilon > 0$. Since α is continuous from the right at a, there exists $a_0 \in [a, b)$ such that $\alpha(a_0) - \alpha(a) < \varepsilon$ and, similarly, for every $k \in \mathbb{N}$ there exists $c_k > b_k$ such that $\alpha(c_k) - \alpha(b_k) < \varepsilon/2^k$. Hence,

$$[a_0, b] \subset \bigcup_{k=1}^{\infty} (a_k, b_k] \subset \bigcup_{k=1}^{\infty} (a_k, c_k)$$

is an open covering of a compact set. By Heine–Borel (Definition 6.14) there exists a finite subcover

$$[a_0, b] \subset \bigcup_{k=1}^{N} (a_k, c_k), \quad \text{hence} \quad (a_0, b] \subset \bigcup_{k=1}^{N} (a_k, c_k],$$

such that by Proposition 12.3 (e)

$$\mu((a_0, b]) \leq \sum_{k=1}^{N} \mu((a_k, c_k]).$$

By the choice of a_0 and c_k,

$$\mu((a_k, c_k]) = \mu((a_k, b_k]) + \alpha(c_k) - \alpha(b_k) \leq \mu((a_k, b_k]) + \frac{\varepsilon}{2^k}.$$

Similarly, $\mu((a, b]) = \mu((a, a_0]) + \mu((a_0, b])$ such that

$$\begin{aligned} \mu((a, b]) &\leq \mu((a_0, b]) + \varepsilon \leq \sum_{k=1}^{N} \left(\mu((a_k, b_k]) + \frac{\varepsilon}{2^k} \right) + \varepsilon \\ &\leq \sum_{k=1}^{N} \mu((a_k, b_k]) + 2\varepsilon \leq \sum_{k=1}^{\infty} \mu((a_k, b_k]) + 2\varepsilon \end{aligned}$$

Since ε was arbitrary,

$$\mu((a, b]) \leq \sum_{k=1}^{\infty} \mu((a_k, b_k]).$$

In view of (12.1), μ_α is countably additive. ∎

Corollary 12.6 *The correspondence $\mu \mapsto \alpha_\mu$ from Example 12.1 (d) defines a bijection between countably additive functions μ on $\mathcal{E}_1$ and the monotonically increasing, right-continuous functions α on $\mathbb{R}$ (up to constant functions, i. e. α and $\alpha + c$ define the same additive function).*

Historical Note. It was the great achievment of Émile Borel (1871–1956) that he really *proved* the countable additivity of the Lebesgue measure. He realized that the countable additivity of μ is a serious mathematical problem far from being evident.

12.1.3 Extension of Countably Additive Functions

Here we must stop the rigorous treatment of measure theory. Up to now, we know only two trivial examples of measures (Example 12.1 (a) and (b)). We give an outline of the steps toward the construction of the Lebesgue measure.

- Construction of an *outer measure* μ^* on $\mathcal{P}(X)$ from a countably additive function μ on an algebra $\mathcal{A}$.
- Construction of the σ-algebra $\mathcal{A}_\mu$ of *measurable sets*.

The extension theory is due to Carathéodory (1914). For a detailed treatment, see [Els02, Section II.4].

Theorem 12.7 (Extension and Uniqueness) *Let μ be a countably additive function on the algebra $\mathcal{A}$.*
(a) *There exists an extension of μ to a measure on the σ-algebra $\sigma(\mathcal{A})$ which coincides with μ on $\mathcal{A}$. We denote the measure on $\sigma(\mathcal{A})$ also by μ. It is defined as the restriction of the outer measure $\mu^*\colon \mathcal{P}(X) \to [0,\infty]$*

$$\mu^*(A) = \inf\left\{\sum_{k=n}^{\infty} \mu(A_n) \mid A \subset \bigcup_{n=1}^{\infty} A_n,\ A_n \in \mathcal{A}, n \in \mathbb{N}\right\}$$

to the μ-measurable sets $\mathcal{A}_{\mu^*}$.
(b) *This extension is unique, if $(X, \mathcal{A}, \mu)$ is σ-finite.*

(For a proof, see [Brö92, (2.6), p.68])

Remark 12.3 (a) A subset $A \subset X$ is said to be *μ-measurable* if for all $Y \subset X$

$$\mu^*(Y) = \mu^*(A \cap Y) + \mu^*(A^{\mathsf{c}} \cap Y).$$

The family of μ-measurable sets form a σ-algebra $\mathcal{A}_{\mu^*}$.
(b) We have $\mathcal{A} \subset \sigma(\mathcal{A}) \subset \mathcal{A}_{\mu^*}$ and $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{A}$.
(c) $(X, \mathcal{A}_{\mu^*}, \mu^*)$ is a measure space, in particular, μ^* is countably additive on the measurable sets $\mathcal{A}_{\mu^*}$ and we redenote it by μ.

12.1.4 The Lebesgue Measure on $\mathbb{R}^n$

Using the facts from the previous subsection we conclude that for any increasing, right continuous function α on $\mathbb{R}$ there exists a measure μ_α on the σ-algebra of Borel sets. We call this measure the *Lebesgue–Stieltjes measure on* $\mathbb{R}$. In case $\alpha(x) = x$ we call it the *Lebesgue measure*. Extending the Lebesgue content on elementary sets of $\mathbb{R}^n$ to the Borel algebra $\mathcal{B}_n$, we obtain the *n-dimensional Lebesgue measure* λ_n on $\mathbb{R}^n$.

Completeness

A measure $\mu\colon \mathcal{A} \to \overline{\mathbb{R}}_+$ on a σ-algebra $\mathcal{A}$ is said to be *complete* if $A \in \mathcal{A}$, $\mu(A) = 0$, and $B \subset A$ implies $B \in \mathcal{A}$. It turns out that the Lebesgue measure λ_n on the Borel sets of $\mathbb{R}^n$ is not complete. Adjoining to $\mathcal{B}_n$ the subsets of measure-zero-sets, we obtain the σ-algebra $\mathcal{A}_{\lambda_n}$ of Legesgue measurable sets $\mathcal{A}_{\lambda_n}$.

$$\mathcal{A}_{\lambda_n} = \sigma\left(\mathcal{B}_n \cup \{X \subseteq \mathbb{R}^n \mid \ \exists\, B \in \mathcal{B}_n\colon X \subset E, \quad \lambda_n(B) = 0\}\right).$$

The Lebesgue measure λ_n on $\mathcal{A}_{\lambda_n}$ is now complete.

Remarks 12.4 (a) The Lebesgue measure is invariant under the *motion group* of $\mathbb{R}^n$. More precisely, let $\mathrm{O}(n) = \{T \in \mathbb{R}^{n\times n} \mid T^\top T = T\,T^\top = E_n\}$ be the group of real orthogonal $n \times n$-matrices ("motions"), then

$$\lambda_n(T(A)) = \lambda_n(A), \quad A \in \mathcal{A}_{\lambda_n}, \quad T \in \mathrm{O}(n).$$

(b) λ_n is *translation invariant*, i. e. $\lambda_n(A) = \lambda_n(x+A)$ for all $x \in \mathbb{R}^n$. Moreover, the invariance of λ_n under translations uniquely characterizes the Lebesgue measure λ_n: If λ is a translation invariant measure on $\mathcal{B}_n$, then $\lambda = c\lambda_n$ for some $c \in \mathbb{R}_+$.
(c) There exist non-measurable subsets in $\mathbb{R}^n$. We construct a subset E of $\mathbb{R}$ that is not Lebesgue measurable.
We write $x \sim y$ if $x - y$ is rational. This is an equivalence relation since $x \sim x$ for all $x \in \mathbb{R}$, $x \sim y$ implies $y \sim x$ for all x and y, and $x \sim y$ and $y \sim z$ implies $x \sim z$. Let E be a subset of $(0, 1)$ that contains exactly one point in every equivalence class. (the assertion that there is such a set E is a direct application of the *axiom of choice*). We claim that E is not measurable. Let $E + r = \{x + r \mid x \in E\}$. We need the following two properties of E:

(a) If $x \in (0, 1)$, then $x \in E + r$ for some rational $r \in (-1, 1)$.
(b) If r and s are distinct rationals, then $(E + r) \cap (E + s) = \varnothing$.

To prove (a), note that for every $x \in (0, 1)$ there exists $y \in E$ with $x \sim y$. If $r = x - y$, then $x = y + r \in E + r$.
To prove (b), suppose that $x \in (E + r) \cap (E + s)$. Then $x = y + r = z + s$ for some $y, z \in E$. Since $y - z = s - r \neq 0$, we have $y \sim z$, and E contains two equivalent points, in contradiction to our choice of E.
Now assume that E is Lebesgue measurable with $\lambda(E) = \alpha$. Define $S = \bigcup(E + r)$ where the union is over all rational $r \in (-1, 1)$. By (b), the sets $E + r$ are pairwise disjoint; since λ is translation invariant, $\lambda(E + r) = \lambda(E) = \alpha$ for all r. Since $S \subset (-1, 2)$, $\lambda(S) \leq 3$. The countable additivity of λ now forces $\alpha = 0$ and hence $\lambda(S) = 0$. But (a) implies $(0, 1) \subset S$, hence $1 \leq \lambda(S)$, and we have a contradiction.
(d) Any countable set has Lebesgue measure zero. Indeed, every single point is a box with edges of length 0; hence $\lambda(\{\mathrm{pt}\}) = 0$. Since λ is countably additive,

$$\lambda(\{x_1, x_2, \ldots, x_n, \ldots\}) = \sum_{n=1}^{\infty} \lambda(\{x_n\}) = 0.$$

In particular, the rational numbers have Lebesgue measure 0, $\lambda(\mathbb{Q}) = 0$.
(e) There are uncountable sets with measure zero. The Cantor set (Cantor: 1845–1918, inventor of set theory) is a prominent example:

$$C = \left\{ \sum_{i=1}^{\infty} \frac{a_i}{3^i} \;\middle|\; a_i \in \{0, 2\} \;\; \forall\, i \in \mathbb{N} \right\};$$

Obviously, $C \subset [0, 1]$; C is compact and can be written as the intersection of a decreasing sequence (C_n) of closed subsets; $C_1 = [0, 1/3] \cup [2/3, 1]$, $\lambda(C_1) = 2/3$, and, recursively,

$$C_{n+1} = \frac{1}{3}C_n \cup \left(\frac{2}{3} + \frac{1}{3}C_n\right) \implies \lambda(C_{n+1}) = \frac{1}{3}\lambda(C_n) + \frac{1}{3}\lambda\left(C_n + \frac{2}{3}\right) = \frac{2}{3}\lambda(C_n).$$

It turns out that $C_n = \left\{ \sum_{i=1}^{\infty} \frac{a_i}{3^i} \;\middle|\; ia_i \in \{0,2\} \;\; \forall\, i = 1,\dots,n \right\}$ Clearly,

$$\lambda(C_{n+1}) = \frac{2}{3}\lambda(C_n) = \cdots = \left(\frac{2}{3}\right)^n \lambda(C_1) = \left(\frac{2}{3}\right)^{n+1}.$$

By Proposition 12.4 (c), $\lambda(C) = \lim\limits_{n\to\infty} \lambda(C_n) = 0$. However, C has the same cardinality as $\{0,2\}^{\mathbb{N}} \cong \{0,1\}^{\mathbb{N}} \cong \mathbb{R}$ which is uncountable.

12.2 Measurable Functions

Let $\mathcal{A}$ be a σ-algebra over X.

Definition 12.4 A real function $f\colon X \to \overline{\mathbb{R}}$ is called *$\mathcal{A}$-measurable* if for all $a \in \mathbb{R}$ the set $\{x \in X \mid f(x) > a\}$ belongs to $\mathcal{A}$.
A complex function $f\colon X \to \mathbb{C}$ is said to be *$\mathcal{A}$-measurable* if both $\operatorname{Re} f$ and $\operatorname{Im} f$ are $\mathcal{A}$-measurable.
A function $f\colon U \to \mathbb{R}$, $U \subset \mathbb{R}^n$, is said to be a *Borel function* if f is $\mathcal{B}_n$-measurable, i. e. f is measurable with respect to the Borel algebra on $\mathbb{R}^n$.
A function $f\colon U \to V$, $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$, is called a *Borel function* if $f^{-1}(B)$ is a Borel set for all Borel sets $B \subset V$. It is part of homework 39.3 (b) to prove that in case $m = 1$ these definitions coincide. Also, $f = (f_1,\dots,f_m)$ is a Borel function if all f_i are.

Note that $\{x \in X \mid f(x) > a\} = f^{-1}((a,+\infty))$. From Proposition 12.8 below it becomes clear that the last two notions are consistent. Note that no measure on $(X,\mathcal{A})$ needs to be specified to define a measurable function.

Example 12.2 (a) Any continuous function $f\colon U \to \mathbb{R}$, $U \subset \mathbb{R}^n$, is a Borel function. Indeed, since f is continuous and $(a,+\infty)$ is open, $f^{-1}((a,+\infty))$ is open as well and hence a Borel set (cf. Proposition 12.2).
(b) The characteristic function χ_A is $\mathcal{A}$-measurable if and only if $A \in \mathcal{A}$ (see homework 35.3).
(c) Let $f\colon U \to V$ and $g\colon V \to W$ be Borel functions. Then $g\circ f\colon U \to W$ is a Borel function, too. Indeed, for any Borel set $C \subset W$, $g^{-1}(C)$ is a Borel set in V since g is a Borel function. Since f is a Borel function $(g\circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$ is a Borel subset of U which shows the assertion.

Proposition 12.8 *Let $f\colon X \to \overline{\mathbb{R}}$ be a function. The following are equivalent*

(a) $\{x \mid f(x) > a\} \in \mathcal{A}$ *for all* $a \in \overline{\mathbb{R}}$ *(i. e. f is $\mathcal{A}$-measurable).*
(b) $\{x \mid f(x) \geq a\} \in \mathcal{A}$ *for all* $a \in \overline{\mathbb{R}}$.
(c) $\{x \mid f(x) < a\} \in \mathcal{A}$ *for all* $a \in \overline{\mathbb{R}}$.
(d) $\{x \mid f(x) \leq a\} \in \mathcal{A}$ *for all* $a \in \overline{\mathbb{R}}$.
(e) $f^{-1}(B) \in \mathcal{A}$ *for all Borel sets* $B \in \mathcal{B}_1$.

Proof. (a) $\to$ (b) follows from the identity

$$[a,+\infty] = \bigcap_{n\in\mathbb{N}} (a - 1/n, +\infty]$$

and the invariance of intersections under preimages, $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$. Since f is $\mathcal{A}$-measurable and $\mathcal{A}$ is a σ-algebra, the countable intersection on the right is in $\mathcal{A}$. (a) $\to$ (d) follows from $\{x \mid f(x) \le a\} = \{x \mid f(x) > a\}^{\mathsf{c}}$. The remaining directions are left to the reader (see also homework 35.5). ∎

Remark 12.5 (a) Let $f, g\colon X \to \overline{\mathbb{R}}$ be $\mathcal{A}$-measurable. Then $\{x \mid f(x) > g(x)\}$ and $\{x \mid f(x) = g(x)\}$ are in $\mathcal{A}$. *Proof.* Since

$$\{x \mid f(x) < g(x)\} = \bigcup_{q \in \mathbb{Q}} \left(\{x \mid f(x) < q\} \cap \{x \mid q < g(x)\}\right),$$

and all sets $\{f < q\}$ and $\{q < g\}$ the right are in $\mathcal{A}$, and on the right there is a countable union, the right hand side is in $\mathcal{A}$. A similar argument works for $\{f > g\}$. Note that the sets $\{f \ge g\}$ and $\{f \le g\}$ are the complements of $\{f < g\}$ and $\{f > g\}$, respectively; hence they belong to $\mathcal{A}$ as well. Finally, $\{f = g\} = \{f \ge g\} \cap \{f \le g\}$. ∎

(b) It is not difficult to see that for any sequence (a_n) of real numbers

$$\overline{\lim_{n\to\infty}}\, a_n = \inf_{n\in\mathbb{N}} \sup_{k\ge n} a_k \quad \text{and} \quad \underline{\lim_{n\to\infty}}\, a_n = \sup_{n\in\mathbb{N}} \inf_{k\ge n} a_k. \tag{12.2}$$

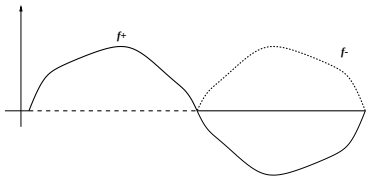
As a consequence we can construct new mesurable functions using $\sup$ and $\lim_{n\to\infty}$. Let (f_n) be a sequence of $\mathcal{A}$-measurable real functions on X. Then $\sup\limits_{n\in\mathbb{N}} f_n$, $\inf\limits_{n\in\mathbb{N}} f_n$, $\overline{\lim\limits_{n\to\infty}}\, f_n$, $\underline{\lim\limits_{n\to\infty}}\, f_n$ are $\mathcal{A}$-measurable. In particular $\lim\limits_{n\to\infty} f_n$ is measurable if the limit exists.
Proof. Note that for all $a \in \mathbb{R}$ we have

$$\{\sup f_n \le a\} = \bigcap_{n\in\mathbb{N}} \{f_n \le a\}.$$

Since all f_n are measurable, so is $\sup f_n$. A similar proof works for $\inf f_n$. By (12.2), $\overline{\lim\limits_{n\to\infty}}\, f_n$ and $\underline{\lim\limits_{n\to\infty}}\, f_n$, are measurable, too. ∎

Proposition 12.9 *Let $f, g\colon X \to \mathbb{R}$ Borel functions on $X \subset \mathbb{R}^n$. Then $\alpha f + \beta g$, $f \cdot g$, and $|f|$ are Borel functions, too.*

Proof. The function $h(x) = (f(x), g(x))\colon X \to \mathbb{R}^2$ is a Borel function since its coordinate functions are so. Since the sum $s(x, y) = x + y$ and the product $p(x, y) = xy$ are continuous functions, the compositions $s{\circ}h$ and $p{\circ}h$ are Borel functions by Example 12.2 (c). Since the constant functions α and β are Borel, so are αf, βg, and finally $\alpha f + \beta g$. Hence, the Borel functions over X form a linear space, moreover a real algebra. In particular $-f$ is Borel and so is $|f| = \max\{f, -f\}$. ∎



Let $(X, \mathcal{A}, \mu)$ be a measure space and $f\colon X \to \overline{\mathbb{R}}$ arbitrary. Let $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$ denote the *positive and negative parts of* f. We have $f = f^+ - f^-$ and $|f| = f^+ + f^-$; moreover $f^+,\ f^- \geq 0$.

Corollary 12.10 *Let f is a Borel function if and only if both f^+ and f^- are Borel.*

12.3 The Lebesgue Integral

We define the Lebesgue integral of a complex function in three steps; first for positive simple functions, then for positive measurable functions and finally for arbitrary measurable functions. In this section $(X, \mathcal{A}, \mu)$ is a measure space.

12.3.1 Simple Functions

Definition 12.5 Let $M \subseteq X$ be a subset. The function

$$\chi_M(x) = \begin{cases} 1, & x \in M, \\ 0, & x \notin M, \end{cases}$$

is called *characteristic function* of M.
An $\mathcal{A}$-measurabel function $f\colon X \to \mathbb{R}$ is called *simple* if f takes only finitely many values $c_1, \ldots, c_n$. We denote the set of simple functions on $(X, \mathcal{A})$ by $\mathcal{S}$; the set of non-negative simple functions is denoted by $\mathcal{S}_+$.

Clearly, if $c_1, \ldots, c_n$ are the distinct values of the simple function f, then

$$f = \sum_{i=1}^{n} c_i \chi_{A_i},$$

where $A_i = \{x \mid f(x) = c_i\}$. It is clear, that f measurable if and only if $A_i \in \mathcal{A}$ for all i. Obviously, $\{A_i \mid i = 1, \ldots, n\}$ is a disjoint family of subsets of X.
It is easy to see that $f, g \in \mathcal{S}$ implies $\alpha f + \beta g \in \mathcal{S}$, $\max\{f, g\} \in \mathcal{S}$, $\min\{f, g\} \in \mathcal{S}$, and $fg \in \mathcal{S}$.

Step 1: Positive Simple Functions

For $f = \sum_{i=1}^{n} c_i \chi_{A_i} \in \mathcal{S}_+$ define

$$\int_X f \, \mathrm{d}\mu = \sum_{i=1}^{n} c_i \, \mu(A_i). \tag{12.3}$$

The convention $0 \cdot (+\infty)$ is used here; it may happen that $c_i = 0$ for some i and $\mu(A_i) = +\infty$.

Remarks 12.6 (a) Since $c_i \geq 0$ for all i, the right hand side is well-defined in $\overline{\mathbb{R}}$.

(b) Given another presentation of f, say, $f(x) = \sum_{j=1}^{m} d_j \chi_{B_j}$, $\sum_{j=1}^{m} d_j \, \mu(B_j)$ gives the same value as (12.3).

The following properties are easily checked.

Lemma 12.11 *For $f, g \in \mathcal{S}_+$, $A \in \mathcal{A}$, $c \in \mathbb{R}_+$ we have*

(1) $\int_X \chi_A \, \mathrm{d}\mu = \mu(A)$.
(2) $\int_X cf \, \mathrm{d}\mu = c \int_X f \, \mathrm{d}\mu$.
(3) $\int_X (f+g) \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu$.
(4) $f \leq g$ *implies* $\int_X f \, \mathrm{d}\mu \leq \int_X g \, \mathrm{d}\mu$.

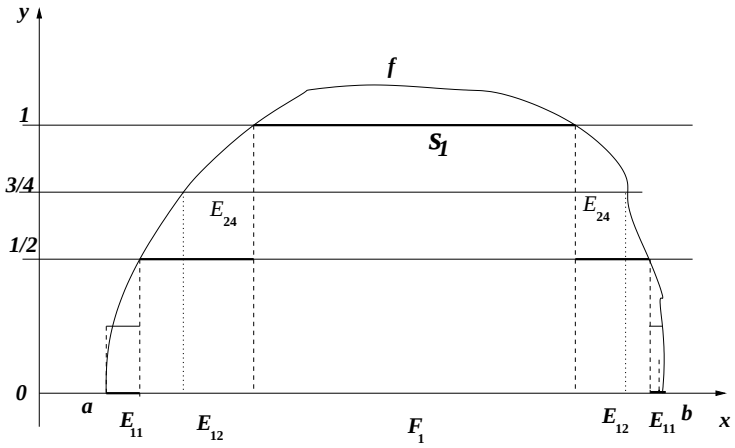
12.3.2 Positive Measurable Functions

The idea is to approximate a positive measurable function with an increasing sequence of positive simple ones.

Theorem 12.12 *Let $f \colon X \to [0, +\infty]$ be measurable. There exist simple functions s_n, $n \in \mathbb{N}$, on X such that*

(a) $0 \leq s_1 \leq s_2 \leq \cdots \leq f$.

(b) $s_n(x) \underset{n\to\infty}{\longrightarrow} f(x)$, *as $n \to \infty$, for every $x \in X$.*



Example. $X = (a, b)$, $n = 1$, $1 \leq i \leq 2$. Then

$$E_{11} = f^{-1}\left(\left[0, \frac{1}{2}\right)\right),$$
$$E_{12} = f^{-1}\left(\left(\frac{1}{2}, 1\right)\right),$$
$$F_2 = f^{-1}\left([1, +\infty]\right).$$

Proof. For $n \in \mathbb{N}$ and for $1 \leq i \leq n2^n$, define

$$E_{ni} = f^{-1}\left(\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right)\right) \quad \text{and} \quad F_n = f^{-1}([n, \infty])$$

and put

$$s_n = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \, \chi_{E_{ni}} + n\chi_{F_n}.$$

Proposition 12.8 shows that E_{ni} and F_n are measurable sets. It is easily seen that the functions s_n satisfy (a). If x is such that $f(x) < +\infty$, then

$$0 \leq f(x) - s_n(x) \leq \frac{1}{2^n} \tag{12.4}$$

as soon as n is large enough, that is, $x \in E_{ni}$ for some $n,\ i \in \mathbb{N}$ and not $x \in F_n$. If $f(x) = +\infty$, then $s_n(x) = n$; this proves (b). ∎

From (12.4) it follows, that $s_n \rightrightarrows f$ uniformly on X if f is bounded.

Step 2: Positive Measurable Real Functions

Definition 12.6 (Lebesgue Integral) Let $f\colon X \to [0, +\infty]$ be measurable. Let (s_n) be an increasing sequence of non-negative simple functions s_n converging to $f(x)$ for all $x \in X$, $\lim\limits_{n\to\infty} s_n(x) = \sup\limits_{n\in\mathbb{N}} s_n(x) = f(x)$. Define

$$\int_X f \,\mathrm{d}\mu = \lim_{n\to\infty} \int_X s_n \,\mathrm{d}\mu = \sup_{n\in\mathbb{N}} \int_X s_n \,\mathrm{d}\mu \tag{12.5}$$

and call this number in $[0, +\infty]$ the *Lebesgue integral of $f(x)$ over X with respect to the measure μ* or *μ-integral of f over X*.

The definition of the Lebesgue integral does not depend on the special choice of the increasing functions $s_n \nearrow f$. One can define

$$\int_X f \,\mathrm{d}\mu = \sup\left\{ \int_X s \,\mathrm{d}\mu \mid s \leq f, \ \text{ and } s \text{ is a simple function} \right\}.$$

Observe, that we apparently have two definitions for $\int_X f \,\mathrm{d}\mu$ if f is a simple function. However these assign the same value to the integral since f is the largest simple function greater than or equal to f.

Proposition 12.13 *The properties* (1) *to* (4) *from Lemma 12.11 hold for any non-negative measurable functions $f, g\colon X \to [0, +\infty]$, $c \in \mathbb{R}_+$.*

(Without proof.)

Step 3: Measurable Real Functions

Let $f\colon X \to \overline{\mathbb{R}}$ be measurable and $f^+(x) = \max(f, 0)$, $f^-(x) = \max(-f(x), 0)$. Then f^+ and f^- are both positive and measurable. Define

$$\int_X f \mathrm{d}\mu = \int_X f^+ \mathrm{d}\mu - \int_X f^- \mathrm{d}\mu$$

if at least one of the integrals on the right is finite. We say that f is *μ-integrable* if both are finite.

Step 4: Measurable Complex Functions

Definition 12.7 (Lebesgue Integral—Continued) A complex, measurable function $f\colon X \to \mathbb{C}$ is called *μ-integrable* if

$$\int_X |\, f \,|\, \mathrm{d}\mu < \infty.$$

If $f = u + \mathrm{i}v$ is μ-integrable, where $u = \operatorname{Re} f$ and $v = \operatorname{Im} f$ are the real and imaginary parts of f, u and v are real measurable functions on X. Define the *μ-integral of f over X* by

$$\int_X f \,\mathrm{d}\mu = \int_X u^+ \,\mathrm{d}\mu - \int_X u^- \,\mathrm{d}\mu + \mathrm{i}\int_X v^+ \,\mathrm{d}\mu - \mathrm{i}\int_X v^- \,\mathrm{d}\mu. \tag{12.6}$$

These four functions u^+, u^-, v^+, and v^- are measurable, real, and non-negative. Since we have $u^+ \le |\, u \,| \le |\, f \,|$ etc., each of these four integrals is finite. Thus, (12.6) defines the integral on the left as a complex number.

We define $\mathscr{L}^1(X, \mu)$ to be the collection of all complex μ-integrable functions f on X. Note that for an integrable functions f, $\int_X f \,\mathrm{d}\mu$ is a finite number.

Proposition 12.14 *Let $f, g\colon X \to \mathbb{C}$ be measurable.*

(a) *f is μ-integrable if and only if $|\, f \,|$ is μ-integrable and we have*

$$\left| \int_X f \,\mathrm{d}\mu \right| \le \int_X |\, f \,|\, \mathrm{d}\mu.$$

(b) *f is μ-integrable if and only if there exists an integrable function h with $|\, f \,| \le h$.*

(c) *If f, g are integrable, so is $c_1 f + c_2 g$ where*

$$\int_X (c_1 f + c_2 g) \,\mathrm{d}\mu = c_1 \int_X f \,\mathrm{d}\mu + c_2 \int_X g \,\mathrm{d}\mu.$$

(d) *If $f \le g$ on X, then $\int_X f \,\mathrm{d}\mu \le \int_X g \,\mathrm{d}\mu$.*

It follows that the set $\mathscr{L}^1(X, \mu)$ of μ-integrable complex-valued functions on X is a linear space. The Lebesgue integral defines a positive linear functional on $\mathscr{L}^1(X, \mu)$. Note that (b) implies that any measurable and bounded function f on a space X with $\mu(X) < \infty$ is integrable.

Step 5: $\int_A f \,\mathrm{d}\mu$

Definition 12.8 Let $A \in \mathcal{A}$, $f\colon X \to \overline{\mathbb{R}}$ or $f\colon X \to \mathbb{C}$ measurable. The function f is called *μ-integrable over A* if $\chi_A f$ is μ-integrable over X. In this case we put,

$$\int_A f \,\mathrm{d}\mu = \int_X \chi_A f \,\mathrm{d}\mu.$$

In particular, Lemma 12.11 (1) now reads $\int_A \mathrm{d}\mu = \mu(A)$.

12.4 Some Theorems on Lebesgue Integrals

12.4.1 The Role Played by Measure Zero Sets

Equivalence Relations

Let X be a set and $R \subset X \times X$. For $a, b \in X$ we write $a \sim b$ if $(a, b) \in R$.

Definition 12.9 (a) The subset $R \subset X \times X$ is said to be an *equivalence relation* if R is *reflexive*, *symmetric* and *transitive*, that is,

(r) $\forall\, x \in X : x \sim x$.
(s) $\forall\, x, y \in X : x \sim y \Longrightarrow y \sim x$.
(t) $\forall\, x, y, z \in X : x \sim y \wedge y \sim z \Longrightarrow x \sim z$.

For $a \in X$ the set $\overline{a} := \{x \in X \mid x \sim a\}$ is called the *equivalence class* of a. We have $\overline{a} = \overline{b}$ if and only if $a \sim b$.
(b) A *partition* $\mathcal{P}$ of X is a disjoint family $\mathcal{P} = \{A_\alpha \mid \alpha \in I\}$ of subsets $A_\alpha \subset X$ such that $\sum\limits_{\alpha \in I} A_\alpha = X$.

The set of equivalence classes is sometimes denoted by $X/\sim$.

Example 12.3 (a) On $\mathbb{Z}$ define $a \sim b$ if $2 \mid (a-b)$. a and b are equivalent if both are odd or both are even. There are two equivalence classes, $\overline{1} = \overline{-5} = 2\mathbb{Z} + 1$ (odd numbers), $\overline{0} = \overline{100} = 2\mathbb{Z}$ even numbers.
(b) Let $W \subset V$ be a subspace of the linear space V. For $x, y \in V$ define $x \sim y$ if $x - y \in W$. This is an equivalence relation, indeed, the relation is reflexive since $x - x = 0 \in W$, it is symmetric since $x - y \in W$ implies $y - x = -(x - y) \in W$, and it is transitive since $x - y, y - z \in W$ implies that there sum $(x - y) + (y - z) = x - z \in W$ such that $x \sim z$. One has $\overline{0} = W$ and $\overline{x} = x + W := \{x + w \mid w \in W\}$. Set set of equivalence classes with respect to this equivalence relation is called the *factor space* or *quotient space* of V with respect to W and is denoted by V/W. The factor space becomes a linear space if we define $\overline{x} + \overline{y} := \overline{x + y}$ and $\lambda\, \overline{x} = \overline{\lambda x}$, $\lambda \in \mathbb{C}$. Addition is indeed well-defined since $x \sim x'$ and $y \sim y'$, say, $x - x' = w_1$, $y - y' = w_2$, $w_1, w_2 \in W$ implies $x + y - (x' + y') = w_1 + w_2 \in W$ such that $\overline{x + y} = \overline{x' + y'}$.
(c) Similarly as in (a), for $m \in \mathbb{N}$ define the equivalence relation $a \equiv b \pmod m$ if $m \mid (a - b)$. We say "a is congruent b modulo m". This defines a partition of the integers into m disjoint equivalence classes $\overline{0}, \overline{1}, \ldots, \overline{m-1}$, where $\overline{r} = \{am + r \mid a \in \mathbb{Z}\}$.
(d) Two triangles in the plane are equivalent if

(1) there exists a *translation* such that the first one is mapped onto the second one.
(2) there exists a *rotation* around $(0, 0)$
(3) there exists a *motion* (rotation or translation or reflexion or composition)

Then (1) – (3) define different equivalence relations on triangles or more generally on subsets of the plane.
(e) Cardinality of sets is an equivalence relation.

Proposition 12.15 *(a) Let $\sim$ be an equivalence relation on X. Then $\mathcal{P} = \{\overline{x} \mid x \in X\}$ defines a partition on X denoted by $\mathcal{P}_\sim$.*
(b) Let $\mathcal{P}$ be a partition on X. Then $x \sim y$ if there exists $A \in \mathcal{P}$ with $x, y \in A$ defines an equivalence relation $\sim_\mathcal{P}$ on X.
(c) $\sim_{\mathcal{P}_\sim} = \sim$ and $\mathcal{P}_{\sim_\mathcal{P}} = \mathcal{P}$.

Let P be a property which a point x may or may not have. For instance, P may be the property "$f(x) > 0$" if f is a given function or "$(f_n(x))$ converges" if (f_n) is a given sequence of functions.

Definition 12.10 If $(X, \mathcal{A}, \mu)$ is a measure space and $A \in \mathcal{A}$, we say "*P holds almost everywhere on A*", abbreviated by "P holds a. e. on A", if there exists $N \in \mathcal{A}$ such that $\mu(N) = 0$ and P holds for every point $x \in A \setminus N$.

This concept of course strongly depends on the measure μ, and sometimes we write a. e. to emphasize the dependence on μ.
(a) Main example. On the set of measurable functions on X, $f\colon X \to \mathbb{C}$ we define an equivalence relation by

$$f \sim g, \quad \text{if} \quad f = g \ \text{ a. e. on } X.$$

This is indeed an equivalence relation since $f(x) = f(x)$ for all $x \in X$, $f(x) = g(x)$ implies $g(x) = f(x)$. Let $f = g$ a. e. on X and $g = h$ a. e. on X, that is, there exist $M, N \in \mathcal{A}$ with $\mu(M) = \mu(N) = 0$ and $f(x) = g(x)$ for all $x \in X \setminus M$, $g(x) = h(x)$ for all $x \in N$. Hence, $f(x) = h(x)$ for all $x \in M \cup N$. Since $0 \le \mu(M \cup N) \le \mu(M) + \mu(N) = 0 + 0 = 0$ by Proposition 12.3 (e), $\mu(M \cup N) = 0$ and finally $f = h$ a. e. on X.
(b) Note that $f = g$ a. e. on X implies

$$\int_X f \,\mathrm{d}\mu = \int_X g \,\mathrm{d}\mu.$$

Indeed, let N denote the zero-set where $f \neq g$. Then

$$\begin{aligned}\left| \int_X f \,\mathrm{d}\mu - \int_X g \,\mathrm{d}\mu \right| \le \int_X |\, f - g \,|\, \mathrm{d}\mu &= \int_N |\, f - g \,|\, \mathrm{d}\mu + \int_{X \setminus N} |\, f - g \,|\, \mathrm{d}\mu \\ &\le \mu(N)(\infty) + \mu(X \setminus N) \cdot 0 = 0.\end{aligned}$$

Here we used that for disjoint sets $A, B \in \mathcal{A}$,

$$\int_{A \cup B} f \,\mathrm{d}\mu = \int_X \chi_{A \cup B} f \,\mathrm{d}\mu = \int_X \chi_A f \,\mathrm{d}\mu + \int_X \chi_B f \,\mathrm{d}\mu = \int_A f \,\mathrm{d}\mu + \int_B f \,\mathrm{d}\mu.$$

Proposition 12.16 *Let $f\colon X \to [0, +\infty]$ be measurable. Then*

$\int_X f \,\mathrm{d}\mu = 0$ *if and only if $f = 0$ a. e. on X.*

Proof. By the above argument in (b), $f = 0$ a. e. implies $\int_X f \,\mathrm{d}\mu = \int_X 0 \,\mathrm{d}\mu = 0$ which proves one direction. The other direction is homework 40.4. ∎

12.4.2 The space $\mathrm{L}^p(X, \mu)$

For any measurable function $f\colon X \to \mathbb{C}$ and any real p, $1 \le p < \infty$ define

$$\|f\|_p = \left(\int_X |\, f \,|^p \, \mathrm{d}\mu \right)^{\frac{1}{p}}. \tag{12.7}$$

This number may be finite or ∞. In the first case, $|\, f \,|^p$ is integrable and we write $f \in \mathscr{L}^1(X, \mu)$.

Proposition 12.17 *Let $p, q \ge 1$ be given such that $\frac{1}{p} + \frac{1}{q} = 1$.*
(a) Let $f, g\colon X \to \mathbb{C}$ be measurable functions such that $f \in \mathscr{L}^p$ and $g \in \mathscr{L}^q$.
Then $fg \in \mathscr{L}^1$ and

$$\int_X |\, fg \,| \, \mathrm{d}\mu \le \|f\|_p \, \|g\|_q \quad \textit{(Hölder inequality)}. \tag{12.8}$$

(b) Let $f, g \in \mathscr{L}^q$. Then $f + g \in \mathscr{L}^q$ and

$$\|f + g\|_q \le \|f\|_q + \|g\|_q, \quad \textit{(Minkowski inequality)}. \tag{12.9}$$

Idea of proof. Hölder follows from Young's inequality (Proposition 1.31, as in the calssical case of Hölder's inequality in $\mathbb{R}^n$, see Proposition 1.32
Minkowski's inequality follows from Hölder's inequality as in Propostion 1.34
Note that Minkowski implies that $f, g \in \mathscr{L}^p$ yields $\|f + g\| < \infty$ such that $f + g \in \mathscr{L}^p$. In particular, $\mathscr{L}^p$ is a linaer space.
Let us check the properties of $\|\cdot\|_p$. For all measurable f, g we have

$$\begin{aligned} \|f\|_p &\ge 0, \\ \|\lambda f\|_p &= |\, \lambda \,| \, \|f\|_p, \\ \|f + g\| &\le \|f\| + \|g\|. \end{aligned}$$

All properties of a norm, see Definition 6.9 at page 179 are satisfied except for the definitness: $\|f\|_p = 0$ imlies $\int_X |\, f \,|^p \, \mathrm{d}\mu = 0$ implies by Proposition 12.16, $|\, f \,|^p = 0$ a. e. implies $f = 0$ a. e. . However, it does not imply $f = 0$. To overcome this problem, we use the equivalece relation $f = g$ a. e. and consider from now on only *equivalence classes* of functions in $\mathscr{L}^p$, that is we identify functions f and g which are equal a. e. .
The space $\mathcal{N} = \{ f\colon X \to \mathbb{C} \mid \ f \text{ is measurable and } \ f = 0 \text{ a. e. } \}$ is a linear subspace of $\mathscr{L}^p(X, \mu)$ for all all p, and $f = g$ a. e. if and only if $f - g \in \mathcal{N}$. Then the factor space $\mathscr{L}^p/\mathcal{N}$, see Example 12.3 (b) is again a linear space.

Definition 12.11 Let $(X, \mathcal{A}, \mu)$ be a measure space. $\mathrm{L}^p(X, \mu)$ denotes the set of equivalence classes of functions of $\mathscr{L}^p(X, \mu)$ with respect to the equivalence relation $f = g$ a. e. that is,

$$\mathrm{L}^p(X, \mu) = \mathscr{L}^p(X, \mu)/\mathcal{N}$$

is the quotient space. $(\mathrm{L}^p(X, \mu), \|\cdot\|_p)$ is a normed space. With this norm $\mathrm{L}^p(X, \mu)$ is complete.

Example 12.4 (a) We have $\chi_{\mathbb{Q}} = 0$ in $\mathrm{L}^p(\mathbb{R}, \lambda)$ since $\chi_{\mathbb{Q}} = 0$ a. e. on $\mathbb{R}$ with respect to the Lebesgue measure (note that $\mathbb{Q}$ is a set of measure zero).
(b) In case of sequence spaces $\mathscr{L}^p(\mathbb{N})$ with respect to the counting measure, $\mathscr{L}^p = \mathrm{L}^p$ since $f = 0$ a. e. implies $f = 0$.
(c) $f(x) = \frac{1}{x^\alpha}$, $\alpha > 0$ is in $\mathrm{L}^2(0,1)$ if and only if $2\alpha < 1$. We identify functions and their equivalence classes.

12.4.3 The Monotone Convergence Theorem

The following theorem about the monotone convergence by Beppo Levi (1875–1961) is one of the most important in the theory of integration. The theorem holds for an *arbitrary* increasing sequence of measurable functions with, possibly, $\int_X f_n \,\mathrm{d}\mu = +\infty$.

Theorem 12.18 (Monotone Convergence Theorem/Lebesgue) *Let (f_n) be a sequence of measurable functions on X and suppose that*

(1) $0 \le f_1(x) \le f_2(x) \le \cdots \le +\infty$ *for all* $x \in X$,
(2) $f_n(x) \underset{n\to\infty}{\longrightarrow} f(x)$, *for every* $x \in X$.

Then f is measurable, and

$$\lim_{n\to\infty} \int_X f_n \,\mathrm{d}\mu = \int_X f \,\mathrm{d}\mu = \int_X \left(\lim_{n\to\infty} f_n\right) \mathrm{d}\mu.$$

(Without proof) Note, that f is measurable is a consequence of Remark 12.5 (b).

Corollary 12.19 (Beppo Levi) *Let $f_n \colon X \to [0, +\infty]$ be measurable for all $n \in \mathbb{N}$, and $f(x) = \sum_{n=1}^{\infty} f_n(x)$ for $x \in X$. Then*

$$\int_X \sum_{n=1}^{\infty} f_n \,\mathrm{d}\mu = \sum_{n=1}^{\infty} \int_X f_n \,\mathrm{d}\mu.$$

Example 12.5 (a) Let $X = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ the σ-algebra of all subsets, and μ the counting measure on $\mathbb{N}$. The functions on $\mathbb{N}$ can be identified with the sequences (x_n), $f(n) = x_n$. Trivially, any function is $\mathcal{A}$-measurable.
What is $\int_{\mathbb{N}} x_n \,\mathrm{d}\mu$? First, let $f \ge 0$. For a simple function g_n, given by $g_n = x_n \chi_{\{n\}}$, we obtain $\int g_n \,\mathrm{d}\mu = x_n \mu(\{n\}) = x_n$. Note that $f = \sum_{n=1}^{\infty} g_n$ and $g_n \ge 0$ since $x_n \ge 0$. By Corollary 12.19,

$$\int_{\mathbb{N}} f \,\mathrm{d}\mu = \sum_{n=1}^{\infty} \int_{\mathbb{N}} g_n \,\mathrm{d}\mu = \sum_{n=1}^{\infty} x_n.$$

Now, let f be arbitrary integrable, i. e. $\int_{\mathbb{N}} |f| \,\mathrm{d}\mu < \infty$; thus $\sum_{n=1}^{\infty} |x_n| < \infty$. Therefore, $(x_n) \in \mathscr{L}^1(\mathbb{N}, \mu)$ if and only if $\sum x_n$ converges *absolutely*. The space of absolutely convergent series is denoted by ℓ_1 or $\ell_1(\mathbb{N})$.

(b) Let $a_{nm} \geq 0$ for all $n, m \in \mathbb{N}$. Then

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{mn} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{mn}.$$

Proof. Consider the measure space $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ from (a). For $n \in \mathbb{N}$ define functions $f_n(m) = a_{mn}$. By Corollary 12.19 we then have

$$\begin{aligned}\int_X \underbrace{\sum_{n=1}^{\infty} f_n(m)}_{f(m)} \,\mathrm{d}\mu &= \int_X f \,\mathrm{d}\mu \underset{\text{(a)}}{=} \sum_{m=1}^{\infty} f(m) = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{mn} \\ &= \sum_{n=1}^{\infty}\int_X f_n(m)\,\mathrm{d}\mu = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{mn}.\end{aligned}$$

■

Proposition 12.20 *Let $f\colon X \to [0, +\infty]$ be measurable. Then*

$$\varphi(A) = \int_A f \,\mathrm{d}\mu, \quad A \in \mathcal{A}$$

defines a measure φ on $\mathcal{A}$.

Proof. Since $f \geq 0$, $\varphi(A) \geq 0$ for all $A \in \mathcal{A}$. Let (A_n) be a countable disjoint family of measurable sets $A_n \in \mathcal{A}$ and let $A = \sum_{n=1}^{\infty} A_n$. By homework 40.1, $\chi_A = \sum_{n=1}^{\infty} \chi_{A_n}$ and therefore

$$\begin{aligned}\varphi(A) = \int_A f \,\mathrm{d}\mu = \int_X \chi_A f \,\mathrm{d}\mu = \int_X \sum_{n=1}^{\infty} \chi_{A_n} f \,\mathrm{d}\mu &\underset{\text{B.Levi}}{=} \sum_{n=1}^{\infty}\int_X \chi_{A_n} f \,\mathrm{d}\mu \\ = \sum_{n=1}^{\infty}\int_{A_n} f \,\mathrm{d}\mu = \sum_{n=1}^{\infty} \varphi(A_n).\end{aligned}$$

■

12.4.4 The Dominated Convergence Theorem

Besides the monotone convergence theorem the present theorem is the most important one. It is due to Henry Lebesgue. The great advantage, compared with Theorem6.6, is that $\mu(X) = \infty$ is allowed, that is, non-compact domains X are included. We only need the *pointwise* convergence of (f_n), not the *uniform* convergence. The main assumtion here is the existence of an integrable upper bound for all f_n.

Theorem 12.21 (Dominated Convergence Theorem of Lebesgue) *Let $f_n\colon X \to \overline{\mathbb{R}}$ or $g, f_n\colon X \to \mathbb{C}$ be measurable functions such that*

(1) $f_n(x) \to f(x)$ *as* $n \to \infty$ *a. e. on* X,

(2) $|\, f_n(x) \,| \le g(x)$ *a. e. on* X,

(3) $\displaystyle\int_X g \,\mathrm{d}\mu < +\infty$.

Then f *is measurable and integrable,* $\int_X |\, f \,| \,\mathrm{d}\mu < \infty$, *and*

$$\lim_{n\to\infty} \int_X f_n \,\mathrm{d}\mu = \int_X f \,\mathrm{d}\mu = \int_X \lim_{n\to\infty} f_n \,\mathrm{d}\mu,$$
$$\lim_{n\to\infty} \int_X |\, f_n - f \,| \,\mathrm{d}\mu = 0. \tag{12.10}$$

Note, that (12.10) shows that (f_n) converges to f in the normed space $\mathrm{L}^1(X,\mu)$.

Example 12.6 (a) Let $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, $A_1 \subset A_2 \subset \cdots$ be an increasing sequence with $\bigcup_{n\in\mathbb{N}} A_n = A$. If $f \in \mathscr{L}^1(A,\mu)$, then $f \in \mathscr{L}^1(A_n)$ for all n and

$$\lim_{n\to\infty} \int_{A_n} f \,\mathrm{d}\mu = \int_A f \,\mathrm{d}\mu. \tag{12.11}$$

Indeed, the sequence $(\chi_{A_n} f)$ is pointwise converging to $\chi_A f$ since $\chi_A(x) = 1$ iff $x \in A$ iff $x \in A_n$ for all $n \ge n_0$ iff $\lim_{n\to\infty} \chi_{A_n}(x) = 1$. Moreover, $|\, \chi_{A_n} f \,| \le |\, \chi_A f \,|$ which is integrable. By Lebesgue's theorem,

$$\lim_{n\to\infty} \int_{A_n} f \,\mathrm{d}\mu = \lim_{n\to\infty} \int_X \chi_{A_n} f = \int_X \chi_A f \,\mathrm{d}\mu = \int_A f \,\mathrm{d}\mu.$$

However, if we do not assume $f \in \mathscr{L}^1(A,\mu)$, the statement is not true (see Remark 12.7 below).

Exhausting theorem. Let (A_n) be an increasing sequencce of measurable sets and $A = \bigcup_{n=1}^{\infty} A_n$. suppose that f is measurable, and $\int_{A_n} f \,\mathrm{d}\mu$ is a bounded sequence. Then $f \in \mathscr{L}^1(A,\mu)$ and (12.11) holds.

(b) Let $f_n(x) = (-1)^n x^n$ on $[0,1]$. The sequence is dominated by the integrable function $1 \ge |\, f_n(x) \,|$ for all $x \in [0,1]$. Hence $\lim_{n\to\infty} \int_{[0,1]} f_n \mathrm{d}\lambda = 0 = \int_{[0,1]} \lim_{n\to\infty} f_n \mathrm{d}\lambda$.

12.4.5 Application of Lebesgue's Theorem to Parametric Integrals

As a direct application of the dominated convergence theorem we now treat parameter dependent integrals see Propostions 7.22 and 7.23

Proposition 12.22 (Continuity) *Let* $U \subset \mathbb{R}^n$ *be an open connected set,* $t_0 \in U$, *and* $f\colon \mathbb{R}^m \times U \to \mathbb{R}$ *be a function. Assume that*

(a) for a.e. $x \in \mathbb{R}^m$, *the function* $t \mapsto f(x,t)$ *is continuous at* t_0,

(b) There exists an integrable function $F\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ *such that for every* $t \in U$,

$$|\, f(x,t) \,| \le F(x) \quad \text{a.e. on } \mathbb{R}^m.$$

Then the function

$$g(t) = \int_{\mathbb{R}^m} f(x,t)\,\mathrm{d}x$$

is continuous at t_0.

Proof. First we note that for any fixed $t \in U$, the function $f_t(x) = f(x,t)$ is integrable on $\mathbb{R}^m$ since it is dominated by the integrable function F. We have to show that for any sequence $t_j \to t_0$, $t_j \in U$, $g(t_j)$ tends to $g(t_0)$ as $n \to \infty$. We set $f_j(x) = f(x,t_j)$ and $f_0(x) = f(x,t_0)$ for all $n \in \mathbb{N}$. By (b) we have

$$f_0(x) = \lim_{j\to\infty} f_j(x), \quad \text{a. e. } x \in \mathbb{R}^m.$$

By (a) and (c), the assumptions of the dominated convergence theorem are satisfied and thus

$$\lim_{j\to\infty} g(t_j) = \lim_{j\to\infty} \int_{\mathbb{R}^m} f_j(x)\,\mathrm{d}x = \int_{\mathbb{R}^m} \lim_{j\to\infty} f_j(x)\,\mathrm{d}x = \int_{\mathbb{R}^m} f_0(x)\,\mathrm{d}x = g(t_0).$$

■

Proposition 12.23 (Differentiation under the Integral Sign) *Let $I \subset \mathbb{R}$ be an open interval and $f\colon \mathbb{R}^m \times I \to \mathbb{R}$ be a function such that*

(a) for every $t \in I$ the function $x \mapsto f(x,t)$ is integrable,
(b) for almost all $x \in \mathbb{R}^m$, the function $t \mapsto f(x,t)$ is finite and continuously differentiable,
(c) There exists an integrable function $F\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ such that for every $t \in I$

$$\left| \frac{\partial f}{\partial t}(x,t) \right| \le F(x), \quad \textit{a.e. } x \in \mathbb{R}^m.$$

Then the function $g(t) = \int_{\mathbb{R}^m} f(x,t)\,\mathrm{d}x$ is differentiable on I with

$$g'(t) = \int_{\mathbb{R}^m} \frac{\partial f}{\partial t}(x,t)\,\mathrm{d}x.$$

The proof uses the previous theorem about the continuity of the parametric integral. A detailed proof is to be found in [Kön90, p. 283].

Example 12.7 (a) Let $f \in \mathscr{L}^1(\mathbb{R})$. Then the Fourier transform $\hat{f}\colon \mathbb{R} \to \mathbb{C}$, $\hat{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tx} f(x)\,\mathrm{d}x$ is continuous on $\mathbb{R}$, see homework 41.3.

(b) Let $K \subset \mathbb{R}^3$ be a compact subset and $\rho\colon K \to \mathbb{R}$ integrable, the Newton potential (with mass density ρ) is given by

$$u(t) = \int_K \frac{\rho(x)}{\|x-t\|}\,\mathrm{d}x, \quad t \notin K.$$

Then $u(t)$ is a harmonic function on $\mathbb{R}^3 \setminus K$.
Similarly, if $K \subset \mathbb{R}^2$ is compact and $\rho \in \mathscr{L}(K)$, the Newton potential is given by

$$u(t) = \int_K \rho(x)\,\log\|x-t\|\,\mathrm{d}x, \quad t \notin K.$$

Then $u(t)$ is a harmonic function on $\mathbb{R}^2 \setminus K$.

12.4.6 The Riemann and the Lebesgue Integrals

Proposition 12.24 *Let f be a bounded function on the finite interval $[a,b]$.*
(a) *f is Riemann integrable on $[a,b]$ if and only if f is continuous a. e. on $[a,b]$.*
(b) *If f is Riemann integrable on $[a,b]$, then f is Lebesgue integrable, too. Both integrals coincide.*
Let $I \subset \mathbb{R}$ be an interval such that f is Riemann integrable on all compact subintervals of I.
(c) *f is Lebesgue integrable on I if and only if $|f|$ is improperly Riemann integrable on I (see Section 5.4); both integrals coincide.*

Remarks 12.7 (a) The characteristic function $\chi_{\mathbb{Q}}$ on $[0,1]$ is Lebesgue but not Riemann integrable; $\chi_{\mathbb{Q}}$ is nowhere continuous on $[0,1]$.
(b) The (improper) Riemann integral

$$\int_1^\infty \frac{\sin x}{x}\,\mathrm{d}x$$

converges (see Example 5.11); however, the Lebesgue integral does not exist since the integral does not converge absolutely. Indeed, for non-negative integers $n \geq 1$ we have with some $c > 0$

$$\int_{n\pi}^{(n+1)\pi} \left|\frac{\sin x}{x}\right| \mathrm{d}x \geq \frac{1}{(n+1)\pi} \int_{n\pi}^{(n+1)\pi} |\sin x|\,\mathrm{d}x = \frac{c}{(n+1)\pi};$$

hence

$$\int_{\pi}^{(n+1)\pi} \left|\frac{\sin x}{x}\right| \mathrm{d}x \geq \frac{c}{\pi} \sum_{k=1}^{n} \frac{1}{k+1}.$$

Since the harmonic series diverges, so does the integral $\int_1^\infty \left|\frac{\sin x}{x}\right| \mathrm{d}x$.

12.4.7 Appendix: Fubini's Theorem

Theorem 12.25 *Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces, let f be an $\mathcal{A}_1 \otimes \mathcal{A}_2$-measurable function and $X = X_1 \times X_2$.*
(a) *If $f\colon X \to [0,+\infty]$, $\varphi(x_1) = \int\limits_{X_2} f(x_1,x_2)\,\mathrm{d}\mu_2$, $\psi(x_2) = \int\limits_{X_1} f(x_1,x_2)\,\mathrm{d}\mu_1$, then*

$$\int_{X_2} \psi(x_2)\,\mathrm{d}\mu_2 = \int\limits_{X_1\times X_2} f\,\mathrm{d}(\mu_1\otimes\mu_2) = \int_{X_1} \varphi(x_1)\,\mathrm{d}\mu_1.$$

(b) *If $f \in \mathscr{L}^1(X, \mu_1\otimes\mu_2)$ then*

$$\int_{X_1\times X_2} f\,\mathrm{d}(\mu_1\otimes\mu_2) = \int_{X_2} \left(\int_{X_1} f(x_1,x_2)\,\mathrm{d}\mu_1\right) \mathrm{d}\mu_2.$$

Here $\mathcal{A}_1 \otimes \mathcal{A}_2$ denotes the smallest σ-algebra over X, which contains all sets $A \times B$, $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$. Define $\mu(A\times B) = \mu_1(A)\mu_2(B)$ and extend μ to a measure $\mu_1 \otimes \mu_2$ on $\mathcal{A}_1 \otimes \mathcal{A}_2$.

Remark 12.8 In (a), as in Levi's theorem, we don't need any assumption on f to change the order of integration since $f \geq 0$. In (b) f is an arbitrary measurable function on $X_1 \times X_2$, however, the integral $\int_X |f|\,\mathrm{d}\mu$ needs to be finite.

Chapter 13

Hilbert Space

Functional analysis is a fruitful interplay between linear algebra and analysis. One defines function spaces with certain properties and certain topologies and considers linear operators between such spaces. The friendliest example of such spaces are Hilbert spaces.
This chapter is divided into two parts—one describes the geometry of a Hilbert space, the second is concerned with linear operators on the Hilbert space.

13.1 The Geometry of the Hilbert Space

13.1.1 Unitary Spaces

Let E be a linear space over $\mathbb{K} = \mathbb{R}$ or over $\mathbb{K} = \mathbb{C}$.

Definition 13.1 An *inner product* on E is a function $\langle \cdot\,,\,\cdot \rangle : E \times E \to \mathbb{K}$ with

(a) $\langle \lambda_1 x_1 + \lambda_2 x_2\,,\, y\rangle = \lambda_1 \langle x_1\,,\, y\rangle + \lambda_2 \langle x_2\,,\, y\rangle$ (Linearity)
(b) $\langle x\,,\, y\rangle = \overline{\langle y\,,\, x\rangle}$. (Hermitian property)
(c) $\langle x\,,\, x\rangle \geq 0$ for all $x \in E$, and $\langle x\,,\, x\rangle = 0$ implies $x = 0$ (Positive definiteness)

A *unitary space* is a linear space together with an inner product.

Let us list some immediate consequences from these axioms: From (a) and (b) it follows that

$$\text{(d)} \quad \langle y\,,\, \lambda_1 x_1 + \lambda_2 x_2\rangle = \overline{\lambda_1}\,\langle y\,,\, x_1\rangle + \overline{\lambda_2}\,\langle y\,,\, x_2\rangle\,.$$

A form on $E \times E$ satisfying (a) and (d) is called a *sesquilinear form*. (a) implies $\langle 0\,,\, y\rangle = 0$ for all $y \in E$. The mapping $x \mapsto \langle x\,,\, y\rangle$ is a linear mapping into $\mathbb{K}$ (a linear functional) for all $y \in E$.
By (c), we may define $\|x\|$, the *norm* of the vector $x \in E$ to be the square root of $\langle x\,,\, x\rangle$; thus

$$\|x\|^2 = \langle x\,,\, x\rangle\,. \tag{13.1}$$

Proposition 13.1 (Cauchy–Schwarz Inequality) *Let $(E, \langle \cdot\,,\,\cdot \rangle)$ be a unitary space. For $x, y \in E$ we have*

$$|\,\langle x\,,\,y\rangle\,| \le \|x\|\;\|y\|\,.$$

Equality holds if and only if $x = \beta y$ for some $\beta \in \mathbb{K}$.

Proof. Choose $\alpha \in \mathbb{C}$, $|\,\alpha\,| = 1$ such that $\alpha\,\langle y\,,\,x\rangle = |\,\langle x\,,\,y\rangle\,|$. For $\lambda \in \mathbb{R}$ we then have (since $\overline{\alpha}\,\langle x\,,\,y\rangle = \alpha\,\langle y\,,\,x\rangle = |\,\langle x\,,\,y\rangle\,|$)

$$\begin{aligned}\langle x - \alpha\lambda y\,,\,x - \alpha\lambda y\rangle &= \langle x\,,\,x\rangle - \alpha\lambda\,\langle y\,,\,x\rangle - \overline{\alpha}\lambda\,\langle x\,,\,y\rangle + \lambda^2\,\langle y\,,\,y\rangle \\ &= \langle x\,,\,x\rangle - 2\lambda\,|\,\langle x\,,\,y\rangle\,| + \lambda^2\,\langle y\,,\,y\rangle \ge 0.\end{aligned}$$

This is a quadratic polynomial $a\lambda^2 + b\lambda + c$ in λ with real coefficients. Since this polynomial takes only non-negative values, its discriminant $b^2 - 4ac$ must be non-positive:

$$4\,|\,\langle x\,,\,y\rangle\,|^2 - 4\,\|x\|^2\;\|y\|^2 \le 0.$$

This implies $|\,\langle x\,,\,y\rangle\,| \le \|x\|\;\|y\|$. ∎

Corollary 13.2 *$\|\cdot\|$ defines a norm on E.*

Proof. It is clear that $\|x\| \ge 0$. From (c) it follows that $\|x\| = 0$ implies $x = 0$. Further, $\|\lambda x\| = \sqrt{\langle \lambda x\,,\,\lambda x\rangle} = \sqrt{|\,\lambda\,|^2\,\langle x\,,\,x\rangle} = |\,\lambda\,|\,\|x\|$. We prove the triangle inequality. Since $2\,\mathrm{Re}\,(z) = z + \overline{z}$ we have by Proposition 1.20 and the Cauchy-Schwarz inequality

$$\begin{aligned}\|x+y\|^2 = \langle x+y\,,\,x+y\rangle &= \langle x\,,\,x\rangle + \langle x\,,\,y\rangle + \langle y\,,\,x\rangle + \langle y\,,\,y\rangle \\ &= \|x\|^2 + \|y\|^2 + 2\,\mathrm{Re}\,\langle x\,,\,y\rangle \\ &\le \|x\|^2 + \|y\|^2 + 2\,|\,\langle x\,,\,y\rangle\,| \\ &\le \|x\|^2 + \|y\|^2 + 2\,\|x\|\,\|y\| = (\|x\| + \|y\|)^2;\end{aligned}$$

hence $\|x+y\| \le \|x\| + \|y\|$. ∎

By the corollary, any unitary space is a normed space with the norm $\|x\| = \sqrt{\langle x\,,\,x\rangle}$.
Recall that any normed vector space is a metric space with the metric $d(x,y) = \|x-y\|$. Hence, the notions of open and closed sets, neighborhoods, converging sequences, Cauchy sequences, continuous mappings, and so on make sense in a unitary space. In particular $\lim\limits_{n\to\infty} x_n = x$ means that the sequence $(\|x_n - x\|)$ of non-negative real numbers tends to 0. Recall from Definition 6.8 that a metric space is said to be complete if every Cauchy sequence converges.

Definition 13.2 A complete unitary space is called a *Hilbert space*.

Example 13.1 Let $\mathbb{K} = \mathbb{C}$.

(a) $E = \mathbb{C}^n$, $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$, $y = (y_1, \ldots, y_n) \in \mathbb{C}^n$. Then $\langle x\,,\,y\rangle = \sum\limits_{k=1}^{n} x_k\overline{y_k}$ defines an inner product, with the euclidean norm $\|x\| = (\sum_{k=1}^{n} |\,x_k\,|)^{\frac{1}{2}}$. $(\mathbb{C}^n, \langle \cdot\,,\,\cdot\rangle)$ is a Hilbert space.

(b) $E = \mathrm{L}^2(X,\mu)$ is a Hilbert space with the inner product $\langle f\,,\,g\rangle = \int_X f\overline{g}\,\mathrm{d}\mu$.
By Proposition 12.17 with $p = q = 2$ we obtain the Cauchy-Schwarz inequality

$$\left|\int_X f\overline{g}\,\mathrm{d}\mu\right| \le \left(\int_X |\,f\,|^2\,\mathrm{d}\mu\right)^{\frac{1}{2}}\left(\int_X |\,g\,|^2\,\mathrm{d}\mu\right)^{\frac{1}{2}}.$$

Using CSI one can prove Minkowski's inequality, that is, $f,g \in \mathrm{L}^2(X,\mu)$ implies $f+g \in \mathrm{L}^2(X,\mu)$. Also, $\langle f\,,\,g\rangle$ is a finite complex number, since $f\overline{g} \in \mathrm{L}^1(X,\mu)$.
Note that the inner product is positive definite since $\int_X |\,f\,|^2\,\mathrm{d}\mu = 0$ implies (by Proposition 12.16) $|\,f\,| = 0$ a. e. and therefore, $f = 0$ in $\mathrm{L}^2(X,\mu)$. To prove the *completeness of* $\mathrm{L}^2(X,\mu)$ is more complicated, we skip the proof.
(c) $E = \ell_2$, i. e.

$$\ell_2 = \{(x_n) \mid x_n \in \mathbb{C}, n \in \mathbb{N},\ \sum_{n=1}^{\infty} |\,x_n\,|^2 < \infty\}.$$

Note that Cauchy–Schwarz's inequality in $\mathbb{R}^k$ (Corollary 1.26) implies

$$\left|\sum_{n=1}^{k} x_n\overline{y_n}\right|^2 \le \sum_{n=1}^{k} |\,x_n\,|^2 \sum_{n=1}^{k} |\,y_n\,|^2 \le \sum_{n=1}^{\infty} |\,x_n\,|^2 \sum_{n=1}^{\infty} |\,y_n\,|^2.$$

Taking the supremum over all $k \in \mathbb{N}$ on the left, we have

$$\left|\sum_{n=1}^{\infty} x_n\overline{y_n}\right|^2 \le \sum_{n=1}^{\infty} |\,x_n\,|^2 \sum_{n=1}^{\infty} |\,y_n\,|^2;$$

hence

$$\langle (x_n)\,,\,(y_n)\rangle = \sum_{n=1}^{\infty} x_n\,\overline{y_n}$$

is an absolutely converging series such that the inner product is well-defined on ℓ_2.

Lemma 13.3 *Let E be a unitary space. For any fixed $y \in E$ the mappings $f,g\colon E \to \mathbb{C}$ given by*

$$f(x) = \langle x\,,\,y\rangle\,, \quad \textit{and} \quad g(x) = \langle y\,,\,x\rangle$$

are continuous functions on E.

Proof. First proof. Let (x_n) be a sequence in E, converging to $x \in E$, that is, $\lim_{n\to\infty} \|x_n - x\| = 0$. Then

$$|\,\langle x_n\,,\,y\rangle - \langle x\,,\,y\rangle\,| = |\,\langle x_n - x\,,\,y\rangle\,| \underset{\text{CSI}}{\le} \|x_n - x\|\,\|y\| \implies 0$$

as $n \to \infty$. This proves continuity of f. The same proof works for g.
Second proof. The Cauchy–Schwarz inequality implies that for $x_1, x_2 \in E$

$$|\,\langle x_1\,,\,y\rangle - \langle x_2\,,\,y\rangle\,| = |\,\langle x_1 - x_2\,,\,y\rangle\,| \le \|x_1 - x_2\|\,\|y\|\,,$$

which proves that the map $x \mapsto \langle x\,,\,y\rangle$ is in fact uniformly continuous (Given $\varepsilon > 0$ choose $\delta = \varepsilon/\|y\|$. Then $\|x_1 - x_2\| < \delta$ implies $|\,\langle x_1\,,\,y\rangle - \langle x_2\,,\,y\rangle\,| < \varepsilon$). The same is true for $x \mapsto \langle y\,,\,x\rangle$. ∎

Definition 13.3 Let H be a unitary space. We call x and y *orthogonal* to each other, and write $x \perp y$, if $\langle x\,,\,y\rangle = 0$. Two subsets $M, N \subset H$ are called *orthogonal* to each other if $x \perp y$ for all $x \in M$ and $y \in N$.
For a subset $M \subset H$ define the *orthogonal complement* $M^\perp$ of M to be the set

$$M^\perp = \{x \in H \mid \langle x\,,\,m\rangle = 0, \quad \text{for all } m \in M\}.$$

For example, $E = \mathbb{R}^n$ with the standard inner product and $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, $v \neq 0$ yields

$$\{v\}^\perp = \{x \in \mathbb{R}^n \mid \sum_{k=1}^{n} x_k\, v_k = 0\}.$$

This is a hyperplane in $\mathbb{R}^n$ which is orthogonal to v.

Lemma 13.4 *Let H be a unitary space and $M \subset H$ be an arbitrary subset. Then, $M^\perp$ is a closed linear subspace of H.*

Proof. (a) Suppose that $x, y \in M^\perp$. Then for $m \in M$ we have

$$\langle \lambda_1 x + \lambda_2 y\,,\,m\rangle = \lambda_1\,\langle x\,,\,m\rangle + \lambda_2\,\langle y\,,\,m\rangle = 0;$$

hence $\lambda_1 x + \lambda_2 y \in M^\perp$. This shows that $M^\perp$ is a linear subspace.
(b) We show that any converging sequence (x_n) of elements of $M^\perp$ has its limit in $M^\perp$. Suppose $\lim\limits_{n\to\infty} x_n = x$, $x_n \in M^\perp$, $x \in H$. Then for all $m \in M$, $\langle x_n\,,\,m\rangle = 0$. Since the inner product is continuous in the first argument (see Lemma 13.3) we obtain

$$0 = \lim_{n\to\infty} \langle x_n\,,\,m\rangle = \langle x\,,\,m\rangle\,.$$

This shows $x \in M^\perp$; hence $M^\perp$ is closed. ∎

13.1.2 Norm and Inner product

Problem. Given a normed linear space $(E, \|\cdot\|)$. Does there exist an inner product $\langle\cdot\,,\,\cdot\rangle$ on E such that $\|x\| = \sqrt{\langle x\,,\,x\rangle}$ for all $x \in E$? In this case we call $\|\cdot\|$ an *inner product norm*.

Proposition 13.5 *(a) A norm $\|\cdot\|$ on a linear space E over $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$ is an inner product norm if and only if the* parallelogram law

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2), \quad x, y \in E \tag{13.2}$$

is satisfied.
(b) If (13.2) *is satisfied, the inner product $\langle\cdot\,,\,\cdot\rangle$ is given by* (13.3) *in the real case $\mathbb{K} = \mathbb{R}$ and by* (13.4) *in the complex case $\mathbb{K} = \mathbb{C}$.*

$$\langle x\,,\,y\rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right), \quad \textit{if} \quad \mathbb{K} = \mathbb{R}. \tag{13.3}$$

$$\langle x\,,\,y\rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 + \mathrm{i}\,\|x+\mathrm{i}y\|^2 - \mathrm{i}\,\|x-\mathrm{i}y\|^2\right), \quad \textit{if} \quad \mathbb{K} = \mathbb{C}. \tag{13.4}$$

These equations are called polarization identities.

Proof. We check the parallelogram and the polarization identity in the real case, $\mathbb{K} = \mathbb{R}$.

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= \langle x+y\,,\,x+y\rangle + \langle x-y\,,\,x-y\rangle \\ &= \langle x\,,\,x\rangle + \langle y\,,\,x\rangle + \langle x\,,\,y\rangle + \langle y\,,\,y\rangle + (\langle x\,,\,x\rangle - \langle y\,,\,x\rangle - \langle x\,,\,y\rangle + \langle y\,,\,y\rangle) \\ &= 2\,\|x\|^2 + 2\,\|y\|^2\,.\end{aligned}$$

Further,

$$\begin{aligned}\|x+y\|^2 - \|x-y\|^2 = (\langle x\,,\,x\rangle + \langle y\,,\,x\rangle + \langle x\,,\,y\rangle + \langle y\,,\,y\rangle) - \\ - (\langle x\,,\,x\rangle - \langle y\,,\,x\rangle - \langle x\,,\,y\rangle + \langle y\,,\,y\rangle) = 4\,\langle x\,,\,y\rangle\,.\end{aligned}$$

■

The proof that the parallelogram identity is sufficient for E being a unitary space is in the appendix to this section.

Example 13.2 We show that $\mathrm{L}^1([0,2])$ with $\|f\|_1 = \int_0^2 |\,f\,|\,\mathrm{d}x$ is not an inner product norm. Indeed, let $f = \chi_{[1,2]}$ and $g = \chi_{[0,1]}$. Then $f+g = |\,f-g\,| = 1$ and $\|f\|_1 = \|g\|_1 = \int_0^1 \mathrm{d}x = 1$ such that

$$\|f+g\|_1^2 + \|f-g\|_1^2 = 2^2 + 2^2 = 8 \neq 4 = 2(\|f\|_1^2 + \|g\|_1^2).$$

The parallelogram identity is not satisfied for $\|\cdot\|_1$ such that $\mathrm{L}^1([0,2])$ is not an inner product space.

13.1.3 Two Theorems of F. Riesz

(born: January 22, 1880 in Austria-Hungary, died: February 28, 1956, founder of functional analysis)

Definition 13.4 Let $(H_1, \langle\cdot\,,\,\cdot\rangle_1)$ and $(H_2, \langle\cdot\,,\,\cdot\rangle_2)$ be Hilbert spaces. Let $H = \{(x_1, x_2) \mid x_1 \in H_1,\, x_2 \in H_2\}$ be the direct sum of the Hilbert spaces H_1 and H_2. Then

$$\langle (x_1,x_2)\,,\,(y_1,y_2)\rangle = \langle x_1\,,\,y_1\rangle_1 + \langle x_2\,,\,y_2\rangle_2$$

defines an inner product on H. With this inner product H becomes a Hilbert space. $H = H_1 \oplus H_2$ is called the (direct) *orthogonal sum* of H_1 and H_2.

Definition 13.5 Two Hilbert spaces H_1 and H_2 are called *isomorphic* if there exists a bijective linear mapping $\varphi\colon H_1 \to H_2$ such that

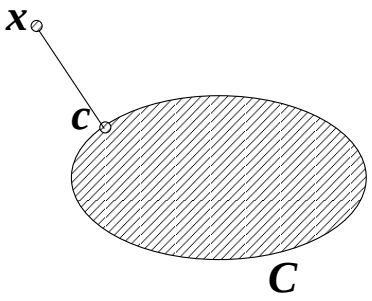
$$\langle \varphi(x)\,,\,\varphi(y)\rangle_2 = \langle x\,,\,y\rangle_1\,, \quad x,y \in H_1.$$

φ is called *isometric isomorphism* or a *unitary map*.

Back to the orthogonal sum $H = H_1 \oplus H_2$. Let $\tilde{H}_1 = \{(x_1,0) \mid x_1 \in H_1\}$ and $\tilde{H}_2 = \{(0,x_2) \mid x_2 \in H_2\}$. Then $x_1 \mapsto (x_1,0)$ and $x_2 \mapsto (0,x_2)$ are isometric isomorphisms from $H_i \to \tilde{H}_i$, $i = 1,2$. We have $\tilde{H}_1 \perp \tilde{H}_2$ and $\tilde{H}_i$, $i = 1,2$ are closed linear subspaces of H.
In this situation we say that H is the *inner* orthogonal sum of the two closed subspaces $\tilde{H}_1$ and $\tilde{H}_2$.

(a) Riesz's First Theorem

Problem. Let H_1 be a closed linear subspace of H. Does there exist another closed linear subspace H_2 such that $H = H_1 \oplus H_2$?
Answer: YES.



Lemma 13.6 (Minimal Distance Lemma) *Let C be a convex and closed subset of the Hilbert space H. For $x \in H$ let*

$$\varrho(x) = \inf\{\|x - y\| \mid y \in C\}.$$

Then there exists a unique element $c \in C$ such that

$$\varrho(x) = \|x - c\|.$$

Proof. Existence. Since $\varrho(x)$ is an infimum, there exists a sequence (y_n), $y_n \in C$, which approximates the infimum, $\lim_{n\to\infty} \|x - y_n\| = \varrho(x)$. We will show, that (y_n) is a Cauchy sequence. By the parallelogram law (see Proposition 13.5) we have

$$\begin{aligned}\|y_n - y_m\|^2 &= \|y_n - x + x - y_m\|^2 \\ &= 2\|y_n - x\|^2 + 2\|x - y_m\|^2 - \|2x - y_n - y_m\|^2 \\ &= 2\|y_n - x\|^2 + 2\|x - y_m\|^2 - 4\left\|x - \frac{y_n + y_m}{2}\right\|^2.\end{aligned}$$

Since C is convex, $(y_n + y_m)/2 \in C$ and therefore $\left\|x - \frac{y_n+y_m}{2}\right\| \geq \varrho(x)$. Hence

$$\leq 2\|y_n - x\|^2 + 2\|x - y_m\|^2 - 4\varrho(x)^2.$$

By the choice of (y_n), the first two sequences tend to $\varrho(x)^2$ as $m, n \to \infty$. Thus,

$$\lim_{m,n\to\infty} \|y_n - y_m\|^2 = 2(\varrho^2(x) + \varrho(x)^2) - 4\varrho(x)^2 = 0,$$

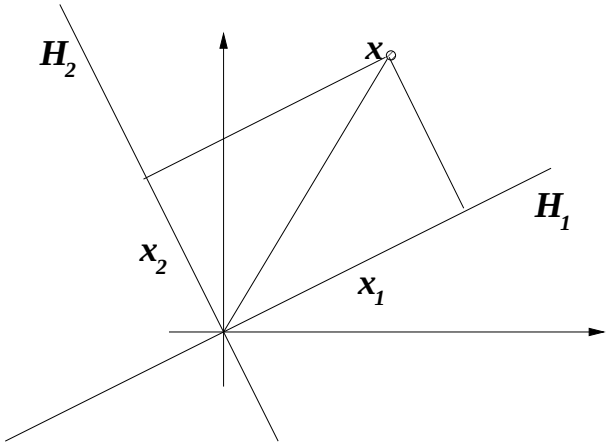
hence (y_n) is a Cauchy sequence. Since H is complete, there exists an element $c \in H$ such that $\lim_{n\to\infty} y_n = c$. Since $y_n \in C$ and C is closed, $c \in C$. By construction, we have $\|y_n - x\| \longrightarrow \varrho(x)$. On the other hand, since $y_n \longrightarrow c$ and the norm is continuous (see homework 42.1. (b)), we have

$$\|y_n - x\| \longrightarrow \|c - x\|.$$

This implies $\varrho(x) = \|c - x\|$.
Uniqueness. Let c, c' two such elements. Then, by the parallelogram law,

$$\begin{aligned}0 \leq \|c - c'\|^2 &= \|c - x + x - c'\|^2 \\ &= 2\|c - x\|^2 + 2\|x - c'\|^2 - 4\left\|x - \frac{c + c'}{2}\right\|^2 \\ &\leq 2(\varrho(x)^2 + \varrho(x)^2) - 4\varrho(x)^2 = 0.\end{aligned}$$

This implies $c = c'$; the point $c \in C$ which realizes the infimum is unique. ∎



Theorem 13.7 (Riesz's first theorem) *Let H_1 be a closed linear subspace of the Hilbert space H. Then we have*

$$H = H_1 \oplus H_1^{\perp},$$

that is, any $x \in H$ has a unique representation $x = x_1 + x_2$ with $x_1 \in H_1$ and $x_2 \in H_1^{\perp}$.

Proof. Existence. Apply Lemma 13.6 to the convex, closed set H_1. There exists a unique $x_1 \in H_1$ such that

$$\varrho(x) = \inf\{\|x - y\| \mid y \in H_1\} = \|x - x_1\| \leq \|x - x_1 - ty_1\|$$

for all $t \in \mathbb{K}$ and $y_1 \in H_1$. homework 42.2 (c) now implies $x_2 = x - x_1 \perp y_1$ for all $y_1 \in H_1$. Hence $x_2 \in H_1^{\perp}$. Therefore, $x = x_1 + x_2$, and the existence of such a representation is shown. *Uniqueness.* Suppose that $x = x_1 + x_2 = x_1' + x_2'$ are two possibilities to write x as a sum of elements of $x_1, x_1' \in H_1$ and $x_2, x_2' \in H_1^{\perp}$. Then

$$x_1 - x_1' = x_2' - x_2 = u$$

belongs to both H_1 and $H_1^{\perp}$ (by linearity of H_1 and H_2). Hence $\langle u\,,\, u\rangle = 0$ which implies $u = 0$. That is, $x_1 = x_1'$ and $x_2 = x_2'$. ∎

Let $x = x_1 + x_2$ be as above. Then the mappings $P_1(x) = x_1$ and $P_2(x) = x_2$ are well-defined on H. They are called *orthogonal projections* of H onto H_1 and H_2, respectively. We will consider projections in more detail later.

Example 13.3 Let H be a Hilbert space, $z \in H$, $z \neq 0$, $H_1 = \mathbb{K}\, z$ the one-dimensional linear subspace spanned by one single vector z. Since any finite dimensional subspace is closed, Riesz's first theorem applies. We want to compute the projections of $x \in H$ with respect to H_1 and $H_1^{\perp}$. Let $x_1 = \alpha z$; we have to determine α such that $\langle x - x_1\,,\, z\rangle = 0$, that is

$$\langle x - \alpha z\,,\, z\rangle = \langle x\,,\, z\rangle - \langle \alpha z\,,\, z\rangle = \langle x\,,\, z\rangle - \alpha\,\langle z\,,\, z\rangle = 0.$$

Hence,

$$\alpha = \frac{\langle x\,,\, z\rangle}{\langle z\,,\, z\rangle} = \frac{\langle x\,,\, z\rangle}{\|z\|^2}.$$

The Riesz's representation with respect to $H_1 = \mathbb{K}z$ and $H_1^{\perp}$ is

$$x = \frac{\langle x\,,\, z\rangle}{\|z\|^2}\, z + \left(x - \frac{\langle x\,,\, z\rangle}{\|z\|^2}\, z\right).$$

(b) Riesz's Representation Theorem

Recall from Section 11 that a *linear functional* on the vector space E is a mapping $F\colon E \to \mathbb{K}$ such that $F(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 F(x_1) + \lambda_2 F(x_2)$ for all $x_1, x_2 \in E$ and $\lambda_1, \lambda_2 \in \mathbb{K}$.
Let $(E, \|\cdot\|)$ be a normed linear space over $\mathbb{K}$. Recall that a linear functional $F\colon E \to \mathbb{K}$ is called *continuous* if $x_n \longrightarrow x$ in E implies $F(x_n) \longrightarrow F(x)$.
The set of all continuous linear functionals F on E form a linear space E' with the same linear operations as in E^*.
Now let $(H, \langle\cdot\,,\,\cdot\rangle)$ be a Hilbert space. By Lemma 13.3, $F_y\colon H \to \mathbb{K}$, $F_y(x) = \langle x\,,\,y\rangle$ defines a continuous linear functional on H. Riesz's representation theorem states that *any* continuous linear functional on H is of this form.

Theorem 13.8 (Riesz's Representation Theorem) *Let F be a continuous linear functional on the Hilbert space H.*
Then there exists a unique element $y \in H$ such that $F(x) = F_y(x) = \langle x\,,\,y\rangle$ for all $x \in H$.

Proof. Existence. Let $H_1 = \ker F$ be the null-space of the linear functional F. H_1 is a linear subspace (since F is linear). H_1 is closed since $H_1 = F^{-1}(\{0\})$ is the preimage of the closed set $\{0\}$ under the continuous map F. By Riesz's first theorem, $H = H_1 \oplus H_1^\perp$.
Case 1. $H_1^\perp = \{0\}$. Then $H = H_1$ and $F(x) = 0$ for all x. We can choose $y = 0$; $F(x) = \langle x\,,\,0\rangle$.
Case 2. $H_1^\perp \neq \{0\}$. Suppose $u \in H_1^\perp$, $u \neq 0$. Then $F(u) \neq 0$ (otherwise, $u \in H_1^\perp \cap H_1$ such that $\langle u\,,\,u\rangle = 0$ which implies $u = 0$). We have

$$F\left(x - \frac{F(x)}{F(u)}\,u\right) = F(x) - \frac{F(x)}{F(u)}\,F(u) = 0.$$

Hence $x - \dfrac{F(x)}{F(u)}\,u \in H_1$. Since $u \in H_1^\perp$ we have

$$0 = \left\langle x - \frac{F(x)}{F(u)}\,u\,,\,u\right\rangle = \langle x\,,\,u\rangle - \frac{F(x)}{F(u)}\,\langle u\,,\,u\rangle$$

$$F(x) = \frac{F(u)}{\langle u\,,\,u\rangle}\,\langle x\,,\,u\rangle = \left\langle x\,,\,\frac{\overline{F(u)}}{\|u\|^2}\,u\right\rangle = F_y(x),$$

where $y = \dfrac{\overline{F(u)}}{\|u\|^2}\,u$.
Uniqueness. Suppose that both $y_1, y_2 \in H$ give the same functional F, i. e. $F(x) = \langle x\,,\,y_1\rangle = \langle x\,,\,y_2\rangle$ for all x. This implies

$$\langle y_1 - y_2\,,\,x\rangle = 0, \quad x \in H.$$

In particular, choose $x = y_1 - y_2$. This gives $\|y_1 - y_2\|^2 = 0$; hence $y_1 = y_2$. ∎

(c) Example

Any continuous linear functionals on $\mathrm{L}^2(X,\mu)$ are of the form $F(f) = \int_X f\,\overline{g}\,\mathrm{d}\mu$ with some $g \in \mathrm{L}^2(X,\mu)$. Any continuous linear functional on ℓ_2 is given by

$$F((x_n)) = \sum_{n=1}^{\infty} x_n \overline{y_n}, \quad \text{with} \quad (y_n) \in \ell_2.$$

13.1.4 Orthogonal Sets and Fourier Expansion

Motivation. Let $E = \mathbb{R}^n$ be the euclidean space with the standard inner product and standard basis $\{\mathrm{e}_1, \ldots, \mathrm{e}_n\}$. Then we have with $x_i = \langle x\,,\,\mathrm{e}_i\rangle$

$$x = \sum_{k=1}^{n} x_k\,\mathrm{e}_k, \quad \|x\|^2 = \sum_{k=1}^{n} |\,x_k\,|^2, \quad \langle x\,,\,y\rangle = \sum_{k=1}^{n} x_k\overline{y_k}.$$

We want to generalize these formulas to arbitrary Hilbert spaces.

(a) Orthonormal Sets

Let $(H, \langle\cdot\,,\,\cdot\rangle)$ be a Hilbert space.

Definition 13.6 Let $\{x_i \mid i \in I\}$ be a family of elements of H.
$\{x_i\}$ is called an *orthogonal set* or *OS* if $\langle x_i\,,\,x_j\rangle = 0$ for $i \neq j$.
$\{x_i\}$ is called an *orthonormal set* or *NOS* if $\langle x_i\,,\,x_j\rangle = \delta_{ij}$ for all $i, j \in I$.

Example 13.4 (a) $H = \ell_2$, $\mathrm{e}_n = (0, 0, \ldots, 0, 1, 0, \ldots)$ with the 1 at the nth component. Then $\{\mathrm{e}_n \mid n \in \mathbb{N}\}$ is an OS in H.
(b) $H = \mathrm{L}^2((0, 2\pi))$ with the Lebesgue measure, $\langle f\,,\,g\rangle = \int_0^{2\pi} f\overline{g}\,\mathrm{d}\lambda$. Then

$$\{1, \sin(nx), \cos(nx) \mid n \in \mathbb{N}\}$$

is an OS in H.

$$\left\{\frac{1}{\sqrt{2\pi}}, \frac{\sin(nx)}{\sqrt{\pi}}, \frac{\cos(nx)}{\sqrt{\pi}} \mid n \in \mathbb{N}\right\}, \quad \left\{\frac{\mathrm{e}^{\mathrm{i}nx}}{\sqrt{2\pi}} \mid n \in \mathbb{N}\right\}$$

to be orthonormal sets of H.

Lemma 13.9 (The Pythagorean Theorem) *Let $\{x_1, \ldots, x_k\}$ be an OS in H, then*

$$\|x_1 + \cdots + x_k\|^2 = \|x_1\|^2 + \cdots + \|x_k\|^2.$$

The easy proof is left to the reader.

Lemma 13.10 *Let $\{x_n\}$ be an OS in H. Then $\sum_{k=1}^{\infty} x_k$ converges if and only if $\sum_{k=1}^{\infty} \|x_k\|^2$ converges.*

The proof is in the appendix.

(b) Fourier Expansion and Completeness

Throughout this paragraph let $\{x_n \mid n \in \mathbb{N}\}$ an NOS in the Hilbert space H.

Definition 13.7 The numbers $\langle x\,,\,x_n\rangle$, $n \in \mathbb{N}$, are called *Fourier coefficients* of $x \in H$ with respect to the NOS $\{x_n\}$.

Example 13.5 Consider the NOS $\left\{\frac{1}{\sqrt{2\pi}},\,\frac{\sin(nx)}{\sqrt{\pi}},\,\frac{\cos(nx)}{\sqrt{\pi}} \mid n \in \mathbb{N}\right\}$ from the previous example on $H = \mathrm{L}^2((0,2\pi))$. Let $f \in H$. Then

$$\left\langle f\,,\,\frac{\sin(nx)}{\sqrt{\pi}}\right\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(t)\sin(nt)\,\mathrm{d}t,$$
$$\left\langle f\,,\,\frac{\cos(nx)}{\sqrt{\pi}}\right\rangle = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(t)\cos(nt)\,\mathrm{d}t,$$
$$\left\langle f\,,\,\frac{1}{\sqrt{2\pi}}\right\rangle = \frac{1}{\sqrt{2\pi}}\int_0^{2\pi} f(t)\,\mathrm{d}t,$$

These are the usual Fourier coefficients—up to a factor. Note that we have another normalization than in Definition 6.3 since the inner product there has the factor $1/(2\pi)$.

Proposition 13.11 (Bessel's Inequality) *For $x \in H$ we have*

$$\sum_{k=1}^{\infty} |\,\langle x\,,\,x_k\rangle\,|^2 \le \|x\|^2\,. \tag{13.5}$$

Proof. Let $n \in \mathbb{N}$ be a positive integer and $y_n = x - \sum_{k=1}^{n}\langle x\,,\,x_k\rangle\;x_k$. Then

$$\langle y_n\,,\,x_m\rangle = \langle x\,,\,x_m\rangle - \sum_{k=1}^{n}\langle x\,,\,x_k\rangle\;\langle x_k\,,\,x_m\rangle = \langle x\,,\,x_m\rangle - \sum_{k=1}^{n}\langle x\,,\,x_k\rangle\;\delta_{km} = 0$$

for $m = 1,\ldots,n$. Hence, $\{y_n, \langle x\,,\,x_1\rangle\,x_1,\ldots,\langle x\,,\,x_n\rangle\,x_n\}$ is an OS. By Lemma 13.9

$$\|x\|^2 = \left\|y_n + \sum_{k=1}^{n}\langle x\,,\,x_k\rangle\;x_k\right\|^2 = \|y_n\|^2 + \sum_{k=1}^{n}|\,\langle x\,,\,x_k\rangle\,|^2\,\|x_k\|^2 \ge \sum_{k=1}^{n}|\,\langle x\,,\,x_k\rangle\,|^2\,,$$

since $\|x_k\|^2 = 1$ for all k. Taking the supremum over all n on the right, the assertion follows. ∎

Corollary 13.12 *For any $x \in H$ the series $\displaystyle\sum_{k=1}^{\infty}\langle x\,,\,x_k\rangle\;x_k$ converges in H.*

Proof. Since $\{\langle x\,,\,x_k\rangle\,x_k\}$ is an OS, by Lemma 13.10 the series converges if and only if the series $\sum_{k=1}^{\infty}\|\langle x\,,\,x_k\rangle\;x_k\|^2 = \sum_{k=1}^{\infty}|\,\langle x\,,\,x_k\rangle\,|^2$ converges. By Bessel's inequality, this series converges. ∎

We call $\sum_{k=1}^{\infty}\langle x\,,\,x_k\rangle\;x_k$ the *Fourier series* of x with respect to the NOS $\{x_k\}$.

Remarks 13.1 (a) In general, the Fourier series of x does *not* converge to x.
(b) The NOS $\{\frac{1}{\sqrt{2\pi}}, \frac{\sin(nx)}{\sqrt{\pi}}, \frac{\cos(nx)}{\sqrt{\pi}}\}$ gives the ordinary Fourier series of a function f which is integrable over $(0, 2\pi)$.

Theorem 13.13 *Let $\{x_k \mid k \in \mathbb{N}\}$ be an NOS in H. The following are equivalent:*

(a) $x = \sum_{k=1}^{\infty} \langle x\,,\, x_k\rangle\; x_k$ *for all $x \in H$, i.e. the Fourier series of x converges to x.*
(b) $\langle z\,,\, x_k\rangle = 0$ *for all $k \in \mathbb{N}$ implies $z = 0$, i. e. the NOS is* maximal.
(c) *For every $x \in H$ we have* $\|x\|^2 = \sum_{k=1}^{\infty} |\,\langle x\,,\, x_k\rangle\,|^2$.
(d) *If $x \in H$ and $y \in H$, then* $\langle x\,,\, y\rangle = \sum_{k=1}^{\infty} \langle x\,,\, x_k\rangle\; \langle x_k\,,\, y\rangle$.

Formula (c) *is called* Parseval's identity.

Definition 13.8 An orthonormal set $\{x_i \mid i \in \mathbb{N}\}$ which satisfies the above (equivalent) properties is called a *complete orthonormal system*, CNOS for short.

Proof. (a) $\to$ (d): Since the inner product is continuous in both components we have

$$\langle x\,,\, y\rangle = \left\langle \sum_{k=1}^{\infty} \langle x\,,\, x_k\rangle\; x_k\,,\, \sum_{n=1}^{\infty} \langle y\,,\, x_n\rangle\; x_n \right\rangle = \sum_{k,n=1}^{\infty} \langle x\,,\, x_k\rangle\, \overline{\langle y\,,\, x_n\rangle}\, \underbrace{\langle x_k\,,\, x_n\rangle}_{\delta_{kn}}$$
$$= \sum_{k=1}^{\infty} \langle x\,,\, x_k\rangle\, \langle x_k\,,\, y\rangle\,.$$

(d) $\to$ (c): Put $y = x$.
(c) $\to$ (b): Suppose $\langle z\,,\, x_k\rangle = 0$ for all k. By (c) we then have

$$\|z\|^2 = \sum_{k=1}^{\infty} |\,\langle z\,,\, x_k\rangle\,|^2 = 0; \quad \text{hence} \quad z = 0.$$

(b) $\to$ (a): Fix $x \in H$ and put $y = \sum_{k=1}^{\infty} \langle x\,,\, x_k\rangle\; x_k$ which converges according to Corollary 13.12. With $z = x - y$ we have for all positive integers $n \in \mathbb{N}$

$$\langle z\,,\, x_n\rangle = \langle x - y\,,\, x_n\rangle = \left\langle x - \sum_{k=1}^{\infty} \langle x\,,\, x_k\rangle\; x_k\,,\, x_n \right\rangle$$
$$\langle z\,,\, x_n\rangle = \langle x\,,\, x_n\rangle - \sum_{k=1}^{\infty} \langle x\,,\, x_k\rangle\; \langle x_k\,,\, x_n\rangle = \langle x\,,\, x_n\rangle - \langle x\,,\, x_n\rangle = 0.$$

This shows $z = 0$ and therefore $x = y$, i. e. the Fourier series of x converges to x. ∎

Example 13.6 (a) $H = \ell_2$, $\{\mathrm{e}_n \mid n \in \mathbb{N}\}$ is an NOS. We show that this NOS is complete. For, let $x = (x_n)$ be orthogonal to every e_n, $n \in \mathbb{N}$; that is, $0 = \langle x\,,\,\mathrm{e}_n\rangle = x_n$. Hence, $x = (0, 0, \dots) = 0$. By (b), $\{\mathrm{e}_n\}$ is a CNOS. How does the Fourier series of x look like? The Fourier coefficients of x are $\langle x\,,\,\mathrm{e}_n\rangle = x_n$ such that

$$x = \sum_{n=1}^{\infty} x_n\,\mathrm{e}_n$$

is the Fourier series of x . The NOS $\{\mathrm{e}_n \mid n \geq 2\}$ is not complete.
(b) $H = \mathrm{L}^2\,((0, 2\pi))$,

$$\left\{\frac{1}{\sqrt{2\pi}},\ \frac{\sin(nx)}{\sqrt{\pi}},\ \frac{\cos(nx)}{\sqrt{\pi}} \mid n \in \mathbb{N}\right\}, \quad \text{and} \quad \left\{\frac{\mathrm{e}^{\mathrm{i}nx}}{\sqrt{2\pi}} \mid n \in \mathbb{Z}\right\}$$

are both CNOSs in H. This was stated in Theorem6.14

(c) Existence of CNOS in a Separable Hilbert Space

Definition 13.9 A metric space E is called *separable* if there exists a countable dense subset of E.

Example 13.7 (a) $\mathbb{R}^n$ is separable. $M = \{(r_1, \dots, r_n) \mid r_1, \dots, r_n \in \mathbb{Q}\}$ is a countable dense set in $\mathbb{R}^n$.
(b) $\mathbb{C}^n$ is separable. $M = \{(r_1 + \mathrm{i}s_1, \dots, r_n + \mathrm{i}s_n) \mid r_1, \dots, r_n, s_1, \dots, s_n \in \mathbb{Q}\}$ is a countable dense subset of $\mathbb{C}^n$.
(c) $\mathrm{L}^2([a, b])$ is separable. The polynomials $\{1, x, x^2, \dots\}$ are linearly independent in $\mathrm{L}^2([a, b])$ and they can be orthonormalized via Schmidt's process. As a result we get a countable CNOS in $\mathrm{L}^2([a, b])$ (Legendre polynomials in case $-a = 1 = b$). However, $\mathrm{L}^2(\mathbb{R})$ contains no polynomial; in this case the *Hermite functions* which are of the form $p_n(x)\,\mathrm{e}^{-x^2}$ with polynomials p_n, form a countable CNOS.
More general, $\mathrm{L}^2(G, \lambda_n)$ is separable for any region $G \subset \mathbb{R}^n$ with respect to the Lebesgue measure.
(d) Any Hilbert space is isomorphic to some $\mathrm{L}^2(X, \mu)$ where μ is the counting measure on X; $X = \mathbb{N}$ gives ℓ_2. X uncountable gives a non-separable Hilbert space.

Proposition 13.14 (Schmidt's Orthogonalization Process) *Let $\{y_k\}$ be an at most countable linearly independent subset of the Hilbert space H. Then there exists an NOS $\{x_k\}$ such that for every n*

$$\mathrm{lin}\,\{y_1, \dots, y_n\} = \mathrm{lin}\,\{x_1, \dots, x_n\}.$$

The NOS can be computed recursively,

$$x_1 := \frac{y_1}{\|y_1\|}, \quad x_{n+1} = (y_{n+1} - \sum_{k=1}^{n} \langle y_{n+1}\,,\,x_k\rangle\,x_k)/\,\|\sim\|$$

Corollary 13.15 *Let $\{\mathrm{e}_k \mid k \in N\}$ be an NOS where $N = \{1, \dots, n\}$ for some $n \in \mathbb{N}$ or $N = \mathbb{N}$. Suppose that $H_1 = \mathrm{lin}\,\{\mathrm{e}_k \mid k \in N\}$ is the linear span of the NOS. Then $x_1 = \sum_{k\in N} \langle x\,,\,\mathrm{e}_k\rangle\,\mathrm{e}_k$ is the orthogonal projection of $x \in H$ onto H_1.*

Proposition 13.16 (a) *A Hilbert space H has an at most countable complete orthonormal system (CNOS) if and only if H is separable.*
(b) *Let H be a separable Hilbert space. Then H is either isomorphic to $\mathbb{K}^n$ for some $n \in \mathbb{N}$ or to ℓ_2.*

13.1.5 Appendix

(a) The Inner Product constructed from an Inner Product Norm

Proof of Proposition 13.5. We consider only the case $\mathbb{K} = \mathbb{R}$. Assume that the parallelogram identity is satisfied. We will show that

$$\langle x\,,\, y\rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right)$$

defines a bilinear form on E.
(a) We show Additivity. First note that the parallelogram identity implies

$$\begin{aligned}
&\|x_1{+}x_2{+}y\|^2 = \frac{1}{2}\|x_1{+}x_2{+}y\|^2 + \frac{1}{2}\|x_1{+}x_2{+}y\|^2 \\
&= \frac{1}{2}\left(2\|x_1{+}y\|^2 + 2\|x_2\|^2 - \|x_1{-}x_2{+}y\|^2\right) + \frac{1}{2}\left(2\|x_2{+}y\|^2 + 2\|x_1\|^2 - \|x_2{-}x_1{+}y\|^2\right) \\
&= \|x_1{+}y\|^2 + \|x_2{+}y\|^2 + \|x_1\|^2 + \|x_2\|^2 - \frac{1}{2}\left(\|x_1{-}x_2{+}y\|^2 + \|x_2{-}x_1{+}y\|^2\right)
\end{aligned}$$

Replacing y by $-y$, we have

$$\|x_1{+}x_2{-}y\|^2 = \|x_1{-}y\|^2 + \|x_2{-}y\|^2 + \|x_1\|^2 + \|x_2\|^2 - \frac{1}{2}\left(\|x_1{-}x_2{-}y\|^2 + \|x_2{-}x_1{-}y\|^2\right).$$

By definition and the above two formulas,

$$\begin{aligned}
\langle x_1 + x_2\,,\, y\rangle &= \frac{1}{4}\left(\|x_1 + x_2 + y\|^2 - \|x_1 + x_2 - y\|^2\right) \\
&= \frac{1}{2}\left(\|x_1 + y\|^2 - \|x_1 - y\|^2 + \|x_2 + y\|^2 - \|x_2 - y\|^2\right) \\
&= \langle x_1\,,\, y\rangle + \langle x_2\,,\, y\rangle\,,
\end{aligned}$$

that is, $\langle\cdot\,,\,\cdot\rangle$ is additive in the first variable. It is obviously symmetric and hence additive in the second variable, too.
(b) We show $\langle \lambda x\,,\, y\rangle = \lambda\langle x\,,\, y\rangle$ for all $\lambda \in \mathbb{R}$, $x, y \in E$. By (a), $\langle 2x\,,\, y\rangle = 2\langle x\,,\, y\rangle$. By induction on n, $\langle nx\,,\, y\rangle = n\langle x\,,\, y\rangle$ for all $n \in \mathbb{N}$. Now let $\lambda = \frac{m}{n}$, $m, n \in \mathbb{N}$. Then

$$\begin{aligned}
n\langle \lambda x\,,\, y\rangle = n\left\langle \frac{m}{n}x\,,\, y\right\rangle = \left\langle n\frac{m}{n}x\,,\, y\right\rangle = m\langle x\,,\, y\rangle \\
\Longrightarrow\ \langle \lambda x\,,\, y\rangle = \frac{m}{n}\langle x\,,\, y\rangle = \lambda\langle x\,,\, y\rangle\,.
\end{aligned}$$

Hence, $\langle \lambda x\,,\, y\rangle = \lambda\langle x\,,\, y\rangle$ holds for all positive rational numbers λ. Suppose $\lambda \in \mathbb{Q}_+$, then

$$0 = \langle x + (-x)\,,\, y\rangle = \langle x\,,\, y\rangle + \langle -x\,,\, y\rangle$$

implies $\langle -x\,,\,y\rangle = -\langle x\,,\,y\rangle$ and, moreover, $\langle -\lambda x\,,\,y\rangle = -\lambda\,\langle x\,,\,y\rangle$ such that the equation holds for all $\lambda \in \mathbb{Q}$. Suppose that $\lambda \in \mathbb{R}$ is given. Then there exists a sequence (λ_n), $\lambda_n \in \mathbb{Q}$ of rational numbers with $\lambda_n \to \lambda$. This implies $\lambda_n x \to \lambda x$ for all $x \in E$ and, since $\|\cdot\|$ is continuous,

$$\langle \lambda x\,,\,y\rangle = \lim_{n\to\infty} \langle \lambda_n x\,,\,y\rangle = \lim_{n\to\infty} \lambda_n\,\langle x\,,\,y\rangle = \lambda\,\langle x\,,\,y\rangle\,.$$

This completes the proof. ∎

(b) Convergence of Orthogonal Series

We reformulate Lemma 13.10

> *Let $\{x_n\}$ be an OS in H. Then $\sum_{k=1}^{\infty} x_k$ converges if and only if $\sum_{k=1}^{\infty} \|x_k\|^2$ converges.*

Note that the convergence of a series $\sum_{i=1}^{\infty} x_i$ of elements x_i of a Hilbert space H is defined to be the limit of the partial sums $\lim_{n\to\infty} \sum_{i=1}^{n} x_i$. In particular, the Cauchy criterion applies since H is complete:

> The series $\sum y_i$ converges if and only if for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that $m, n \geq n_0$ imply $\left\| \sum_{i=m}^{n} y_i \right\| < \varepsilon$.

Proof. By the above discussion, $\sum_{k=1}^{\infty} x_k$ converges if and only if $\|\sum_{k=m}^{n} x_k\|^2$ becomes small for sufficiently large $m, n \in \mathbb{N}$. By the Pythagorean theorem this term equals

$$\sum_{k=m}^{n} \|x_k\|^2\,;$$

hence the series $\sum x_k$ converges, if and only if the series $\sum \|x_k\|^2$ converges. ∎

13.2 Bounded Linear Operators in Hilbert Spaces

13.2.1 Bounded Linear Operators

Let $(E_1, \|\cdot\|_1)$ and $(E_2, \|\cdot\|_2)$ be normed linear space. Recall that a linear map $T\colon E_1 \to E_2$ is called *continuous* if $x_n \longrightarrow x$ in E_1 implies $T(x_n) \longrightarrow T(x)$ in E_2.

Definition 13.10 (a) A linear map $T\colon E_1 \to E_2$ is called *bounded* if there exist a positive real number $C > 0$ such that

$$\|T(x)\|_2 \leq C\,\|x\|_1\,, \quad \text{for all} \quad x \in E_1. \tag{13.6}$$

(b) Suppose that $T\colon E_1 \to E_2$ is a bounded linear map. Then the *operator norm* is the smallest number C satisfying (13.6) for all $x \in E_1$, that is

$$\|T\| = \inf\{C > 0 \mid \ \forall\, x \in E_1\colon \ \|T(x)\|_2 \le C\,\|x\|_1\}\,.$$

One can show that

$$\begin{aligned}
&\text{(a)} \quad \|T\| = \sup\left\{\frac{\|T(x)\|_2}{\|x\|_1} \mid x \in E_1,\ x \neq 0\right\},\\
&\text{(b)} \quad \|T\| = \sup\{\|T(x)\|_2 \mid \|x\|_1 \le 1\}\\
&\text{(c)} \quad \|T\| = \sup\{\|T(x)\|_2 \mid \|x\|_1 = 1\}
\end{aligned}$$

Indeed, we may restrict ourselves to unit vectors since

$$\frac{\|T(\alpha x)\|_2}{\|\alpha x\|_1} = \frac{|\,\alpha\,|\,\|T(x)\|_2}{|\,\alpha\,|\,\|x\|_1} = \frac{\|T(x)\|_2}{\|x\|_1}.$$

This shows the equivalence of (a) and (c). Since $\|T(\alpha x)\|_2 = |\,\alpha\,|\,\|T(x)\|_2$, the suprema (b) and (c) are equal. From The last equality follows from the fact that the least upper bound is the infimum over all upper bounds. From (a) and (d) it follows,

$$\|T(x)\|_2 \le \|T\|\,\|x\|_1\,. \tag{13.7}$$

Also, if $E_1 \overset{S}{\Rightarrow} E_2 \overset{T}{\Rightarrow} E_3$ are bounded linear mappings, then $T\circ S$ is a bounded linear mapping with

$$\|T\circ S\| \le \|T\|\ \|S\|\,.$$

Indeed, for $x \neq 0$ one has by (13.7)

$$\|T(S(x))\|_3 \le \|T\|\ \|S(x)\|_2 \le \|T\|\ \|S\|\,\|x\|_1\,.$$

Hence, $\|(T\circ S)(x)\|_3 \,/\, \|x\|_1 \le \|T\|\ \|S\|$.

Proposition 13.17 *For a linear map $T\colon E_1 \to E_2$ of a normed space E_1 into a normed space E_2 the following are equivalent:*

(a) *T is bounded.*
(b) *T is continuous.*
(c) *T is continuous at one point of E_1.*

Proof. (a) $\to$ (b). This follows from the fact

$$\|T(x_1) - T(x_2)\| = \|T(x_1 - x_2)\| \le \|T\|\,\|x_1 - x_2\|\,,$$

and T is even uniformly continuous on E_1. (b) trivially implies (c).
(c) $\to$ (a). Suppose T is continuous at x_0. To each $\varepsilon > 0$ one can find $\delta > 0$ such that $\|x - x_0\| < \delta$ implies $\|T(x) - T(x_0)\| < \varepsilon$. Let $y = x - x_0$. In other words $\|y\| < \delta$ implies

$$\|T(y + x_0) - T(x_0)\| = \|T(y)\| < \varepsilon.$$

Suppose $z \in E_1$, $\|z\| \le 1$. Then $\|\delta/2z\| \le \delta/2 < \delta$; hence $\|T(\delta/2z)\| < \varepsilon$. By linearity of T, $\|T(z)\| < 2\varepsilon/\delta$. This shows $\|T\| \le 2\varepsilon/\delta$. ∎

Definition 13.11 Let E and F be normed linear spaces. Let $\mathscr{L}(E,F)$ denote the set of all bounded linear maps from E to F. In case $E = F$ we simply write $\mathscr{L}(E)$ in place of $\mathscr{L}(E,F)$.

Proposition 13.18 *Let E and F be normed linear spaces. Then $\mathscr{L}(E,F)$ is a normed linear space if we define the linear structure by*

$$(S+T)(x) = S(x) + T(x), \quad (\lambda T)(x) = \lambda\, T(x)$$

for $S, T \in \mathscr{L}(E,F)$, $\lambda \in \mathbb{K}$. The operator norm $\|T\|$ makes $\mathscr{L}(E,F)$ a normed linear space.

Note that $\mathscr{L}(E,F)$ is complete if and only if F is complete.

Example 13.8 (a) Recall that $\mathscr{L}(\mathbb{K}^n, \mathbb{K}^m)$ is a normed vector space with $\|A\| \leq \left(\sum_{i,j} |\, a_{ij}\,|^2\right)^{\frac{1}{2}}$, where $A = (a_{ij})$ is the matrix representation of the linear operator A, see Proposition 7.1
(b) The space $E' = \mathscr{L}(E, \mathbb{K})$ of continuous linear functionals on E.
(c) $H = \mathrm{L}^2((0,1))$, $g \in \mathrm{C}([0,1])$,

$$T_g(f)(t) = g(t) f(t)$$

defines a bounded linear operator on H. (see homework)
(d) $H = \mathrm{L}^2((0,1))$, $k(s,t) \in \mathrm{L}^2([0,1]\times[0,1])$. Then

$$(Kf)(t) = \int_0^1 k(s,t) f(s)\,\mathrm{d}s, \quad f \in H = \mathrm{L}^2([0,1])$$

defines a continuous linear operator $K \in \mathscr{L}(H)$. We have

$$\begin{aligned} |\,(Kf)(t)\,|^2 = \left| \int_0^1 k(s,t) f(s)\,\mathrm{d}s \right|^2 &\leq \left(\int_0^1 |\,k(s,t)\,|\;|\,f(s)\,|\;\mathrm{d}s \right)^2 \\ &\underset{\text{C-S-I}}{\leq} \int_0^1 |\,k(s,t)\,|^2\;\mathrm{d}s \int_0^1 |\,f(s)\,|^2\;\mathrm{d}s \\ &= \int_0^1 |\,k(s,t)\,|^2\;\mathrm{d}s\, \|f\|_H^2\,. \end{aligned}$$

Hence,

$$\begin{aligned} \|K(f)\|_H^2 &\leq \int_0^1 \left(\int_0^1 |\,k(s,t)\,|^2\;\mathrm{d}s \right) \mathrm{d}t\, \|f\|_H^2 \\ \|K(f)\|_H &\leq \|k\|_{\mathrm{L}^2([0,1]\times[0,1])}\; \|f\|_H\,. \end{aligned}$$

This shows $Kf \in H$ and further, $\|K\| \leq \|k\|_{\mathrm{L}^2([0,1]^2)}$. K is called an *integral operator*; K is compact, i. e. it maps the unit ball into a set whose closure is compact.
(e) $H = \mathrm{L}^2(\mathbb{R})$, $a \in \mathbb{R}$,

$$(V_a f)(t) = f(t-a), \quad t \in \mathbb{R},$$

defines a bounded linear operator called the shift operator. Indeed,

$$\|V_a f\|_2^2 = \int_{\mathbb{R}} |\,f(t-a)\,|^2\;\mathrm{d}t = \int_{\mathbb{R}} |\,f(t)\,|^2\;\mathrm{d}t = \|f\|_2^2\,;$$

since all quotients $\|V_a(x)\| \,/\, \|x\| = 1$, $\|V_a\| = 1$.
(f) $H = \ell_2$. We define the *right-shift* S by

$$S(x_1, x_2, \dots) = (0, x_1, x_2, \dots).$$

Obviously, $\|S(x)\| = \|x\| = \left(\sum_{n=1}^{\infty} |\, x_n \,|^2\right)^{\frac{1}{2}}$. Hence, $\|S\| = 1$.
(g) Let $E_1 = \mathrm{C}^1([0,1])$ and $E_2 = \mathrm{C}([0,1])$. Define the *differentiation* operator $(Tf)(t) = f'(t)$. Let $\|f\|_1 = \|f\|_2 = \sup\limits_{t\in[0,1]} |\, f(t) \,|$. Then T is linear but not bounded. Indeed, let $f_n(t) = 1 - t^n$. Then $\|f_n\|_1 = 1$ and $Tf_n(t) = nt^{n-1}$ such that $\|tf_n\|_2 = n$. Thus, $\|Tf_n\|_2 \,/\, \|f_n\|_1 = n \to +\infty$ as $n \to \infty$. T is unbounded.
However, if we put $\|f\|_1 = \sup\limits_{t\in[0,1]} |\, f(t) \,| + \sup\limits_{t\in[0,1]} |\, f'(t) \,|$ and $\|f\|_2$ as before, then T is bounded since

$$\|Tf\|_2 = \sup_{t\in[0,1]} |\, f'(t) \,| \le \|f\|_1 \implies \|T\| \le 1.$$

13.2.2 The Adjoint Operator

In this subsection H is a Hilbert space and $\mathscr{L}(H)$ the space of bounded linear operators on H. Let $T \in \mathscr{L}(H)$ be a bounded linear operator and $y \in H$. Then $F(x) = \langle T(x)\,,\, y\rangle$ defines a continuous linear functional on H. Indeed,

$$|\, F(x) \,| = |\, \langle T(x)\,,\, y\rangle \,| \underset{\text{CSI}}{\le} \|T(x)\|\, \|y\| \le \underbrace{\|T\|\, \|y\|}_{C} \|x\| \le C\, \|x\|\,.$$

Hence, F is bounded and therefore continuous. in particular,

$$\|F\| \le \|T\|\, \|y\|$$

By Riesz's representation theorem, there exists a unique vector $z \in H$ such that

$$\langle T(x)\,,\, y\rangle = F(x) = \langle x\,,\, z\rangle\,.$$

Note that by the above inequality

$$\|z\| = \|F\| \le \|T\|\, \|y\|\,. \tag{13.8}$$

Suppose y_1 is another element of H which corresponds to $z_1 \in H$ with

$$\langle T(x)\,,\, y_1\rangle = \langle x\,,\, z_1\rangle\,.$$

Finally, let $u \in H$ be the element which corresponds to $y + y_1$,

$$\langle T(x)\,,\, y + y_1\rangle = \langle x\,,\, u\rangle\,.$$

Since the element u which is given by Riesz's representation theorem is unique, we have $u = z + z_1$. Similarly,

$$\langle T(x)\,,\, \lambda y\rangle = F(x) = \langle x\,,\, \lambda z\rangle$$

shows that λz corresponds to λy.

Definition 13.12 The above correspondence $y \mapsto z$ is linear. Define the linear operator T^* by $z = T^*(y)$. By definition,

$$\langle T(x)\,,\, y\rangle = \big\langle x\,,\, T^*(y)\big\rangle\,, \quad x, y \in H. \tag{13.9}$$

T^* is called the *adjoint operator to* T.

Proposition 13.19 *Let $T, T_1, T_2 \in \mathscr{L}(H)$. Then T^* is a bounded linear operator with $\big\|T^*\big\| = \|T\|$. We have*

(a) $(T_1 + T_2)^* = T_1^* + T_2^*$ *and*

(b) $(\lambda\, T)^* = \overline{\lambda}\, T^*$.

(c) $(T_1 T_2)^* = T_2^*\, T_1^*$.

(d) *If T is invertible in $\mathscr{L}(H)$, so is T^*, and we have $(T^*)^{-1} = (T^{-1})^*$.*

(e) $(T^*)^* = T$.

Proof. Inequality (13.8) shows that

$$\big\|T^*(y)\big\| \le \|T\|\,\|y\|\,, \quad y \in H.$$

By definition, this implies

$$\big\|T^*\big\| \le \|T\|$$

and T^* is bounded. Since

$$\big\langle T^*(x)\,,\, y\big\rangle = \overline{\langle y\,,\, T^*(x)\rangle} = \overline{\langle T(y)\,,\, x\rangle} = \langle x\,,\, T(y)\rangle\,,$$

we get $(T^*)^* = T$. We conclude $\|T\| = \big\|T^{**}\big\| \le \big\|T^*\big\|$; such that $\big\|T^*\big\| = \|T\|$.
(a). For $x, y \in H$ we have

$$\begin{aligned}\langle (T_1 + T_2)(x)\,,\, y\rangle &= \langle T_1(x) + T_2(x)\,,\, y\rangle = \langle T_1(x)\,,\, y\rangle + \langle T_2(x)\,,\, y\rangle \\ &= \big\langle x\,,\, T_1^*(y)\big\rangle + \big\langle x\,,\, T_2^*(y)\big\rangle = \big\langle x\,,\, (T_1^* + T_2^*)(y)\big\rangle\,;\end{aligned}$$

which proves (a).
(c) and (d) are left to the reader. ∎

A mapping $*\colon A \to A$ such that the above properties (a), (b), and (c) are satisfied is called an *involution*. An algebra with involution is called a $*$*-algebra*.
We have seen that $\mathscr{L}(H)$ is a (non-commutative) $*$-algebra. An example of a commutative $*$-algebra is $\mathrm{C}(K)$ with the involution $f^*(x) = \overline{f(x)}$.

Example 13.9 (Example 13.8 continued)
(a) $H = \mathbb{C}^n$, $A = (a_{ij}) \in \mathrm{M}(n \times n, \mathbb{C})$. Then $A^* = (b_{ij})$ has the matrix elements $b_{ij} = \overline{a_{ji}}$.
(b) $H = \mathrm{L}^2([0,1])$, $T_g^* = T_{\overline{g}}$.
(c) $H = \mathrm{L}^2(\mathbb{R})$, $V_a(f)(t) = f(t-a)$ (Shift operator), $V_a^* = V_{-a}$.

(d) $H = \ell_2$. The *right-shift* S is defined by $S((x_n)) = (0, x_1, x_2, \dots)$. We compute the adjoint S^*.

$$\langle S(x)\,,\, y\rangle = \sum_{n=2}^{\infty} x_{n-1}y_n = \sum_{n=1}^{\infty} x_n y_{n+1} = \langle (x_1, x_2, \dots)\,,\, (y_2, y_3, \dots)\rangle\,.$$

Hence, $S^*((y_n)) = (y_2, y_3, \dots)$ is the *left-shift*.

13.2.3 Classes of Bounded Linear Operators

Let H be a *complex* Hilbert space.

(a) Self-Adjoint and Normal Operators

Definition 13.13 An operator $A \in \mathscr{L}(H)$ is called

(a) *self-adjoint*, if $A^* = A$,

(b) *normal* , if $A^*A = A\,A^*$,

A self-adjoint operator A is called *positive*, if $\langle Ax\,,\, x\rangle \geq 0$ for all $x \in H$. We write $A \geq 0$. If A and B are self-adjoint, we write $A \geq B$ if $A - B \geq 0$.

A crucial role in proving the simplest properties plays the *polarization identity* which generalizes the polarization identity from Subsection 13.1.2. However, this exist only in *complex* Hilbert spaces.

$$4\,\langle A(x)\,,\, y\rangle = \langle A(x+y)\,,\, x+y\rangle - \langle A(x-y)\,,\, x-y\rangle + \\ + \mathrm{i}\,\langle A(x+\mathrm{i}y)\,,\, x+\mathrm{i}y\rangle - \mathrm{i}\,\langle A(x-\mathrm{i}y)\,,\, x-\mathrm{i}y\rangle\,.$$

We use the identity as follows

$$\langle A(x)\,,\, x\rangle = 0 \quad \text{for all } x \in H \text{ implies} \quad A = 0.$$

Indeed, by the polarization identity, $\langle A(x)\,,\, y\rangle = 0$ for all $x, y \in H$. In particular $y = A(x)$ yields $A(x) = 0$ for all x; thus, $A = 0$.

Remarks 13.2 (a) A is normal if and only if $\|A(x)\| = \left\|A^*(x)\right\|$ for all $x \in H$. Indeed, if A is normal, then for all $x \in H$ we have $\left\langle A^*A(x)\,,\, x\right\rangle = \left\langle A\,A^*(x)\,,\, x\right\rangle$ which imply $\|A(x)\|^2 = \langle A(x)\,,\, A(x)\rangle = \left\langle A^*(x)\,,\, A^*(x)\right\rangle = \left\|A^*(x)\right\|^2$. On the other hand, the polarization identity and $\left\langle A^*A(x)\,,\, x\right\rangle = \left\langle A\,A^*(x)\,,\, x\right\rangle$ implies $\left\langle (A^*A - A\,A^*)(x)\,,\, x\right\rangle = 0$ for all x; hence $A^*A - A\,A^* = 0$ which proves the claim.
(b) Sums and real scalar multiples of self-adjoint operators are self-adjoint.
(c) The product AB of self-adjoint operators is self-adjoint if and only if A and B commute with each other, $AB = BA$.
(d) A is self-adjoint if and only if $\langle Ax\,,\, x\rangle$ is real for all $x \in H$.
Proof. Let $A^* = A$. Then $\langle Ax\,,\, x\rangle = \langle x\,,\, Ax\rangle = \overline{\langle Ax\,,\, x\rangle}$ is real; for the opposite direction $\langle A(x)\,,\, x\rangle = \langle x\,,\, A(x)\rangle$ and the polarization identity yields $\langle A(x)\,,\, y\rangle = \langle x\,,\, A(y)\rangle$ for all x, y; hence $A^* = A$. ∎

(b) Unitary and Isometric Operators

Definition 13.14 Let $T \in \mathscr{L}(H)$. Then T is called

(a) *unitary*, if $T^*T = I = T\,T^*$.

(b) *isometric*, if $\|T(x)\| = \|x\|$ for all $x \in H$.

Proposition 13.20 (a) *T is isometric if and only if $T^*T = I$ and if and only if $\langle T(x)\,,\,T(y)\rangle = \langle x\,,\,y\rangle$ for all $x, y \in H$.*
(b) *T is unitary, if and only if T is isometric and surjective.*
(c) *If S, T are unitary, so are ST and T^{-1}. The unitary operators of $\mathscr{L}(H)$ form a group.*

Proof. (a) T isometric yields $\langle T(x)\,,\,T(x)\rangle = \langle x\,,\,x\rangle$ and further $\big\langle (T^*\,T - I)(x)\,,\,x\big\rangle = 0$ for all x. The polarization identity implies $T^*\,T = I$. This implies $\big\langle (T^*\,T - I)(x)\,,\,y\big\rangle = 0$, for all $x, y \in H$. Hence, $\langle T(x)\,,\,T(y)\rangle = \langle x\,,\,y\rangle$. Inserting $y = x$ shows T is isometric.
(b) Suppose T is unitary. $T^*T = I$ shows T is isometric. Since $T\,T^* = I$, T is surjective.
Suppose now, T is isometric and surjective. Since T is isometric, $T(x) = 0$ implies $x = 0$; hence, T is bijective with an inverse operator T^{-1}. Insert $y = T^{-1}(z)$ into $\langle T(x)\,,\,T(y)\rangle = \langle x\,,\,y\rangle$. This gives

$$\langle T(x)\,,\,z\rangle = \big\langle x\,,\,T^{-1}(z)\big\rangle\,, \quad x, z \in H.$$

Hence $T^{-1} = T^*$ and therefore $T^*T = T\,T^* = I$.
(c) is easy (see homework 45.4). ∎

Note that an isometric operator is injective with norm 1 (since $\|T(x)\| \,/\, \|x\| = 1$ for all x). In case $H = \mathbb{C}^n$, the unitary operators on $\mathbb{C}^n$ form the *unitary group* $\mathrm{U}(n)$. In case $H = \mathbb{R}^n$, the unitary operators on H form the *orthogonal group* $\mathrm{O}(n)$.

Example 13.10 (a) $H = \mathrm{L}^2(\mathbb{R})$. The shift operator V_a is unitary since $V_aV_b = V_{a+b}$. The multiplication operator $T_g f = gf$ is unitary if and only if $|\,g\,| = 1$. T_g is self-adjoint (resp. positive) if and only if g is real (resp. positive).
(b) $H = \ell_2$, the right-shift $S((x_n)) = (0, x_1, x_2, \dots)$ is isometric but not unitary since S is not surjective. S^* is not isometric since $S^*(1, 0, \dots) = 0$; hence S^* is not injective.
(c) **Fourier transform.** For $f \in \mathrm{L}^1(\mathbb{R})$ define

$$(\mathcal{F}f)(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tx} f(x)\,\mathrm{d}x.$$

Let $\mathcal{S}(\mathbb{R}) = \{f \in \mathrm{C}^\infty(\mathbb{R}) \mid \sup_{t\in\mathbb{R}} \big|\, t^n f^{(k)}(t)\,\big| < \infty,\ \forall\, n, k \in \mathbb{Z}_+\}$. $\mathcal{S}(\mathbb{R})$ is called the *Schwartz space* after Laurent Schwartz. We have $\mathcal{S}(\mathbb{R}) \subseteq \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, for example, $f(x) = \mathrm{e}^{-x^2} \in \mathcal{S}(\mathbb{R})$. We will show later that $\mathcal{F}\colon \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is bijective and norm preserving, $\|\mathcal{F}(f)\|_{\mathrm{L}^2(\mathbb{R})} = \|f\|_{\mathrm{L}^2(\mathbb{R})}$, $f \in \mathcal{S}(\mathbb{R})$. $\mathcal{F}$ has a unique extension to a unitary operator on $\mathrm{L}^2(\mathbb{R})$. The inverse Fourier transform is

$$(\mathcal{F}^{-1}f)(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} f(x)\,\mathrm{d}x, \quad f \in \mathcal{S}(\mathbb{R}).$$

13.2.4 Orthogonal Projections

(a) Riesz's First Theorem—revisited

Let H_1 be a closed linear subspace. By Theorem 13.7 any $x \in H$ has a unique decomposition $x = x_1 + x_2$ with $x_1 \in H_1$ and $x_2 \in H_1^\perp$. The map $P_{H_1}(x) = x_1$ is a linear operator from H to H, (see homework 44.1). P_{H_1} is called the *orthogonal projection* from H onto the closed subspace H_1. Obviously, H_1 is the image of P_{H_1}; in particular, P_{H_1} is surjective if and only if $H_1 = H$. In this case, $P_H = I$ is the identity. Since

$$\|P_{H_1}(x)\|^2 = \|x_1\|^2 \leq \|x_1\|^2 + \|x_2\|^2 = \|x\|^2$$

we have $\|P_{H_1}\| \leq 1$. If $H_1 \neq \{0\}$, there exists a non-zero $x_1 \in H_1$ such that $\|P_{H_1}(x_1)\| = \|x_1\|$. This shows $\|P_{H_1}\| = 1$.
Here is the algebraic characterization of orthogonal projections.

Proposition 13.21 *A linear operator $P \in \mathscr{L}(H)$ is an orthogonal projection if and only if $P^2 = P$ and $P^* = P$.*
In this case $H_1 = \{x \in H \mid P(x) = x\}$.

Proof. "$\rightarrow$". Suppose that $P = P_{H_1}$ is the projection onto H_1. Since P is the identity on H_1, $P^2(x) = P(x_1) = x_1 = P(x)$ for all $x \in H$; hence $P^2 = P$.
Let $x = x_1 + x_2$ and $y = y_1 + y_2$ be the unique decompositions of x and y in elements of H_1 and $H_1^\perp$, respectively. Then

$$\langle P(x)\,,\, y\rangle = \langle x_1\,,\, y_1 + y_2\rangle = \langle x_1\,,\, y_1\rangle + \underbrace{\langle x_1\,,\, y_2\rangle}_{=0} = \langle x_1\,,\, y_1\rangle = \langle x_1 + x_2\,,\, y_1\rangle = \langle x\,,\, P(y)\rangle\,,$$

that is, $P^* = P$.
"$\leftarrow$". Suppose $P^2 = P = P^*$ and put $H_1 = \{x \mid P(x) = x\}$. First note, that for $P \neq 0$, $H_1 \neq \{0\}$ is non-trivial. Indeed, since $P(P(x)) = P(x)$, the image of P is part of the eigenspace of P to the eigenvalues 1, $P(H) \subset H_1$. Since for $z \in H_1$, $P(z) = z$, $H_1 \subset P(H)$ and thus $H_1 = P(H)$.
Since P is continuous and $\{0\}$ is closed, $H_1 = (P - I)^{-1}(\{0\})$ is a closed linear subspace of H. By Riesz's first theorem, $H = H_1 \oplus H_1^\perp$. We have to show that $P(x) = x_1$ for all x.
Since $P^2 = P$, $P(P(x)) = P(x)$ for all x; hence $P(x) \in H_1$. We show $x - P(x) \in H_1^\perp$ which completes the proof. For, let $z \in H_1$, then

$$\langle x - P(x)\,,\, z\rangle = \langle x\,,\, z\rangle - \langle P(x)\,,\, z\rangle = \langle x\,,\, z\rangle - \langle x\,,\, P(z)\rangle = \langle x\,,\, z\rangle - \langle x\,,\, z\rangle = 0.$$

Hence $x = P(x) + (I - P)(x)$ is the unique Riesz decomposition of x with respect to H_1 and $H_1^\perp$. ∎

Example 13.11 (a) Let $\{x_1, \ldots, x_n\}$ be an NOS in H. Then

$$P(x) = \sum_{k=1}^{n} \langle x\,,\, x_k\rangle\; x_k, \quad x \in H,$$

defines the orthogonal projection $P\colon H \to H$ onto $\operatorname{lin}\{x_1, \ldots, x_n\}$. Indeed, since $P(x_m) = \sum_{k=1}^n \langle x_m\,,\, x_k\rangle\, x_k = x_m$, $P^2 = P$ and since

$$\langle P(x)\,,\, y\rangle = \sum_{k=1}^n \langle x\,,\, x_k\rangle\ \langle x_k\,,\, y\rangle = \left\langle x\,,\, \sum_{k=1}^n \overline{\langle x_k\,,\, y\rangle}\, x_k \right\rangle = \langle x\,,\, P(y)\rangle\,.$$

Hence, $P^* = P$ and P is a projection.
(b) $H = \mathrm{L}^2([0,1]\cup[2,3])$, $g \in \mathrm{C}([0,1]\cup[2,3])$. For $f \in H$ define $T_g f = gf$. Then $T_g = (T_g)^*$ if and only if $g(t)$ is real for all t. T_g is an orthogonal projection if $g(t)^2 = g(t)$ such that $g(t) = 0$ or $g(t) = 1$. Since g is continuous, there are only four solutions: $g_1 = 0$, $g_2 = 1$, $g_3 = \chi_{[0,1]}$, and $g_4 = \chi_{[2,3]}$.
In case of g_3, the subspace H_1 can be identified with $\mathrm{L}^2([0,1])$ since $f \in H_1$ iff $T_g f = f$ iff $gf = f$ iff $f(t) = 0$ for all $t \in [2,3]$.

(b) Properties of Orthogonal Projections

Throughout this paragraph let P_1 and P_2 be orthogonal projections on the closed subspaces H_1 and H_2, respectively.

Lemma 13.22 *The following are equivalent.*

(a) $P_1 + P_2$ *is an orthogonal projection.*
(b) $P_1 P_2 = 0$.
(c) $H_1 \perp H_2$.

Proof. (a) → (b). Let $P_1 + P_2$ be a projection. Then

$$(P_1 + P_2)^2 = P_1^2 + P_1P_2 + P_2P_1 + P_2^2 = P_1 + P_2 + P_1P_2 + P_2P_1 \overset{!}{=} P_1 + P_2,$$

hence $P_1P_2 + P_2P_1 = 0$. Multiplying this from the left by P_1 and from the right by P_1 yields

$$P_1P_2 + P_1P_2P_1 = 0 = P_1P_2P_1 + P_2P_1.$$

This implies $P_1P_2 = P_2P_1$ and finally $P_1P_2 = P_2P_1 = 0$.
(b) → (c). Let $x_1 \in H_1$ and $x_2 \in H_2$. Then

$$0 = \langle P_1P_2(x_2)\,,\, x_1\rangle = \langle P_2(x_2)\,,\, P_1(x_1)\rangle = \langle x_2\,,\, x_1\rangle\,.$$

This shows $H_1 \perp H_2$.
(c) → (b). Let $x, z \in H$ be arbitrary. Then

$$\langle P_1P_2(x)\,,\, z\rangle = \langle P_2(x)\,,\, P_1(z)\rangle = \langle x_2\,,\, z_1\rangle = 0;$$

Hence $P_1P_2(x) = 0$ and therefore $P_1P_2 = 0$. The same argument works for $P_2P_1 = 0$.
(b) → (a). Since $P_1P_2 = 0$ implies $P_2P_1 = 0$ (via $H_1 \perp H_2$),

$$\begin{aligned}(P_1 + P_2)^* &= P_1^* + P_2^* = P_1 + P_2,\\ (P_1 + P_2)^2 &= P_1^2 + P_1P_2 + P_2P_1 + P_2^2 = P_1 + 0 + 0 + P_2.\end{aligned}$$

■

Lemma 13.23 *The following are equivalent*

(a) P_1P_2 *is an orthogonal projection.*
(b) $P_1P_2 = P_2P_1$.

In this case, P_1P_2 *is the orthogonal projection onto* $H_1 \cap H_2$.

Proof. (b) $\rightarrow$ (a). $(P_1P_2)^* = P_2^*\,P_1^* = P_2P_1 = P_1P_2$, by assumption. Moreover, $(P_1P_2)^2 = P_1P_2P_1P_2 = P_1P_1P_2P_2 = P_1P_2$ which completes this direction.
(a) $\rightarrow$ (b). $P_1P_2 = (P_1P_2)^* = P_2^*P_1^* = P_2P_1$.
Clearly, $P_1P_2(H) \subseteq H_1$ and $P_2P_1(H) \subseteq H_2$; hence $P_1P_2(H) \subseteq H_1 \cap H_2$. On the other hand $x \in H_1 \cap H_2$ implies $P_1P_2x = x$. This shows $P_1P_2(H) = H_1 \cap H_2$. ∎

The proof of the following lemma is quite similar to that of the previous two lemmas, so we omit it (see homework 40.5).

Lemma 13.24 *The following are equivalent.*

(a) $H_1 \subseteq H_2$, (d) $P_1 \leq P_2$,
(b) $P_1P_2 = P_1$, (c) $P_2P_1 = P_1$,
(e) $P_2 - P_1$ *is an orth. projection,* (f) $\|P_1(x)\| \leq \|P_2(x)\|\,, \quad x \in H.$

Proof. We show (d) $\Rightarrow$ (c). From $P_1 \leq P_2$ we conclude that $I - P_2 \leq I - P_2$. Note that both $I - P_1$ and $I - P_2$ are again orthogonal projections on $H_1^\perp$ and $H_2^\perp$, respectively. Thus for all $x \in H$:

$$\begin{aligned}\|(I-P_2)P_1(x)\|^2 &= \langle (I-P_2)P_1(x)\,,\,(I-P_2)P_1(x)\rangle \\ &= \big\langle (I-P_2)^*(I-P_2)P_1(x)\,,\,P_1(x)\big\rangle = \langle (I-P_2)P_1(x)\,,\,P_1(x)\rangle \\ &\leq \langle (I-P_1)P_1(x)\,,\,P_1(x)\rangle = \langle P_1(x)-P_1(x)\,,\,P_1(x)\rangle = \langle 0\,,\,P_1(x)\rangle = 0.\end{aligned}$$

Hence, $\|(I-P_2)P_1(x)\|^2 = 0$ which implies $(I-P_2)P_1 = 0$ and therefore $P_1 = P_2P_1$. ∎

13.2.5 Spectrum and Resolvent

Let $T \in \mathscr{L}(H)$ be a bounded linear operator.

(a) Definitions

Definition 13.15 (a) The *resolvent set* of T, denoted by $\rho(T)$, is the set of all $\lambda \in \mathbb{C}$ such that there exists a bounded linear operator $R_\lambda(T) \in \mathscr{L}(H)$ with

$$R_\lambda(T)(T - \lambda I) = (T - \lambda I)R_\lambda(T) = I,$$

i. e. there $T - \lambda I$ has a bounded (continuous) inverse $R_\lambda(T)$. We call $R_\lambda(T)$ the *resolvent* of T at λ.
(b) The set $\mathbb{C} \setminus \rho(T)$ is called the *spectrum* of T and is denoted by $\sigma(T)$.
(c) $\lambda \in C$ is called an *eigenvalue* of T if there exists a nonzero vector x, called *eigenvector*, with $(T - \lambda I)x = 0$. The set of all eigenvalues is the *point spectrum* $\sigma_{\mathrm{p}}(T)$

Remark 13.3 (a) Note that the point spectrum is a subset of the spectrum, $\sigma_{\mathrm{p}}(T) \subseteq \sigma(T)$. Suppose to the contrary, the eigenvalue λ with eigenvector y belongs to the resolvent set. Then there exists $R_\lambda(T) \in \mathscr{L}(H)$ with

$$y = R_\lambda(T)(T - \lambda I)(y) = R_\lambda(T)(0) = 0$$

which contradicts the definition of an eigenvector; hence eigenvalues belong to the spectrum.
(b) $\lambda \in \sigma_{\mathrm{p}}(T)$ is equivalent to $T - \lambda I$ not being injective. It may happen that $T - \lambda I$ is not surjective, which also implies $\lambda \in \sigma(T)$ (see Example 13.12 (b) below).

Example 13.12 (a) $H = \mathbb{C}^n$, $A \in \mathrm{M}(n \times n, \mathbb{C})$. Since in finite dimensional spaces $T \in \mathscr{L}(H)$ is injective if and only if T is surjective, $\sigma(A) = \sigma_{\mathrm{p}}(A)$.
(b) $H = \mathrm{L}^2([0,1])$. $(Tf)(x) = xf(x)$. We have

$$\boldsymbol{\sigma_{\mathrm{p}}(T) = \varnothing.}$$

Indeed, suppose λ is an eigenvalue and $f \in \mathscr{L}^2([0,1])$ an eigenfunction to T, that is $(T - \lambda I)(f) = 0$; hence $(x - \lambda)f(x) \equiv 0$ a. e. on $[0,1]$. Since $x - \lambda$ is nonzero a. e. , $f = 0$ a. e. on $[0,1]$. That is $f = 0$ in H which contradicts the definition of an eigenvector. We have

$$\boldsymbol{\mathbb{C} \setminus [0,1] \subseteq \rho(T).}$$

Suppose $\lambda \notin [0,1]$. Since $x - \lambda \neq 0$ for all $x \in [0,1]$, $g(x) = \dfrac{1}{x - \lambda}$ is a continuous (hence bounded) function on $[0,1]$. Hence,

$$(R_\lambda f)(x) = \frac{1}{x - \lambda} f(x)$$

defines a bounded linear operator which is inverse to $T - \lambda I$ since

$$(T - \lambda I)\left(\frac{1}{x - \lambda} f(x)\right) = (x - \lambda)\left(\frac{1}{x - \lambda} f(x)\right) = f(x).$$

We have

$$\boldsymbol{\sigma(T) = [0,1].}$$

Suppose to the contrary that there exists $\lambda \in \rho(T) \cap [0,1]$. Then there exists $R_\lambda \in \mathscr{L}(H)$ with

$$R_\lambda(T - \lambda I) = I. \tag{13.10}$$

By homework 39.5 (a), the norm of the multiplication operator T_g is less that or equal to $\|g\|_\infty$ (the supremum norm of g). Choose $f_\varepsilon = \chi_{(\lambda - \varepsilon, \lambda + \varepsilon)}$. Since $\chi_M = \chi_M^2$,

$$\|(T - \lambda I)f_\varepsilon\| = \left\|(x - \lambda)\chi_{U_\varepsilon(\lambda)}(x) f_\varepsilon(x)\right\| \leq \sup_{x \in [0,1]} \left|\, (x - \lambda)\chi_{U_\varepsilon(\lambda)}(x) \,\right| \|f_\varepsilon\|.$$

However,

$$\sup_{x\in[0,1]} \big|\,(x-\lambda)\chi_{U_\varepsilon(\lambda)}(x)\,\big| = \sup_{x\in U_\varepsilon(\lambda)} |\,x-\lambda\,| = \varepsilon.$$

This shows

$$\|(T-\lambda I)f_\varepsilon\| \le \varepsilon\,\|f_\varepsilon\|\,.$$

Inserting f_ε into (13.10) we obtain

$$\|f_\varepsilon\| = \|R_\lambda(T-\lambda I)f_\varepsilon\| \le \|R_\lambda\|\,\|(T-\lambda I)f_\varepsilon\| \le \|R_\lambda\|\,\varepsilon\,\|f_\varepsilon\|$$

which implies $\|R_\lambda\| \ge 1/\varepsilon$. This contradicts the boundedness of R_λ since $\varepsilon > 0$ was arbitrary.

(b) Properties of the Spectrum

Lemma 13.25 *Let $T \in \mathscr{L}(H)$. Then*

$$\sigma(T^*) = \sigma(T)^*, \quad \textit{(complex conjugation)} \quad \rho(T^*) = \rho(T)^*.$$

Proof. Suppose that $\lambda \in \rho(T)$. Then there exists $R_\lambda(T) \in \mathscr{L}(H)$ such that

$$\begin{aligned} R_\lambda(T)(T-\lambda I) = (T-\lambda I)R_\lambda(T) &= I \\ (R_\lambda(T)(T-\lambda I))^* = ((T-\lambda I)R_\lambda)^* &= I \\ (T^*-\overline{\lambda}I)R_\lambda(T)^* = R_\lambda(t)^*(T^*-\overline{\lambda}I) &= I. \end{aligned}$$

This shows $R_{\overline{\lambda}}(T^*) = R_\lambda(T)^*$ is again a bounded linear operator on H. Hence, $\rho(T^*) \subseteq (\rho(T))^*$. Since $*$ is an involution ($T^{**} = T$), the opposite inclusion follows. Since $\sigma(T)$ is the complement of the resolvent set, the claim for the spectrum follows as well. ∎

For λ, μ, T and S we have

$$\begin{aligned} R_\lambda(T) - R_\mu(T) &= (\lambda-\mu)R_\lambda(T)R_\mu(T) = (\lambda-\mu)R_\mu(T)R_\lambda(T), \\ R_\lambda(T) - R_\lambda(S) &= R_\lambda(T)(S-T)R_\lambda(S). \end{aligned}$$

Proposition 13.26 (a) *$\rho(T)$ is open and $\sigma(T)$ is closed.*
(b) *If $\lambda_0 \in \rho(T)$ and $|\,\lambda-\lambda_0\,| < \|R_{\lambda_0}(T)\|^{-1}$ then $\lambda \in \rho(T)$ and*

$$R_\lambda(T) = \sum_{n=0}^{\infty} (\lambda-\lambda_0)^n R_{\lambda_0}(T)^{n+1}.$$

(c) *If $|\,\lambda\,| > \|T\|$, then $\lambda \in \rho(T)$ and*

$$R_\lambda(T) = -\sum_{n=0}^{\infty} \lambda^{-n-1} T^n.$$

Proof. (a) follows from (b).
(b) For brevity, we write R_{λ_0} in place of $R_{\lambda_0}(T)$. With $q = |\lambda - \lambda_0|\,\|R_{\lambda_0}(T)\|$, $q \in (0,1)$ we have

$$\sum_{n=0}^{\infty} |\lambda - \lambda_0|^n \|R_{\lambda_0}\|^{n+1} = \sum_{n=0}^{\infty} q^n \|R_{\lambda_0}\| = \frac{\|R_{\lambda_0}\|}{1-q} \quad \text{converges.}$$

By homework 38.4, $\sum x_n$ converges if $\sum \|x_n\|$ converges. Hence,

$$B = \sum_{n=0}^{\infty} (\lambda - \lambda_0)^n R_{\lambda_0}^{n+1}$$

converges in $\mathscr{L}(H)$ with respect to the operator norm. Moreover,

$$\begin{aligned}(T-\lambda I)B &= (T-\lambda_0 I)B - (\lambda - \lambda_0)B \\ &= \sum_{n=0}^{\infty} (\lambda-\lambda_0)^n (T-\lambda_0 I) R_{\lambda_0}^{n+1} - \sum_{n=0}^{\infty} (\lambda-\lambda_0)^{n+1} R_{\lambda_0}^{n+1} \\ &= \sum_{n=0}^{\infty} (\lambda-\lambda_0)^n R_{\lambda_0}^{n} - \sum_{n=0}^{\infty} (\lambda-\lambda_0)^{n+1} R_{\lambda_0}^{n+1} \\ &= (\lambda-\lambda_0)^0 R_{\lambda_0}^0 = I.\end{aligned}$$

Similarly, one shows $B(T-\lambda I) = I$. Thus, $R_\lambda(T) = B$.
(c) Since $|\lambda| > \|T\|$, the series converges with respect to operator norm, say

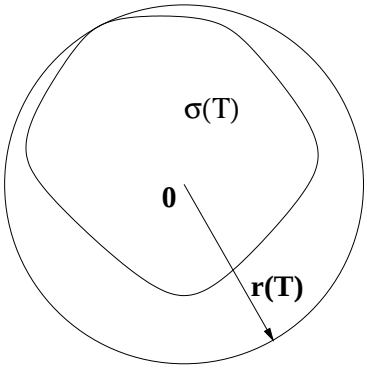
$$C = -\sum_{n=0}^{\infty} \lambda^{-n-1} T^n.$$

We have

$$(T-\lambda I)C = -\sum_{n=0}^{\infty} \lambda^{-n-1} T^{n+1} + \sum_{n=0}^{\infty} \lambda^{-n} T^n = \lambda^0 T^0 = I.$$

Similarly, $C(T-\lambda I) = I$; hence $R_\lambda(T) = C$. ∎

Remarks 13.4 (a) By (b), $R_\lambda(T)$ is a holomorphic (i. e. complex differentiable) function in the variable λ with values in $\mathscr{L}(H)$. One can use this to show that the spectrum is non-empty, $\sigma(T) \neq \varnothing$.
(b) If $\|T\| < 1$, $T - I$ is invertible with inverse $-\sum_{n=0}^{\infty} T^n$.



(c) Proposition 13.26 (c) means: If $\lambda \in \sigma(T)$ then $|\lambda| \leq \|T\|$. However, there is, in general, a smaller disc around 0 which contains the spectrum. By definition, the *spectral radius* $r(T)$ of T is the smallest non-negative number such that the spectrum is completely contained in the disc around 0 with radius $r(T)$:

$$r(T) = \sup\{|\lambda| \mid \lambda \in \sigma(T)\}.$$

(d) $\lambda \in \sigma(T)$ implies $\lambda^n \in \sigma(T^n)$ for all non-negative integers. Indeed, suppose $\lambda^n \in \rho(T^n)$, that is $B(T^n - \lambda^n) = (T^n - \lambda^n)B = I$ for some bounded B. Hence,

$$B\sum_{k=0}^{n} T^k \lambda^{n-1-k}(T-\lambda) = (T-\lambda)CB = I;$$

thus $\lambda \in \rho(T)$.
We shall refine the above statement and give a better upper bound for $\{|\,\lambda\,| \mid \lambda \in \sigma(T)\}$ than $\|T\|$.

Proposition 13.27 *Let $T \in \mathscr{L}(H)$ be a bounded linear operator. Then the spectral radius of T is*

$$r(T) = \lim_{n\to\infty} \|T^n\|^{\frac{1}{n}}. \tag{13.11}$$

The proof is in the appendix.

13.2.6 The Spectrum of Self-Adjoint Operators

Proposition 13.28 *Let $T = T^*$ be a self-adjoint operator in $\mathscr{L}(H)$. Then $\lambda \in \rho(T)$ if and only if there exists $C > 0$ such that*

$$\|(T-\lambda I)x\| \geq C\,\|x\|\,.$$

Proof. Suppose that $\lambda \in \rho(T)$. Then there exists (a non-zero) bounded operator $R_\lambda(T)$ such that

$$\|x\| = \|R_\lambda(T)(T-\lambda I)x\| \leq \|R_\lambda(T)\|\,\|(T-\lambda I)x\|\,.$$

Hence,

$$\|(T-\lambda I)x\| \geq \frac{1}{\|R_\lambda(T)\|}\,\|x\|\,, \quad x \in H.$$

We can choose $C = 1/\,\|R_\lambda(T)\|$ and the condition of the proposition is satisfied.
Suppose, the condition is satisfied. We prove the other direction in 3 steps, i. e. $T - \lambda_0 I$ has a bounded inverse operator which is defined on the whole space H.
Step 1. $T - \lambda I$ is injective. Suppose to the contrary that $(T-\lambda)x_1 = (T-\lambda)x_2$. Then

$$0 = \|(T-\lambda)(x_1 - x_2)\| \geq C\,\|x_1 - x_2\|\,,$$

and $\|x_1 - x_2\| = 0$ follows. That is $x_1 = x_2$. Hence, $T - \lambda I$ is injective.
Step 2. $H_1 = (T-\lambda I)H$, the range of $T - \lambda I$ is closed. Suppose that $y_n = (T-\lambda I)x_n$, $x_n \in H$, converges to some $y \in H$. We want to show that $y \in H_1$. Clearly (y_n) is a Cauchy sequence such that $\|y_m - y_n\| \to 0$ as $m, n \to \infty$. By assumption,

$$\|y_m - y_n\| = \|(T-\lambda I)(x_n - x_m)\| \geq C\,\|x_n - x_m\|\,.$$

Thus, (x_n) is a Cauchy sequence in H. Since H is complete, $x_n \to x$ for some $x \in H$. Since $T - \lambda I$ is continuous,

$$y_n = (T-\lambda I)x_n \underset{n\to\infty}{\longrightarrow} (T-\lambda I)x.$$

Hence, $y = (T - \lambda I)x$ and H_1 is a closed subspace.
Step 3. $H_1 = H$. By Riesz first theorem, $H = H_1 \oplus H_1^\perp$. We have to show that $H_1^\perp = \{0\}$. Let $u \in H_1^\perp$, that is, since $T^* = T$,

$$0 = \langle (T - \lambda I)x \,,\, u\rangle = \left\langle x \,,\, (T - \overline{\lambda} I)u \right\rangle, \quad \text{for all} \quad x \in H.$$

This shows $(T - \overline{\lambda} I)u = 0$, hence $T(u) = \overline{\lambda} u$. This implies

$$\langle T(u) \,,\, u\rangle = \overline{\lambda} \langle u \,,\, u\rangle .$$

However, $T = T^*$ implies that the left side is real, by Remark 13.2 (d). Hence $\overline{\lambda} = \lambda$ is real. We conclude, $(T - \lambda I)u = 0$. By injectivity of $T - \lambda I$, $u = 0$. That is $H_1 = H$.
We have shown that there exists a linear operator $S = (T - \lambda I)^{-1}$ which is inverse to $T - \lambda I$ and defined on the whole space H. Since

$$\|y\| = \|(T - \lambda I)S(y)\| \geq C \|S(y)\| ,$$

S is bounded with $\|S\| \leq 1/C$. Hence, $S = R_\lambda(T)$. ∎

Note that for any bounded real function $f(x, y)$ we have

$$\sup_{x,y} f(x, y) = \sup_x (\sup_y f(x, y)) = \sup_y (\sup_x f(x, y)).$$

In particular, $\|x\| = \sup_{\|y\|\leq 1} |\langle x \,,\, y\rangle|$ since $y = x/\|x\|$ yields the supremum and CSI gives the upper bound. Further, $\|T(x)\| = \sup_{\|y\|\leq 1} |\langle T(x) \,,\, y\rangle|$ such that

$$\|T\| = \sup_{\|x\|\leq 1} \sup_{\|y\|\leq 1} |\langle T(x) \,,\, y\rangle| = \sup_{\|x\|\leq 1,\ \|y\|\leq 1} |\langle T(x) \,,\, y\rangle| = \sup_{\|y\|\leq 1} \sup_{\|x\|\leq 1} |\langle T(x) \,,\, y\rangle|$$

In case of self-adjoint operators we can generalize this.

Proposition 13.29 *Let $T = T^* \in \mathscr{L}(H)$. Then we have*

$$\|T\| = \sup_{\|x\|\leq 1} |\langle T(x) \,,\, x\rangle| . \tag{13.12}$$

Proof. Let $C = \sup_{\|x\|\leq 1} |\langle T(x) \,,\, x\rangle|$. By Cauchy–Schwarz inequality, $|\langle T(x) \,,\, x\rangle| \leq \|T\| \|x\|^2$ such that $C \leq \|T\|$.
For any real positive $\alpha > 0$ we have:

$$\begin{aligned}
\|T(x)\|^2 = \langle T(x) \,,\, T(x)\rangle = \left\langle T^2(x) \,,\, x\right\rangle &= \frac{1}{4}\left(\left\langle T(\alpha x + \alpha^{-1}T(x)) \,,\, \alpha x + \alpha^{-1}T(x)\right\rangle - \right. \\
&\quad \left. - \left\langle T(\alpha x - \alpha^{-1}T(x)) \,,\, \alpha x - \alpha^{-1}T(x)\right\rangle\right) \\
&\leq \frac{1}{4}\left(C \left\|\alpha x + \alpha^{-1}T(x)\right\|^2 + C \left\|\alpha x - \alpha^{-1}T(x)\right\|^2\right) \\
&\underset{\text{P.I.}}{=} \frac{C}{4}\left(2\|\alpha x\|^2 + 2\left\|\alpha^{-1}T(x)\right\|^2\right) = \frac{C}{2}\left(\alpha^2 \|x\|^2 + \alpha^{-2}\|T(x)\|^2\right) .
\end{aligned}$$

Inserting $\alpha^2 = \|T(x)\| \,/\, \|x\|$ we obtain

$$= \frac{C}{2} \left(\|T(x)\| \, \|x\| + \|x\| \, \|T(x)\| \right)$$

which implies $\|T(x)\| \leq C \, \|x\|$. Thus, $\|T\| = C$. ∎

Let $m = \inf_{\|x\|=1} \langle T(x) \, , \, x \rangle$ and $M = \sup_{\|x\|=1} \langle T(x) \, , \, x \rangle$ denote the *lower* and *upper* bound of T. Then we have

$$\sup_{\|x\| \leq 1} | \, \langle T(x) \, , \, x \rangle \, | = \max\{ | \, m \, | \, , M \} = \|T\| \, ,$$

and

$$m \, \|x\|^2 \leq \langle T(x) \, , \, x \rangle \leq M \, \|x\|^2 \, , \quad \text{for all} \quad x \in H.$$

Corollary 13.30 *Let $T = T^* \in \mathscr{L}(H)$ be a self-adjoint operator. Then*

$$\sigma(T) \subset [m, M].$$

Proof. Suppose that $\lambda_0 \notin [m, M]$. Then

$$C := \inf_{\mu \in [m,M]} | \, \lambda_0 - \mu \, | > 0.$$

Since $m = \inf_{\|x\|=1} \langle T(x) \, , \, x \rangle$ and $M = \sup_{\|x\|=1} \langle T(x) \, , \, x \rangle$ we have for $\|x\| = 1$

$$\|(T - \lambda_0 I)x\| = \|x\| \, \|(T - \lambda_0 I)x\| \underset{\text{CSI}}{\geq} | \, \langle (T - \lambda_0 I)x \, , \, x \rangle \, | = \left| \underbrace{\langle T(x) \, , \, x \rangle}_{\in [m,M]} - \lambda_0 \underbrace{\|x\|^2}_{1} \right| \geq C.$$

This implies

$$\|(T - \lambda_0 I)x\| \geq C \, \|x\| \quad \text{for all} \quad x \in H.$$

By Proposition 13.28, $\lambda_0 \in \rho(T)$. ∎

Example 13.13 (a) Let $H = \mathrm{L}^2[0,1]$, $g \in \mathrm{C}[0,1]$ a real-valued function, and $(T_g f)(t) = g(t) f(t)$. Let $m = \inf_{t \in [0,1]} g(t)$, $M = \sup_{t \in [0,1]} g(t)$. One proves that m and M are the lower and upper bounds of T_g such that $\sigma(T_g) \subseteq [m, M]$. Since g is continuous, by the intermediate value theorem, $\sigma(T_g) = [m, M]$.
(b) Let $T = T^* \in \mathscr{L}(H)$ be self-adjoint. Then all eigenvalues of T are real and eigenvectors to different eigenvalues are orthogonal to each other. *Proof.* The first statement is clear from Corollary 13.30. Suppose that $T(x) = \lambda x$ and $T(y) = \mu y$ with $\lambda \neq \mu$. Then

$$\lambda \, \langle x \, , \, y \rangle = \langle T(x) \, , \, y \rangle = \langle x \, , \, T(y) \rangle = \overline{\mu} \, \langle x \, , \, y \rangle = \mu \, \langle x \, , \, y \rangle \, .$$

Since $\lambda \neq \mu$, $\langle x \, , \, y \rangle = 0$. ∎

The statement about orthogonality holds for arbitrary normal operators.

Appendix: Compact Self-Adjoint Operator in Hilbert Space

Proof of Proposition 13.27. From the theory of power series, Theorem 2.34 we know that the series

$$-z\sum_{n=0}^{\infty}\|T^n\|\ z^n \tag{13.13}$$

converges if $|\,z\,|<R$ and diverges if $|\,z\,|>R$, where

$$R=\frac{1}{\overline{\lim\limits_{n\to\infty}}\ \sqrt[n]{\|T^n\|}}. \tag{13.14}$$

Inserting $z=1/\lambda$ and using homework 38.4, we have

$$-\sum_{n=0}^{\infty}\lambda^{-n-1}T^n$$

diverges if $|\,\lambda\,|<\overline{\lim\limits_{n\to\infty}}\ \sqrt[n]{\|T^n\|}$ (and converges if $|\,\lambda\,|>\overline{\lim\limits_{n\to\infty}}\ \sqrt[n]{\|T^n\|}$). The reason for the divergence of the power series is, that the spectrum $\sigma(T)$ and the circle with radius $\overline{\lim\limits_{n\to\infty}}\ \sqrt[n]{\|T^n\|}$ have points in common; hence

$$r(T)=\overline{\lim_{n\to\infty}}\ \sqrt[n]{\|T^n\|}.$$

On the other hand, by Remark 13.4 (d), $\lambda\in\sigma(T)$ implies $\lambda^n\in\sigma(T^n)$; hence, by Remark 13.4 (c),

$$|\,\lambda^n\,|\le\|T^n\|\implies|\,\lambda\,|\le\sqrt[n]{\|T^n\|}.$$

Taking the supremum over all $\lambda\in\sigma(T)$ on the left and the $\underline{\lim}$ over all n on the right, we have

$$r(T)\le\underline{\lim_{n\to\infty}}\ \sqrt[n]{\|T^n\|}\le\overline{\lim_{n\to\infty}}\ \sqrt[n]{\|T^n\|}=r(T).$$

Hence, the sequence $\sqrt[n]{\|T^n\|}$ converges to $r(T)$ as n tends to ∞. ∎

Compact operators generalize finite rank operators. Integral operators on compact sets are compact.

Definition 13.16 A linear operator $T\in\mathscr{L}(H)$ is called *compact* if the closure $\overline{T(U_1)}$ of the unit ball $U_1=\{x\mid\|x\|\le 1\}$ is compact in H. In other words, for every sequence (x_n), $x_n\in U_1$, there exists a subsequence such that $T(x_{n_k})$ converges.

Proposition 13.31 *For $T\in\mathscr{L}(H)$ the following are equivalent:*

(a) *T is compact.*
(b) *T^* is compact.*
(c) *For all sequences (x_n) with $(\langle x_n\,,\,y\rangle)\longrightarrow\langle x\,,\,y\rangle$ converges for all y we have $T(x_n)\longrightarrow T(x)$.*
(d) *There exists a sequence (T_n) of operators of finite rank such that $\|T-T_n\|\longrightarrow 0$.*

Definition 13.17 Let T be an operator on H and H_1 a closed subspace of H. We call H_1 an *reducing subspace* if both H_1 and $H_1^\perp$ are T-invariant, i. e. $T(H_1) \subset H_1$ and $T(H_1^\perp) \subset H_1^\perp$.

Proposition 13.32 *Let $T \in \mathscr{L}(H)$ be normal.*
(a) *The eigenspace $\ker(T-\lambda I)$ is a reducing subspace for T and $\ker(T-\lambda I) = \ker(T-\lambda I)^*$.*
(b) *If λ, μ are distinct eigenvalues of T, $\ker(T-\lambda I) \perp \ker(T-\mu I)$.*

Proof. (a) Since T is normal, so is $T-\lambda$. Hence $\|(T-\lambda)(x)\| = \|(T-\lambda)^*(x)\|$. Thus, $\ker(T-\lambda) = \ker(T-\lambda)^*$. In particular, $T^*(x) = \overline{\lambda}x$ if $x \in \ker(T-\lambda I)$.
We show invariance. Let $x \in \ker(T-\lambda)$; then $T(x) = \lambda x \in \ker(T-\alpha I)$. Similarly, $x \in \ker(T-\lambda I)^\perp$, $y \in \ker(T-\lambda I)$ imply

$$\langle T(x)\,,\, y\rangle = \langle x\,,\, T^*(y)\rangle = \langle x\,,\, \overline{\lambda}y\rangle = 0.$$

Hence, $\ker(T-\lambda I)^\perp$ is T-invariant, too.
(b) Let $T(x) = \lambda x$ and $T(y) = \mu y$. Then (a) and $T^*(y) = \overline{\mu}y$ ... imply

$$\lambda\,\langle x\,,\, y\rangle = \langle T(x)\,,\, y\rangle = \langle x\,,\, T^*(y)\rangle = \langle x\,,\, \overline{\mu}y\rangle = \mu\,\langle x\,,\, y\rangle\,.$$

Thus $(\lambda-\mu)\,\langle x\,,\, y\rangle = 0$; since $\lambda \neq \mu$, $x \perp y$. ∎

Theorem 13.33 (Spectral Theorem for Compact Self-Adjoint Operators) *Let H be an infinite dimensional separable Hilbert space and $T \in \mathscr{L}(H)$ compact and self-adjoint.*
Then there exists a real sequence (λ_n) with $\lambda_n \underset{n\to\infty}{\longrightarrow} 0$ and an CNOS $\{\mathrm{e}_n \mid n \in \mathbb{N}\} \cup \{f_k \mid k \in N \subset \mathbb{N}\}$ such that

$$T(\mathrm{e}_n) = \lambda_n \mathrm{e}_n, \quad n \in \mathbb{N} \qquad T(f_k) = 0, \quad k \in N.$$

Moreover,

$$T(x) = \sum_{n=1}^{\infty} \lambda_n\,\langle x\,,\, \mathrm{e}_n\rangle\,\mathrm{e}_n, \quad x \in H. \tag{13.15}$$

Remarks 13.5 (a) Since $\{\mathrm{e}_n\} \cup \{f_k\}$ is a CNOS, any $x \in H$ can be written as its Fourier series

$$x = \sum_{n=1}^{\infty} \langle x\,,\, \mathrm{e}_n\rangle\,\mathrm{e}_n + \sum_{k\in N} \langle x\,,\, f_k\rangle\, f_k.$$

Applying T using $T(\mathrm{e}_n) = \lambda_n \mathrm{e}_n$ we have

$$T(x) = \sum_{n=1}^{\infty} \langle x\,,\, \mathrm{e}_n\rangle\,\lambda_n\mathrm{e}_n + \sum_{k\in N} \langle x\,,\, f_k\rangle \underbrace{T(f_k)}_{=0}$$

which establishes (13.15). The main point is the existence of a CNOS of eigenvectors $\{\mathrm{e}_n\} \cup \{f_k\}$.
(b) In case $H = \mathbb{C}^n$ ($\mathbb{R}^n$) the theorem says that any hermitean (symmetric) matrix A is diagonalizable with only real eigenvalues.

Chapter 14

Complex Analysis

Here are some useful textbooks on Complex Analysis: [FL88] (in German), [Kno78] (in German), [Nee97], [Rüh83] (in German), [Hen88].
The main part of this chapter deals with holomorphic functions which is another name for a function which is complex differentiable in an open set. On the one hand, we are already familiar with a huge class of holomorphic functions: polynomials, the exponential function, sine and cosine functions. On the other hand holomorphic functions possess quite amazing properties completely unusual from the vie point of *real* analysis. The properties are very strong. For example, it is easy to construct a real function which is 17 times differentiable but not 18 times. A complex differentiable function (in a small region) is automatically infinitely often differentiable.
Good references are Ahlfors [Ahl78], a little harder is Conway [Con78], easier is Howie [How03].

14.1 Holomorphic Functions

14.1.1 Complex Differentiation

We start with some notations.

U_r	$\{z \mid \lvert z \rvert < r\}$	open ball of radius r around 0
$U_R(a)$	$\{z \mid \lvert z-a \rvert < R\}$	open ball of radius R around a
$\overline{U_r}$	$\{z \mid \lvert z \rvert \leq r\}$	closed ball of radius r around 0
$\overset{\circ}{U}_r$	$\{z \mid 0 < \lvert z \rvert < r\}$	punctured ball of radius r
S_r	$\{z \mid \lvert z \rvert = r\}$	circle of radius r around 0

Definition 14.1 Let $U \subset \mathbb{C}$ be an open subset of $\mathbb{C}$ and $f\colon U \to \mathbb{C}$ be a complex function.
(a) If $z_0 \in U$ and the limit

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} =: f'(z_0)$$

exists, we call f *complex differentiable* at z_0 and $f'(z_0)$ the derivative of f at z_0. We call $f'(z_0)$ the *derivative* of f at z_0.

(b) If f is complex differentiable for every $z_0 \in U$, we say that f is *holomorphic* in U. We call f' the derivative of f on U.
(c) f is *holomorphic* at z_0 if it complex differentiable in a certain neighborhood of z_0.

To be quite explicit, $f'(z_0)$ exists if to every $\varepsilon > 0$ there exists some $\delta > 0$ such that $z \in U_\delta(z_0)$ implies

$$\left| \frac{f(z) - f(z_0)}{z - z_0} - f'(z_0) \right| < \varepsilon.$$

Remarks 14.1 (a) Differentiability of f at z_0 forces f to be continuous at z_0. Indeed, f is differentiable at z_0 with derivative $f'(z_0)$ if and only if, there exists a function $r(z, z_0)$ such that

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + (z - z_0)r(z, z_0),$$

where $\lim_{z \to z_0} r(z, z_0) = 0$, In particular, taking the limit $z \to z_0$ in the above equation we get

$$\lim_{z \to z_0} f(z) = f(z_0),$$

which proves continuity at z_0.
Complex conjugation is a uniformly continuous function on $\mathbb{C}$ since $|\,\overline{z} - \overline{z}_0\,| = |\,z - z_0\,|$ for all $z, z_0 \in \mathbb{C}$.
(b) The derivative satisfies the well-known sum, product, and quotient rules, that is, if both f and g are holomorphic in U, so are $f + g$, fg, and f/g, provided $g \neq 0$ in U and we have

$$(f + g)' = f' + g', \quad (fg)' = f'g + fg', \quad \left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}.$$

Also, the chain rule holds; if $U \xrightarrow{f} V \xrightarrow{g} \mathbb{C}$ are holomorphic, so is $g \circ f$ and

$$(g \circ f)'(z) = g'(f(z))\, f'(z).$$

The proofs are exactly the same as in the real case. Since the constant functions $f(z) = c$ and the identity $f(z) = z$ are holomorphic in $\mathbb{C}$, so is every polynomial with complex coefficients and, moreover, every rational function (quotient of two polynomials) $f\colon U \to \mathbb{C}$, provided the denominator has no zeros in U. So, we already know a large class of holomorphic functions. Another bigger class are the convergent power series.

Example 14.1 $f(z) = |\,z\,|^2$ is complex differentiable at 0 with $f'(0) = 0$. f is not differentiable at $z_0 = 1$. Indeed,

$$\lim_{h \to 0} \frac{f(h + 0) - f(0)}{h} = \lim_{h \to 0} \frac{|\,h\,|^2}{h} = \lim_{h \to 0} \overline{h} = 0.$$

On the other hand. Let $\varepsilon \in \mathbb{R}$

$$\lim_{\varepsilon \to 0} \frac{|\,1 + \varepsilon\,|^2 - 1}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{2\varepsilon + \varepsilon^2}{\varepsilon} = 2$$

whereas

$$\lim_{\varepsilon \to 0} \frac{|\,1 + \mathrm{i}\varepsilon\,|^2 - 1}{\mathrm{i}\varepsilon} = \lim_{\varepsilon \to 0} \frac{1 + \varepsilon^2 - 1}{\mathrm{i}\varepsilon} = 0.$$

This shows that $f'(1)$ does not exist.

14.1.2 Power Series

Recall from Subsection 2.3.9 that a power series $\sum c_n z^n$ has a radius of convergence

$$R = \frac{1}{\overline{\lim\limits_{n\to\infty}} \sqrt[n]{|\, c_n \,|}}.$$

That is, the series converges absolutely for all z with $|\, z \,| < R$; the series diverges for $|\, z \,| > R$, the behaviour for $|\, z \,| = R$ depends on the (c_n). Moreover, it converges uniformly on every closed ball $\overline{U_r}$ with $0 < r < R$, see Proposition 6.4.
We already know that a real power series can be differentiated elementwise, see Corollary 6.11. We will see, that power series are holomorphic inside its radius of convergence.

Proposition 14.1 *Let $a \in \mathbb{C}$ and*

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n \tag{14.1}$$

be a power series with radius of convergence R. Then $f\colon U_R(a) \to \mathbb{C}$ is holomorphic and the derivative is

$$f'(z) = \sum_{n=1}^{\infty} n c_n (z-a)^{n-1}. \tag{14.2}$$

Proof. If the series (14.1) converges in $U_R(a)$, the root test shows that the series (14.2) also converges there. Without loss of generality, take $a = 0$. Denote the sum of the series (14.2) by $g(z)$, fix $w \in U_R(0)$ and choose r so that $|\, w \,| < r < R$. If $z \neq w$, we have

$$\frac{f(z) - f(w)}{z - w} - g(w) = \sum_{n=0}^{\infty} c_n \left(\frac{z^n - w^n}{z - w} - n w^{n-1} \right).$$

The expression in the brackets is 0 if $n = 1$. For $n \geq 2$ it is (by direct computation of the following term)

$$= (z-w) \sum_{k=1}^{n-1} k\, w^{k-1}\, z^{n-k-1} = \sum_{k=1}^{n-1} \left(k w^{k-1}\, z^{n-k} - k w^k z^{n-k-1} \right), \tag{14.3}$$

which gives a telescope sum if we shift $k := k+1$ in the first summand. If $|\, z \,| < r$, the absolute value of the sum (14.3) is less than

$$\frac{n(n-1)}{2} r^{n-2},$$

so

$$\left| \frac{f(z) - f(w)}{z - w} - g(w) \right| \leq |\, z - w \,| \sum_{n=2}^{\infty} n^2\, |\, c_n \,|\, r^{n-2}. \tag{14.4}$$

Since $r < R$, the last series converges. Hence the left side of (14.4) tends to 0 as $z \to w$. This says that $f'(w) = g(w)$, and completes the proof. ∎

Corollary 14.2 *Since $f'(z)$ is again a power series with the same radius of convergence R, the proposition can be applied to $f'(z)$. It follows that f has derivatives of all orders and that each derivative has a power series expansion around a*

$$f^{(k)}(z) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)c_n\,(z-a)^{n-k} \tag{14.5}$$

Inserting $z = a$ implies

$$f^{(k)}(a) = k!c_k, \quad k = 0, 1, \ldots.$$

This shows that the coefficients c_n in the power series expansion $f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n$ of f with midpoint a are unique.

Example 14.2 The exponential function $\mathrm{e}^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ is holomorphic on the whole complex plane with $(\mathrm{e}^z)' = \mathrm{e}^z$; similarly, the trigonometric functions $\sin z$ and $\cos z$ are holomorphic in $\mathbb{C}$ since

$$\sin z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}, \quad \cos z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}.$$

We have $(\sin z)' = \cos z$ and $(\cos z)' = -\sin z$.

Definition 14.2 A complex function which is defined on $\mathbb{C}$ and which is holomorphic on the entire complex plane is called an *entire* function.

14.1.3 Cauchy–Riemann Equations

Let us identify the complex field $\mathbb{C}$ and the two dimensional real plane $\mathbb{R}^2$ via $z = x + \mathrm{i}y$, that is, every complex number z corresponds to an ordered pair (x, y) of real numbers. In this way, a complex function $w = f(z)$ corresponds to a function $U \to \mathbb{R}^2$ where $U \subset \mathbb{C}$ is open. We have $w = u + \mathrm{i}v$ where $u = u(x, y)$ and $v = v(x, y)$ are the real and the imaginary parts of the function f; $u = \operatorname{Re} w$ and $v = \operatorname{Im} w$. Problem: What is the relation between complex differentiability and the differentiability of f as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$?

Proposition 14.3 *Let*

$$f\colon U \to \mathbb{C}, \quad U \subset \mathbb{C} \quad \textit{open}, \quad a \in U$$

be a function. Then the following are equivalent:

(a) *f is complex differentiable at a.*

(b) *$f(x, y) = u(x, y) + \mathrm{i}v(x, y)$ is real differentiable at a as a function $f\colon U \subset \mathbb{R}^2 \to \mathbb{R}^2$, and the Cauchy–Riemann equations are satisfied at a:*

$$\frac{\partial u}{\partial x}(a) = \frac{\partial v}{\partial y}(a), \qquad \frac{\partial u}{\partial y}(a) = -\frac{\partial v}{\partial x}(a).$$
$$u_x = v_y, \qquad u_y = -v_x. \tag{14.6}$$

In this case,

$$f' = u_x + \mathrm{i}v_x = v_y - \mathrm{i}u_y.$$

Proof. (a) $\to$ (b): Suppose that $z = h + \mathrm{i}k$ is a complex number such that $a + z \in U$; put $f'(a) = b_1 + \mathrm{i}b_2$. By assumption,

$$\lim_{z\to 0} \frac{|\, f(a+z) - f(a) - zf'(a)\,|}{|\, z\,|} = 0.$$

We shall write this in the real form with real variables h and k. Note that

$$\begin{aligned} zf'(a) &= (h+\mathrm{i}k)(b_1+\mathrm{i}b_2) = hb_1 - kb_2 + \mathrm{i}(hb_2 + kb_1) \\ &= \begin{pmatrix} hb_1 - kb_2 \\ hb_2 + kb_2 \end{pmatrix} = \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix}. \end{aligned}$$

This implies, with the identification $z = (h, k)$,

$$\lim_{z\to 0} \frac{\left\| f(a+z) - f(a) - \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} \right\|}{|\, z\,|} = 0.$$

That is (see Subsection 7.2), f is real differentiable at a with the Jacobian matrix

$$f'(a) = Df(a) = \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix}. \tag{14.7}$$

By Proposition 7.6, the Jacobian matrix is exactly the matrix of the partial derivatives, that is

$$Df(a) = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}.$$

Comparing this with (14.7), we obtain $u_x(a) = v_y(a) = \operatorname{Re} f'(a)$ and $u_y(a) = -v_x(a) = \operatorname{Im} f'(a)$. This completes the proof of the first direction.
(b) $\to$ (a). Since $f = (u, v)$ is differentiable at $a \in U$ as a real function, there exists a linear mapping $Df(a) \in \mathscr{L}(\mathbb{R}^2)$ such that

$$\lim_{(h,k)\to 0} \frac{\left\| f(a+(h,k)) - f(a) - Df(a) \begin{pmatrix} h \\ k \end{pmatrix} \right\|}{\|(h,k)\|} = 0.$$

By Proposition 7.6,

$$Df(a) = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}.$$

The Cauchy–Riemann equations show that Df takes the form

$$Df(a) = \begin{pmatrix} b_1 & -b_2 \\ b_2 & b_1 \end{pmatrix},$$

where $u_x = b_1$ and $v_x = b_2$. Writing

$$Df(a)\begin{pmatrix} h \\ k \end{pmatrix} = \begin{pmatrix} hb_1 - kb_2 \\ hb_2 + kb_2 \end{pmatrix} = z(b_1 + \mathrm{i}b_2)$$

in the complex form with $z = h + \mathrm{i}k$ gives f is complex differentiable at a with $f'(a) = b_1 + \mathrm{i}b_2$. ■

Example 14.3 (a) We already know that $f(z) = z^2$ is complex differentiable. Hence, the Cauchy–Riemann equations must be fulfilled. From

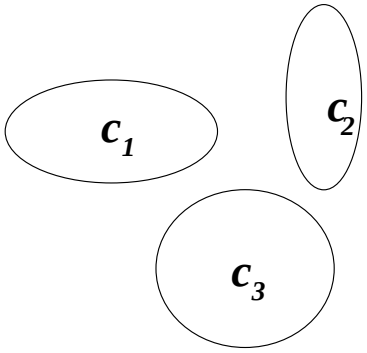
$$f(z) = z^2 = (x + \mathrm{i}y)^2 = x^2 - y^2 + 2\mathrm{i}xy, \quad u(x, y) = x^2 - y^2, \quad v(x, y) = 2xy$$

we conclude

$$u_x = 2x, \quad u_y = -2y, \quad v_x = 2y, \quad v_y = 2x.$$

The Cauchy–Riemann equations are satisfied.
(b) $f(z) = |\, z\,|^2$. Since $f(z) = x^2 + y^2$, $u(x, y) = x^2 + y^2$, $v(x, y) = 0$. The Cauchy-Riemann equations yield $u_x = 2x = 0 = v_y$ and $u_y = 2y = 0 = -v_x$ such that $z = 0$ is the only solution of the CRE. $z = 0$ is the only point where f is differentiable.
$f(z) = \overline{z}$ is nowhere differentiable since $u(x, y) = x$, $v(x, y) = -y$; thus

$$1 = u_x \neq v_y = -1.$$



A function $f\colon U \to \mathbb{C}$, $U \subset \mathbb{C}$ open, is called *locally constant in* U, if for every point $a \in U$ there exists a ball V with $a \in V \subset U$ such that f is constant on V.
Clearly, on every connectedness component of U, f is constant. In fact, one can *define* U to be connected if for every holomorphic $f\colon U \to \mathbb{C}$, f is constant.

Corollary 14.4 *Let $U \subset \mathbb{C}$ be open and $f\colon U \to \mathbb{C}$ be a holomorphic function on U.*
(a) *If $f'(z) = 0$ for all $z \in U$, then f is locally constant in U.*
(b) *If f takes real values only, then f is locally constant.*
(c) *If f has a continuous second derivative, $u = \operatorname{Re} f$ and $v = \operatorname{Im} f$ are harmonic functions, i. e., they satisfy the Laplace equation $\Delta(u) = u_{xx} + u_{yy} = 0$ and $\Delta(v) = 0$.*

Proof. (a) Since $f'(z) = 0$ for all $z \in U$, the Cauchy–Riemann equations imply $u_x = u_y = v_x = v_y = 0$ in U. From real analysis, it is known that u and v are locally constant in U (apply Corollary 7.12 with $\operatorname{grad} f(a + \theta x) = 0$).
(b) Since f takes only real values, $v(x, y) = 0$ for all $(x, y) \in U$. This implies $v_x = v_y = 0$ on U. By the Cauchy–Riemann equations, $u_x = u_y = 0$ and f is locally constant by (a).

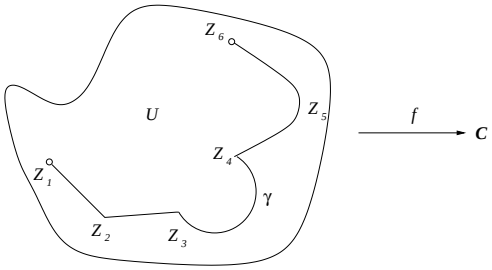
(c) $u_x = v_y$ implies $u_{xx} = v_{yx}$ and differentiating $u_y = -v_x$ with respect to y yields $u_{yy} = -v_{xy}$. Since both u and v are twice continuously differentiable (since so is f), by Schwarz' Lemma, the sum is $u_{xx} + u_{yy} = v_{yx} - v_{xy} = 0$. The same argument works for $v_{xx} + v_{yy} = 0$. ∎

Remarks 14.2 (a) We will see soon that the additional differentiability assumption in (c) is superfluous.
(b) Note, that an inverse statement to (c) is easily proved: If $Q = (a, b) \times (c, d)$ is an open rectangle and $u\colon Q \to \mathbb{R}$ is harmonic, then there exists a holomorphic function $f\colon Q \to \mathbb{C}$ such that $u = \operatorname{Re} f$.

14.2 Cauchy's Integral Formula

14.2.1 Integration

The major objective of this section is to prove the converse to Proposition 14.1: Every in D holomorphic function is representable as a power series in D. The quickest route to this is via Cauchy's Theorem and Cauchy's Integral Formula. The required integration theory will be developed. It is a useful tool to study holomorphic functions.
Recall from Section 5.4 the definition of the Riemann integral of a bounded complex valued function $\varphi\colon [a, b] \to \mathbb{C}$. It was defined by integrating both the real and the imaginary parts of φ. In what follows, a *path* is always a piecewise continuously differentiable curve.



Definition 14.3 Let $U \subset \mathbb{C}$ be open and $f\colon U \to \mathbb{C}$ a continuous function on U. Suppose that $\gamma\colon [t_1, t_n] \to U$ is a path in U. The *integral of f along γ* is defined as the line integral

$$\int_\gamma f(z)\,\mathrm{d}z := \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} f(\gamma(t))\gamma'(t)\,\mathrm{d}t, \quad (14.8)$$

where γ is continuously differentiable on $[t_{k-1}, t_k]$ for all $k = 1, \ldots, n$.

By the change of variable rule, the integral of f along γ does not depend on the parametrization γ of the path $\{\gamma(t) \mid t \in [t_0, t_1]\}$. However, if we exchange the initial and the end point of $\gamma(t)$, we obtain a negative sign.

Remarks 14.3 (Properties of the complex integral) (a) The integral of f along γ is linear over $\mathbb{C}$:

$$\int_\gamma (\alpha f_1 + \beta f_2)\,\mathrm{d}z = \alpha \int_\gamma f_1\,\mathrm{d}z + \beta \int_\gamma f_2\,\mathrm{d}z, \quad \int_{\gamma_-} f(z)\,\mathrm{d}z = -\int_{\gamma_+} f(z)\,\mathrm{d}z,$$

where γ_- has the opposite orientation of γ_+.
(b) If γ_1 and γ_2 are two paths so that γ_1 and γ_2 join to form γ, then we have

$$\int_\gamma f(z)\,\mathrm{d}z = \int_{\gamma_1} f(z)\,\mathrm{d}z + \int_{\gamma_2} f(z)\,\mathrm{d}z.$$

(c) From the definition and the triangle inequality, it follows that for a continuously differentiable path γ

$$\left| \int_\gamma f(z)\,\mathrm{d}z \right| \le M\,\ell,$$

where $|\,f(z)\,| \le M$ for all $z \in \gamma$ and ℓ is the length of γ, $\ell = \int_a^b |\,\gamma'(t)\,|\,\mathrm{d}t$. $t \in [t_0, t_1]$. Note that the integral on the right is the length of the curve $\gamma(t)$.
(d) The integral of f over γ generalizes the real integral $\int_a^b f(t)\,\mathrm{d}t$. Indeed, let $\gamma(t) = t$, $t \in [a, b]$, then

$$\int_\gamma f(z)\,\mathrm{d}z = \int_a^b f(t)\,\mathrm{d}t.$$

(e) Let γ be the circle $\mathrm{S}_r(a)$ of radius r with center a. We can parametrize the *positively oriented circle* as $\gamma(t) = a + r\mathrm{e}^{\mathrm{i}t}$, $t \in [0, 2\pi]$. Then

$$\int_\gamma f(z)\,\mathrm{d}z = \mathrm{i}r \int_0^{2\pi} f\left(a + r\mathrm{e}^{\mathrm{i}t}\right) \mathrm{e}^{\mathrm{i}t}\,\mathrm{d}t.$$

Example 14.4 (a) Let $\gamma_1(t) = \mathrm{e}^{\mathrm{i}t}$, $t \in [0, \pi]$, be the half of the unit circle from 1 to -1 via i and $\gamma_2(t) = -t$, $t \in [-1, 1]$ the segment from 1 to -1. Then $\gamma_1'(t) = \mathrm{i}\mathrm{e}^{\mathrm{i}t}$ and $\gamma_2(t)' = -1$. Hence,

$$\begin{aligned}
\int_{\gamma_1} \overline{z}^2\,\mathrm{d}z &= \mathrm{i}\int_0^\pi \overline{\mathrm{e}^{2\mathrm{i}t}}\mathrm{e}^{\mathrm{i}t}\,\mathrm{d}t = \mathrm{i}\int_0^\pi \mathrm{e}^{-2\mathrm{i}t+\mathrm{i}t}\,\mathrm{d}t = \mathrm{i}\int_0^\pi \mathrm{e}^{-\mathrm{i}t}\,\mathrm{d}t \\
&= \frac{\mathrm{i}}{-\mathrm{i}}\mathrm{e}^{-\mathrm{i}t}\Big|_0^\pi = -(-1-1) = 2. \\
\int_{\gamma_2} \overline{z}^2\,\mathrm{d}z &\underset{\text{see (b)}}{=} -\int_{-1}^1 t^2\,\mathrm{d}t = -\frac{2}{3}.
\end{aligned}$$

In particular, the integral of $\overline{z}^2$ is not path independent.
(b)

$$\int_{\mathrm{S}_r} \frac{\mathrm{d}z}{z^n} = \int_0^{2\pi} \frac{\mathrm{i}r\mathrm{e}^{\mathrm{i}t}}{r^n\mathrm{e}^{\mathrm{i}nt}}\,\mathrm{d}t = \mathrm{i}r^{-n+1}\int_0^{2\pi} \mathrm{e}^{-(n-1)\mathrm{i}t}\,\mathrm{d}t = \begin{cases} 0, & n \ne 1, \\ 2\pi\mathrm{i}, & n = 1. \end{cases}$$

14.2.2 Cauchy's Theorem

Cauchy's theorem is the main part in the proof that every in a holomorphic function can be written as a power series with midpoint a. As a consequence of Corollary 14.2, holomorphic functions have derivatives of all orders.
We start with a very weak form. The additional assumption is, that f has an antiderivative.

Lemma 14.5 *Let $f\colon U \to \mathbb{C}$ be continuous, and suppose that f has an antiderivative F which is holomorphic on U, $F' = f$. If γ is any path in U joining z_0 and z_1 from U, we have*

$$\int_\gamma f(z)\,\mathrm{d}z = F(z_2) - F(z_1).$$

In particular, if γ is a closed path in U

$$\int_\gamma f(z)\,\mathrm{d}z = 0.$$

Proof. It suffices to prove the statement for a continuously differentiable curve $\gamma(t)$. Put $h(t) = F(\gamma(t))$. By the chain rule

$$h'(t) = \frac{\mathrm{d}}{\mathrm{d}t}F(\gamma(t)) = F'(\gamma(t))\gamma'(t) = f(\gamma(t))\gamma'(t)$$

By definition of the integral and the fundamental theorem of calculus (see Subsection 5.5),

$$\int_\gamma f(z)\,\mathrm{d}z = \int_a^b f(\gamma(t))\gamma'(t)\,\mathrm{d}t = \int_a^b h'(t)\,\mathrm{d}t = h(t)|_a^b = h(b) - h(a) = F(z_1) - F(z_0).$$

■

Example 14.5 (a)

$$\int_{2+3\mathrm{i}}^{1-\mathrm{i}} z^3\,\mathrm{d}z = \frac{(1-\mathrm{i})^4}{4} - \frac{(2+3\mathrm{i})^4}{4}.$$

(b) $\int_1^{\pi\mathrm{i}} \mathrm{e}^z\,\mathrm{d}z = -1 - \mathrm{e}$.

Theorem 14.6 (Cauchy's Theorem) *Let U be a simply connected region in $\mathbb{C}$ and let $f(z)$ be holomorphic in U. Suppose that $\gamma(t)$ is a path in U joining z_0 and z_1 in U. Then $\int_\gamma f(z)\,\mathrm{d}z$ depends on z_0 and z_1 only and not on the choice of the path. In particular, $\int_\gamma f(z)\,\mathrm{d}z = 0$ for any closed path in U.*

Proof. We give the proof under the weak additional assumption that f' not only exists but is continuous in U. In this case, the partial derivatives u_x, u_y, v_x, and v_y are continuous and we can apply the integrability criterion Proposition 8.3 which was a consequence of Green's theorem, see Theorem 10.3. Note that we need U to be simply connected in contrast to Lemma 14.5.

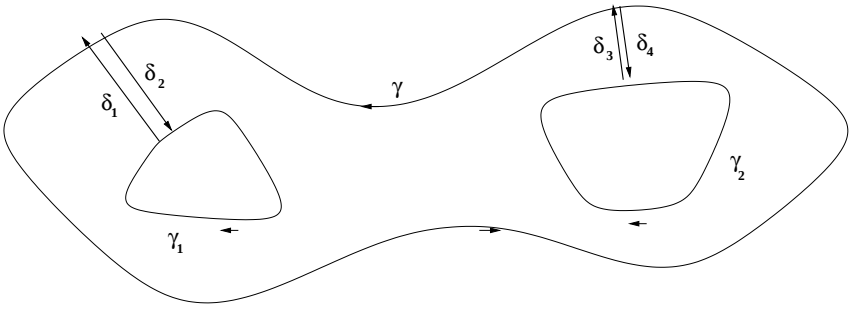
Without this additional assumption (f' is continuous), the proof is lengthy (see [FB93, Lan89, Jän93]) and starts with triangular or rectangular paths and is generalized then to arbitrary paths. We have

$$\int_\gamma f(z)\,\mathrm{d}z = \int_\gamma (u+\mathrm{i}v)(\,\mathrm{d}x+\mathrm{i}\,\mathrm{d}y) = \int_\gamma (u\,\mathrm{d}x - v\,\mathrm{d}y) + \mathrm{i}\int_\gamma (v\,\mathrm{d}x + u\,\mathrm{d}y).$$

We have path independence of the line integral $\int_\gamma P\,\mathrm{d}x + Q\,\mathrm{d}y$ if and only if, the integrability condition $Q_x = P_y$ is satisfied if and only if $P\,\mathrm{d}x + Q\,\mathrm{d}y$ is a closed form.
In our case, the real part is path independent if and only if $-v_x = u_y$. The imaginary part is path independent if and only if $u_x = v_y$. These are exactly the Cauchy–Riemann equations which are satisfied since f is holomorphic. ∎

Remarks 14.4 (a) The proposition holds under the following weaker assumption: f is continuous in the closure $\overline{U}$ and holomorphic in U, U is a simply connected region, and $\gamma = \partial U$ is a path.
(b) The statement is wrong without the assumption "U is simply connected". Indeed, consider the circle of radius r with center a, that is $\gamma(t) = a + r\mathrm{e}^{\mathrm{i}t}$. Then $f(z) = 1/(z-a)$ is singular at a and we have

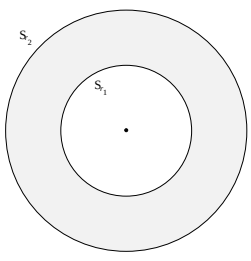
$$\int_{\mathrm{S}_r(a)} \frac{\mathrm{d}z}{z-a} = \mathrm{i}r\int_0^{2\pi} \frac{\mathrm{e}^{\mathrm{i}t}}{r\mathrm{e}^{\mathrm{i}t}}\,\mathrm{d}t = \mathrm{i}\int_0^{2\pi} \mathrm{d}t = 2\pi\mathrm{i}.$$



(c) For a non-simply connected region G one cuts G with pairwise inverse to each other paths (in the picture: δ_1, δ_2, δ_3 and δ_4). The resulting region $\tilde{G}$ is now simply connected such that $\int_{\partial G} f(z)\,\mathrm{d}z = 0$ by (a).

Since the integrals along δ_i, $i = 1,\dots,4$, cancel, we have

$$\int_{\gamma+\gamma_1+\gamma_2} f(z)\,\mathrm{d}z = 0.$$



In particular, if f is holomorphic in $\{z \mid 0 < |z-a| < R\}$ and $0 < r_1 < r_2 < R$, then

$$\int_{\mathrm{S}_{r_1}(a)} f(z)\,\mathrm{d}z = \int_{\mathrm{S}_{r_2}(a)} f(z)\,\mathrm{d}z$$

if both circles are positively oriented.

Proposition 14.7 *Let U be a simply connected region, $z_0 \in U$, $U_0 = U \setminus \{z_0\}$. Suppose that f is holomorphic in U_0 and bounded in a certain neighborhood of z_0.*
Then

$$\int_\gamma f(z)\,\mathrm{d}z = 0$$

for every non-selfintersecting closed path γ in U_0.

Proof. Suppose that $|\, f(z)\,| \le C$ for $|\, z - z_0\,| < \varepsilon_0$. For any ε with $0 < \varepsilon < \varepsilon_0$ we then have by Remark 14.3 (c)

$$\left| \int_{\mathrm{S}_\varepsilon(z_0)} f(z)\,\mathrm{d}z \right| \le 2\pi\varepsilon\, C.$$

By Remark 14.4 (c), $\displaystyle\int_\gamma f(z)\,\mathrm{d}z = \int_{\mathrm{S}_{\varepsilon_0}(z_0)} f(z)\,\mathrm{d}z = \int_{\mathrm{S}_\varepsilon(z_0)} f(z)\,\mathrm{d}z$. Hence

$$\left| \int_\gamma f(z)\,\mathrm{d}z \right| = \left| \int_{\mathrm{S}_\varepsilon(z_0)} f(z)\,\mathrm{d}z \right| \le 2\pi\varepsilon\, C.$$

Since this is true for all small $\varepsilon > 0$, $\int_\gamma f(z)\,\mathrm{d}z = 0$. ∎

We will see soon that under the conditions of the proposition, f can be made holomorphic at z_0, too.

14.2.3 Cauchy's Integral Formula

Theorem 14.8 (Cauchy's Integral Formula) *Let U be a region. Suppose that f is holomorphic in U, and γ a non-selfintersecting positively oriented path in U such that γ is the boundary of $U_0 \subset U$; in particular, U_0 is simply connected.*
Then for every $a \in U_0$ we have

$$f(a) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{f(z)\,\mathrm{d}z}{z - a} \tag{14.9}$$

Proof. $a \in U_0$ is fixed. For $z \in U$ we define

$$F(z) = \begin{cases} \frac{f(z)-f(a)}{z-a}, & z \ne a \\ 0, & z = a. \end{cases}$$

Then $F(z)$ is holomorphic in $U \setminus \{a\}$ and bounded in a neighborhood of a since $f'(a)$ exists and therefore,

$$\left| \frac{f(z) - f(a)}{z - a} - f'(a) \right| < \varepsilon$$

as z approaches a. Using Proposition 14.7 and Remark14.4 (b) we have $\int_\gamma F(z)\,\mathrm{d}z = 0$, that is

$$\int_\gamma \frac{f(z)-f(a)}{z-a}\,\mathrm{d}z = 0,$$

such that

$$\int_\gamma \frac{f(z)\,\mathrm{d}z}{z-a} = \int_\gamma \frac{f(a)\,\mathrm{d}z}{z-a} = f(a)\int_\gamma \frac{\mathrm{d}z}{z-a} = 2\pi\mathrm{i}\, f(a).$$

■

Remark 14.5 The values of a holomorphic function f inside a path γ are completely determined by the values of f on γ.

Example 14.6 Evaluate

$$I_r := \int_{\mathrm{S}_r(a)} \frac{\sin z}{z^2+1}\,\mathrm{d}z$$

in cases $a = 1+\mathrm{i}$ and $r = \frac{1}{2}, 2, 3$.

Solution. We use the partial fraction decomposition of z^2+1 to obtain linear terms in the denominator.

$$\frac{1}{z^2+1} = \frac{1}{2\mathrm{i}}\left(\frac{1}{z-\mathrm{i}} - \frac{1}{z+\mathrm{i}}\right).$$

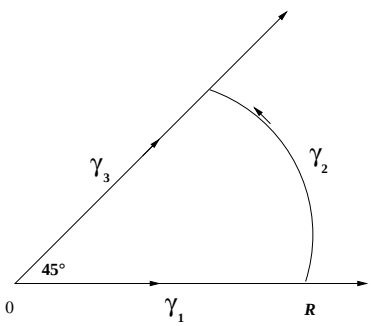
Hence, with $f(z) = \sin z$ we have in case $r = 3$

$$\begin{aligned} I_3 = \int_{\mathrm{S}_r(a)} \frac{\sin z\,\mathrm{d}z}{z^2+1} &= \frac{1}{2\mathrm{i}}\int_{\mathrm{S}_r(a)} \frac{\sin z}{z-\mathrm{i}}\,\mathrm{d}z - \frac{1}{2\mathrm{i}}\int_{\mathrm{S}_r(a)} \frac{\sin z}{z+\mathrm{i}}\,\mathrm{d}z \\ &= \pi(f(\mathrm{i}) - f(-\mathrm{i})) = 2\pi\sin(\mathrm{i}) = \pi\mathrm{i}(\mathrm{e}-1/\mathrm{e}). \end{aligned}$$

In case $r = 2$, the function $\frac{\sin z}{z+\mathrm{i}}$ is holomorphic inside the circle of radius 2 with center a. Hence,

$$I_2 = \pi\sin(\mathrm{i}) = I_3/2.$$

In case $r = \frac{1}{2}$, both integrand are holomorphic, such that $I_{\frac{1}{2}} = 0$.



Example 14.7 Consider the function $f(z) = \mathrm{e}^{\mathrm{i}z^2}$ which is an entire function. Let $\gamma_1(t) = t$, $t \in [0,R]$, be the segment from 0 to R on the real line; let $\gamma_2(t) = R\mathrm{e}^{\mathrm{i}t}$, $t \in [0,\pi/4]$, be the sector of the circle of radius R with center 0; and let finally $\gamma_3(t) = t\mathrm{e}^{\mathrm{i}\pi/4}$, $t \in [0,R]$, be the segment from 0 to $R\mathrm{e}^{\mathrm{i}\pi/4}$. By Cauchy's Theorem,

$$I_1 + I_2 - I_3 = \int_{\gamma_1+\gamma_2-\gamma_3} f(z)\,\mathrm{d}z = 0.$$

Obviously, since $\left(\mathrm{e}^{\mathrm{i}\pi/4}\right)^2 = \mathrm{e}^{\mathrm{i}\pi/2} = \mathrm{i}$

$$I_1 = \int_0^R \mathrm{e}^{\mathrm{i}t^2}\,\mathrm{d}t,$$
$$I_3 = \mathrm{e}^{\mathrm{i}\pi/4}\int_0^R \mathrm{e}^{-t^2}\,\mathrm{d}t$$

We shall show that $|\, I_2(R)\,| \underset{n\to\infty}{\longrightarrow} 0$ as R tends to ∞. We have

$$|\, I_2(R)\,| = \left|\int_0^{\pi/4} \mathrm{e}^{\mathrm{i}\left(R^2\mathrm{e}^{2\mathrm{i}t}\right)} R\mathrm{i}\mathrm{e}^{\mathrm{i}t}\,\mathrm{d}t\right| \le R\int_0^{\pi/4}\left|\,\mathrm{e}^{\mathrm{i}R^2(\cos 2t+\mathrm{i}\sin 2t)}\,\right|\,\mathrm{d}t \le R\int_0^{\pi/4}\mathrm{e}^{-R^2\sin 2t}\,\mathrm{d}t.$$

Note that $\sin t$ is a concave function on $[0,\pi/2]$, that is, the graph of the sine function is above the graph of the corresponding linear function through $(0,0)$ and $(\pi/2,1)$; thus, $\sin t \ge 2t/\pi$, $t\in[0,\pi/2]$. We have

$$|\, I_2(R)\,| \le R\int_0^{\pi/4}\mathrm{e}^{-R^2 4t/\pi}\,\mathrm{d}t = -\frac{\pi}{4R}\left(\mathrm{e}^{-R^2}-1\right) = \frac{\pi}{4R}\left(1-\mathrm{e}^{-R^2}\right).$$

We conclude that $|\, I_2(R)\,|$ tends to 0 as $R\to\infty$. By Cauchy's Theorem $I_1 + I_2 - I_3 = 0$ for all R, we conclude

$$\lim_{R\to\infty} I_1(R) = \int_0^\infty \mathrm{e}^{\mathrm{i}t^2}\,\mathrm{d}t = \mathrm{e}^{\mathrm{i}\pi/4}\int_0^\infty \mathrm{e}^{-t^2}\,\mathrm{d}t = \lim_{R\to\infty} I_3(R).$$

The integral on the right is $\sqrt{\pi}/2$ (see below); hence $\mathrm{e}^{\mathrm{i}t^2} = \cos(t^2) + \mathrm{i}\sin(t^2)$ implies

$$\int_0^\infty \cos(t^2)\,\mathrm{d}t = \frac{\sqrt{2\pi}}{4} = \int_0^\infty \sin(t^2)\,\mathrm{d}t.$$

These are the so called *Fresnel* integrals. We show that $I = \int_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}/2$. (This was already done in Homework 41) For, we compute the double integral using Fubini's theorem:

$$\int_0^\infty\int_0^\infty \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y = \int_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x\int_0^\infty \mathrm{e}^{-y^2}\,\mathrm{d}y = I^2.$$

Passing to polar coordinates yields $\mathrm{d}x\mathrm{d}y = r\,\mathrm{d}r$, $x^2+y^2 = r^2$ such that

$$\iint_{(\mathbb{R}_+)^2} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y = \lim_{R\to\infty}\int_0^{\pi/2}\mathrm{d}\varphi\int_0^R \mathrm{e}^{-r^2} r\,\mathrm{d}r.$$

The change of variables $r^2 = t$, $\mathrm{d}t = 2r\,\mathrm{d}r$ yields

$$I^2 = \frac{1}{2}\frac{\pi}{2}\int_0^\infty \mathrm{e}^{-t}\,\mathrm{d}t = \frac{\pi}{4} \Longrightarrow I = \frac{\sqrt{\pi}}{2}.$$

This proves the claim. In addition, the change of variables $\sqrt{x} = s$ also yields

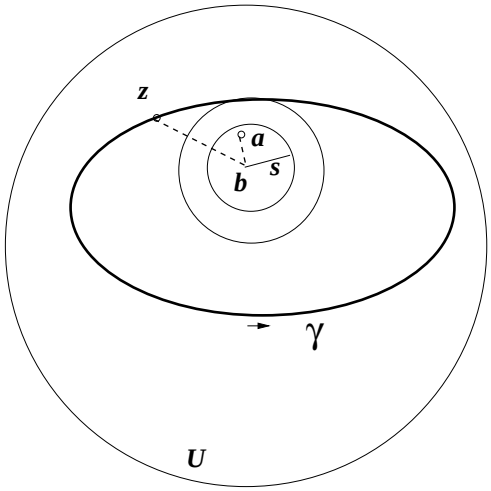
$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{\mathrm{e}^{-x}}{\sqrt{x}}\,\mathrm{d}x = \sqrt{\pi}.$$

Theorem 14.9 *Let γ be a path in an open set U and g be a continuous function on U. If a is not on γ, define*

$$h(a) = \int_\gamma \frac{g(z)}{z-a}\,\mathrm{d}z.$$

Then h is holomorphic on the complement of γ in U and has derivatives of all orders. They are given by

$$h^{(n)}(a) = n! \int_\gamma \frac{g(z)}{(z-a)^{n+1}}\,\mathrm{d}z.$$



Proof. Let $b \in U$ and not on γ. Then there exists some $r > 0$ such that $|\, z-b \,| \geq r$ for all points z on γ. Let $0 < s < r$. We shall see that h has a power series expansion in the ball $U_s(b)$. We write

$$\begin{aligned}\frac{1}{z-a} &= \frac{1}{z-b-(a-b)} = \frac{1}{z-b}\,\frac{1}{1-\frac{a-b}{z-b}} \\ &= \frac{1}{z-b}\left(1 + \frac{a-b}{z-b} + \left(\frac{a-b}{z-b}\right)^2 + \cdots\right).\end{aligned}$$

This geometric series converges absolutely and uniformly for $|\, a-b \,| \leq s$ because

$$\left|\frac{a-b}{z-b}\right| \leq \frac{s}{r} < 1.$$

Since g is continuous and γ is a compact set, $g(z)$ is bounded on γ such that by Theorem 6.6, the series $\sum_{n=0}^\infty g(z)\left(\frac{a-b}{z-b}\right)^n$ can be integrated term by term, and we find

$$\begin{aligned}h(a) &= \int_\gamma \sum_{n=0}^\infty g(z)\frac{(a-b)^n}{(z-b)^{n+1}}\,\mathrm{d}z \\ &= \sum_{n=0}^\infty (a-b)^n \int_\gamma \frac{g(z)}{(z-b)^{n+1}}\,\mathrm{d}z \\ &= \sum_{n=0}^\infty c_n (a-b)^n,\end{aligned}$$

where

$$c_n = \int_\gamma \frac{g(z)\,\mathrm{d}z}{(z-b)^{n+1}}.$$

This proves that h can be expanded into a power series in a neighborhood of b. By Proposition 14.1 and Corollary 14.2, f has derivatives of all orders in a neighborhood of b. By the

formula in Corollary 14.2

$$h^{(n)}(b) = n!c_n = n! \int_\gamma \frac{g(z)\,\mathrm{d}z}{(z-b)^{n+1}}.$$

■

Remark 14.6 There is an easy way to deduce the formula. Formally, we can exchange the differentiation $\frac{\mathrm{d}}{\mathrm{d}a}$ and $\int_\gamma$:

$$h'(a) = \frac{\mathrm{d}}{\mathrm{d}a} \int_\gamma \frac{g(z)}{z-a}\,\mathrm{d}z = \int_\gamma \frac{\mathrm{d}}{\mathrm{d}a}\left(g(z)(z-a)^{-1}\right)\,\mathrm{d}z = \int_\gamma \frac{g(z)}{(z-a)^2}\,\mathrm{d}z.$$

$$h''(a) = \frac{\mathrm{d}}{\mathrm{d}a} \int_\gamma \left(g(z)(z-a)^{-2}\right)\,\mathrm{d}z = 2\int_\gamma \frac{g(z)}{(z-a)^3}\,\mathrm{d}z.$$

Theorem 14.10 *Suppose that f is holomorphic in U and $U_r(a) \subset U$, then f has a power series expansion in $U_r(a)$*

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n.$$

In particular, f has derivatives of all orders, and we have the following **coefficient formula**

$$c_n = \frac{f^{(n)}(a)}{n!} = \frac{1}{2\pi\mathrm{i}} \int_{\mathrm{S}_r(a)} \frac{f(z)\,\mathrm{d}z}{(z-a)^{n+1}}. \tag{14.10}$$

Proof. In view of Cauchy's Integral Formula (Theorem 14.8) we obtain

$$f(a) = \frac{1}{2\pi\mathrm{i}} \int_{\mathrm{S}_R(a)} \frac{f(z)\,\mathrm{d}z}{z-a}$$

Inserting $g(z) = f(z)/(2\pi\mathrm{i})$ (f is continuous) into Theorem 14.9, we see that f can be expanded into a power series with center a and, therefore, it has derivatives of all orders at a,

$$f^{(n)}(a) = \frac{n!}{2\pi\mathrm{i}} \int_{\mathrm{S}_R(a)} \frac{f(z)\,\mathrm{d}z}{(z-a)^{n+1}}.$$

■

14.2.4 Applications of the Coefficient Formula

Proposition 14.11 (Growth of Taylor Coefficients) *Suppose that f is holomorphic in U and is bounded by $M > 0$ in $U_r(a) \subset U$; that is, $|\, z-a \,| < r$ implies $|\, f(z) \,| \le M$.*
Let $\sum_{n=0}^{\infty} c_n (z-a)^n$ be the power series expansion of f at a.
Then we have

$$|\, c_n \,| \le \frac{M}{r^n}. \tag{14.11}$$

Proof. By the coefficient formula (14.10) and Remark 14.3 (c) we have noting that $\left| \dfrac{f(z)}{(z-a)^{n+1}} \right| \leq \dfrac{M}{r^{n+1}}$ for $z \in \mathrm{S}_r a$

$$|\, c_n \,| \leq \left| \frac{1}{2\pi \mathrm{i}} \int\limits_{\mathrm{S}_r(a)} \frac{f(z)}{(z-a)^{n+1}} \, \mathrm{d}z \right| \leq \frac{1}{2\pi} \, \frac{M}{r^{n+1}} \, \ell(\mathrm{S}_r(a)) = \frac{M}{2\pi r^{n+1}} 2\pi r = \frac{M}{r^n}.$$

■

Theorem 14.12 (Liouville's Theorem) *A bounded entire function is constant.*

Proof. Suppose that $|\, f(z) \,| \leq M$ for all $z \in \mathbb{C}$. Since f is given by a power series $f(z) = \sum_{n=0}^{\infty} c_n z^n$ with radius of convergence $R = \infty$, the previous proposition gives

$$|\, c_n \,| \leq \frac{M}{r^n}$$

for *all* $r > 0$. This shows $c_n = 0$ for all $n \neq 0$; hence $f(z) = c_0$ is constant. ■

Remarks 14.7 (a) Note that we explicitly assume f to be holomorphic on the entire complex plane. For example, $f(z) = \mathrm{e}^{1/z}$ is holomorphic and bounded outside every ball $U_\varepsilon(0)$. However, f is not constant.
(b) Note that $f(z) = \sin z$ is an entire function which is not constant. Hence, $\sin z$ is unbounded as a complex function.

Theorem 14.13 (Fundamental Theorem of Algebra) *A polynomial $p(z)$ with complex coefficients of degree $\deg p \geq 1$ has a complex root.*

Proof. Suppose to the contrary that $p(z) \neq 0$ for all $z \in \mathbb{C}$. It is known, see Example 3.3, that $\lim\limits_{|\, z \,| \to \infty} |\, p(z) \,| = +\infty$. In particular there exists $R > 0$ such that

$$|\, z \,| \geq R \Longrightarrow \; |\, p(z) \,| \geq 1.$$

That is, $f(z) = 1/p(z)$ is bounded by 1 if $|\, z \,| \geq R$. On the other hand, f is a continuous function and $\{z \mid |\, z \,| \leq R\}$ is a compact subset of $\mathbb{C}$. Hence, $f(z) = 1/p(z)$ is bounded on $\overline{U_R}$, too. That is, f is bounded on the entire plane. By Liouville's theorem, f is constant and so is p. This contradicts our assumption $\deg p \geq 1$. Hence, p has a root in $\mathbb{C}$. ■

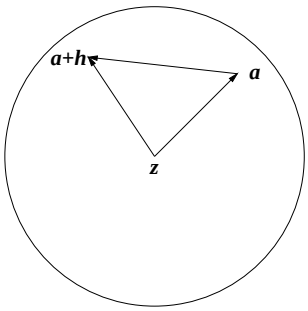
Now, there is an inverse-like statement to Cauchy's Theorem.

Theorem 14.14 (Morera's Theorem) *Let $f \colon U \to \mathbb{C}$ be a continuous function where $U \subset \mathbb{C}$ is open. Suppose that the integral of f along each closed triangular path $[z_1, z_2, z_3]$ in U is 0. Then f is holomorphic in U.*

Proof. Fix $z_0 \in U$. We show that f has an anti-derivative in a small neighborhood $U_\varepsilon(z_0) \subset U$. For $a \in U_\varepsilon(z_0)$ define

$$F(a) = \int_{z_0}^{a} f(z)\,\mathrm{d}z.$$

Note that $F(a)$ takes the same value for all polygonal paths from z_0 to a by assumption of the theorem. We have



$$\left| \frac{F(a+h)-F(a)}{h} - f(a) \right| = \left| \frac{1}{h} \int_a^{a+h} (f(z)-f(a))\,\mathrm{d}z \right|,$$

where the integral on the right is over the segment form a to $a+h$ and we used $\int_a^{a+h} c\,\mathrm{d}z = ch$. By Remark 14.3 (c), the right side is less than or equal to

$$\leq \frac{1}{|\,h\,|} \sup_{z \in U_h(a)} |\,f(z)-f(a)\,|\,|\,h\,| = \sup_{z \in U_h(a)} |\,f(z)-f(a)\,|.$$

Since f is continuous the above term tends to 0 as h tends to 0. This shows that F is differentiable at a with $F'(a) = f(a)$. Since F is holomorphic in U, by Theorem 14.10 it has derivatives of all orders; in particular f is holomorphic. ∎

Corollary 14.15 *Suppose that (f_n) is a sequence of holomorphic functions on U, uniformly converging to f on U. Then f is holomorphic on U.*

Proof. Since f_n are continuous and uniformly converging, f is continuous on U. Let γ be any closed triangular path in U. Since (f_n) converges uniformly, we may exchange integration and limit:

$$\int_\gamma f(z)\,\mathrm{d}z = \int_\gamma \lim_{n\to\infty} f_n(z)\,\mathrm{d}z = \lim_{n\to\infty} \int_\gamma f_n(z)\,\mathrm{d}z = \lim_{n\to\infty} 0 = 0$$

since each f_n is holomorphic. By Morera's theorem, f is holomorphic in U. ∎

Summary

Let U be a region and $f\colon U \to \mathbb{C}$ be a function on U. The following are equivalent:

(a) f is holomorphic in U.

(b) $f = u + \mathrm{i}v$ is real differentiable and the Cauchy–Riemann equations $u_x = v_y$ and $u_v = -v_x$ are satisfied in U.

(c) If U is simply connected, f is continuous and for every closed triangular path $\gamma = [z_1, z_2, z_3]$ in U, $\int_\gamma f(z)\,\mathrm{d}z = 0$ (Morera condition).

(d) f possesses locally an antiderivative, that is, for every $a \in U$ there is a ball $U_\varepsilon(a) \subset U$ and a holomorphic function F such that $F'(z) = f(z)$ for all $z \in U_\varepsilon(a)$.

(e) f is continuous and for every ball $U_r(a)$ with $\overline{U_r(a)} \subset U$ we have

$$f(b) = \frac{1}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_r(a)} \frac{f(z)}{z-b}\,\mathrm{d}z, \qquad \forall\, b \in U_r(a).$$

(f) For $a \in U$ there exists a ball with center a such that f can be expanded in that ball into a power series.

(g) For every ball B which is completely contained in U, f can be expanded into a power series in B.

14.2.5 Power Series

Since holomorphic functions are locally representable by power series, it is quite useful to know how to operate with power series. In case that a holomorphic function f is represented by a power series, we say that f is an *analytic* function. In other words, every holomorphic function is analytic and vice versa. Thus, by Theorem 14.10, any holomorphic function is analytic and vice versa.

(a) Uniqueness

If both $\sum c_n z^n$ and $\sum b_n z^n$ converge in a ball around 0 and define the same function then $c_n = b_n$ for all $n \in \mathbb{N}_0$.

(b) Multiplication

If both $\sum c_n z^n$ and $\sum b_n z^n$ converge in a ball $U_r(0)$ around 0 then

$$\sum_{n=0}^{\infty} c_n z^n \cdot \sum_{n=0}^{\infty} b_n z^n = \sum_{n=0}^{\infty} d_n z^n, \quad |\,z\,| < r,$$

where $d_n = \sum_{k=0}^{n} c_{n-k} b_k$.

(c) The Inverse $1/f$

Let $f(z) = \sum\limits_{n=0}^{\infty} c_n z^n$ be a convergent power series and

$$c_0 \neq 0.$$

Then $f(0) = c_0 \neq 0$ and, by continuity of f, there exists $r > 0$ such that the power series converges in the ball $U_r(0)$ and is non-zero there. Hence, $1/f(z)$ is holomorphic in $U_r(0)$ and therefore it can be expanded into a converging power series in $U_r(0)$, see summary (f). Suppose

that $1/f(z) = g(z) = \sum_{n=0}^{\infty} b_n z^n$, $|z| < r$. Then $f(z)g(z) = 1 = 1 + 0z + 0z^2 + \cdots$, the uniqueness and (b) yields

$$1 = c_0 b_0, \quad 0 = c_0 b_1 + c_1 b_0, \quad 0 = c_0 b_2 + c_1 b_1 + c_2 b_0, \cdots$$

This system of equations can be solved recursively for b_n, $n \in \mathbb{N}_0$, for example, $b_0 = 1/c_0$, $b_1 = -c_1 b_0 / c_0$.

(d) Double Series

Suppose that

$$f_k(z) = \sum_{n=0}^{\infty} c_{kn}(z-a)^n, \quad k \in \mathbb{N}$$

are converging in $U_r(a)$ power series. Suppose further that the series

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

converges locally uniformly in $U_r(a)$ as well. Then

$$\sum_{k=1}^{\infty} f_k(z) = \sum_{n=0}^{\infty} \left(\sum_{k=1}^{\infty} c_{kn} \right) (z-a)^n.$$

In particular, one can form the sum of a locally uniformly convergent series $\sum f_k(z)$ of power series coefficientwise. Note that a series of functions $\sum_{k=1}^{\infty} f_k(z)$ *converges locally uniformly* at b if there exists $\varepsilon > 0$ such that the series converges uniformly in $U_\varepsilon(b)$.

Note that any locally uniformly converging series of holomorphic functions defines a holomorphic function (Theorem of Weierstraß). Indeed, since the series converges uniformly, line integral and summation can be exchanged: Let $\gamma = [z_0 z_1 z_2]$ be any closed triangular path inside U, then by Cauchy's theorem

$$\int_\gamma f(z)\,\mathrm{d}z = \int_\gamma \sum_{k=1}^{\infty} f_k(z)\,\mathrm{d}z = \sum_{k=1}^{\infty} \int_\gamma f_k(z)\,\mathrm{d}z = \sum_{k=1}^{\infty} 0 = 0.$$

By Morera's theorem, $f(z) = \sum_{k=1}^{\infty} f(z)$ is holomorphic.

(e) Change of Center

Let $f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n$ be convergent in $U_r(a)$, $r > 0$, and $b \in U_r(a)$. Then f can be expanded into a power series with center b

$$f(z) = \sum_{n=0}^{\infty} b_n (z-b)^n, \quad b_n = \frac{f^{(n)}(b)}{n!},$$

with radius of convergence at least $r - |b-a|$. Also, the coefficients can be obtained by reordering for powers of $(z-b)^k$ using the binomial formula

$$(z-a)^n = (z-b+b-a)^n = \sum_{k=0}^{n} \binom{n}{k} (z-b)^k (b-a)^{n-k}.$$

(f) Composition

We restrict ourselves to the case

$$f(z) = a_0 + a_1 z + a_2 z^2 + \cdots$$
$$g(z) = b_1 z + b_2 z^2 + \cdots$$

where $g(0) = 0$ and therefore, the image of g is a small neighborhood of 0, and we assume that the first power series f is defined there; thus, $f(g(z))$ is defined and holomorphic in a certain neighborhood of 0, see Remark 14.1. Hence

$$h(z) = f(g(z)) = c_0 + c_1 z + c_2 z^2 + \cdots$$

where the coefficients $c_n = h^{(n)}(0)/n!$ can be computed using the chain rule, for example, $c_0 = f(g(0)) = a_0$, $c_1 = f'(g(0))g'(0) = a_1 b_1$.

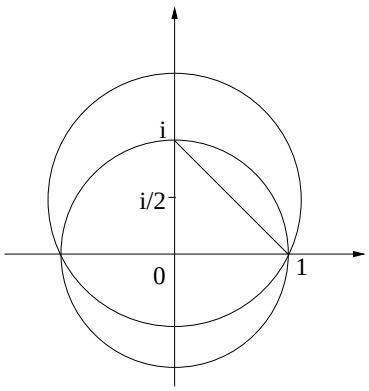
(g) The Composition Inverse f^{-1}

Suppose that $f(z) = \sum_{n=1}^{\infty} a_n z^n$, $a_1 \neq 0$, has radius of convergence $r > 0$. Then there exists a power series $g(z) = \sum_{n=1}^{\infty} b_n z^n$ converging on $U_\varepsilon(0)$ such that $f(g(z)) = z = g(f(z))$ for all $z \in U_\varepsilon(0)$. Using (f) and the uniqueness, the coefficients b_n can be computed recursively.

Example 14.8 (a) The function

$$f(z) = \frac{1}{1+z^2} + \frac{1}{3-z}$$

is holomorphic in $\mathbb{C} \setminus \{\mathrm{i}, -\mathrm{i}, 3\}$. Expanding f into a power series with center 1, the closest singularity to 1 is $\pm\mathrm{i}$. Since the disc of convergence cannot contain $\pm\mathrm{i}$, the radius of convergence is $|\,1-\mathrm{i}\,| = \sqrt{2}$. Expanding the power series around $a = 2$, the closest singularity of f is 3; hence, the radius of convergence is now $|\,3-2\,| = 1$.



(b) Change of center. We want to expand $f(z) = \dfrac{1}{1-z}$ which is holomorphic in $\mathbb{C} \setminus \{1\}$ into a power series around $b = \mathrm{i}/2$. For arbitrary b with $|\,b\,| < 1$ we have

$$\frac{1}{1-z} = \frac{1}{1-b-(z-b)} = \frac{1}{1-b} \cdot \frac{1}{1-\frac{z-b}{1-b}}$$
$$= \sum_{n=0}^{\infty} \frac{1}{(1-b)^{n+1}} (z-b)^n = \tilde{f}(z).$$

By the root test, the radius of convergence of this series is $|\,1-b\,|$. In case $b = \mathrm{i}/2$ we have $r = |\,1-\mathrm{i}/2\,| = \sqrt{1+1/4} = \sqrt{5}/2$. Note that the power series $1 + z + z^2 + \cdots$ has radius of convergence 1 and a priori defines an analytic (= holomorphic) function in the open unit ball. However, changing the center we obtain an analytic continuation of f to a larger region. This example shows that (under certain assumptions) analytic functions can be extended into a larger region by changing the center of the series.

14.3 Local Properties of Holomorphic Functions

We omit the proof of the Open Mapping Theorem and refer to [Jän93, Satz 11, Satz 13] see also [Con78].

Theorem 14.16 (Open Mapping Theorem) *Let G be a region and f a non-constant holomorphic function on G. Then for every open subset $U \subset G$, $f(U)$ is open.*

The main idea is to show that any at some point a holomorphic function f with $f(a) = 0$ looks like a power function, that is, there exists a positive integer k such that $f(z) = h(z)^k$ in a small neighborhood of a, where h is holomorphic at a with a zero of order 1 at a.

Theorem 14.17 (Maximum Modulus Theorem) *Let f be holomorphic in the region U and $a \in U$ is a point such that $|f(a)| \geq |f(z)|$ for all $z \in U$. Then f must be a constant function.*

Proof. First proof. Let $V = f(U)$ and $b = f(a)$. By assumption, $|b| \geq |w|$ for all $w \in V$. b cannot be an inner point of V since otherwise, there is some c in the neighborhood of b with $|c| > |b|$ which contradict the assumption. Hence b is in the boundary $\partial V \cap V$. In particular, V is not open. Hence, the Open Mapping Theorem says that f is constant.
Second Proof. We give a direct proof using Cauchy's Integral formula. For simplicity let $a = 0$ and let $U_r(0) \subseteq U$ be a small ball in U. By Cauchy's theorem with $\gamma = \mathrm{S}_r(0)$, $z = \gamma(t) = r\mathrm{e}^{\mathrm{i}t}$, $t \in [0, 2\pi]$, $\mathrm{d}z = r\mathrm{i}\mathrm{e}^{\mathrm{i}t}\,\mathrm{d}t$ we get

$$f(0) = \frac{1}{2\pi\mathrm{i}} \int_0^{2\pi} \frac{f\left(r\mathrm{e}^{\mathrm{i}t}\right) r\mathrm{i}\mathrm{e}^{\mathrm{i}t}}{r\mathrm{e}^{\mathrm{i}t}}\,\mathrm{d}t = \frac{1}{2\pi} \int_0^{2\pi} f(r\mathrm{e}^{\mathrm{i}t})\,\mathrm{d}t.$$

In other words, $f(0)$ is the arithmetic mean of the values of f on any circle with center 0. Let $M = |f(a)| \geq |f(z)|$ be the maximal modulus of f on U. Suppose, there exists z_0 with $z_0 = r_0\mathrm{e}^{\mathrm{i}t_0}$ with $r_0 < r$ and $|f(z_0)| < M$. Since f is continuous, there exists a whole neighborhood of $t \in U_\varepsilon(t_0)$ with $\left|f(r\mathrm{e}^{\mathrm{i}t})\right| < M$. However, in this case

$$M = |f(0)| = \left|\frac{1}{2\pi}\int_0^{2\pi} f(r\mathrm{e}^{\mathrm{i}t})\,\mathrm{d}t\right| \leq \frac{1}{2\pi}\int_0^{2\pi} \left|f(r\mathrm{e}^{\mathrm{i}t})\right|\,\mathrm{d}t < M$$

which contradicts the mean value property. Hence, $|f(z)| = M$ is constant in any sufficiently small neighborhood of 0. Let $z_1 \in U$ be any point in U. We connect 0 and z_1 by a path in U. Let d be its distance from the boundary ∂U. Let z continuously moving from 0 to z_1 and cosider the chain of balls with center z and radius $d/2$. By the above, $|f(z)| = M$ in any such ball, hence $|f(z)| = M$ in U. It follows from homework 47.2 that f is constant. ∎

Remark 14.8 In other words, if f is holomorphic in G and $U \subset G$, then $\sup_{z\in U} |f(z)|$ is attained on the boundary ∂U. Note that both theorems are not true in the real setting: The image of the sine function of the open set $(0, 2\pi)$ is $[-1, 1]$ which is not open. The maximum of $f(x) = 1-x^2$ over $(-1, 1)$ is not attained on the boundary since $f(-1) = f(1) = 0$ while $f(0) = 1$. However $|z^2 - 1|$ on the complex unit ball attains its maximum in $z = \pm\mathrm{i}$—on the boundary.

Recall from topology:

- An *accumulation point* of a set $M \subset \mathbb{C}$ is a point of $z \in \mathbb{C}$ such that for every $\varepsilon > 0$, $U_\varepsilon(z)$ contains infinitely many elements of M. Accumulation points are in the closure of M, not necessarily in the boundary of M. The set of accumulation points of M is closed (Indeed, suppose that a is in the closure of the set of accumulation points of M. Then every neighborhood of a meets the set of accumulation points of M; in particular, every neighborhood has infinitely many element of M. Hence a itself is an accumulation point of M).
- M is *connected* if every locally constant function is constant. M is not connected if M is the disjoint union of two non-empty subsets A and B, both are open *and* closed in M.

For example, $M = \{1/n \mid n \in \mathbb{N}\}$ has no accumulation point in $\mathbb{C}\setminus\{0\}$ but it has one accumulation point, 0, in $\mathbb{C}$.

Proposition 14.18 *Let U be a region and let $f\colon U \to \mathbb{C}$ be holomorphic in U.*
If the set $Z(f)$ of the zeros of f has an accumulation point in U, then f is identically 0 in U.

Example 14.9 Consider the holomorphic function $f(z) = \sin\frac{1}{z}$, $f\colon U \to \mathbb{C}$, $U = \mathbb{C}\setminus\{0\}$ and with the set of zeros $Z(f) = \{z_n = \frac{1}{n\pi} \mid n \in \mathbb{Z}\}$. The only accumulation point 0 of $Z(f)$ does not belong to U. The proposition does not apply.

Proof. Suppose $a \in U$ is an accumulation point of $Z(f)$. Expand f into a power series with center a:

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n, \quad |\,z-a\,| < r.$$

Since a is an accumulation point of the zeros, there exists a sequence (z_n) of zeros converging to a. Since f is continuous at a, $\lim_{n\to\infty} f(z_n) = 0 = f(a)$. This shows $c_0 = 0$. The same argument works with the function

$$f_1(z) = c_1 + c_2(z-a) + c_3(z-a)^3 + \cdots = \frac{f(z)}{z-a},$$

which is holomorphic in the same ball with center a and has a as an accumulation point of zeros. Hence, $c_1 = 0$. In the same way we conclude that $c_2 = c_3 = \cdots = c_n = \cdots = 0$. This shows that f is identically 0 on $U_r(a)$. That is, the set

$$A = \{a \in U \mid a \text{ is an accumulation point of } Z(f)\,\}$$

is an open set. Also,

$$B = \{a \in U \mid a \text{ is not an accumulation point of } Z(f)\,\}$$

is open (with every non-accumulation point z, there is a whole neighborhood of z not containing accumulation points of $Z(f)$). Now, U is the disjoint union of A and B, both are open as well as closed in U. Hence, the characteristic function on A is a locally constant function on U. Since U is connected, either $U = A$ or $U = B$. Since by assumption A is non-empty, $A = U$, that is f is identically 0 on U. ∎

Theorem 14.19 (Uniqueness Theorem) *Suppose that f and g are both holomorphic functions on U and U is a region. Then the following are equivalent:*

(a) $f = g$
(b) *The set $D = \{z \in U \mid f(z) = g(z)\}$ where f and g are equal has an accumulation point in U.*
(c) *There exists $z_0 \in U$ such that $f^{(n)}(z_0) = g^{(n)}(z_0)$ for all non-negative integers $n \in \mathbb{N}_0$.*

Proof. (a) $\leftrightarrow$ (b). Apply the previous proposition to the function $f - g$.
(a) implies (c) is trivial. Suppose that (c) is satisfied. Then, the power series expansion of $f - g$ at z_0 is identically 0. In particular, the set $Z(f - g)$ contains a ball $B_\varepsilon(z_0)$ which has an accumulation point. Hence, $f - g = 0$. ∎

The following proposition is an immediate consequence of the uniqueness theorem.

Proposition 14.20 (Uniqueness of Analytic Continuation) *Suppose that $M \subset U \subset \mathbb{C}$ where U is a region and M has an accumulation point in U. Let g be a function on M and suppose that f is a holomorphic function on U which extents g, that is $f(z) = g(z)$ on M.*
Then f is unique.

Remarks 14.9 (a) The previous proposition shows a quite amazing property of a holomorphic function: It is completely determined by "very few values". This is in a striking contrast to C^∞-functions on the real line. For example, the "hat function"

$$h(x) = \begin{cases} \mathrm{e}^{-\frac{1}{1-x^2}} & |\,x\,| < 1, \\ 0 & |\,x\,| \geq 1 \end{cases}$$

is identically 0 on $[2, 3]$ (a set with accumulation points), however, h is not identically 0. This shows that h is not holomorphic.
(b) For the uniqueness theorem, it is an essential point that U is connected.
(c) It is now clear that the real function e^x, $\sin x$, and $\cos x$ have a unique analytic continuation into the complex plane.
(d) The algebra $\mathcal{O}(U)$ of holomorphic functions on a region U is a *domain*, that is, $fg = 0$ implies $f = 0$ or $g = 0$. Indeed, suppose that $f(z_0) \neq 0$, then $f(z) \neq 0$ in a certain neighborhood of z_0 (by continuity of f). Then $g = 0$ on that neighborhood. Since an open set has always an accumulation point in itself, $g = 0$.

14.4 Singularities

We consider functions which are holomorphic in a punctured ball $\overset{\circ}{U}_r(a)$. From information about the behaviour of the function near the center a, a number of interesting and useful results will be derived. In particular, we will use these results to evaluate certain unproper integrals over the real line which cannot be evaluated by methods of calculus.

14.4.1 Classification of Singularities

Throughout this subsection U is a region, $a \in U$, and $f\colon U \setminus \{a\} \to \mathbb{C}$ is holomorphic.

Definition 14.4 (a) Let f be holomorphic in $U \setminus \{a\}$ where U is a region and $a \in U$. Then a is said to be an *isolated singularity*.
(b) The point a is called a *removable singularity* if there exists a holomorphic function $g\colon U_r(a) \to \mathbb{C}$ such that $g(z) = f(z)$ for all z with $0 < |\, z - a \,| < r$.

Example 14.10 The functions $\dfrac{\sin z}{z}$, $\frac{1}{z}$, and $\mathrm{e}^{1/z}$ all have isolated singularities at 0. However, only $f(z) = \dfrac{\sin z}{z}$ has a removable singularity. The holomorphic function $g(z)$ which coincides with f on $\mathbb{C} \setminus \{0\}$ is $g(z) = 1 - z^2/3! + z^4/5! - + \cdots$. Hence, redefining $f(0) := g(0) = 1$ makes f holomorphic in $\mathbb{C}$. We will see later that the other two singularities are not removable. It is convenient to denote the the new function g with one more point in its domain (namely a) also by f.

Proposition 14.21 (Riemann—1851) *Suppose that $f\colon U \setminus \{a\} \to \mathbb{C}$, $a \in U$, is holomorphic. Then a is a removable singularity of f if and only if there exists a punctured neighborhood $\overset{\circ}{U}_r(a)$ where f is bounded.*

Proof. The necessity of the condition follows from the fact, that a holomorphic function g is continuous and the continuous function $|\, g(z) \,|$ defined on the compact set $\overline{U_{r/2}(a)}$ is bounded; hence f is bounded.
For the sufficiency we assume without loss of generality, $a = 0$ (if a is non-zero, consider the function $\tilde{f}(z) = f(z - a)$ instead). The function

$$h(z) = \begin{cases} z^2 f(z), & z \neq 0, \\ 0, & z = 0 \end{cases}$$

is holomorphic in $\overset{\circ}{U}_r(0)$. Moreover, h is differentiable at 0 since f is bounded in a neighborhood of 0 and

$$h'(0) = \lim_{z \to 0} \frac{h(z) - h(0)}{z} = \lim_{z \to 0} z f(z) = 0.$$

Thus, h can be expanded into a power series at 0,

$$h(z) = c_0 + c_1 z + c_2 z^3 + c_3 z^3 + \cdots = c_2 z^2 + c_3 z^3 + \cdots$$

with $c_0 = c_1 = 0$ since $h(0) = h'(0) = 0$. For non-zero z we have

$$f(z) = \frac{h(z)}{z^2} = c_2 + c_3 z + c_4 z^2 + \cdots .$$

The right side defines a holomorphic function in a neighborhood of 0 which coincides with f for $z \neq 0$. The setting $f(0) = c_2$ removes the singularity at 0. ∎

Definition 14.5 (a) An isolated singularity a of f is called a *pole* of f if there exists a positive integer $m \in \mathbb{N}$ and a holomorphic function $g\colon U_r(a) \to \mathbb{C}$ such that

$$f(z) = \frac{g(z)}{(z-a)^m}$$

The smallest number m such that $(z-a)^m f(z)$ has a removable singularity at a is called the *order* of the pole.

(b) An isolated singularity a of f which is neither removable nor a pole is called an *essential singularity*.

(c) If f is holomorphic at a and there exists a positive integer m and a holomorphic function g such that $f(z) = (z-a)^m g(z)$, and $g(a) \neq 0$, a is called a *zero of order* m of f.

Note that $m = 0$ corresponds to removable singularities. If $f(z)$ has a zero of order m at a, $1/f(z)$ has a pole of order m at a and vice versa.

Example 14.11 The function $f(z) = 1/z^2$ has a pole of order 2 at $z = 0$ since $z^2 f(z) = 1$ has a removable singularity at 0 and $zf(z) = 1/z$ not. The function $f(z) = (\cos z - 1)/z^3$ has a pole of order 1 at 0 since $(\cos z - 1)/z^3 = -/(2z) + z/4! \mp \cdots$.

14.4.2 Laurent Series

In a neighborhood of an isolated singularity a holomorphic function cannot expanded into a power series, however, in a so called Laurent series.

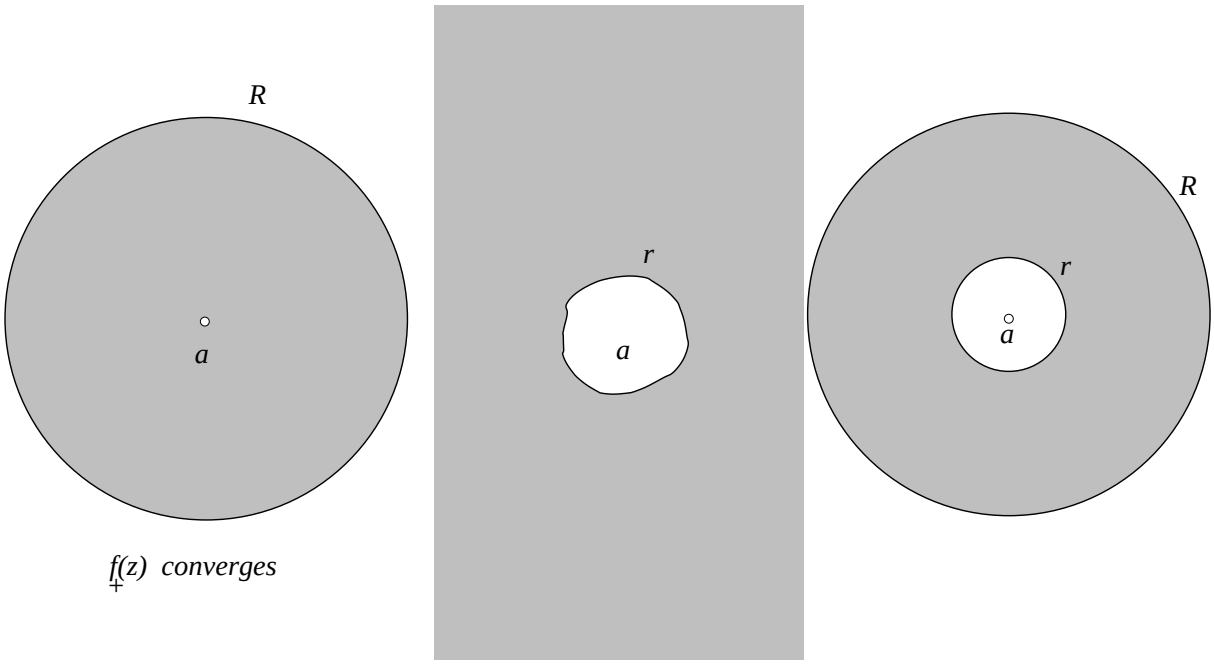
Definition 14.6 A *Laurent series* with center a is a series of the form

$$\sum_{n=-\infty}^{\infty} c_n (z-a)^n$$

or more precisely the pair of series

$$f_-(z) = \sum_{n=1}^{\infty} c_{-n}(z-a)^{-n} \quad \text{and} \quad f_+(z) = \sum_{n=0}^{\infty} c_n (z-a)^n.$$

The Laurent series is said to be convergent if both series converge.



Remark 14.10 (a) $f_-(z)$ is "a power series in $\dfrac{1}{z-a}$." Thus, we can derive facts about the convergence of Laurent series from the convergence of power series. In fact, suppose that $1/r$ is the radius of convergence of the power series $\sum_{n=1}^{\infty} c_{-n}\zeta^n$ and R is the radius of convergence of the series $\sum_{n=0}^{\infty} c_n z^n$, then the Laurent series $\sum_{n\in\mathbb{Z}} c_n z^n$ converges in the annulus $A_{r,R} = \{z \mid r < z < R\}$ and defines there a holomorphic function.
(a) The power series $f_+(z) = \sum_{n\geq 0} c_n(z-a)^n$ converges in the inner part of the ball $U_R(a)$ whereas the series with negative powers, called the *principal part* of the Laurent series, $f_-(z) = \sum_{n<0} c_n(z-a)^n$ converges in the exterior of the ball $U_r(a)$. Since both series must converge, $f(z)$ convergence in intersection of the two domains which is the annulus $A_{r,R}(a)$.

The easiest way to determine the type of an isolated singularity is to use Laurent series which are, roughly speaking, power series with both positive and negative powers z^n.

Proposition 14.22 *Suppose that f is holomorphic in the open annulus $A_{r,R}(a) = \{z \mid r < |z-a| < R\}$. Then $f(z)$ has an expansion in a convergent Laurent series for $z \in A_{r,R}$*

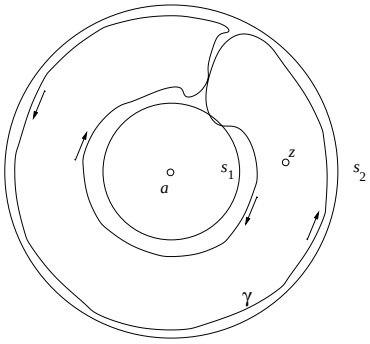
$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n + \sum_{n=1}^{\infty} c_{-n}\frac{1}{(z-a)^n} \tag{14.12}$$

with coefficients

$$c_n = \frac{1}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_\rho(a)} \frac{f(z)}{(z-a)^{n+1}}\,\mathrm{d}z, \quad n \in \mathbb{Z}, \tag{14.13}$$

where $r < \rho < R$. The series converges uniformly on every annulus $A_{s_1,s_2}(a)$ with $r < s_1 \leq s_2 < R$.

Proof.



Let z be in the annulus A_{s_1,s_2} and let γ be the closed path around z in the annulus consisting of the two circles $-\mathrm{S}_{s_1}(a)$, $\mathrm{S}_{s_2}(a)$ and the two "bridges." By Cauchy's integral formula,

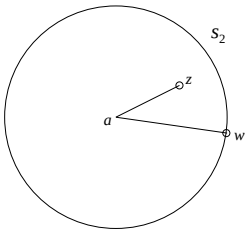
$$f(z) = \frac{1}{2\pi\mathrm{i}} \int_\gamma \frac{f(w)}{w-z}\mathrm{d}w = f_1(z) + f_2(z) = \\ = \frac{1}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_{s_2}(a)} \frac{f(w)}{w-z}\mathrm{d}w - \frac{1}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_{s_1}(a)} \frac{f(w)}{w-z}\mathrm{d}w$$

We consider the two functions

$$f_1(z) = \frac{1}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_{s_2}(a)} \frac{f(w)}{w-z}\mathrm{d}w, \qquad f_2(z) = -\frac{1}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_{s_1}(a)} \frac{f(w)}{w-z}\mathrm{d}w$$

separately.

In what follows, we will see that $f_1(z)$ is a power series $\sum_{n=0}^\infty c_n(z-a)^n$ and $f_2(z) = \sum_{n=1}^\infty c_{-n}\frac{1}{(z-a)^n}$. The first part is completely analogous to the proof of Theorem 14.9.



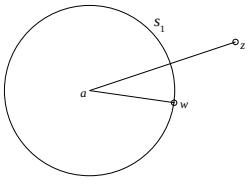
Case 1. $w \in \mathrm{S}_{s_2}(a)$. Then $|\,z-a\,| < |\,w-a\,|$ and $|\,q\,| = \left|\dfrac{z-a}{w-a}\right| < 1$ such that

$$\frac{1}{w-z} = \frac{1}{w-a}\,\frac{1}{\left(1-\frac{z-a}{w-a}\right)} = \frac{1}{w-a}\sum_{n=0}^\infty q^n = \sum_{n=0}^\infty \frac{(z-a)^n}{(w-a)^{n+1}}.$$

Since $f(w)$ is bounded on $\mathrm{S}_{s_2}(a)$, the geometric series has a converging numerical upper bound. Hence, the series converges uniformly with respect to w; we can exchange integration and summation:

$$f_1(z) = \frac{1}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_{s_2}(a)} \sum_{n=0}^\infty f(w)\frac{(z-a)^n}{(w-a)^{n+1}}\mathrm{d}w = \sum_{n=0}^\infty \frac{(z-a)^n}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_{s_2}(a)} \frac{f(w)\mathrm{d}w}{(w-a)^{n+1}} = \sum_{n=0}^\infty c_n(z-a)^n,$$

where $c_n = \frac{1}{2\pi\mathrm{i}} \int\limits_{\mathrm{S}_{s_2}(a)} \frac{f(w)\mathrm{d}w}{(w-a)^{n+1}}$ are the coefficients of the power series $f_1(z)$.



Case 2. $w \in \mathrm{S}_{s_1}(a)$. Then $|\,z-a\,| > |\,w-a\,|$ and $\left|\dfrac{w-a}{z-a}\right| < 1$ such that

$$\frac{1}{w-z} = -\frac{1}{z-a}\,\frac{1}{\left(1-\frac{w-a}{z-a}\right)} = -\frac{1}{z-a}\sum_{n=0}^\infty q^n = -\sum_{n=0}^\infty \frac{(w-a)^n}{(z-a)^{n+1}}.$$

Since $f(w)$ is bounded on $\mathrm{S}_{s_1}(a)$, the geometric series has a converging numerical upper bound. Hence, the series converges uniformly with respect to w; we can exchange integration and

summation:

$$f_2(z) = \frac{1}{2\pi \mathrm{i}} \int\limits_{\mathrm{S}_{s_1}(a)} \sum_{n=0}^{\infty} f(w) \frac{(w-a)^n}{(z-a)^{n+1}} \mathrm{d}w = \sum_{n=0}^{\infty} \frac{1}{2\pi \mathrm{i}(z-a)^{n+1}} \int\limits_{\mathrm{S}_{s_1}(a)} f(w)\,(w-a)^n \mathrm{d}w$$
$$= \sum_{n=1}^{\infty} c_{-n}(z-a)^{-n},$$

where $c_{-n} = \frac{1}{2\pi \mathrm{i}} \int\limits_{\mathrm{S}_{s_2}(a)} f(w)\,(w-a)^{n-1} \mathrm{d}w$ are the coefficients of the series $f_2(z)$.

Since the integrand $\dfrac{f(w)}{(w-a)^k}$, $k \in \mathbb{Z}$, is holomorphic in both annuli $A_{s_1,\rho}$ and A_{ρ,s_2}, by Remark 14.4 (c)

$$\int\limits_{\mathrm{S}_{s_2}(a)} \frac{f(w)\mathrm{d}w}{(w-a)^k} = \int\limits_{\mathrm{S}_{\rho}(a)} \frac{f(w)\mathrm{d}w}{(w-a)^k}, \quad \text{and} \quad \int\limits_{\mathrm{S}_{s_1}(a)} \frac{f(w)\mathrm{d}w}{(w-a)^k} = \int\limits_{\mathrm{S}_{\rho}(a)} \frac{f(w)\mathrm{d}w}{(w-a)^k},$$

that is, in the coefficient formulas we can replace both circles $\mathrm{S}_{s_1}(a)$ and $\mathrm{S}_{s_2}(a)$ by a common circle $\mathrm{S}_{\rho}(a)$. Since a power series converge uniformly on every compact subset of the disc of convergence, the last assertion follows. ∎

Remark 14.11 The Laurent series of f on $A_{r,R}(a)$ is unique. Its coefficients c_n, $n \in \mathbb{Z}$ are uniquely determined by (14.13). Another value of ρ with $r < \rho < R$ yields the same values c_n by Remark 14.4 (c).

Example 14.12 Find the Laurent expansion of $f(z) = \dfrac{2}{z^2 - 4z + 3}$ in the three annuli with midpoint 0

$$0 < |\,z\,| < 1, \quad 1 < |\,z\,| < 3, \quad 3 < |\,z\,|\,.$$

Using partial fraction decomposition, $f(z) = \dfrac{1}{1-z} + \dfrac{1}{z-3}$, we find in the case
(a) $|\,z\,| < 1$

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n, \quad \frac{1}{3-z} = \frac{1}{3}\left(\frac{1}{1-\frac{z}{3}}\right) = \frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{z}{3}\right)^n.$$

Hence,

$$f(z) = \sum_{n=0}^{\infty}\left(1 - \frac{1}{3^{n+1}}\right) z^n, \quad |\,z\,| < 1.$$

In the case $|\,z\,| > 1$,

$$\frac{1}{1-z} = \frac{1}{z}\,\frac{1}{1-\frac{1}{z}} = \sum_{n=0}^{\infty} \frac{1}{z^{n+1}}$$

and as in (a) for $|\,z\,| < 3$

$$\frac{1}{3-z} = \frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{z}{3}\right)^n$$

such that

$$f(z) = \sum_{n=0}^{\infty} \frac{-1}{z^n} + \sum_{n=0}^{\infty} \frac{-1}{3^{n+1}} z^n.$$

(c) In case $|\, z \,| > 3$ we have

$$\frac{1}{z-3} = \frac{1}{z\left(1 - \frac{3}{z}\right)} = \sum_{n=0}^{\infty} \frac{3^n}{z^{n+1}}$$

such that

$$f(z) = \sum_{n=1}^{\infty} (3^{n-1} - 1) \frac{1}{z^n}.$$

We want to study the behaviour of f in a neighborhood of an essential singularity a. It is characterized by the following theorem. For the proof see Conway, [Con78, p. 300].

Theorem 14.23 (Great Picard Theorem (1879)) *Suppose that $f(z)$ is holomorphic in the annulus $G = \{z \mid 0 < |\, z - a \,| < r\}$ with an essential singularity at a.*
Then there exists a complex number w_1 with the following property. For any complex number $w \neq w_1$, there are infinitely many $z \in G$ with $f(z) = w$.

In other words, in every neighborhood of a the function $f(z)$ takes all complex values with possibly one omission. In case of $f(z) = \mathrm{e}^{1/z}$ the number 0 is omitted; in case of $f(z) = \sin \frac{1}{z}$ no complex number is omitted.
We will prove a much weaker form of this statement.

Proposition 14.24 (Casorati–Weierstraß) *Suppose that $f(z)$ is holomorphic in the annulus $G = \{z \mid 0 < |\, z - a \,| < r\}$ with an essential singularity at a.*
Then the image of any neighborhood of a in G is dense in $\mathbb{C}$, that is for every $w \in \mathbb{C}$ and any $\varepsilon > 0$ and $\delta > 0$ there exists $z \in G$ such that $|\, z - a \,| < \delta$ and $|\, f(z) - w \,| < \varepsilon$.

Proof. For simplicity, assume that $\delta < r$. Assume to the contrary that there exists $w \in \mathbb{C}$ and $\varepsilon > 0$ such that $|\, f(z) - w \,| \geq \varepsilon$ for all $z \in \overset{\circ}{U}_\delta(a)$. Then the function

$$g(z) = \frac{1}{f(z) - w}, \quad z \in \overset{\circ}{U}_\delta(a)$$

is bounded (by $1/\varepsilon$) in some neighborhood of a; hence, by Proposition 14.21, a is a removable singularity of $g(z)$. We conclude that

$$f(z) = \frac{1}{g(z)} + w$$

has a removable singularity at a if $g(a) \neq 0$. If, on the other hand, $g(z)$ has a zero at a of order m, that is

$$g(z) = \sum_{n=m}^{\infty} c_n (z-a)^n, \quad c_m \neq 0,$$

the function $(z-a)^m f(z)$ has a removable singularity at a. Thus, f has a pole of order m at a. Both conclusions contradict our assumption that f has an essential singularity at a. ∎

The Laurent expansion establishes an easy classification of the singularity of f at a. We summarize the main facts about isolated singularities.

Proposition 14.25 *Suppose that $f(z)$ is holomorphic in the punctured disc $U = \overset{\circ}{U}_R(a)$ and possesses there the Laurent expansion $f(z) = \sum_{n=-\infty}^{\infty} c_n(z-a)^n$.*

Then the singularity at a

(a) *is removable if $c_n = 0$ for all $n < 0$. In this case, $|\, f(z)\,|$ is bounded in U.*

(b) *is a pole of order m if $c_{-m} \neq 0$ and $c_n = 0$ for all $n < -m$. In this case, $\lim_{z\to a} |\, f(z)\,| = +\infty$.*

(c) *is an essential singularity if $c_n \neq 0$ for infinitely many $n < 0$. In this case, $|\, f(z)\,|$ has no finite or infinite limit as $z \to a$.*

The easy proof is left to the reader. Note that Casorati–Weierstraß implies that $|\, f(z)\,|$ has no limit at a.

Example 14.13 $f(z) = \mathrm{e}^{1/z}$ has in $\mathbb{C}\setminus\{0\}$ the Laurent expansion

$$\mathrm{e}^{\frac{1}{z}} = \sum_{n=0}^{\infty} \frac{1}{n! z^n}, \quad |\, z\,| > 0.$$

Since $c_{-n} \neq 0$ for all n, f has an essential singularity at 0.

14.5 Residues

Throughout $U \subset \mathbb{C}$ is an open connected subset of $\mathbb{C}$.

Definition 14.7 Suppose that $f\colon \overset{\circ}{U}_r(a) \to \mathbb{C}$, is holomorphic, $0 < r_1 < r$ and let

$$f(z) = \sum_{n\in\mathbb{Z}} c_n(z-a)^n, \quad c_n = \frac{1}{2\pi\mathrm{i}} \int_{\mathrm{S}_{r_1}(a)} \frac{f(z)\,\mathrm{d}z}{(z-a)^{n+1}}$$

be the Laurent expansion of f in the annulus $\{z \mid 0 < |\, z-a\,| < r\}$.
Then the coefficient

$$c_{-1} = \frac{1}{2\pi\mathrm{i}} \int_{\mathrm{S}_{r_1}(a)} f(z)\,\mathrm{d}z$$

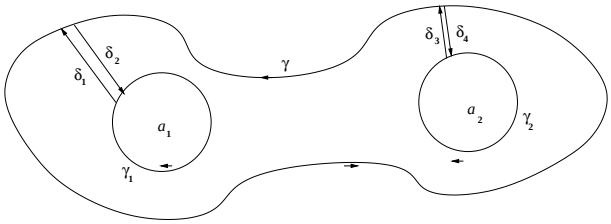
is called the *residue* of f at a and is denoted by $\operatorname*{Res}_a f(z)$ or $\operatorname*{Res}_a f$.

Remarks 14.12 (a) If f is holomorphic at a, $\operatorname*{Res}_a f = 0$ by Cauchy's theorem.
(b) The integral $\int_\gamma f(z)\,\mathrm{d}z$ depends only on the coefficient c_{-1} in the Laurent expansion of $f(z)$ around a. Indeed, every summand $c_n(z-a)^n$, $n \neq 0$, has an antiderivative in $U \setminus \{a\}$ such that the integral over a closed path is 0.
(c) $\operatorname*{Res}_a f + \operatorname*{Res}_a g = \operatorname*{Res}_a f + g$ and $\operatorname*{Res}_a \lambda f = \lambda \operatorname*{Res}_a f$.

Theorem 14.26 (Residue Theorem) *Suppose that $f\colon U \setminus \{a_1, \ldots, a_m\} \to \mathbb{C}$, $a_1, \ldots, a_m \in U$, is holomorphic. Further, let γ be a non-selfintersecting positively oriented closed curve in U such that the points $a_1, \ldots, a_m$ are in the inner part of γ. Then*

$$\int_\gamma f(z)\,\mathrm{d}z = 2\pi\mathrm{i} \sum_{k=1}^{m} \operatorname*{Res}_{a_k} f \tag{14.14}$$

Proof.



As in Remark 14.4 we can replace $\int_\gamma$ by the sum of integrals over small circles, one around each singularity. As before, we obtain

$$\int_\gamma f(z)\,\mathrm{d}z = \sum_{k=1}^{m} \int_{\mathrm{S}_\varepsilon(a_k)} f(z)\,\mathrm{d}z,$$

where all circles are positively oriented. Applying the definition of the residue we obtain the assertion. ∎

Remarks 14.13 (a) The residue theorem generalizes the Cauchy's Theorem, see Theorem 14.6. Indeed, if $f(z)$ possesses an analytic continuation to the points $a_1, \ldots, a_m$, all the residues are zero and therefore $\int_\gamma f(z)\,\mathrm{d}z = 0$.
(b) If $g(z)$ is holomorphic in the region U, $g(z) = \sum_{n=0}^{\infty} c_n(z-a)^n$, $c_0 = g(a)$, then

$$f(z) = \frac{g(z)}{z-a}, \quad z \in U \setminus \{a\}$$

is holomorphic in $U \setminus \{a\}$ with a Laurent expansion around a:

$$f(z) = \frac{c_0}{z-a} + c_1 + c_2(z-a)^2 + \cdots,$$

where $c_0 = g(a) = \operatorname*{Res}_a f$. The residue theorem gives

$$\int_{\mathrm{S}_r(a)} \frac{g(z)}{z-a}\,\mathrm{d}z = 2\pi\mathrm{i} \operatorname*{Res}_a f = 2\pi\mathrm{i}\, c_0 = 2\pi\mathrm{i} g(a).$$

We recovered Cauchy's integral formula.

14.5.1 Calculating Residues

(a) Pole of order 1

As in the previous Remark, suppose that f has a pole of order 1 at a and $g(z) = (z-a)f(z)$ is the corresponding holomorphic function in $U_r(a)$. Then

$$\operatorname*{Res}_a f = g(a) = \lim_{z\to a, z\neq a} g(z) = \lim_{z\to a}(z-a)f(z). \tag{14.15}$$

(b) Pole of order m

Suppose that f has a pole of order m at a. Then

$$f(z) = \frac{c_{-m}}{(z-a)^m} + \frac{c_{-m+1}}{(z-a)^{m-1}} + \cdots + \frac{c_{-1}}{z-a} + c_0 + c_1(z-a) + \cdots, \quad 0 < |\,z-a\,| < r \tag{14.16}$$

is the Laurent expansion of f around a. Multiplying (14.16) by $(z-a)^m$ yields a holomorphic function

$$(z-a)^m f(z) = c_{-m} + c_{-m+1}(z-a) + \cdots c_{-1}(z-a)^{m-1} + \cdots, \quad |\,z-a\,| < r.$$

Differentiating this $(m-1)$ times, all terms having coefficient $c_{-m},\ c_{-m+1}, \ldots, c_{-2}$ vanish and we are left with the power series

$$\frac{\mathrm{d}^{m-1}}{\mathrm{d}z^{m-1}}\left((z-a)^m f(z)\right) = (m-1)!c_{-1} + m(m-1)\cdots 2\, c_0(z-a) + \cdots$$

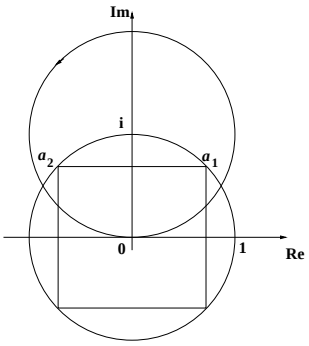
Inserting $z = a$ on the left, we obtain c_{-1}. However, on the left we have to take the limit $z \to a$ since f is not defined at a.
Thus, if f has a pol of order m at a,

$$\operatorname*{Res}_a f(z) = \frac{1}{(m-1)!}\lim_{z\to a}\frac{\mathrm{d}^{m-1}}{\mathrm{d}z^{m-1}}\left((z-a)^m f(z)\right). \tag{14.17}$$

(c) Quotients of Holomorphic Functions

Suppose that $f = \frac{p}{q}$ where p and q are holomorphic at a and q has a zero of order 1 at a, that is $q(a) = 0 \neq q'(a)$. Then, by (a)

$$\operatorname*{Res}_a \frac{p}{q} = \lim_{z\to a}(z-a)\frac{p(z)}{q(z)} = \lim_{z\to a}\frac{p(z)}{\frac{q(z)-q(a)}{z-a}} = \frac{\lim\limits_{z\to a} p(z)}{\lim\limits_{z\to a}\frac{q(z)-q(a)}{z-a}} = \frac{p(a)}{q'(a)}. \tag{14.18}$$



Example 14.14 Compute $\int_{S_1(i)} \frac{dz}{1+z^4}$. The only singularities of $f(z) = 1/(1+z^4)$ inside the disc $\{z \mid |z-i| < 1\}$ are $a_1 = e^{\pi i/4} = (1+i)/\sqrt{2}$ and $a_2 = e^{3\pi i/4} = (-1+i)/\sqrt{2}$. Indeed, $|a_1 - i|^2 = 2 - \sqrt{2} < 1$. We apply the Residue Theorem and (c) and obtain

$$\int_{S_1(i)} \frac{dz}{1+z^4} = 2\pi i \left(\operatorname*{Res}_{a_1} f + \operatorname*{Res}_{a_2} f \right) = 2\pi i \left(\frac{1}{4a_1^3} + \frac{1}{4a_2^3} \right) = 2\pi i \frac{-a_1 - a_2}{4} = \frac{\sqrt{2}\pi}{2}.$$

14.6 Real Integrals

14.6.1 Rational Functions in Sine and Cosine

Suppose we have to compute the integral of such a function over a full period $[0, 2\pi]$. The idea is to replace t by $z = e^{it}$ on the unit circle, $\cos t$ and $\sin t$ by $(z + 1/z)/2$ and $(z - 1/z)/(2i)$, respectively, and finally $dt = dz/(iz)$.

Proposition 14.27 *Supppose that $R(x, y)$ is a rational function in two variables and $R(\cos t, \sin t)$ is defined for all $t \in [0, 2\pi]$. Then*

$$\int_0^{2\pi} R(\cos t, \sin t)\, dt = 2\pi i \sum_{a \in U_1(0)} \operatorname*{Res}_a f(z), \tag{14.19}$$

where

$$f(z) = \frac{1}{iz} R\left(\frac{1}{2}\left(z + \frac{1}{z} \right), \frac{1}{2i}\left(z - \frac{1}{z} \right) \right)$$

and the sum is over all isolated singularities of $f(z)$ in the open unit ball.

Proof. By the residue theorem,

$$\int_{S_1(0)} f(z)\, dz = 2\pi i \sum_{a \in U_1(0)} \operatorname*{Res}_a f$$

Let $z = e^{it}$ for $t \in [0, 2\pi]$. Rewriting the integral on the left using $dz = e^{it} i\, dt = iz\, dt$

$$\int_{S_1(0)} f(z)\, dz = \int_0^{2\pi} R(\cos t, \sin t)\, dt$$

completes the proof. ∎

Example 14.15 For $|a| < 1$,

$$\int_0^{2\pi} \frac{dt}{1 - 2a\cos t + a^2} = \frac{2\pi}{1 - a^2}.$$

For $a = 0$, the statement is trivially true; suppose now $a \neq 0$. Indeed, the complex function corresponding to the integrand is

$$f(z) = \frac{1}{\mathrm{i}z(1 + a^2 - az - a/z)} = \frac{1}{\mathrm{i}(-az^2 - a + (1 + a^2)z)} = \frac{\mathrm{i}/a}{(z - a)\left(z - \frac{1}{a}\right)}.$$

In the unit disc, $f(z)$ has exactly one pole of order 1, namely $z = a$. By (14.15), the formula in Subsection 14.5.1,

$$\operatorname*{Res}_a f = \lim_{z \to a}(z - a)f(z) = \frac{\frac{\mathrm{i}}{a}}{a - \frac{1}{a}} = \frac{\mathrm{i}}{a^2 - 1};$$

the assertion follows from the proposition:

$$\int_0^{2\pi} \frac{\mathrm{d}t}{1 - 2a\cos t + a^2} = 2\pi\mathrm{i}\,\frac{\mathrm{i}}{a^2 - 1} = \frac{2\pi}{1 - a^2}.$$

Specializingi $R = 1$ and $r = a \in \mathbb{R}$ in Homework 49.1, we obtain the same formula.

14.6.2 Integrals of the form $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$

(a) The Principal Value

We often compute improper integrals of the form $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$. Using the residue theorem, we calculate limits

$$\lim_{R \to \infty} \int_{-R}^{R} f(x)\,\mathrm{d}x, \tag{14.20}$$

which is called the *principal value* (or Cauchy mean value) of the integral over $\mathbb{R}$ and we denote it by

$$\operatorname{Vp} \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x.$$

The existence of the "coupled" limit (14.20) in general does not imply the existence of the improper integral

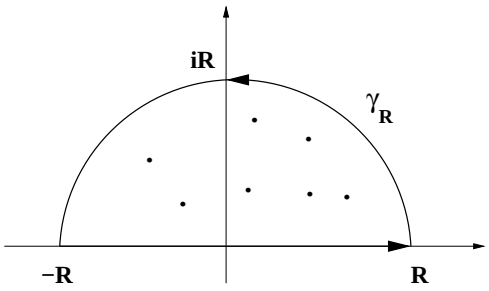
$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \lim_{r \to -\infty} \int_r^0 f(x)\,\mathrm{d}x + \lim_{s \to \infty} \int_0^s f(x)\,\mathrm{d}x.$$

For example, $\operatorname{Vp} \int_{-\infty}^{\infty} x\,\mathrm{d}x = 0$ whereas $\int_{-\infty}^{\infty} x\,\mathrm{d}x$ does not exist since $\int_0^{\infty} x\,\mathrm{d}x = +\infty$. In general, the existence of the improper integral implies the existence of the principal value. If f is an even function or $f(x) \geq 0$, the existence of the principal value implies the existence of the improper integral.

(b) Rational Functions

The main idea to evaluate the integral $\int_{\mathbb{R}} f(x)\,\mathrm{d}x$ is as follows. Let $\mathbb{H} = \{z \mid \operatorname{Im}(z) > 0\}$ be the upper half-plane and $f\colon \mathbb{H} \setminus \{a_1, \ldots, a_m\} \to \mathbb{C}$ be holomorphic. Choose $R > 0$ large

enough such that $|\, a_k \,| < R$ for all $k = 1, \ldots, m$, that is, all isolated singularities of f are in the upper-plane-half-disc of radius R around 0.



Consider the path as in the picture which consists of the segment from $-R$ to R on the real line and the half-circle γ_R of radius R. By the residue theorem,

$$\int_{-R}^{R} f(x)\,\mathrm{d}x + \int_{\gamma_R} f(z)\,\mathrm{d}z = 2\pi\mathrm{i}\sum_{k=1}^{m} \operatorname*{Res}_{a_k} f(z).$$

If

$$\lim_{R\to\infty} \int_{\gamma_R} f(z)\,\mathrm{d}z = 0 \tag{14.21}$$

the above formula implies

$$\lim_{R\to\infty} \int_{-R}^{R} f(x)\,\mathrm{d}x = 2\pi\mathrm{i}\sum_{k=1}^{m} \operatorname*{Res}_{a_k}(z).$$

Knowing the existence of the improper integral $\displaystyle\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$ one has

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 2\pi\mathrm{i}\sum_{k=1}^{m} \operatorname*{Res}_{a_k}(z).$$

Suppose that $f = \frac{p}{q}$ is a rational function such that q has no real zeros and $\deg q \geq \deg p + 2$. Then (14.21) is satisfied. Indeed, since only the two leading terms of p and q determine the the limit behaviour of $f(z)$ for $|\, z \,| \to \infty$, there exists $C > 0$ with $\left| \dfrac{p(z)}{q(z)} \right| \leq \dfrac{C}{R^2}$ on γ_R. Using the estimate $M\ell(\gamma)$ from Remark 14.3 (c) we get

$$\left| \int_{\gamma_R} \frac{p(z)}{q(z)}\,\mathrm{d}z \right| \leq \frac{C}{R^2}\ell(\gamma_R) = \frac{\pi C}{R} \underset{R\to\infty}{\longrightarrow} 0.$$

By the same reason namely $|\, p(x)/q(x) \,| \leq C/x^2$, for large x, the improper real integral exists (comparison test) and converges absolutely. Thus, we have shown the following proposition.

Proposition 14.28 *Suppose that p and q are polynomials with $\deg q \geq \deg p + 2$. Further, q has no real zeros and $a_1, \ldots, a_m$ are all poles of the rational function $f(z) = \frac{p(z)}{q(z)}$ in the open upper half-plane $\mathbb{H}$.*
Then

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 2\pi\mathrm{i}\sum_{k=1}^{m} \operatorname*{Res}_{a_k} f.$$

Example 14.16 (a) $\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{1+x^2}$. The only zero of $q(z) = z^2 + 1$ in $\mathbb{H}$ is $a_1 = \mathrm{i}$ and $\deg(1 + z^2) = 2 \geq \deg(1) + 2$ such that

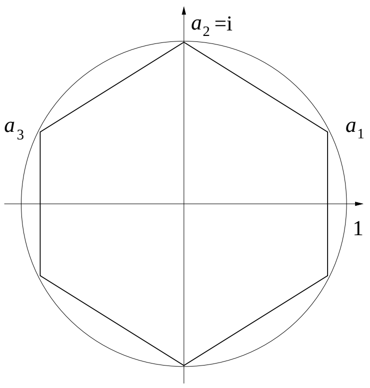
$$\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{1+x^2} = 2\pi\mathrm{i}\operatorname*{Res}_{\mathrm{i}} \frac{1}{1+z^2} = 2\pi\mathrm{i}\left.\frac{1}{1+z}\right|_{z=\mathrm{i}} = \pi.$$

(b) It follows from Example 14.14 that

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{1+x^4} = 2\pi\mathrm{i}\left(\operatorname*{Res}_{a_1} f + \operatorname*{Res}_{a_2} f\right) = \frac{\pi\sqrt{2}}{2}.$$

(c) We compute the integral

$$\int_0^{\infty} \frac{\mathrm{d}t}{1+t^6} = \frac{1}{2}\int_{-\infty}^{\infty} \frac{\mathrm{d}t}{1+t^6}.$$



The zeros of $q(z) = z^6 + 1$ in the upper half-plane are $a_1 = \mathrm{e}^{\mathrm{i}\pi/6}$, $a_2 = \mathrm{e}^{\mathrm{i}\pi/2}$ and $a_3 = \mathrm{e}^{5\mathrm{i}\pi/6}$. They are all of multiplicity 1 such that Formula (14.18) applies:

$$\operatorname*{Res}_{a_k} \frac{1}{q(z)} = \frac{1}{q'(a_k)} = \frac{1}{6a_k^5} = -\frac{a_k}{6}.$$

By Proposition 14.28 and noting that $a_1 + a_3 = \mathrm{i}$,

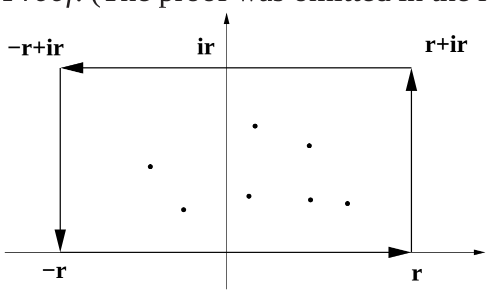
$$\int_0^{\infty} \frac{\mathrm{d}t}{1+t^6} = \frac{1}{2} 2\pi\mathrm{i}\frac{-1}{6}\left(\mathrm{e}^{\mathrm{i}\pi/6} + \mathrm{i} + \mathrm{e}^{5\mathrm{i}\pi/6}\right) = -\frac{\pi\mathrm{i}}{6}2\mathrm{i} = \frac{\pi}{3}.$$

(c) Functions of Type $g(z)\,\mathrm{e}^{\mathrm{i}\alpha z}$

Proposition 14.29 *Suppose that p and q are polynomials with $\deg q \geq \deg p + 1$. Further, q has no real zeros and $a_1, \dots, a_m$ are all poles of the rational function $g = \frac{p}{q}$ in the open upper half-plane $\mathbb{H}$. Put $f(z) = g(z)\,\mathrm{e}^{\mathrm{i}\alpha z}$, where $\alpha \in \mathbb{R}$ is positive $\alpha > 0$.*
Then

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 2\pi\mathrm{i}\sum_{k=1}^{m} \operatorname*{Res}_{a_k} f.$$

Proof. (The proof was omitted in the lecture.)



Instead of a semi-circle it is more appropriate to consider a rectangle now.

According to the residue theorem,

$$\int_{-r}^{r} f(x)\,\mathrm{d}x + \int_{r}^{r+\mathrm{i}r} f(z)\,\mathrm{d}z + \int_{r+\mathrm{i}r}^{r-\mathrm{i}r} f(z)\,\mathrm{d}z + \int_{-r+\mathrm{i}r}^{-r} f(z)\,\mathrm{d}z = 2\pi\mathrm{i}\sum_{k=1}^{m} \operatorname*{Res}_{a_k} f.$$

Since $\deg q \geq \deg p + 1$, $\lim_{z\to\infty} \left|\frac{p(z)}{q(z)}\right| = 0$. Thus, $s_r = \sup\limits_{|z|\geq r} \left|\frac{p(z)}{q(z)}\right|$ exists and tends to 0 as $r \to \infty$.

Consider the second integral I_2 with $z = r + \mathrm{i}t$, $t \in [0, r]$, $\mathrm{d}z = \mathrm{i}\,\mathrm{d}t$. On this segment we have the following estimate

$$\left| \frac{p(z)}{q(z)} \mathrm{e}^{\mathrm{i}\alpha(r+\mathrm{i}t)} \right| \leq s_r\, \mathrm{e}^{-\alpha t}$$

which implies

$$|\, I_2 \,| \leq s_r \int_0^r \mathrm{e}^{-\alpha t}\, \mathrm{d}t = \frac{s_r}{\alpha} \left(1 - \mathrm{e}^{-\alpha r}\right) \leq \frac{s_r}{\alpha}.$$

A similar estimate holds for the fourth integral from $-r + \mathrm{i}r$ to $-r$. In case of the third integral one has $z = t + \mathrm{i}r$, $t \in [-r, r]$, $\mathrm{d}z = \mathrm{d}t$ such that

$$|\, I_3 \,| \leq \int_{-r}^{r} s_r \left| \mathrm{e}^{\mathrm{i}\alpha(t+r\mathrm{i})} \right| \mathrm{d}t = s_r \mathrm{e}^{-\alpha r} \int_{-r}^{r} \mathrm{d}t = 2r s_r \mathrm{e}^{-\alpha r}.$$

Since $2r\mathrm{e}^{-\alpha r}$ is bounded and $s_r \to 0$ as $r \to \infty$, all three integrals I_2, I_3, and I_4 tend to 0 as $r \to \infty$. This completes the proof. ∎

Example 14.17 For $a > 0$,

$$\int_0^\infty \frac{\cos t}{t^2 + a^2}\, \mathrm{d}t = \frac{\pi}{2a} \mathrm{e}^{-a}.$$

Obviously,

$$\int_0^\infty \frac{\cos t}{t^2 + a^2}\, \mathrm{d}t = \frac{1}{2} \operatorname{Re} \left(\int_{-\infty}^{\infty} \frac{\mathrm{e}^{\mathrm{i}t}}{t^2 + a^2}\, \mathrm{d}t \right).$$

The function $f(z) = \dfrac{\mathrm{e}^{\mathrm{i}t}}{t^2 + a^2}$ has a single pole of order 1 in the upper half-plane at $z = a\mathrm{i}$. By formula (14.18)

$$\operatorname*{Res}_{a\mathrm{i}} \frac{\mathrm{e}^{\mathrm{i}z}}{z^2 + a^2} = \left. \frac{\mathrm{e}^{\mathrm{i}z}}{2z} \right|_{z=a\mathrm{i}} = \frac{\mathrm{e}^{-a}}{2a\mathrm{i}}.$$

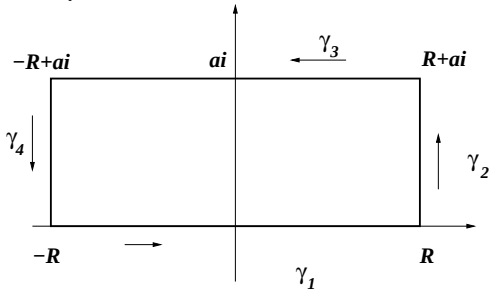
Proposition 14.29 gives the result.

(d) A Fourier transformations

Lemma 14.30 *For $a \in \mathbb{R}$,*

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}x^2 - \mathrm{i}ax}\, \mathrm{d}x = \mathrm{e}^{-\frac{1}{2}a^2}. \tag{14.22}$$

Proof.



Let $f(z) = \mathrm{e}^{-\frac{1}{2}z^2}$, $z \in \mathbb{C}$ and γ the closed rectangular path $\gamma = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$ as in the picture. Since f is an entire function, by Cauchy's theorem $\int_\gamma f(z)\, \mathrm{d}z = 0$. Note that γ_2 is parametrized as $z = R + t\mathrm{i}$, $t \in [0, a]$, $\mathrm{d}z = \mathrm{i}\,\mathrm{d}t$, such that

$$\int_{\gamma_2} f(z)\,\mathrm{d}z = \int_0^a \mathrm{e}^{-\frac{1}{2}(R^2+\mathrm{i}t)^2}\mathrm{i}\,\mathrm{d}t = \int_0^a \mathrm{e}^{-\frac{1}{2}(R^2+2R\mathrm{i}t-t^2)}\mathrm{i}\,\mathrm{d}t$$
$$\left|\int_{\gamma_2} f(z)\,\mathrm{d}z\right| \le \int_0^a \mathrm{e}^{-\frac{1}{2}R^2+\frac{1}{2}t^2}\,\mathrm{d}t = \mathrm{e}^{-\frac{1}{2}R^2}\int_0^a \mathrm{e}^{\frac{1}{2}t^2}\,\mathrm{d}t = C\mathrm{e}^{-\frac{1}{2}R^2}.$$

Since $\mathrm{e}^{-\frac{1}{2}R^2}$ tends to 0 as $R \to \infty$, the above integral tends to 0, as well; hence $\lim_{R\to\infty}\int_{\gamma_2} f(z)\,\mathrm{d}z = 0$. Similarly, one can show that $\lim_{R\to\infty}\int_{\gamma_4} f(z)\,\mathrm{d}z = 0$. Since $\int_\gamma f(z)\,\mathrm{d}z = 0$, we have $\int_{\gamma_1+\gamma_3} f(z)\,\mathrm{d}z = 0$, that is

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x+a\mathrm{i})\,\mathrm{d}x.$$

Using

$$\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x = \sqrt{2\pi},$$

which follows from Example 14.7, page 374, or from homework 41.3, we have

$$\sqrt{2\pi} = \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}x^2}\,\mathrm{d}x = \int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}(x^2+2\mathrm{i}ax-a^2)}\,\mathrm{d}x = \mathrm{e}^{\frac{1}{2}a^2}\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}x^2-\mathrm{i}ax}\,\mathrm{d}x$$
$$\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{e}^{-\frac{1}{2}x^2-\mathrm{i}ax}\,\mathrm{d}x = \mathrm{e}^{-\frac{1}{2}a^2}.$$

■

Chapter 15

Partial Differential Equations I — an Introduction

15.1 Classification of PDE

15.1.1 Introduction

There is no general theory known concerning the solvability of all PDE. Such a theory is extremely unlikely to exist, given the rich variety of physical, geometric, probabilistic phenomena which can be modelled by PDE. Instead, research focuses on various particular PDEs that are important for applications in mathematics and physics.

Definition 15.1 A *partial differential equation* (abbreviated as PDE) is an equation of the form

$$F(x, y, \dots, u, u_x, u_y, \dots, u_{xx}, u_{xy}, \dots) = 0 \tag{15.1}$$

where F is a given function of the independent variables $x, y, \dots$ of the unknown function u and a finite number of its partial derivatives.
We call u a *solution* of (15.1) if after substitution of $u(x, y, \dots)$ and its partial derivatives (15.1) is identically satisfied in some region Ω in the space of the independent variables $x, y, \dots$. The *order* of a PDE is the order of the highest derivative that occurs.

A PDE is called *linear* if it is linear in the unknown function u and their derivatives u_x, u_y, u_{xy}, $\dots$, with coefficients depending only on the variables $x, y, \dots$. In other words, a linear PDE can be written in the form

$$G(u, u_x, u_y, \dots, u_{xx}, u_{xy}, \dots) = f(x, y, \dots), \tag{15.2}$$

where the function f on the right depends only on the variables $x, y, \dots$ and G is linear in all components with coefficients depending on $x, y, \dots$. More precisely, the formal *differential operator* $L(u) = G(u, u_x, u_y, \dots, u_{xx}, u_{xy}, \dots)$ which associates to each function $u(x, y, \dots)$ a new function $L(u)(x, y, \dots)$ is a linear operator. The linear PDE (15.2) ($L(u) = f$) is called *homogeneous* if $f = 0$ and *inhomogeneous* otherwise. For example, $\cos(xy^2)u_{xxy} - y^2 u_x +$

$u \sin x + \tan(x^2 + y^2) = 0$ is a linear inhomogeneous PDE of order 3, the corresponding homogeneous linear PDE is $\cos(xy^2)u_{xxy} - y^2 u_x + u \sin x = 0$.
A PDE is called *quasi-linear* if it is linear in all partial derivatives of order m (the order of the PDE) with coefficients which depend on the variables $x, y, \cdots$ and partial derivatives of order less than m; for example $u_x u_{xx} + u^2 = 0$ is quasi-linear, $u_{xy} u_{xx} + 1 = 0$ not. *Semi-linear* equations are those quasi-linear equation in which the coefficients of the highest order terms does not depend on u and its partial derivatives; $\sin x u_{xx} + u^2 = 0$ is semi-linear; $u_x u_{xx} + u^2 = 0$ not. Sometimes one considers *systems* of PDEs involving one or more unknown functions and their derivatives.

15.1.2 Examples

(1) The **Laplace equation** in n dimensions for a function $u(x_1, \ldots, x_n)$ is the linear second order equation

$$\Delta u = u_{x_1 x_1} + \cdots + u_{x_n x_n} = 0.$$

The solutions u are called *harmonic* (or *potential*) functions. In case $n = 2$ we associate with a harmonic function $u(x, y)$ its "conjugate" harmonic function $v(x, y)$ such that the first-order system of Cauchy–Riemann equations

$$u_x = v_y, \quad u_y = -v_x$$

is satisfied. A real solution (u, v) gives rise to the analytic function $f(z) = u + \mathrm{i}v$. The **Poisson equation** is

$$\Delta u = f, \quad \text{for a given function} \quad f \colon \Omega \to \mathbb{R}.$$

The Laplace equation models equilibrium states while the Poisson equation is important in electrostatics. Laplace and Poisson equation always describe stationary processes (there is no time dependence).

(2) The **heat equation.** Here one coordinate t is distinguished as the "time" coordinate, while the remaining coordinates $x_1, \ldots, x_n$ represent spatial coordinates. We consider

$$u \colon \Omega \times \mathbb{R}^+ \to \mathbb{R}, \quad \Omega \text{ open in } \mathbb{R}^n,$$

where $\mathbb{R}^+ = \{t \in \mathbb{R} \mid t > 0\}$ is the positive time axis and pose the equation

$$k u_t = \Delta u, \quad \text{where} \quad \Delta u = u_{x_1 x_1} + \cdots + u_{x_n x_n}.$$

The heat equation models heat conduction and other diffusion processes.

(3) The **wave equation.** With the same notations as in (2), here we have the equation

$$u_{tt} - a^2 \Delta u = 0.$$

It models wave and oscillation phenomena.

(4) The **Korteweg–de Vries equation**

$$u_t - 6u\,u_x + u_{xxx} = 0$$

models the propagation of waves in shallow waters.

(5) The **Monge–Ampère equation**

$$u_{xx}u_{yy} - u_{xy}^2 = f$$

with a given function f, is used for finding surfaces with prescribed curvature.

(6) The **Maxwell equations** for the electric field strength $E = (E_1, E_2, E_3)$ and the magnetic field strength $B = (B_1, B_2, B_3)$ as functions of (t, x_1, x_2, x_3):

$$\begin{aligned} \operatorname{div} B &= 0, && \text{(magnetostatic law)},\\ B_t + \operatorname{curl} E &= 0, && \text{(magnetodynamic law)},\\ \operatorname{div} E &= 4\pi\rho, && \text{(electrostatic law, } \rho = \text{charge density)},\\ E_t - \operatorname{curl} B &= -4\pi j && \text{(electrodynamic law, } j = \text{current density)} \end{aligned}$$

(7) The **Navier–Stokes equations** for the velocity $v(x,t) = (v^1, v^2, v^3)$ and the pressure $p(x,t)$ of an incompressible fluid of density ρ and viscosity η:

$$\rho v_t^j + \rho \sum_{i=1}^{3} v^i\, v_{x_i}^j - \eta \Delta v^j = -p_{x_j}, \quad j = 1, 2, 3,$$

$$\operatorname{div} v = 0.$$

(8) The **Schrödinger equation**

$$\mathrm{i}\hbar u_t = -\frac{\hbar^2}{2m}\Delta u + V(x,u)$$

(m = mass, V = given potential, $u\colon \Omega \to \mathbb{C}$) from quantum mechanics is formally similar to the heat equation, in particular in the case $V = 0$. The factor i, however, leads to crucial differences.

Classification

We have seen so many rather different-looking PDEs, and it is hopeless to develop a theory that can treat all these diverse equations. In order to proceed we want to look for criteria to classify PDEs. Here are some possibilities

(I) *Algebraically.*

(a) Linear equations are (1), (2), (3), (6) which is of first order, and (8)

(b) semi-linear equations are (4) and (7)

(c) a non-linear equation is (5)

naturally, linear equations are simple than non-linear ones. We shall therefore mostly study linear equations.

(II) The *order* of the equation. The Cauchy–Riemann equations and the Maxwell equations are linear first order equations. (1), (2), (3), (5), (7), (8) are of second order; (4) is of third order. Equations of higher order rarely occur. The most important PDEs are second order PDEs.

(III) *Elliptic, parabolic, hyperbolic.* In particular, for the second order equations the following classification turns out to be useful: Let $x = (x_1, \dots, x_n) \in \Omega$ and

$$F(x, u, u_{x_i}, u_{x_i x_j}) = 0$$

be a second-order PDE. We introduce auxiliary variables p_i, p_{ij}, $i, j = 1, \dots, n$, and study the function $F(x, u, p_i, p_{ij})$. The equation is called *elliptic* in Ω if the matrix

$$F_{p_{ij}}(x, u(x), u_{x_i}(x), u_{x_i x_j}(x))_{i,j=1,\dots,n}$$

of the first derivatives of F with respect to the variables p_{ij} is positive definite or negative definite for all $x \in \Omega$.

this may depend on the function u. The Laplace equation is the prominent example of an elliptic equation. Example (5) is elliptic if $f(x) > 0$.
The equation is called *hyperbolic* if the above matrix has precisely one negative and $d-1$ positive eigenvalues (or conversely, depending on the choice of the sign). Example (3) is hyperbolic and so is (5) if $f(x) < 0$.
Finally, the equation is *parabolic* if one eigenvalue of the above matrix is 0 and all the other eigenvalues have the same sign. More precisely, the equation can be written in the form

$$u_t = F(t, x, u, u_{x_i}, u_{x_i x_j})$$

with an elliptic F.

(IV) According to *solvability*. We consider the second-order PDE $F(x, u, u_{x_i}, u_{x_i x_j}) = 0$ for $u\colon \Omega \to \mathbb{R}$, and wish to impose additional conditions upon the solution u, typically prescribing the values of u or of certain first derivatives of u on the boundary $\partial\Omega$ or part of it.

Ideally such a boundary problem satisfies the three conditions of Hadamard for a *well-posed problem*

(a) Existence of a solution u for the given boundary values;

(b) Uniqueness of the solution;

(c) Stability, meaning continuous dependence on the boundary values.

Example 15.1 In the following examples $\Omega = \mathbb{R}^2$ and $u = u(x, y)$.
(a) Find all solutions $u \in \mathrm{C}^2(\mathbb{R}^2)$ with $u_{xx} = 0$. We first integrate with respect to x and find that u_x is independent on x, say $u_x = a(y)$. We again integrate with respect to x and obtain $u(x, y) = xa(y) + b(y)$ with arbitrary functions a and b. Note that the ODE $u'' = 0$ has the general solution $ax + b$ with coefficients a, b. Now the coefficients are *functions* on y.
(b) Solve $u_{xx} + u = 0$, $u \in \mathrm{C}^2(\mathbb{R}^2)$. The solution of the corresponding ODE $u'' + u = 0$, $u = u(x)$, $u \in \mathrm{C}^2(\mathbb{R})$, is $a\cos x + b\sin x$ such that the general solution of the corresponding PDE in 2 variables x and y is $a(y)\cos x + b(y)\sin x$ with arbitrary functions a and b.
(c) Solve $u_{xy} = 0$, $u \in \mathrm{C}^2(\mathbb{R}^2)$. First integrate $\frac{\partial}{\partial y}(u_x) = 0$ with respect to y. we obtain $u_x = \tilde{f}(x)$. Integration with respect to x yields $u = \int \tilde{f}(x)\,\mathrm{d}x + g(y) = f(x) + g(y)$, where f is differentiable and g is arbitrary.

15.2 First Order PDE — The Method of Characteristics

We solve first order PDE by the method of *chracteristics*. It applies to quasi-linear equations

$$a(x, y, u)u_x + b(x, y, u)u_y = c(x, y, u) \tag{15.3}$$

as well as to the linear equation

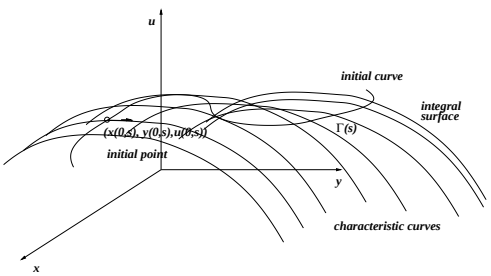
$$a(x, y)u_x + b(x, y)u_y = c_0(x, y)u + c_1(x, y). \tag{15.4}$$

We restrict ourselves to the linear equation with an *initial condition* given as a parametric curve in the xyu-space

$$\Gamma = \Gamma(s) = (x_0(s), y_0(s), u_0(s)), \quad s \in (a, b) \subseteq \mathbb{R}. \tag{15.5}$$

The curve Γ will be called the *initial curve*. The initial condition then reads

$$u(x_0(s), y_0(s)) = u_0(s), \quad s \in (a, b).$$



The geometric idea behind this method is the following. The solution $u = u(x, y)$ can be thought as surface in $\mathbb{R}^3 = \{(x, y, u) \mid x, y, u \in \mathbb{R}^3\}$. Starting from a point on the initial curve, we construct a *chracteristic curve* in the surface u. If we do so for any point of the initial curve, we obtain a one-parameter family of characteristic curves; glueing all these curves we get the solution surface u.

The linear equation (15.4) can be rewritten as

$$(a, b, c_0\,u + c_1)\cdot(u_x, u_y, -1) = 0. \tag{15.6}$$

Recall that $(u_x, u_y, -1)$ is the normal vector to the surface $(x, y, u(x, y))$, that is, the tangent equation to u at (x_0, y_0, u_0) is

$$u - u_0 = u_x\,(x - x_0) + u_y(y - y_0) \Leftrightarrow (x - x_0, y - y_0, u - u_0)\cdot(u_x, u_y, -1) = 0.$$

It follows from (15.6) that $(a, b, c_0u + c_1)$ is a vector in the tangent plane. Finding a curve $(x(t), y(t), u(t))$ with exactely this tangent vector

$$(a(x(t), y(t)), b(x(t), y(t)), c_0(x(t), y(t))u(t) + c_1(x(t), y(t)))$$

is equivalent to solve the ODE

$$x'(t) = a(x(t), y(t)), \tag{15.7}$$
$$y'(t) = b(x(t), y(t)), \tag{15.8}$$
$$u'(t) = c_0(x(t), y(t))u(t) + c_1(x(t), y(t))). \tag{15.9}$$

This system is called the *characteristic equations*. The solutions are called *characteristic curves* of the equation. Note that the above system is autonomous, i. e. there is no explicit dependence on the parameter t.
In order to determine characteristic curves we need an initial condition. We shall require the initial point to lie on the initial curve $\Gamma(s)$. Since each curve $(x(t), y(t), u(t))$ emanates from a different point $\Gamma(s)$, we shall explicitely write the curves in the form $(x(t, s), y(t, s), u(t, s))$. The initial conditions are written as

$$x(0, s) = x_0(s), \quad y(0, s) = y_0(s), \quad u(0, s) = u_0(s).$$

Notice that we selected the parameter t such that the characteristic curve is located at the initial curve at $t = 0$. Note further that the parametrization $(x(t, s), y(t, s), u(t, s))$ represents a surface in $\mathbb{R}^3$.
The method of characteristics also applies to quasi-linear equations.
To summarize the method: In the first step we identify the initial curve Γ. In the second step we select a point s on Γ as initial point and solve the characterictic equations using the point we selected on Γ as an initial point. After preforming the steps for all points on Γ, we obtain a portion of the solution surface, also called *integral surface*. That consists of the union of the characteristic curves.

Example 15.2 Solve the equation

$$u_x + u_y = 2$$

subject to the initial condition $u(x, 0) = x^2$. The characteristic equations and the parametric initial conditions are

$$x_t(t, s) = 1, \qquad y_t(t, s) = 1, \qquad u_t(t, s) = 2,$$
$$x(0, s) = s, \qquad y(0, s) = 0, \qquad u(0, s) = s^2.$$

It is easy to solve the characteristic equations:

$$x(t, s) = t + f_1(s), \qquad y(t, s) = t + f_2(s), \qquad u(t, s) = 2t + f_3(s).$$

Inserting the initial conditions, we find

$$x(t,s) = t + s, \qquad y(t,s) = t, \qquad u(t,s) = 2t + s^2.$$

We have obtained a parametric representation of the integral surface. To find an explicit representation we have to invert the transformation $(x(t,s), y(t,s))$ in the form $(t(x,y), s(x,y))$, namely, we have to solve for s and t. In the current example, we find $t = y$, $s = x - y$. Thus the explicit formula for the integral surface is

$$u(x,y) = 2y + (x-y)^2.$$

Remark 15.1 (a) This simple example might lead us to think that each initial value problem for a first-order PDE possesses a unique solution. But this is not the case Is the problem(15.3) together with the initial condition (15.5) well-posed? Under which conditions does there exists a unique integral surface that contains the initial curve?
(b) Notice that even if the PDE is linear, the characteristic equations are non-linear. It follows that one can expect at most a local existanece theorem for a first ordwer PDE.
(c) The inversion of the parametric presentation of the integral surface might hide further difficulties. Recall that the implicit function theorem implies that the inversion locally exists if the Jacobian $\frac{\partial(x,y)}{\partial(t,s)} \neq 0$. An explicit computation of the Jacobian at a point s of the initial curve gives

$$J = \frac{\partial x}{\partial t}\frac{\partial y}{\partial s} - \frac{\partial x}{\partial s}\frac{\partial y}{\partial t} = ay_0' - bx_0' = \begin{vmatrix} a & b \\ x_0' & y_0' \end{vmatrix}.$$

Thus, the Jacobian vanishes at some point if and only if the vectors (a,b) and (x_0', y_0') are linearly dependent. The geometrical meaning of $J = 0$ is that the projection of Γ into the xy plane is tangent to the projection of the characteristic curve into the xy plane. To ensure a unique solution near the initial curve we must have $J \neq 0$. This condition is called the *transersality condition*.

Example 15.3 (Well-posed and Ill-posed Problems) (a) Solve $u_x = 1$ subject to the initial condition $u(0,y) = g(y)$. The characteristic equations and the inition conditions are given by

$$\begin{aligned} x_t &= 1, & y_t(t,s) &= 0, & u_t(t,s) &= 1, \\ x(0,s) &= 0, & y(0,s) &= s, & u(0,s) &= g(s). \end{aligned}$$

The parametric integral surface is $(x(t,s), y(t,s), u(t,s)) = (t, s, t+g(s))$ such that the explicit solution is $u(x,y) = x + g(y)$.
(b) If we keep $u_x = 1$ but modify the initial condition into $u(x,0) = h(x)$, the picture changess dramatically.

$$\begin{aligned} x_t &= 1, & y_t(t,s) &= 0, & u_t(t,s) &= 1, \\ x(0,s) &= s, & y(0,s) &= 0, & u(0,s) &= h(s). \end{aligned}$$

In this case the parametric solution is

$$(x(t,s), y(t,s), u(t,s)) = (t+s, 0, t+h(s)).$$

Now, however, the transformation $x = t + s$, $y = 0$ cannot be inverted. Geometrically, the projection of the initial curve is the x axis, but this is also the projection of the characteristic curve. In the speial case $h(x) = x + c$ for some constant c, we obtain $u(t, s) = t + s + c$. Then it is not necessary to invert (x, y) since we find at once $u = x + c + f(y)$ for every differentiable function f with $f(0) = 0$. We have infinitely many solutions — **uniqueness fails**.
(c) However, for any other choice of h **Existence fails** — the problem has no solution at all. Note that the Jacobian is

$$J = \begin{vmatrix} a & b \\ x_0' & y_0' \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 0 \end{vmatrix} = 0.$$

Remark 15.2 Because of the special role played by the *projecions* of the characteristics on the xy plane, we also use the term characteristics to denote them. In case of the linear PDE (15.4) the ODE for the projection is

$$x'(t) = \frac{\mathrm{d}x}{\mathrm{d}t} = a(x(t), y(t)), \quad y'(t) = \frac{\mathrm{d}y}{\mathrm{d}t} = b(x(t), y(t)), \tag{15.10}$$

which yields $y'(x) = \dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{b(x, y)}{a(x, y)}$.

15.3 Classification of Semi-Linear Second-Order PDEs

15.3.1 Quadratic Forms

We recall some basic facts about quadratic forms and symmetric matrices.

Proposition 15.1 (Sylvester's Law of Inertia) *Suppose that $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix.*
(a) *Then there exist an invertible matrix $B \in \mathbb{R}^{n\times n}$, $r, s, t \in \mathbb{N}_0$ with $r + s + t = n$ and a diagonal matrix $D = \operatorname{diag}(d_1, d_2, \ldots, d_{r+s}, 0, \ldots, 0)$ with $d_i > 0$ for $i = 1, \ldots, r$ and $d_i < 0$ for $i = r + 1, \ldots, r + s$ and*

$$BAB^\top = \operatorname{diag}(d_1, \ldots, d_{r+s}, 0, \ldots, 0).$$

We call (r, s, t) the signature of A.
(b) *The signature does not depend on the change of coordinates, i. e. If there exist another regular matrix B' and a diagonal matrix D' with $D' = B'A(B')^\top$ then the signature of D' and D coincide.*

15.3.2 Elliptic, Parabolic and Hyperbolic

Consider the semi-linear second-order PDE in n variables $x_1, \ldots, x_n$ in a region $\Omega \subset \mathbb{R}^n$

$$\sum_{i,j=1}^{n} a_{ij}(x)\, u_{x_i x_j} + F(x, u, u_{x_1}, \ldots, u_{x_n}) = 0 \tag{15.11}$$

with continuous coefficients $a_{ij}(x)$. Since we assume $u \in \mathrm{C}^2(\Omega)$, by Schwarz's lemma we assume without loss of generality that $a_{ij} = a_{ji}$. Using the terminology of the introduction

(Classification (III), see page 404) we find that the matrix $A(x) := (a_{ij}(x))_{i,j=1,\dots,n}$, coincides with the matrix $(F_{p_{ij}})_{i,j}$ defined therein.

Definition 15.2 We call the PDE (15.11) *elliptic* at x_0 if the matrix $A(x_0)$ is positive definite or negative definite. We call it *parabolic* at x_0 if $A(x_0)$ is positive or negative semidefinite with exactly one eigenvalue 0. We call it *hyperbolic* if $A(x_0)$ has the signature $(n-1,1,0)$, i. e. A is indefinite with $n-1$ positive eigenvalues and one negative eigenvalue and no zero eigenvalue (or vice versa).

15.3.3 Change of Coordinates

First we study how the coefficients a_{ij} will change if we impose a non-singular transformation of coordinates $y = \varphi(x)$;

$$y_l = \varphi_l(x_1, \dots, x_n), \quad l = 1, \dots, n;$$

The transformation is called *non-singular* if the Jacobian $\frac{\partial(\varphi_1,\dots,\varphi_n)}{\partial(x_1,\dots,x_n)}(x_0) \neq 0$ is non-zero at any point $x_0 \in \Omega$. By the Inverse Mapping Theorem, the transformation possesses locally an inverse transformation denoted by $x = \psi(y)$

$$x_l = \psi_l(y_1, \dots, y_n), \quad l = 1, \dots, n.$$

Putting

$$\tilde{u}(y) := u(\psi(y)), \quad \text{then} \quad u(x) = \tilde{u}(\varphi(x))$$

and if moreover $\varphi_l \in \mathrm{C}^2(\Omega)$ we have by the chain rule

$$\begin{aligned} u_{x_i} &= \sum_{l=1}^{n} \tilde{u}_{y_l} \frac{\partial \varphi_l}{\partial x_i}, \\ u_{x_i x_j} = (u_{x_i})_{x_j} &= \sum_{k,l=1}^{n} \tilde{u}_{y_l y_k} \frac{\partial \varphi_l}{\partial x_i} \frac{\partial \varphi_k}{\partial x_j} + \sum_{l=1}^{n} \tilde{u}_{y_l} \frac{\partial^2 \varphi_l}{\partial x_i \partial x_j}. \end{aligned} \tag{15.12}$$

Inserting (15.12) into (15.11) one has

$$\sum_{k,l=1}^{n} \tilde{u}_{y_l y_k} \sum_{i,j=1}^{n} a_{ij} \frac{\partial \varphi_l}{\partial x_i} \frac{\partial \varphi_k}{\partial x_j} + \sum_{l=1}^{n} \tilde{u}_{y_l} \sum_{i,j=1}^{n} a_{ij} \frac{\partial^2 \varphi_l}{\partial x_i \partial x_j} + \tilde{F}(y, \tilde{u}, \tilde{u}_{y_1}, \dots, \tilde{u}_{y_n}) = 0. \tag{15.13}$$

We denote by $\tilde{a}_{lk}$ the new coefficients of the partial second derivatives of $\tilde{u}$,

$$\tilde{a}_{lk} = \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial \varphi_l}{\partial x_i} \frac{\partial \varphi_k}{\partial x_j}, \tag{15.14}$$

and write (15.13) in the same form as (15.11)

$$\sum_{k,l=1}^{n} \tilde{a}_{lk}(y) \tilde{u}_{y_l y_k} + \tilde{F}(y, \tilde{u}, \tilde{u}_{y_1}, \dots, \tilde{u}_{y_n}) = 0.$$

Equation (15.14) later plays a crucial role in simplifying PDE (15.11). Namely, if we want some of the coefficients $\tilde{a}_{lk}$ to be 0, the right hand side of (15.14) has to be 0. Writing

$$b_{lj} = \frac{\partial \varphi_l}{\partial x_j}, \quad l, j = 1, \ldots, n, \quad B = (b_{lj}),$$

the new coefficient matrix $\tilde{A}(y) = (\tilde{a}_{lk}(y))$ reads as follows

$$\tilde{A} = B \cdot A \cdot B^\top.$$

By Proposition 15.1, A and $\tilde{A}$ have the same signature. We have shown the following proposition.

Proposition 15.2 *The type of a semi-linear second order PDE is invariant under the change of coordinates.*

Notation. We call the operator L with

$$L(u) = \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + F(x, u, u_{x_1}, \ldots, u_{x_n})$$

differential operator and denote by L_2

$$L_2(u) = \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j}$$

the sum of its the highest order terms; L_2 is a linear operator.

Definition 15.3 The second-order PDE $L(u) = 0$ has *normal form* if

$$L_2(u) = \sum_{j=1}^{m} u_{x_j x_j} - \sum_{j=m+1}^{r} u_{x_j x_j}$$

with some positive integers $m \leq r \leq n$.

Remarks 15.3 (a) It happens that the type of the equation depends on the point $x_0 \in \Omega$. For example, the *Trichomi equation*

$$y u_{xx} + u_{yy} = 0$$

is of mixed type. More precisely, it is elliptic if $y > 0$, parabolic if $y = 0$ and hyperbolic if $y < 0$.
(b) The Laplace equation is elliptic, the heat equation is parabolic, the wave equation is hyperbolic.
(c) The classification is not complete in case $n \geq 3$; for example, the quadratic form can be of type $(n-2, 1, 1)$.
(d) Case $n = 2$. The PDE

$$a u_{xx} + 2b u_{xy} + c u_{yy} + F(x, y, u, u_x, u_y) = 0$$

with coefficients $a = a(x, y)$, $b = b(x, y)$ and $c = c(x, y)$ is elliptic, parabolic or hyperbolic at (x_0, y_0) if and only if $ac - b^2 > 0$, $ac - b^2 = 0$ or $ac - b^2 < 0$ at (x_0, y_0), respectively.

15.3.4 Characteristics

Suppose we are given the semi-linear second-order PDE in $\Omega \subset \mathbb{R}^n$

$$\sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + F(x, u, u_{x_1}, \ldots, u_{x_n}) = 0 \tag{15.15}$$

with continuous a_{ij}; $a_{ij}(x) = a_{ji}(x)$.
We define the concept of characteristics which plays an important role in the theory of PDEs, not only of second-order PDEs.

Definition 15.4 Suppose that $\sigma \in \mathrm{C}^1(\Omega)$ is continuously differentiable, $\operatorname{grad} \sigma \neq 0$, and
(a) for some point x_0 of the hypersurface $\mathcal{F} = \{x \in \Omega \mid \sigma(x) = c\}$, $c \in \mathbb{R}$, we have

$$\sum_{i,j=1}^{n} a_{ij}(x_0) \, \frac{\partial \sigma(x_0)}{\partial x_i} \, \frac{\partial \sigma(x_0)}{\partial x_j} = 0. \tag{15.16}$$

Then $\mathcal{F}$ is said to be *characteristic* at x_0.
(b) If $\mathcal{F}$ is characteristic at every point of $\mathcal{F}$, $\mathcal{F}$ is called a *characteristic hypersurface* or simply a *characteristic* of the PDE (15.11). Equation (15.16) is called the *characteristic equation* of (15.11).

In case $n = 2$ we speak of *characteristic lines*.
If all hypersurfaces $\sigma(x) = c$, $a < c < b$, are characteristic, this family of hypersurfaces fills the region Ω such that for any point $x \in \Omega$ there is *exactly one* hypersurface with $\sigma(x) = c$. This c can be chosen to be one new coordinate. Setting

$$y_1 = \sigma(x)$$

we see from (15.14) that $\tilde{a}_{11} = 0$. That is, the knowledge of one or more characteristic hypersurfaces can simplify the PDE.

Example 15.4 (a) The characteristic equation of $u_{xy} = 0$ is $\sigma_x \sigma_y = 0$ such that $\sigma_x = 0$ and $\sigma_y = 0$ define the characteristic lines; the parallel lines to the coordinate axes, $y = c_1$ and $x = c_2$, are the characteristics.
(b) Find type and characteristic lines of

$$x^2 u_{xx} - y^2 u_{yy} = 0, \quad x \neq 0, \, y \neq 0.$$

Since $\det = x^2(-y^2) - 0 = -x^2 y^2 < 0$, the equation is hyperbolic. The characteristic equation, in the most general case, is

$$a\sigma_x^2 + 2b\sigma_x \, \sigma_y + c\sigma_y^2 = 0.$$

Since $\operatorname{grad} \sigma \neq 0$, $\sigma(x, y) = c$ is locally solvable for $y = y(x)$ such that $y' = -\sigma_x/\sigma_y$. Another way to obtain this is as follows: Differentiating the equation $\sigma(x, y) = c$ yields $\sigma_x \, \mathrm{d}x + \sigma_y \, \mathrm{d}y = 0$ or $\mathrm{d}y/\,\mathrm{d}x = -\sigma_x/\sigma_y$. Inserting this into the previous equation we obtain a quadratic ODE

$$a(y')^2 - 2by' + c = 0,$$

with solutions

$$y' = \frac{b \pm \sqrt{b^2 - ac}}{a}, \quad \text{if} \quad a \neq 0.$$

We can see, that the elliptic equation has no characteristic lines, the parabolic equation has one family of characteristics, the hyperbolic equation has two families of characteristic lines.

Hyperbolic case. In general, if $c_1 = \varphi_1(x, y)$ is the first family of characteristic lines and $c_2 = \varphi_2(x, y)$ is the second family of characteristic lines,

$$\xi = \varphi_1(x, y), \quad \eta = \varphi_2(x, y)$$

is the appropriate change of variable which gives $\tilde{a} = \tilde{c} = 0$. The transformed equation then reads

$$2\tilde{b}\,\tilde{u}_{\xi\eta} + F(\xi, \eta, \tilde{u}, \tilde{u}_\xi, \tilde{u}_\eta) = 0.$$

Parabolic case. Since $\det A = 0$, there is only one real family of characteristic hyperplanes, say $c_1 = \varphi_1(x, y)$. We impose the change of variables

$$z = \varphi_1(x, y), \quad y = y.$$

Since $\det \tilde{A} = 0$, the coefficients $\tilde{b}$ vanish (together with $\tilde{a}$). The transformed equation reads

$$\tilde{c}\,\tilde{u}_{yy} + F(z, y, \tilde{u}, \tilde{u}_z, \tilde{u}_y) = 0.$$

The above two equations are called the *characteristic forms* of the PDE (15.11)
In our case the characteristic equation is

$$x^2(y')^2 - y^2 = 0, \quad y' = \pm y/x.$$

This yields

$$\frac{\mathrm{d}y}{y} = \pm\frac{\mathrm{d}x}{x}, \qquad \log|\,y\,| = \pm\log|\,x\,| + c_0.$$

We obtain the two families of characteristic lines

$$y = c_1 x, \quad y = \frac{c_2}{x}.$$

Indeed, in our example

$$\xi = \frac{y}{x} = c_1, \quad \eta = xy = c_2$$

gives

$$\begin{array}{lllll} \eta_x = y, & \eta_y = x, & \eta_{xx} = 0, & \eta_{yy} = 0, & \eta_{xy} = 1, \\ \xi_x = -\dfrac{y}{x^2}, & \xi_y = \dfrac{1}{x}, & \xi_{xx} = 2\dfrac{y}{x^3}, & \xi_{yy} = 0, & \xi_{xy} = -\dfrac{1}{x^2}. \end{array}$$

In our case (15.12) reads

$$\begin{aligned} u_{xx} &= \tilde{u}_{\xi\xi}\,\xi_x^2 + 2\tilde{u}_{\xi\eta}\,\xi_x\eta_x + \tilde{u}_{\eta\eta}\,\eta_x^2 + \tilde{u}_\xi\,\xi_{xx} + \tilde{u}_\eta\,\eta_{xx} \\ u_{yy} &= \tilde{u}_{\xi\xi}\,\xi_y^2 + 2\tilde{u}_{\xi\eta}\,\xi_y\eta_y + \tilde{u}_{\eta\eta}\,\eta_y^2 + \tilde{u}_\xi\,\xi_{yy} + \tilde{u}_\eta\,\eta_{yy} \end{aligned}$$

Noting $x^2 = \eta/\xi$, $y^2 = \xi\eta$ and inserting the values of the partial derivatives of ξ and η we get

$$u_{xx} = \tilde{u}_{\xi\xi}\frac{y^2}{x^4} - 2\frac{y^2}{x^2}\tilde{u}_{\xi\eta} + \tilde{u}_{\eta\eta}y^2 + 2\frac{y}{x^3}\tilde{u}_{\xi},$$
$$u_{yy} = \tilde{u}_{\xi\xi}\frac{1}{x^2} + 2\tilde{u}_{\xi\eta} + \tilde{u}_{\eta\eta}x^2.$$

Hence

$$x^2u_{xx} - y^2u_{yy} = -4y^2\tilde{u}_{\xi\eta} + 2\frac{y}{x}\tilde{u}_{\xi} = 0$$
$$\tilde{u}_{\xi\eta} - \frac{1}{2}\frac{1}{xy}\tilde{u}_{\xi} = 0.$$

Since $\eta = xy$, we obtain the characteristic form of the equation to be

$$\tilde{u}_{\xi\eta} - \frac{1}{2\eta}\tilde{u}_{\xi} = 0.$$

Using the substitution $v = \tilde{u}_{\xi}$, we obtain $v_{\eta} - \frac{1}{2\eta}v = 0$ which corresponds to the ODE $v' - \frac{1}{2\eta}v = 0$. Hence, $v(\eta, \xi) = c(\xi)\sqrt{\eta}$. Integration with respect to ξ gives $\tilde{u}(\xi, \eta) = A(\xi)\sqrt{\eta} + B(\eta)$. Transforming back to the variables x and y, the general solution is

$$u(x, y) = A\left(\frac{y}{x}\right)\sqrt{xy} + B(xy).$$

(c) The one-dimensional wave equation $u_{tt} - a^2u_{xx} = 0$. The characteristic equation $\sigma_t^2 = a^2\sigma_x^2$ yields

$$-\sigma_t/\sigma_x = \mathrm{d}x/\,\mathrm{d}t = \dot{x} = \pm a.$$

The characteristics are $x = at + c_1$ and $x = -at + c_2$. The change of variables $\xi = x - at$ and $\eta = x + at$ yields $\tilde{u}_{\xi\eta} = 0$ which has general solution $\tilde{u}(\xi, \eta) = f(\xi) + g(\eta)$. Hence, $u(x, t) = f(x - at) + g(x + at)$ is the general solution, see also homework 23.2. .
(d) The wave equation in n dimensions has characteristic equation

$$\sigma_t^2 - a^2\sum_{i=1}^{n}\sigma_{x_i}^2 = 0.$$

This equation is satisfied by the characteristic cone

$$\sigma(x, t) = a^2(t - t^{(0)})^2 - \sum_{i=1}^{n}(x_i - x_i^{(0)})^2 = 0,$$

where the point $(x^{(0)}, t^{(0)})$ is the peak of the cone. Indeed,

$$\sigma_t = 2a^2(t - t^{(0)}), \quad \sigma_{x_1} = -2(x_i - x_i^{(0)})$$

implies $\sigma_t^2 - a^2\sum_{i=1}^{n}(x_i - x_i^{(0)})^2 = 0$.

Further, there are other characteristic surfaces: the hyperplanes

$$\sigma(x,t) = at + \sum_{i=1}^{n} b_i x_i = 0,$$

where $\|b\| = 1$.
(e) The heat equation has characteristic equation $\sum_{i=1}^{n} \sigma_{x_i}^2 = 0$ which implies $\sigma_{x_i} = 0$ for all $i = 1, \ldots, n$ such that $t = c$ is the only family of characteristic surfaces (the coordinate hyperplanes).
(f) The Poisson and Laplace equations have the same characteristic equation; however we have one variable less (no t) and obtain $\operatorname{grad}\sigma = 0$ which is impossible. The Poisson and Laplace equations don't have characteristic surfaces.

15.3.5 The Vibrating String

(a) The Infinite String on $\mathbb{R}$

We consider the Cauchy problem for an infinite string (no boundary values):

$$\begin{aligned} u_{tt} - a^2 u_{xx} &= 0, \\ u(x,0) = u_0(x), \quad u_t(x,0) &= u_1(x), \end{aligned}$$

where u_0 and u_1 are given.
Inserting the initial values into the general solution (see Example 15.4 (c)) $u(x,t) = f(x-at) + g(x+at)$ we get

$$u_0(x) = f(x) + g(x), \quad u_1(x) = -af'(x) + ag'(x).$$

Differentiating the first one yields $u_0'(x) = f'(x) + g'(x)$ such that

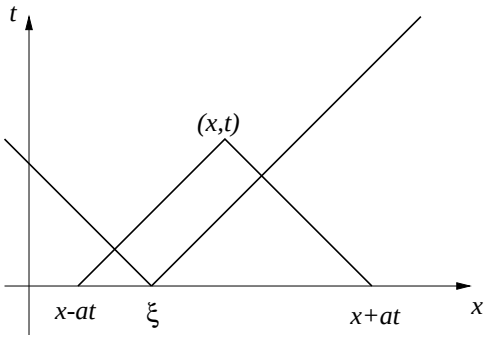
$$f'(x) = \frac{1}{2}u_0'(x) - \frac{1}{2a}u_1(x), \quad g'(x) = \frac{1}{2}u_0'(x) + \frac{1}{2a}u_1(x).$$

Integrating these equations we obtain

$$f(x) = \frac{1}{2}u_0(x) - \frac{1}{2a}\int_0^x u_1(y)\,\mathrm{d}y + A, \quad g(x) = \frac{1}{2}u_0(x) + \frac{1}{2a}\int_0^x u_1(y)\,\mathrm{d}y + B,$$

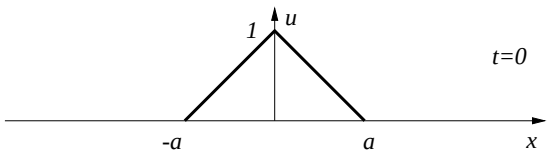
where A and B are constants such that $A + B = 0$ (since $f(x) + g(x) = u_0(x)$). Finally we have

$$\begin{aligned} u(x,t) &= f(x-at) + g(x+at) \\ &= \frac{1}{2}\left(u_0(x+at) + u_0(x-at)\right) - \frac{1}{2a}\int_0^{x-at} u_1(y)\,\mathrm{d}y + \frac{1}{2a}\int_0^{x+at} u_1(y)\,\mathrm{d}y \\ &= \frac{1}{2}\left(u_0(x+at) + u_0(x-at)\right) + \frac{1}{2a}\int_{x-at}^{x+at} u_1(y)\,\mathrm{d}y. \end{aligned} \tag{15.17}$$



It is clear from (15.17) that $u(x,t)$ is uniquely determined by the values of the initial functions u_0 and u_1 in the interval $[x-at, x+at]$ whose end points are cut out by the characteristic lines through the point (x,t). This interval represents the *domain of dependence* for the solution at point $u(x,t)$ as shown in the figure.

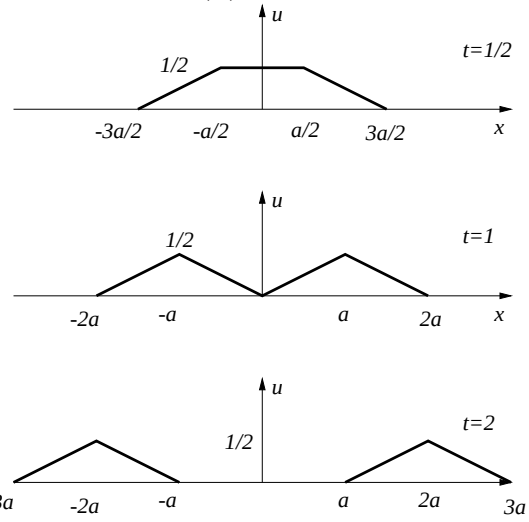
Conversely, the initial values at point $(\xi,0)$ of the x-axis *influence* $u(x,t)$ at points (x,t) in the wedge-shaped region bounded by the characteristics through $(\xi,0)$, i. e. , for $\xi-at < x < \xi+at$. This indicates that our "signal" or "disturbance" only moves with speed a.



We want to give some interpretation of the solution (15.17). Suppose $u_1 = 0$ and

$$u_0(x) = \begin{cases} 1 - \frac{|x|}{a}, & |x| \le a, \\ 0, & |x| > a. \end{cases}$$

In this example we consider the vibrating string which is plucked at time $t = 0$ as in the above picture (given $u_0(x)$). The initial velocity is zero ($u_1 = 0$).



In the pictures one can see the behaviour of the string. The initial peek is divided into two smaller peeks with the half displacement, one moving to the right and one moving to the left with speed a.

Formula (15.17) is due to d'Alembert (1746). Usually one assumes $u_0 \in \mathrm{C}^2(\mathbb{R})$ and $u_1 \in \mathrm{C}^1(\mathbb{R})$. In this case, $u \in \mathrm{C}^2(\mathbb{R}^2)$ and we are able to evaluate the *classical* Laplacian $\Delta(u)$ which gives a continuous function. On the other hand, the right hand side of (15.17) makes sense for arbitrary continuous function u_1 and arbitrary u_0. If we want to call these $u(x,t)$ *a generalized solution* of the Cauchy problem we have to alter the meaning of $\Delta(u)$. In particular, we need more general notion of functions and derivatives. This is our main objective of the next section.

(b) The Finite String over $[0, l]$

We consider the initial boundary value problem (IBVP)

$$u_{tt} = a^2 u_{xx}, \quad u(0,x) = u_0(x), \quad u_t(0,x) = u_1(x),\ x \in [0,l], u(0,t) = u(l,t) = 0, \quad t \in \mathbb{R}.$$

Suppose we are given functions $u_0 \in \mathrm{C}^2([0,l])$ and $u_1 \in \mathrm{C}^1([0,l])$ on $[0,l]$ with

$$u_0(0) = u_0(l) = 0, \quad u_1(0) = u_1(l) = 0, \quad u_0''(0) = u_0''(l) = 0.$$

To solve the IBVP, we define new functions $\tilde{u}_0$ and $\tilde{u}_1$ on $\mathbb{R}$ as follows: first extend both functions to $[-l,l]$ as *odd* functions, that is, $\tilde{u}_i(-x) = -u_i(x)$, $i = 0,1$. Then extend $\tilde{u}_i$ as a $2l$-periodic function to the entire real line. The above assumptions ensure that $\tilde{u}_0 \in \mathrm{C}^2(\mathbb{R})$ and $\tilde{u}_1 \in \mathrm{C}^1(\mathbb{R})$. Put

$$u(x,t) = \frac{1}{2}\left(\tilde{u}_0(x+at) + \tilde{u}_0(x-at)\right) + \frac{1}{2a}\int_{x-at}^{x+at} \tilde{u}_1(y)\,\mathrm{d}y.$$

Then $u(x,t)$ solves the IVP.

Chapter 16

Distributions

16.1 Introduction — Test Functions and Distributions

In this section we introduce the notion of distributions. Distributions are generalized functions. The class of distributions has a lot of very nice properties: they are differentiable up to arbitrary order, one can exchange limit procedures and differentiation, Schwarz' lemma holds. Distributions play an important role in the theory of PDE, in particular, the notion of a fundamental solution of a differential operator can be made rigorous within the theory of distributions only.

Generalized functions were first used by P. Dirac to study quantum mechanical phenomena. Systematically he made use of the so called δ-function (better: δ-distribution). The mathematical foundations of this theory are due to S. L. Sobolev (1936) and L. Schwartz (1950, 1915 – 2002).

Since then many mathematicians made progress in the theory of distributions. Motivation comes from problems in mathematical physics and in the theory of partial differential equations.

Good accessible (German) introductions are given in the books of W. Walter [Wal74] and O. Forster [For81, § 17]. More detailed explanations of the theory are to be found in the books of H. Triebel (in English and German), V. S. Wladimirow (in russian and german) and Gelfand/Schilow (in Russian and German, part I, II, and III), [Tri92, Wla72, GS69, GS64].

16.1.1 Motivation

Distributions generalize the notion of a function. They are linear functionals on certain spaces of test functions. Using distributions one can express rigorously the density of a mass point, charge density of a point, the single-layer and the double-layer potentials, see [Arn04, pp. 92]. Roughly speaking, a generalized function is given at a point by the "mean values" in the neighborhood of that point.

The main idea to associate to each "sufficiently nice" function f a linear functional T_f (a distribution) on an appropriate function space $\mathcal{D}$ is described by the following formula.

$$\langle T_f\,,\,\varphi\rangle = \int_{\mathbb{R}} f(x)\varphi(x)\,\mathrm{d}x, \quad \varphi \in \mathcal{D}. \tag{16.1}$$

On the left we adopt the notation of a dual pairing of vector spaces from Definition 11.1. In general the bracket $\langle T\,,\,\varphi\rangle$ denotes the evaluation of the functional T on the test function φ. Sometimes it is also written as $T(\varphi)$. It does not denote an inner product; the left and the right arguments are from completely different spaces.
What we really want of T_f is

(a) The correspondence should be **one-to-one**, i. e., different functionals T_f and T_g correspond to different functions f and g. To achieve this, we need the function space $\mathcal{D}$ sufficiently large.

(b) The class of functions f should **contain** at least the **continuous functions**. However, if $f(x) = x^n$, the function $f(x)\varphi(x)$ must be integrable over $\mathbb{R}$, that is $x^n\varphi(x) \in \mathrm{L}^1(\mathbb{R})$. Since polynomials are not in $\mathrm{L}^1(\mathbb{R})$, the functions φ must be "very small" for large $|\,x\,|$. Roughly speaking, there are two possibilities to this end. First, take only those functions φ which are identically zero outside a compact set (which depends on φ). This leads to the test functions $\mathcal{D}(\mathbb{R})$. Then T_f is well-defined if f is integrable over every compact subset of $\mathbb{R}$. These functions f are called *locally integrable*.

Secondly, we take $\varphi(x)$ to be rapidly decreasing as $|\,x\,|$ tends to ∞. More precisely, we want

$$\sup_{x\in\mathbb{R}} |\,x^n\,\varphi(x)\,| < \infty$$

for all non-negative integers $n \in \mathbb{Z}_+$. This concept leads to the notion of the so called *Schwartz space* $\mathscr{S}(\mathbb{R})$.

(c) We want to **differentiate** f arbitrarily often, even in case that f has discontinuities. The only thing we have to do is to give the expression

$$\int_{\mathbb{R}} f'(x)\varphi(x)\,\mathrm{d}x, \quad \varphi \in \mathcal{D}$$

a meaning. Using integration by parts and the fact that $\varphi(+\infty) = \varphi(-\infty) = 0$, the above expression equals $-\int_{\mathbb{R}} f(x)\varphi'(x)\,\mathrm{d}x$. That is, instead differentiating f, we differentiate the test function φ. In this way, the functional $T_{f'}$ makes sense as long as $f\varphi'$ is integrable. Since we want to differentiate f arbitrarily often, we need the test function φ to be arbitrarily differentiable, $\varphi \in \mathrm{C}^\infty(\mathbb{R})$.

Note that conditions (b) and (c) make the space of test functions sufficiently small.

16.1.2 Test Functions $\mathcal{D}(\mathbb{R}^n)$ and $\mathcal{D}(\Omega)$

We want to solve the problem $f\varphi$ to be integrable for all polynomials f. We use the first approach and consider only functions φ which are 0 outside a bounded set. If nothing is stated otherwise, $\Omega \subseteq \mathbb{R}^n$ denotes an open, connected subset of $\mathbb{R}^n$.

(a) The Support of a Function and the Space of Test Functions

Let f be a function, defineds on Ω. The set

$$\operatorname{supp} f := \overline{\{x \in \Omega \mid f(x) \neq 0\}} \subset \mathbb{R}^n$$

is called the *support* of f, denoted by $\operatorname{supp} f$.

Remark 16.1 (a) $\operatorname{supp} f$ is always closed; it is the smallest closed subset M such that $f(x) = 0$ for all $x \in \mathbb{R}^n \setminus M$.
(b) A point $x_0 \notin \operatorname{supp} f$ if and only if there exists $\varepsilon > 0$ such that $f \equiv 0$ in $U_\varepsilon(x_0)$. This in particular implies that for $f \in \mathrm{C}^\infty(\mathbb{R}^n)$ we have $f^{(k)}(x_0) = 0$ for all $k \in \mathbb{N}$.
(c) $\operatorname{supp} f$ is compact if and only if it is bounded.

Example 16.1 (a) $\operatorname{supp} \sin = \mathbb{R}$.
(b) Let let $f\colon (-1,1) \to \mathbb{R}$, $f(x) = x(1-x)$. Then $\operatorname{supp} f = [-1,1]$.
(c)The characteristic function χ_M has support $\overline{M}$.
(d) Let h be the hat function on $\mathbb{R}$ – note that $\operatorname{supp} h = [-1,1]$ and $f(x) = 2h(x) - 3h(x-10)$. Then $\operatorname{supp} f = [-1,1] \cup [-11,-9]$.

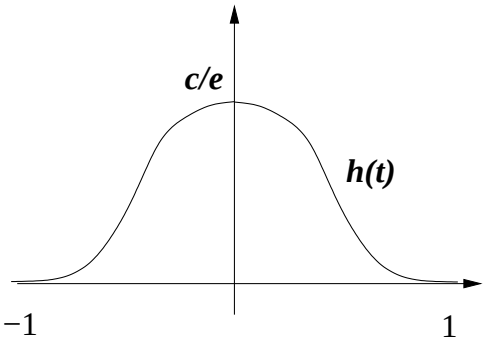
Definition 16.1 (a) The space $\boldsymbol{\mathcal{D}(\mathbb{R}^n)}$ consists of all infinitely differentiable functions f on $\mathbb{R}^n$ with compact support.

$$\mathcal{D}(\mathbb{R}^n) = \mathrm{C}_0^\infty(\mathbb{R}^n) = \{f \in \mathrm{C}^\infty(\mathbb{R}^n) \mid \operatorname{supp} f \quad \text{is compact}\}.$$

(b) Let Ω be a region in $\mathbb{R}^n$. Define $\boldsymbol{\mathcal{D}(\Omega)}$ as follows

$$\mathcal{D}(\Omega) = \{f \in \mathrm{C}^\infty(\Omega) \mid \operatorname{supp} f \text{ is compact in } \mathbb{R}^n \text{ and } \operatorname{supp} f \subset \Omega\}.$$

We call $\mathcal{D}(\Omega)$ the space of *test functions* on Ω.



First of all let us make sure the existence of such functions. On the real axis consider the "hat" function (also called bump function)

$$h(t) = \begin{cases} c\,\mathrm{e}^{-\frac{1}{1-t^2}}, & |\,t\,| < 1, \\ 0, & |\,t\,| \geq 1. \end{cases}$$

The constant c is chosen such that $\int_{\mathbb{R}} h(t)\,\mathrm{d}t = 1$. The function h is continuous on $\mathbb{R}$. It was already shown in Example 4.5 that $h^{(k)}(-1) = h^{(k)}(1) = 0$ for all $k \in \mathbb{N}$. Hence $h \in \mathcal{D}(\mathbb{R})$ is a test function with $\operatorname{supp} h = [-1,1]$. Accordingly, the function

$$h(x) = \begin{cases} c_n \mathrm{e}^{-\frac{1}{1-\|x\|^2}}, & \|x\| < 1, \\ 0, & \|x\| \geq 1. \end{cases}$$

is an element of $\mathcal{D}(\mathbb{R}^n)$ with support being the closed unit ball $\operatorname{supp} h = \{x \mid \|x\| \le 1\}$. The constant c_n is chosen such that $\int\limits_{\mathbb{R}^n} h(x)\,\mathrm{d}x = \int\limits_{U_1(0)} h(x)\,\mathrm{d}x = 1$.

For $\varepsilon > 0$ we introduce the notation

$$h_\varepsilon(x) = \frac{1}{\varepsilon^n}\, h\left(\frac{x}{\varepsilon}\right).$$

Then $\operatorname{supp} h_\varepsilon = \overline{U}_\varepsilon(0)$ and

$$\int_{\mathbb{R}^n} h_\varepsilon(x)\,\mathrm{d}x = \frac{1}{\varepsilon^n}\int_{U_\varepsilon(0)} h\left(\frac{x}{\varepsilon}\right)\,\mathrm{d}x = \int_{U_1(0)} h(y)\,\mathrm{d}y = 1.$$

So far, we have constructed only one function $h(x)$ (as well as its scaled relatives $h_\varepsilon(x)$) which is C^∞ and has compact support. Using this single hat-function h_ε we are able

(a) to restrict the support of an arbitrary integrable function f to a given domain by replacing f by $f h_\varepsilon(x-a)$ which has a support in $U_\varepsilon(a)$,

(b) to make f smooth.

(b) Mollification

In this way, we have an amount of C^∞ functions with compact support which is large enough for our purposes (especially, to recover the function f from the functional T_f). Using the function h_ε, S. L. Sobolev developed the following *mollification* method.

Definition 16.2 (a) Let $f \in \mathrm{L}^1(\mathbb{R}^n)$ and $g \in \mathcal{D}(\mathbb{R}^n)$, define the *convolution product* $f * g$ by

$$(f*g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)\,\mathrm{d}y = \int_{\mathbb{R}^n} f(x-y)g(y)\,\mathrm{d}y = (g*f)(x).$$

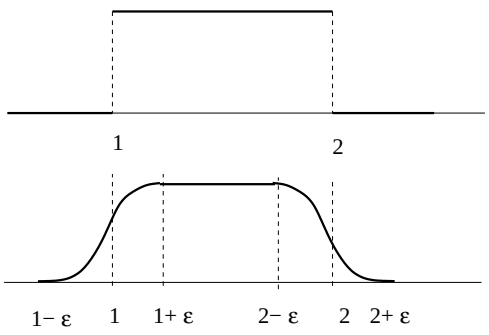
(b) We define the *mollified function* f_ε of f by

$$f_\varepsilon = f * h_\varepsilon.$$

Note that

$$f_\varepsilon(x) = \int\limits_{\mathbb{R}^n} h_\varepsilon(x-y)f(y)\,\mathrm{d}y = \int\limits_{U_\varepsilon(x)} h_\varepsilon(x-y)f(y)\,\mathrm{d}y. \tag{16.2}$$

Roughly speaking, $f_\varepsilon(x)$ is the mean value of f in the ε-neighborhood of x. If f is continuous at x_0 then $f_\varepsilon(x_0) = f(\xi)$ for some $\xi \in U_\varepsilon(x_0)$. This follows from Proposition 5.18.



In particular let $f = \chi_{[1,2]}$ the characteristic function of the interval $[1,2]$. The mollification f_ε looks as follows

$$f_\varepsilon(x) = \begin{cases} 0, & x < 1-\varepsilon, \\ *, & 1-\varepsilon < x < 1+\varepsilon, \\ 1, & 1+\varepsilon < x < 2-\varepsilon, \\ *, & 2-\varepsilon < x < 2+\varepsilon, \\ 0, & 2+\varepsilon < x, \end{cases}$$

where $*$ denotes a value between 0 and 1.

Remarks 16.2 (a) For $f \in \mathrm{L}^1(\mathbb{R}^n)$, $f_\varepsilon \in \mathrm{C}^\infty(\mathbb{R}^n)$,
(b) $f_\varepsilon \to f$ in $\mathrm{L}^1(\mathbb{R}^n)$ as $\varepsilon \to 0$.
(c) $\mathrm{C}_0(\mathbb{R}^n) \subset \mathrm{L}^1(\mathbb{R}^n)$ is dense (with respect to the L^1-norm). In other words, for any $f \in \mathrm{L}^1(\mathbb{R}^n)$ and $\varepsilon > 0$ there exists $g \in \mathrm{C}(\mathbb{R}^n)$ with $\operatorname{supp} g$ is compact and $\int_{\mathbb{R}^n} |\, f - g \,|\, \mathrm{d}x < \varepsilon$.

(Sketch of Proof). (A) Any integrable function can be approximated by integrable functions with compact support. This follows from Example 12.6.
(B) Any integrable function with compact support can be approximated by simple functions (which are finite linear combinations of characteristic functions) with compact support.
(C) Any characteristic function with compact support can be approximated by characteristic functions χ_Q where Q is a finite union of boxes.
(D) Any χ_Q where Q is a closed box can be approximated by a sequence f_n of continuous functions with compact support:

$$f_n(x) = \max\{0, n\, d(x, Q)\}, \quad n \in \mathbb{N},$$

where $d(x, Q)$ denotes the distance of x from Q. Note that f_n is 1 in Q and 0 outside $U_{1/n}(Q)$.

(d) $\mathrm{C}_0^\infty(\mathbb{R}^n) \subset \mathrm{L}^1(\mathbb{R}^n)$ is dense.

(b) Convergence in $\mathcal{D}$

Notations. For $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}_0^n$ (a multi-index), $\alpha = (\alpha_1, \ldots, \alpha_n)$ we write

$$\begin{aligned} |\, \alpha \,| &= \alpha_1 + \alpha_2 + \cdots + \alpha_n, \\ \alpha! &= \alpha_1! \cdots \alpha_n! \\ x^\alpha &= x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \\ D^\alpha u(x) &= \frac{\partial^{|\, \alpha \,|} u(x)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}. \end{aligned}$$

It is clear that $\mathcal{D}(\mathbb{R}^n)$ is a linear space. We shall introduce an appropriate notion of convergence.

Definition 16.3 A sequence $(\varphi_n(x))$ of functions of $\mathcal{D}(\mathbb{R}^n)$ *converges* to $\varphi \in \mathcal{D}(\mathbb{R}^n)$ if there exists a compact set $K \subseteq \mathbb{R}^n$ such that

(a) $\operatorname{supp} \varphi_n \subseteq K$ for all $n \in \mathbb{N}$ and

(b)

$$D^\alpha \varphi_n \rightrightarrows D^\alpha \varphi, \quad \text{uniformly on } K \text{ for all multi-indices } \alpha\,.$$

We denote this type of convergence by $\varphi_n \underset{\mathcal{D}}{\longrightarrow} \varphi$.

Example 16.2 Let $\varphi \in \mathcal{D}$ be a fixed test function and consider the sequence $(\varphi_n(x))$ given by
(a) $\left(\frac{\varphi(x)}{n}\right)$. This sequence converges to 0 in $\mathcal{D}$ since $\operatorname{supp}\varphi_n = \operatorname{supp}\varphi$ for all n and the convergence is uniform for all $x \in \mathbb{R}^n$ (in fact, it suffices to consider $x \in \operatorname{supp}\varphi$).
(b) $\left(\frac{\varphi(x/n)}{n}\right)$. The sequence does not converge to 0 in $\mathcal{D}$ since the supports $\operatorname{supp}(\varphi_n) = n\operatorname{supp}(\varphi)$, $n \in \mathbb{N}$, are not in any common compact subset.
(c) $\left(\frac{\varphi(nx)}{n}\right)$ has no limit if $\varphi \neq 0$, see homework 49.2.

Note that $\mathcal{D}(\mathbb{R}^n)$ is not a metric space, more precisely, there is no metric on $\mathcal{D}(\mathbb{R}^n)$ such that the metric convergence and the above convergence coincide.

16.2 The Distributions $\mathcal{D}'(\mathbb{R}^n)$

Definition 16.4 A *distribution* (generalized function) is a continuous linear functional on the space $\mathcal{D}(\mathbb{R}^n)$ of test functions.
Here, a linear functional T on $\mathcal{D}$ is said to be *continuous* if and only if for all sequences (φ_n), $\varphi_n, \varphi \in \mathcal{D}$, with $\varphi_n \underset{\mathcal{D}}{\longrightarrow} \varphi$ we have $\langle T\,,\,\varphi_n\rangle \to \langle T\,,\,\varphi\rangle$ in $\mathbb{C}$.
The set of distributions is denoted by $\mathcal{D}'(\mathbb{R}^n)$ or simply by $\mathcal{D}'$.

The evaluation of a distribution $T \in \mathcal{D}'$ on a test function $\varphi \in \mathcal{D}$ is denoted by $\langle T\,,\,\varphi\rangle$. Two distributions T_1 and T_2 are equal if and only if $\langle T_1\,,\,\varphi\rangle = \langle T_2\,,\,\varphi\rangle$ for all $\varphi \in \mathcal{D}$.

Remark 16.3 (Characterization of continuity.) (a) A linear functional T on $\mathcal{D}(\mathbb{R}^n)$ is continuous if and only if $\varphi_n \underset{\mathcal{D}}{\longrightarrow} 0$ implies $\langle T\,,\,\varphi_n\rangle \to 0$ in $\mathbb{C}$. Indeed, T continuous, trivially implies the above statement. Suppose now, that $\varphi_n \underset{\mathcal{D}}{\longrightarrow} \varphi$. Then $(\varphi_n - \varphi) \underset{\mathcal{D}}{\longrightarrow} 0$; thus $\langle T\,,\,\varphi_n - \varphi\rangle \to 0$ as $n \to \infty$. Since T is linear, this shows $\langle T\,,\,\varphi_n\rangle \to \langle T\,,\,\varphi\rangle$ and T is continuous.
(b) A linear functional T on $\mathcal{D}$ is continuous if and only if for all compact sets K there exist a constant $C > 0$ and $l \in \mathbb{Z}_+$ such that for all

$$|\langle T\,,\,\varphi\rangle| \le C \sup_{x\in K,\,|\alpha|\le l} |D^\alpha\varphi(x)|\,, \qquad \forall\,\varphi \in \mathcal{D} \quad \text{with} \quad \operatorname{supp}\varphi \subset K. \tag{16.3}$$

We show that the criterion (16.3) in implies continuity of T. Indeed, let $\varphi \underset{\mathcal{D}}{\longrightarrow} 0$. Then there exists compact subset $K \subset \mathbb{R}^n$ such that $\operatorname{supp}\varphi_n \subseteq K$ for all n. By the criterion, there is a $C > 0$ and an $l \in \mathbb{Z}_+$ with $|\langle T\,,\,\varphi_n\rangle| \le C \sup |D^\alpha\varphi_n(x)|$, where the supremum is taken over all $x \in K$ and multiindices α with $|\alpha| \le l$. Since $D^\alpha\varphi_n \rightrightarrows 0$ on K for all α, we particularly have $\sup|D^\alpha\varphi_n(x)| \longrightarrow 0$ as $n \to \infty$. This shows $\langle T\,,\,\varphi_n\rangle \to 0$ and T is continuous.
For the proof of the converse direction, see [Tri92, p. 52]

16.2.1 Regular Distributions

A large subclass of distributions of $\mathcal{D}'$ is given by ordinary functions via the correspondence $f \leftrightarrow T_f$ given by $\langle T_f\,,\,\varphi\rangle = \int_{\mathbb{R}} f(x)\varphi(x)\,\mathrm{d}x$. We are looking for a class which is as large as

possible.

Definition 16.5 Let Ω be an open subset of $\mathbb{R}^n$. A function $f(x)$ on Ω is said to be *locally integrable* over Ω if $f(x)$ is integrable over every compact subset $K \subseteq \Omega$; we write in this case $f \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$.

Remark 16.4 The following are equivalent:

(a) $f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$.

(b) For any $R > 0$, $f \in \mathrm{L}^1(U_R(0))$.

(c) For any $x_0 \in \mathbb{R}^n$ there exists $\varepsilon > 0$ such that $f \in \mathrm{L}^1(U_\varepsilon(x_0))$.

Lemma 16.1 *If f is locally integrable $f \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$, T_f is a distribution, $T_f \in \mathcal{D}'(\Omega)$.*
A distribution T which is of the form $T = T_f$ with some locally integrable function f is called regular.

Proof. First, T_f is linear functional on $\mathcal{D}$ since integration is a linear operation. Secondly, if $\varphi_m \underset{\mathcal{D}}{\longrightarrow} 0$, then there exists a compact set K with $\operatorname{supp}\varphi_m \subset K$ for all m. We have the following estimate:

$$\left| \int_{\mathbb{R}^n} f(x)\varphi_m(x)\,\mathrm{d}x \right| \leq \sup_{x\in K} |\,\varphi_m(x)\,| \int_K |\,f(x)\,|\,\mathrm{d}x = C \sup_{x\in K} |\,\varphi_m(x)\,|\,,$$

where $C = \int_K |\,f\,|\,\mathrm{d}x$ exists since $f \in \mathrm{L}^1_{\mathrm{loc}}$. The expression on the right tends to 0 since $\varphi_m(x)$ uniformly tends to 0. Hence $\langle T_f\,,\,\varphi_m\rangle \to 0$ and T_f belongs to $\mathcal{D}'$. ∎

Example 16.3 (a) $\mathrm{C}(\Omega) \subset \mathrm{L}^1_{\mathrm{loc}}(\Omega)$, $\mathrm{L}^1(\Omega) \subseteq \mathrm{L}^1_{\mathrm{loc}}(\Omega)$.
(b) $f(x) = \dfrac{1}{x}$ is in $\mathrm{L}^1_{\mathrm{loc}}((0,1))$; however, $f \notin \mathrm{L}^1((0,1))$ and $f \notin \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R})$ since f is not integrable over $[-1,1]$.

Lemma 16.2 (Du Bois–Reymond, Fund. Lemma of the Calculus of Variation) *Let $\Omega \subseteq \mathbb{R}^n$ be a region. Suppose that $f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$ and $\langle T_f\,,\,\varphi\rangle = 0$ for all $\varphi \in \mathcal{D}(\Omega)$.*
Then $f = 0$ almost everywhere in Ω.

Proof. For simplicity we consider the case $n = 1$, $\Omega = (-\pi,\pi)$. Fix ε with $0 < \varepsilon < \pi$. Let $\varphi_n(x) = \mathrm{e}^{-\mathrm{i}nx}\,h_\varepsilon(x)$, $n \in \mathbb{Z}$. Then $\operatorname{supp}\varphi_n \subset [-\pi,\pi]$. Since both, e^x and h_ε are C^∞-functions, $\varphi_n \in \mathcal{D}(\Omega)$ and

$$c_n = \langle T_f\,,\,\varphi_n\rangle = \int_{-\pi}^{\pi} f(x)\mathrm{e}^{-\mathrm{i}nx}h_\varepsilon(x)\,\mathrm{d}x = 0, \quad n \in \mathbb{Z};$$

and all Fourier coefficients of $f\,h_\varepsilon \in \mathrm{L}^2[-\pi,\pi]$ vanish. From Theorem 13.13 (b) it follows that fh_ε is 0 in $\mathrm{L}^2(-\pi,\pi)$. By Proposition 12.16 it follows that fh_ε is 0 a.e. in $(-\pi,\pi)$. Since $h_\varepsilon > 0$ on $(-\pi,\pi)$, $f = 0$ a.e. on $(-\pi,\pi)$. ∎

Remark 16.5 The previous lemma shows, if f_1 and f_2 are locally integrable and $T_{f_1} = T_{f_2}$ then $f_1 = f_2$ a.e.; that is, the correspondence is one-to-one. In this way we can identify $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n) \subseteq \mathcal{D}'(\mathbb{R}^n)$ the locally integrable functions as a subspace of the distributions.

16.2.2 Other Examples of Distributions

Definition 16.6 Every non-regular distribution is called *singular*. The most important example of singular distribution is the *δ-distribution* δ_a defined by

$$\langle \delta_a \, , \, \varphi \rangle = \varphi(a), \quad a \in \mathbb{R}^n, \quad \varphi \in \mathcal{D}.$$

It is immediate that δ_a is a linear functional on $\mathcal{D}$. Suppose that $\varphi_n \underset{\mathcal{D}}{\longrightarrow} 0$ then $\varphi_n(x) \to 0$ pointwise. Hence, $\delta_a(\varphi_n) = \varphi_n(a) \to 0$; the functional is continuous on $\mathcal{D}$ and therefore a distribution. We will also use the notation $\delta(x-a)$ in place of δ_a and δ or $\delta(x)$ in place of δ_0.
Proof that δ_a is singular. If $\delta_a \in \mathcal{D}'$ were regular there would exist a function $f \in \mathrm{L}^1_{\mathrm{loc}}$ such that $\delta_a = T_f$, that is $\varphi(a) = \int_{\mathbb{R}^n} f(x)\varphi(x)\,\mathrm{d}x$. *First proof.* Let $\Omega \subset \mathbb{R}^n$ be an open such that $a \notin \Omega$. Let $\varphi \in \mathcal{D}(\Omega)$, that is, $\operatorname{supp}\varphi \subset \Omega$. In particular $\varphi(a) = 0$. That is, $\int_\Omega f(x)\varphi(x)\,\mathrm{d}x = 0$ for all $\varphi \in \mathcal{D}(\Omega)$. By Du Bois-Reymond's Lemma, $f = 0$ a.e. in Ω. Since Ω was arbitrary, $f = 0$ a.e. in $\mathbb{R}^n \setminus \{a\}$ and therefore, $f = 0$ a.e. in $\mathbb{R}^n$. It follows that $T_f = 0$ in $\mathcal{D}'(\mathbb{R}^n)$, however $\delta_a \neq 0$ – a contradiction.
Second Proof for $a = 0$. Since $f \in \mathrm{L}^1_{\mathrm{loc}}$ there exists $\epsilon > 0$ such that

$$d := \int_{U_\epsilon(0)} |\, f(x) \,|\, \mathrm{d}x < 1.$$

Putting $\varphi(x) = h(x/\varepsilon)$ with the bump function h we have $\operatorname{supp}\varphi = \overline{U_\varepsilon(0)}$ and $\sup_{x\in\mathbb{R}^n} |\, \varphi(x) \,| = \varphi(0) > 0$ such that

$$\left| \int_{\mathbb{R}^n} f(x)\varphi(x)\,\mathrm{d}x \right| \leq \sup |\, \varphi(x) \,| \int_{U_\varepsilon(0)} |\, f(x) \,|\, \mathrm{d}x = \varphi(0)d < \varphi(0).$$

This contradicts $\left| \int_{\mathbb{R}^n} f(x)\varphi(x)\,\mathrm{d}x \right| = |\, \varphi(0) \,| = \varphi(0)$.
In the same way one can show that the assignment

$$\langle T \, , \, \varphi \rangle = D^\alpha \varphi(a), \quad a \in \mathbb{R}^n, \quad \varphi \in \mathcal{D}$$

defines an element of $\mathcal{D}'$ which is singular.
The distribution

$$\langle T \, , \, \varphi \rangle = \int_{\mathbb{R}^n} f(x) D^\alpha \varphi(x)\,\mathrm{d}x, \quad f \in \mathrm{L}^1_{\mathrm{loc}},$$

may be regular or singular which depends on the properties of f.
Locally integrable functions as well as δ_a describe mass, force, or charge densities. That is why L. Schwartz named the generalized functions "distributions."

16.2.3 Convergence and Limits of Distributions

There are a lot of possibilities to approximate the distribution δ by a sequence of $\mathrm{L}^1_{\mathrm{loc}}$ functions.

Definition 16.7 A sequence (T_n), $T_n \in \mathcal{D}'(\mathbb{R}^n)$, is said to be *convergent* to $T \in \mathcal{D}'(\mathbb{R}^n)$ if and only if for all $\varphi \in \mathcal{D}(\mathbb{R}^n)$

$$\lim_{n\to\infty} T_n(\varphi) = T(\varphi).$$

Similarly, let T_ε, $\varepsilon > 0$, be a family of distributions in $\mathcal{D}'$, we say that $\lim_{\varepsilon\to 0} T_\varepsilon = T$ if $\lim_{\varepsilon\to 0} T_\varepsilon(\varphi) = T(\varphi)$ for all $\varphi \in \mathcal{D}$.

Note that $\mathcal{D}'(\mathbb{R}^n)$ with the above notion of convergence is complete, see [Wal02, p. 39].

Example 16.4 Let $f(x) = \frac{1}{2}\chi_{[-1,1]}$ and $f_\varepsilon = \frac{1}{\varepsilon} f\left(\frac{x}{\varepsilon}\right)$ be the scaling of f. Note that $f_\varepsilon = 1/(2\varepsilon)\chi_{[-\varepsilon,\varepsilon]}$. We will show tht $f_\varepsilon \to \delta$ in $\mathcal{D}'(\mathbb{R})$. Indeed, for $\varphi \in \mathcal{D}(\mathbb{R})$, by the Mean Value Theorem of integration,

$$T_{f_\varepsilon}(\varphi) = \frac{1}{2\varepsilon}\int_{\mathbb{R}} \chi_{[-\varepsilon,\varepsilon]}\varphi \,\mathrm{d}x = \frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon} \varphi(x)\,\mathrm{d}x = \frac{1}{2\varepsilon} 2\varepsilon\varphi(\xi) = \varphi(\xi), \quad \xi \in [-\varepsilon,\varepsilon],$$

for some ξ. Since φ is continuous at 0, $\varphi(\xi)$ tends to $\varphi(0)$ as $\varepsilon \to 0$ such that

$$\lim_{\varepsilon\to 0} T_{f_\varepsilon}(\varphi) = \varphi(0) = \delta(\varphi).$$

This proves the calaim.

The following lemma generalizes this example.

Lemma 16.3 *Suppose that $f \in \mathrm{L}^1(\mathbb{R})$ with $\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$. For $\varepsilon > 0$ define the scaled function $f_\varepsilon(x) = \frac{1}{\varepsilon} f\left(\frac{x}{\varepsilon}\right)$.*
Then $\lim_{\varepsilon\to 0+0} T_{f_\varepsilon} = \delta$ in $\mathcal{D}'(\mathbb{R})$.

Proof. By the change of variable theorem, $\int_{\mathbb{R}} f_\varepsilon(x)\,\mathrm{d}x = 1$ for all $\varepsilon > 0$. To prove the claim we have to show that for all $\varphi \in \mathcal{D}$

$$\int_{\mathbb{R}} f_\varepsilon(x)\varphi(x)\,\mathrm{d}x \longrightarrow \varphi(0) = \int_{\mathbb{R}} f_\varepsilon(x)\varphi(0)\,\mathrm{d}x \quad \text{as } \varepsilon \to 0;$$

or, equivalently,

$$\left| \int_{\mathbb{R}} f_\varepsilon(x)(\varphi(x) - \varphi(0))\,\mathrm{d}x \right| \longrightarrow 0, \quad \text{as } \varepsilon \to 0.$$

Using the new coordinate y with $x = \varepsilon y$, $\mathrm{d}x = \varepsilon\,\mathrm{d}y$ the above integral equals

$$\left| \int_{\mathbb{R}} \varepsilon f_\varepsilon(\varepsilon y)(\varphi(\varepsilon y) - \varphi(0))\,\mathrm{d}y \right| = \left| \int_{\mathbb{R}} f(y)(\varphi(\varepsilon y) - \varphi(0))\,\mathrm{d}y \right|.$$

Since φ is continuous at 0, for every fixed y, the family of functions $(\varphi(\varepsilon y) - \varphi(0))$ tends to 0 as $\varepsilon \to 0$. Hence, the family of functions $g_\varepsilon(y) = f(y)(\varphi(\varepsilon y) - \varphi(0))$ pointwise tends to

0. Further, g_ε has an integrable upper bound, $2C\,|\,f\,|$, where $C = \sup_{x\in\mathbb{R}} |\,\varphi(x)\,|$. By Lebesgue's theorem about the dominated convergence, the limit of the integrals is 0:

$$\lim_{\varepsilon\to 0}\int_{\mathbb{R}} |\,f(y)\,|\;|\,\varphi(\varepsilon y)-\varphi(0)\,|\;\mathrm{d}y = \int_{\mathbb{R}} |\,f(y)\,| \lim_{\varepsilon\to 0} |\,\varphi(\varepsilon y)-\varphi(0)\,|\;\mathrm{d}y = 0.$$

This proves the claim. ∎

The following sequences of locally integrable functions approximate δ as $\varepsilon \to 0$.

$$f_\varepsilon(x) = \frac{1}{\pi\varepsilon x^2}\,\sin^2\frac{x}{\varepsilon}, \qquad f_\varepsilon(x) = \frac{1}{\pi}\,\frac{\varepsilon}{x^2+\varepsilon^2}, \tag{16.4}$$
$$f_\varepsilon(x) = \frac{1}{2\varepsilon\sqrt{\pi}}\,\mathrm{e}^{-\frac{x^2}{4\varepsilon^2}}, \qquad f_\varepsilon(x) = \frac{1}{\pi x}\,\sin\frac{x}{\varepsilon}$$

The first three functions satisfy the assumptions of the Lemma, the last one not since $\left|\,\frac{\sin x}{x}\,\right|$ is not in $\mathrm{L}^1(\mathbb{R})$. Later we will see that the above lemma even holds if $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ as an improper Riemann integral.

16.2.4 The distribution $\mathscr{P}\,\frac{1}{x}$

Since the function $\frac{1}{x}$ is not locally integrable in a neighborhood of 0, $1/x$ is not a regular distribution. However, we can define a substitute that coincides with $1/x$ for all $x \neq 0$.
Recall that the *principal value* (or Cauchy mean value) of an improper Riemann integral is defined as follows. Suppose $f(x)$ has a singularity at $c \in [a,b]$ then

$$\mathrm{Vp}\int_a^b f(x)\,\mathrm{d}x := \lim_{\varepsilon\to 0}\left(\int_a^{c-\varepsilon} + \int_{c+\varepsilon}^b\right) f(x)\,\mathrm{d}x.$$

For example, $\mathrm{Vp}\int_{-1}^{1}\frac{\mathrm{d}x}{x^{2n+1}} = 0$, $n \in \mathbb{N}$.
For $\varphi \in \mathcal{D}$ define

$$F(\varphi) = \mathrm{Vp}\int_{-\infty}^{\infty}\frac{\varphi(x)}{x}\;\mathrm{d}x = \lim_{\varepsilon\to 0}\left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon}^{\infty}\right)\frac{\varphi(x)}{x}\;\mathrm{d}x.$$

Then F is obviously linear. We have to show, that $F(\varphi)$ is finite and continuous on $\mathcal{D}$. Suppose that $\operatorname{supp}\varphi \subseteq [-R,R]$. Define the auxiliary function

$$\psi(x) = \begin{cases} \frac{\varphi(x)-\varphi(0)}{x}, & x \neq 0 \\ \varphi'(0), & x = 0. \end{cases}$$

Since φ is differentiable at 0, $\psi \in \mathrm{C}(\mathbb{R})$. Since $1/x$ is odd, $\int_{-\varepsilon}^{\varepsilon}\mathrm{d}x/x = 0$ and we get

$$\begin{aligned} F(\varphi) &= \lim_{\varepsilon\to 0}\left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon}^{\infty}\right)\frac{\varphi(x)}{x}\,\mathrm{d}x = \lim_{\varepsilon\to 0}\left(\int_{-R}^{-\varepsilon} + \int_{\varepsilon}^{R}\right)\frac{\varphi(x)-\varphi(0)}{x}\,\mathrm{d}x \\ &= \lim_{\varepsilon\to 0}\left(\int_{-R}^{-\varepsilon} + \int_{\varepsilon}^{R}\right)\psi(x)\,\mathrm{d}x = \int_{-R}^{R}\psi(x)\,\mathrm{d}x. \end{aligned}$$

Since ψ is continuous, the above integral exists.
We now prove continuity of F. By Taylor's theorem, $\varphi(x) = \varphi(0) + x\varphi'(\xi_x)$ for some value ξ_x between x and 0. Therefore

$$\begin{aligned} |\, F(\varphi)\,| &= \left| \lim_{\varepsilon\to 0} \left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon}^{\infty} \right) \frac{\varphi(x)}{x}\,\mathrm{d}x \right| \\ &= \left| \lim_{\varepsilon\to 0} \left(\int_{-R}^{-\varepsilon} + \int_{\varepsilon}^{R} \right) \frac{\varphi(0) + x\varphi'(\xi_x)}{x}\,\mathrm{d}x \right| \\ &\le \int_{-R}^{R} |\,\varphi'(\xi_x)\,|\,\mathrm{d}x \le 2R \sup_{x\in\mathbb{R}} |\,\varphi'(x)\,|\,. \end{aligned}$$

This shows that the condition (16.3) in Remark 16.3 is satisfied with $C = 2R$ and $l = 1$ such that F is a continuous linear functional on $\mathcal{D}(\mathbb{R})$, $F \in \mathcal{D}'(\mathbb{R})$. We denote this distribution by $\mathscr{P}\frac{1}{x}$.
In quantum physics one needs the so called Sokhotsky's formulas, [Wla72, p.76]

$$\begin{aligned} \lim_{\varepsilon\to 0+0} \frac{1}{x + \varepsilon\mathrm{i}} &= -\pi\mathrm{i}\delta(x) + \mathscr{P}\,\frac{1}{x}, \\ \lim_{\varepsilon\to 0+0} \frac{1}{x - \varepsilon\mathrm{i}} &= \pi\mathrm{i}\delta(x) + \mathscr{P}\,\frac{1}{x}. \end{aligned}$$

Idea of proof: Show the sum and the difference of the above formulas instead.

$$\lim_{\varepsilon\to 0+0} \frac{2x}{x^2 + \varepsilon^2} = 2\mathscr{P}\,\frac{1}{x}, \qquad \lim_{\varepsilon\to 0+0} \frac{-2\mathrm{i}\varepsilon}{x^2 + \varepsilon^2} = -2\pi\mathrm{i}\delta.$$

The second limit follows from (16.4).

16.2.5 Operation with Distributions

The distributions are distinguished by the fact that in many calculations they are much easier to handle than functions. For this purpose it is necessary to define operations on the set $\mathcal{D}'$. We already know how to add distributions and how to multiply them with complex numbers since $\mathcal{D}'$ is a linear space. Our guiding principle to define multiplication, derivatives, tensor products, convolution, Fourier transform is always the same: for regular distributions, i.e. locally integrable functions, we want to recover the old well-known operation.

(a) Multiplication

There is no notion of a product T_1T_2 of distributions. However, we can define $a\cdot T = T\cdot a$, $T \in \mathcal{D}'(\mathbb{R}^n)$, $a \in \mathrm{C}^\infty(\mathbb{R}^n)$. What happens in case of a regular distribution $T = T_f$?

$$\langle aT_f\,,\,\varphi\rangle = \int_{\mathbb{R}^n} a(x)f(x)\varphi(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} f(x)\,a(x)\varphi(x)\,\mathrm{d}x = \langle T_f\,,\,a\,\varphi\rangle\,. \tag{16.5}$$

Obviously, $a\varphi \in \mathcal{D}(\mathbb{R}^n)$ since $a \in \mathrm{C}^\infty(\mathbb{R}^n)$ and φ has compact support; thus, $a\varphi$ has compact support, too. Hence, the right hand side of (16.5) defines a linear functional on $\mathcal{D}(\mathbb{R}^n)$. We have to show continuity. Suppose that $\varphi_n \underset{\mathcal{D}}{\longrightarrow} 0$ then $a\varphi_n \underset{\mathcal{D}}{\longrightarrow} 0$. Then $\langle T\,,\,a\varphi_n\rangle \to 0$ since T is continuous.

Definition 16.8 For $a \in \mathrm{C}^\infty(\mathbb{R}^n)$ and $T \in \mathcal{D}'(\mathbb{R}^n)$ we define $aT \in \mathcal{D}'(\mathbb{R}^n)$ by

$$\langle aT\,,\,\varphi\rangle = \langle T\,,\,a\varphi\rangle$$

and call aT the *product* of a and T.

Example 16.5 (a) $x\mathscr{P}\frac{1}{x} = 1$. Indeed, for $\varphi \in \mathcal{D}(\mathbb{R}^n)$,

$$\left\langle x\mathscr{P}\frac{1}{x}\,,\,\varphi\right\rangle = \left\langle \mathscr{P}\frac{1}{x}\,,\,x\varphi(x)\right\rangle = \mathrm{Vp}\int_{-\infty}^{\infty}\frac{x\varphi(x)}{x}\,\mathrm{d}x = \int_{\mathbb{R}}\varphi(x)\,\mathrm{d}x = \langle 1\,,\,\varphi\rangle\,.$$

(b) If $f(x) \in \mathrm{C}^\infty(\mathbb{R}^n)$ then

$$\langle f(x)\delta_a\,,\,\varphi\rangle = \langle \delta_a\,,\,f(x)\varphi(x)\rangle = f(a)\varphi(a) = f(a)\,\langle\delta_a\,,\,\varphi\rangle\,.$$

This shows $f(x)\delta_a = f(a)\delta_a$.
(c) Note that multiplication of distribution is no longer associative:

$$(\delta\cdot x)\,\mathscr{P}\frac{1}{x} \underset{\text{(b)}}{=} 0\cdot\mathscr{P}\frac{1}{x} = 0, \quad \delta\left(x\cdot\mathscr{P}\frac{1}{x}\right) \underset{\text{(a)}}{=} \delta\cdot 1 = \delta.$$

(b) Differentiation

Consider $n = 1$. Suppose that $f \in \mathrm{L}^1_{\mathrm{loc}}$ is continuously differentiable. Suppose further that $\varphi \in \mathcal{D}$ with $\operatorname{supp}\varphi \subset (-a,a)$ such that $\varphi(-a) = \varphi(a) = 0$. We want to define $(T_f)'$ to be $T_{f'}$. Using integration by parts we find

$$\begin{aligned}\langle T_{f'}\,,\,\varphi\rangle &= \int_{-a}^{a} f'(x)\varphi(x)\,\mathrm{d}x = f(x)\varphi(x)|_{-a}^{a} - \int_{-a}^{a} f(x)\varphi'(x)\,\mathrm{d}x\\ &= -\int_{-a}^{a} f(x)\varphi'(x)\,\mathrm{d}x = -\langle T_f\,,\,\varphi'\rangle\,,\end{aligned}$$

where we used that $\varphi(-a) = \varphi(a) = 0$. Hence, it makes sense to define $\langle T_f'\,,\,\varphi\rangle = -\langle T_f\,,\,\varphi'\rangle$. This can easily be generalized to arbitrary partial derivatives $D^\alpha T_f$.

Definition 16.9 For $T \in \mathcal{D}'(\mathbb{R}^n)$ and a multi-index $\alpha \in \mathbb{N}_0^n$ define $D^\alpha T \in \mathcal{D}'(\mathbb{R}^n)$ by

$$\langle D^\alpha T\,,\,\varphi\rangle = (-1)^{|\alpha|}\,\langle T\,,\,D^\alpha\varphi\rangle\,.$$

We have to make sure that $D^\alpha T$ is indeed a distribution. The linearity of $D^\alpha T$ is obvious. To prove continuity let $\varphi_n \underset{\mathcal{D}}{\longrightarrow} 0$. By definition, this implies $D^\alpha\varphi_n \underset{\mathcal{D}}{\longrightarrow} 0$. Since T is continuous, $\langle T\,,\,D^\alpha\varphi_n\rangle \to 0$. This shows $\langle D^\alpha T\,,\,\varphi_n\rangle \to 0$; hence $D^\alpha T$ is a continuous linear functional on $\mathcal{D}$.
Note, that exactly the fact $D^\alpha T \in \mathcal{D}'$ needs the complicated looking notion of convergence in $\mathcal{D}$, $D^\alpha\varphi_n \to D^\alpha\varphi$. Further, a distribution has partial derivatives of all orders.

Lemma 16.4 *Let $a \in \mathrm{C}^\infty(\mathbb{R}^n)$ and $T \in \mathcal{D}'(\mathbb{R}^n)$. Then*
(a) *Differentiation $D^\alpha\colon \mathcal{D}' \to \mathcal{D}'$ is continuous in $\mathcal{D}'$, that is, $T_n \to T$ in $\mathcal{D}'$ implies $D^\alpha T_n \to D^\alpha T$ in $\mathcal{D}'$.*
(b)

$$\frac{\partial}{\partial x_i}(a\,T) = \frac{\partial a}{\partial x_i}\,T + a\,\frac{\partial T}{\partial x_i}, \quad i = 1, \ldots, n \quad \textit{(product rule)}.$$

(c) *For any two multi-indices α and β*

$$D^{\alpha+\beta}T = D^\alpha(D^\beta T) = D^\beta(D^\alpha T) \quad \textit{(Schwarz's Lemma)}.$$

Proof. (a) Suppose that $T_n \to T$ in $\mathcal{D}'$, that is $\langle T_n\,,\,\psi\rangle \to \langle T\,,\,\psi\rangle$ for all $\psi \in \mathcal{D}$. In particular, for $\psi = D^\alpha\varphi$, $\varphi \in \mathcal{D}$, we get

$$(-1)^{|\,\alpha\,|}\,\langle D^\alpha T_n\,,\,\varphi\rangle = \langle T_n\,,\,D^\alpha\varphi\rangle \longrightarrow \langle T\,,\,D^\alpha\varphi\rangle = (-1)^{|\,\alpha\,|}\,\langle D^\alpha T\,,\,\varphi\rangle\,.$$

Since this holds for all $\varphi \in \mathcal{D}$, the assertion follows.
(b) This follows from

$$\begin{aligned}\left\langle a\frac{\partial T}{\partial x_i}\,,\,\varphi\right\rangle &= \left\langle \frac{\partial T}{\partial x_i}\,,\,a\,\varphi\right\rangle = -\left\langle T\,,\,\frac{\partial}{\partial x_i}(a\varphi)\right\rangle \\ &= -\,\langle T\,,\,a_{x_i}(x)\varphi\rangle - \left\langle T\,,\,a\,\frac{\partial\varphi}{\partial x_i}\right\rangle = -\,\langle a_{x_i}T\,,\,\varphi\rangle - \left\langle aT\,,\,\frac{\partial\varphi}{\partial x_i}\right\rangle \\ &= -\,\langle a_{x_i}T\,,\,\varphi\rangle + \left\langle \frac{\partial}{\partial x_i}(aT)\,,\,\varphi\right\rangle = \left\langle -a_{x_i}T + \frac{\partial}{\partial x_i}(aT)\,,\,\varphi\right\rangle.\end{aligned}$$

"Cancelling" φ on both sides proves the claim.
(c) The easy proof uses $D^{\alpha+\beta}\varphi = D^\alpha(D^\beta\varphi)$ for $\varphi \in \mathcal{D}$. ∎

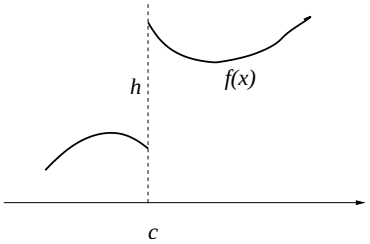
Example 16.6 (a) Let $a \in \mathbb{R}^n$, $f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$, $\varphi \in \mathcal{D}$. Then

$$\begin{aligned}\langle D^\alpha\delta_a\,,\,\varphi\rangle &= (-1)^{|\,\alpha\,|}\,\langle\delta_a\,,\,D^\alpha\varphi\rangle = (-1)^{|\,\alpha\,|}D^\alpha\varphi(a) \\ \langle D^\alpha f\,,\,\varphi\rangle &= (-1)^{|\,\alpha\,|}\int_{\mathbb{R}^n} f\,D^\alpha\varphi\,\mathrm{d}x.\end{aligned}$$

(b) Recall that the so-called *Heaviside function* $H(x)$ is defined as the characteristic function of the half-line $(0, +\infty)$. We compute its derivative in $\mathcal{D}'$:

$$\langle T_H'\,,\,\varphi(x)\rangle = -\int_{\mathbb{R}} H(x)\varphi'(x)\,\mathrm{d}x = -\int_0^\infty \varphi(x)'\,\mathrm{d}x = -\varphi(x)|_0^\infty = \varphi(0) = \langle\delta\,,\,\varphi\rangle\,.$$

This shows $T_H' = \delta$.



(c) More generally, let $f(x)$ be differentiable on $G = \mathbb{R}\setminus\{c\} = (-\infty, c)\cup(c, \infty)$ with a discontinuity of the first kind at c.

The derivative of T_f in $\mathcal{D}'$ is

$$T_f' = T_{f'} + h\delta_c, \quad \text{where} \quad h = f(c+0) - f(c-0),$$

is the difference between the right-handed and left-handed limits of f at c. Indeed, for $\varphi \in \mathcal{D}$ we have

$$\begin{aligned}\langle T_f' \,,\, \varphi\rangle &= \left(-\int_{-\infty}^{c} - \int_{c}^{\infty}\right) f(x)\varphi'(x)\,\mathrm{d}x \\ &= -f(c-0)\varphi(c) + f(c+0)\varphi(c) + \int_G f'(x)\varphi(x)\,\mathrm{d}x \\ &= \big\langle (f(c+0) - f(c-0))\delta_c + T_{f'(x)} \,,\, \varphi\big\rangle \\ &= \langle h\,\delta_c + T_{f'} \,,\, \varphi\rangle\,.\end{aligned}$$

(d) We prove that $f(x) = \log|\,x\,|$ is in $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{R})$ (see homework 50.4, and 51.5) and compute its derivative in $\mathcal{D}'(\mathbb{R})$.

Proof. Since $f(x)$ is continuous on $\mathbb{R}\setminus\{0\}$, the only critical point is 0. Since the integral (improper Riemann or Lebesgue) $\int_0^1 \log x\,\mathrm{d}x = -\int_{-\infty}^0 \mathrm{e}^t\,\mathrm{d}t = -1$ exists f is locally integrable at 0 and therefore defines a regular distribution. We will show that $f'(x) = \mathscr{P}\frac{1}{x}$. We use the fact that $\int_{-\infty}^{\infty} = \int_{-\infty}^{-\varepsilon} + \int_{-\varepsilon}^{\varepsilon} + \int_{\varepsilon}^{\infty}$ for all positive $\varepsilon > 0$. Also, the limit $\varepsilon \to 0$ of the right hand side gives the $\int_{-\infty}^{\infty}$. By definition of the derivative,

$$\begin{aligned}\langle \log'|\,x\,|\,,\,\varphi(x)\rangle &= -\langle \log|\,x\,|\,,\,\varphi'(x)\rangle = \int_{-\infty}^{\infty} \log|\,x\,|\;\varphi'(x)\,\mathrm{d}x \\ &= -\left(\left(\int_{-\infty}^{-\varepsilon} + \int_{-\varepsilon}^{\varepsilon} + \int_{\varepsilon}^{\infty}\right)\log|\,x\,|\;\varphi'(x)\,\mathrm{d}x\right).\end{aligned}$$

Since $\left|\int_{-1}^{1} \log|\,x\,|\,\varphi'(x)\,\mathrm{d}x\right| < \infty$, the middle integral $\int_{-\varepsilon}^{\varepsilon} \log|\,x\,|\,\varphi'(x)\,\mathrm{d}x$ tends to 0 as $\varepsilon \to 0$ (Apply Lebesgue's theorem to the family of functions $g_\varepsilon(x) = \chi_{[-\varepsilon,\varepsilon]}(x)\log|\,x\,|\,\varphi'(x)$ which pointwise tends to 0 and is dominated by the integrable function $\log|\,x\,|\,\varphi'(x)$). We consider the third integral. Integration by parts and $\varphi(+\infty) = 0$ gives

$$\int_{\varepsilon}^{\infty} \log x\,\varphi'(x)\,\mathrm{d}x = \log x\,\varphi(x)\big|_{\varepsilon}^{\infty} - \int_{\varepsilon}^{\infty} \frac{\varphi(x)}{x}\,\mathrm{d}x = \log\varepsilon\varphi(\varepsilon) - \int_{\varepsilon}^{\infty} \frac{\varphi(x)}{x}\,\mathrm{d}x.$$

Similarly,

$$\int_{-\infty}^{-\varepsilon} \log(-x)\,\varphi'(x)\,\mathrm{d}x = -\log\varepsilon\,\varphi(-\varepsilon) - \int_{-\infty}^{-\varepsilon} \frac{\varphi(x)}{x}\,\mathrm{d}x.$$

The sum of the first two (non-integral) terms tends to 0 as $\varepsilon \to 0$ since $\varepsilon\log\varepsilon \to 0$. Indeed,

$$\log\varepsilon\,\varphi(\varepsilon) - \log\varepsilon\,\varphi(-\varepsilon) = \log\varepsilon\;\frac{\varphi(\varepsilon) - \varphi(-\varepsilon)}{2\varepsilon}\,2\varepsilon \longrightarrow 2\lim_{\varepsilon\to 0}\varepsilon\log\varepsilon\;\varphi'(0) = 0.$$

Hence,

$$\langle f' \,,\, \varphi\rangle = \lim_{\varepsilon\to 0}\left(\int_{-\infty}^{-\varepsilon} + \int_{\varepsilon}^{\infty}\right)\frac{\varphi(x)}{x}\,\mathrm{d}x = \left\langle \mathscr{P}\,\frac{1}{x} \,,\, \varphi\right\rangle.$$

■

(c) Convergence and Fourier Series

Lemma 16.5 *Suppose that (f_n) converges locally uniformly to some function f, that is, $f_n \rightrightarrows f$ uniformly on every compact set; assume further that f_n is locally integrable for all n, $f_n \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$.*
(a) Then $f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$ and $T_{f_n} \to T_f$ in $\mathcal{D}'(\mathbb{R}^n)$.
(b) $D^\alpha T_{f_n} \to D^\alpha T_f$ in $\mathcal{D}'(\mathbb{R}^n)$ for all multi-indices α.

Proof. (a) Let K be a compact subset of $\mathbb{R}^n$, we will show that $f \in \mathrm{L}^1(K)$. Since f_n converge uniformly on K to 0, by Theorem 6.6 f is integrable and $\lim_{n\to\infty} \int_K f_n(x)\,\mathrm{d}x = \int_K f\,\mathrm{d}x.$ such that $f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$.
We show that $T_{f_n} \to T_f$ in $\mathcal{D}'$. Indeed, for any $\varphi \in \mathcal{D}$ with compact support K, again by Theorem 6.6 and uniform convergence of $f_n\varphi$ on K,

$$\begin{aligned}\lim_{n\to\infty} T_{f_n}(\varphi) &= \lim_{n\to\infty} \int_K f_n(x)\varphi(x)\,\mathrm{d}x \\ &= \int_K \Big(\lim_{n\to\infty} f_n(x)\Big)\,\varphi(x)\,\mathrm{d}x = \int_K f(x)\varphi(x)\,\mathrm{d}x = T_f(\varphi);\end{aligned}$$

we are allowed to exchange limit and integration since $(f_n(x)\varphi(x))$ uniformly converges on K. Since this is true for all $\varphi \in \mathcal{D}$, it follows that $T_{f_n} \to T_f$.
(b) By Lemma 16.4 (a), differentiation is a continuous operation in $\mathcal{D}'$. Thus $D^\alpha T_{f_n} \to D^\alpha T_f$. ∎

Example 16.7 (a) Suppose that $a, b > 0$ and $m \in \mathbb{N}$ are given such that $|\,c_n\,| \le a\,|\,n\,|^m + b$ for all $n \in \mathbb{Z}$. Then the Fourier series

$$\sum_{n\in\mathbb{Z}} c_n \mathrm{e}^{\mathrm{i}nx},$$

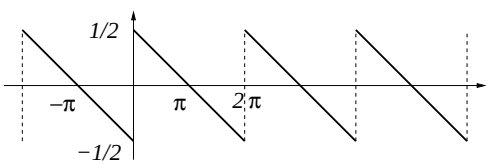
converges in $\mathcal{D}'(\mathbb{R})$.
First consider the series

$$\frac{c_0 x^{m+2}}{(m+2)!} + \sum_{n\in\mathbb{Z}, n\neq 0} \frac{c_n}{(n\mathrm{i})^{m+2}}\,\mathrm{e}^{\mathrm{i}nx}. \tag{16.6}$$

By assumption,

$$\left|\,\frac{c_n}{(n\mathrm{i})^{m+2}}\,\mathrm{e}^{\mathrm{i}nx}\,\right| = \left|\,\frac{c_n}{(n\mathrm{i})^{m+2}}\,\right| \le \frac{a\,|\,n\,|^m + b}{|\,n\,|^{m+2}} \le \frac{\tilde{a}}{|\,n\,|^2}.$$

Since $\sum_{n\neq 0} \frac{\tilde{a}}{|\,n\,|^2} < \infty$, the series (16.6) converges uniformly on $\mathbb{R}$ by the criterion of Weierstraß (Theorem 6.2). By Lemma 16.5, the series (16.6) converges in $\mathcal{D}'$, too and can be differentiated term-by-term. The $(m+2)^{\mathrm{nd}}$ derivative of (16.6) is exactly the given Fourier series.



The 2π-periodic function $f(x) = \frac{1}{2} - \frac{x}{2\pi}$, $x \in [0, 2\pi)$ has discontinuities of the first kind at $2\pi n$, $n \in \mathbb{Z}$; the jump has height 1 since $f(0+0) - f(0-0) = \frac{1}{2} + \frac{1}{2} = 1$.

Therefore in $\mathcal{D}'$

$$f'(x) = -\frac{1}{2\pi} + \sum_{n\in\mathbb{Z}} \delta(x - 2\pi n).$$

The Fourier series of $f(x)$ is

$$f(x) \sim \frac{1}{2\pi \mathrm{i}} \sum_{n\neq 0} \frac{1}{n}\, \mathrm{e}^{\mathrm{i}nx}.$$

Note that f and the Fourier series g on the right are equal in $\mathrm{L}^2(0, 2\pi)$. Hence $\int_0^{2\pi} |\, f - g\,|^2 = 0$. This implies $f = g$ a. e. on $[0, 2\pi]$; moreover $f = g$ a. e. on $\mathbb{R}$. Thus $f = g$ in $\mathrm{L}^1_{\mathrm{loc}}(R)$ and f coincides with g in $\mathcal{D}'(\mathbb{R})$.

$$f(x) = \frac{1}{2\pi \mathrm{i}} \sum_{n\neq 0} \frac{1}{n}\, \mathrm{e}^{\mathrm{i}nx} \quad \text{in} \quad \mathcal{D}'(\mathbb{R}).$$

By Lemma 16.5 the series can be differentiated elementwise up to arbitrary order. Applying Example 16.6 we obtain:

$$f'(x) + \frac{1}{2\pi} = \sum_{n\in\mathbb{Z}} \delta(x - 2\pi n) = \frac{1}{2\pi} \sum_{n\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}nx} \quad \text{in} \quad \mathcal{D}'(\mathbb{R}).$$

(b) A solution of $x^m u(x) = 0$ in $\mathcal{D}'$ is

$$u(x) = \sum_{n=0}^{m-1} c_n \delta^{(n)}(x), \quad c_n \in \mathbb{C}.$$

Since for every $\varphi \in \mathcal{D}$ and $n = 0, \ldots, m-1$ we have

$$\left\langle x^m \delta^{(n)}(x)\,,\, \varphi \right\rangle = (-1)^n \left\langle \delta\,,\, (x^m \varphi(x))^{(n)} \right\rangle = (x^m \varphi(x))^{(n)}\big|_{x=0} = 0;$$

thus, the given u satisfies $x^m u = 0$. One can show, that this is the general solution, see [Wla72, p. 84].

(c) The general solution of the ODE $u^{(m)} = 0$ in $\mathcal{D}'$ is a polynomial of degree $m - 1$.

Proof. We only prove that $u' = 0$ implies $u = c$ in $\mathcal{D}'$. The general statement follows by induction on m.

Suppose that $u' = 0$. That is, for all $\psi \in \mathcal{D}$ we have $0 = \langle u'\,,\, \psi\rangle = -\langle u\,,\, \psi'\rangle$. In particular, for $\varphi, \varphi_1 \in \mathcal{D}$ we have

$$\psi(x) = \int_{-\infty}^{x} (\varphi(t) - \varphi_1(t) I)\, \mathrm{d}t, \quad \text{where} \quad I = \langle 1\,,\, \varphi\rangle\,,$$

belongs to $\mathcal{D}$ since both φ and φ_1 do; φ_1 plays an auxiliary role. Since $\langle u\,,\, \psi'\rangle = 0$ and $\psi' = \varphi - I\,\varphi_1$ we obtain

$$\begin{aligned} 0 = \langle u\,,\, \psi'\rangle &= \langle u\,,\, \varphi - \varphi_1 I\rangle = \langle u\,,\, \varphi\rangle - \langle u\,,\, \varphi_1\rangle \langle 1\,,\, \varphi\rangle \\ &= \langle u\,,\, \varphi\rangle - \langle 1 \langle u\,,\, \varphi_1\rangle\,,\, \varphi\rangle \\ &= \langle u - 1 \langle u\,,\, \varphi_1\rangle\,,\, \varphi\rangle = \langle u - c\,1\,,\, \varphi\rangle\,, \end{aligned}$$

where $c = \langle u\,,\, \varphi_1\rangle$. Since this is true for all test functions $\varphi \in \mathcal{D}(\mathbb{R}^n)$, we obtain $0 = u - c$ or $u = c$ which proves the assertion. ∎

16.3 Tensor Product and Convolution Product

16.3.1 The Support of a Distribution

Let $T \in \mathcal{D}'$ be a distribution. We say that T *vanishes at* x_0 if there exists $\varepsilon > 0$ such that $\langle T\,,\,\varphi\rangle = 0$ for all functions $\varphi \in \mathcal{D}$ with $\operatorname{supp}\varphi \subseteq U_\varepsilon(x_0)$. Similarly, we say that two distributions T_1 and T_2 *are equal at* x_0, $T_1(x_0) = T_2(x_0)$, if $T_1 - T_2$ vanishes at x_0. Note that $T_1 = T_2$ if and only if $T_1 = T_2$ at $a \in \mathbb{R}^n$ for all points $a \in \mathbb{R}^n$.

Definition 16.10 Let $T \in \mathcal{D}'$ be a distribution. The *support* of T, denoted by $\operatorname{supp} T$, is the set of all points x such that T does not vanish at x, that is

$$\operatorname{supp} T = \{x \mid \ \forall\,\varepsilon > 0\ \ \exists\,\varphi \in \mathcal{D}(U_\varepsilon(x))\colon\ \langle T\,,\,\varphi\rangle \neq 0\}.$$

Remarks 16.6 (a) If f is continuous, then $\operatorname{supp} T_f = \operatorname{supp} f$; for an arbitrary locally integrable function we have, in general $\operatorname{supp} T_f \subset \operatorname{supp} f$. The support of a distribution is closed. Its complement is the largest open subset G of $\mathbb{R}^n$ such that $T\restriction_G = 0$.
(b) $\operatorname{supp}\delta_a = \{a\}$, that is, δ_a vanishes at all points $b \neq a$. $\operatorname{supp} T_H = [0, +\infty)$, $\operatorname{supp} T_{\chi_{\mathbb{Q}}} = \varnothing$.

16.3.2 Tensor Products

(a) Tensor product of Functions

Let $f\colon \mathbb{R}^n \to \mathbb{C}$, $g\colon \mathbb{R}^m \to \mathbb{C}$ be functions. Then the *tensor product* $f \otimes g\colon \mathbb{R}^{n+m} \to \mathbb{C}$ is defined via $f\otimes g(x, y) = f(x)g(y)$, $x \in \mathbb{R}^n, y \in \mathbb{R}^m$.
If $\varphi_k \in \mathcal{D}(\mathbb{R}^n)$ and $\psi_k \in \mathcal{D}(\mathbb{R}^m)$, $k = 1, \ldots, r$, we call the function $\varphi(x, y) = \sum_{k=1}^r \varphi_k(x)\psi_k(y)$ which is defined on $\mathbb{R}^{n+m}$ the *tensor product* of the functions φ_k and ψ_k. It is denoted by $\sum_k \varphi_k \otimes \psi_k$. The set of such tensors $\sum_{k=1}^r \varphi_k \otimes \psi_k$ is denoted by $\mathcal{D}(\mathbb{R}^n) \otimes \mathcal{D}(\mathbb{R}^m)$. It is a linear space.
Note first that under the above assumptions on φ_k and ψ_k the tensor product $\varphi = \sum_k \varphi_k \otimes \psi_k \in \mathrm{C}^\infty(\mathbb{R}^{n+m})$. Let $K_1 \subset \mathbb{R}^n$ and $K_2 \subset \mathbb{R}^m$ denote the common compact supports of the families $\{\varphi_k\}$ and $\{\psi_k\}$, respectively. Then $\operatorname{supp}\varphi \subset K_1 \times K_2$. Since both K_1 and K_2 are compact, its product $K_1 \times K_2$ is agian compact. Hence, $\varphi(x, y) \in \mathcal{D}(\mathbb{R}^{n+m})$. Thus, $\mathcal{D}(\mathbb{R}^n) \otimes \mathcal{D}(\mathbb{R}^m) \subset \mathcal{D}(\mathbb{R}^{n+m})$. Moreover, $\mathcal{D}(\mathbb{R}^n) \otimes \mathcal{D}(\mathbb{R}^m)$ is a dense subspace in $\mathcal{D}(\mathbb{R}^{n+m})$. That is, for any $\eta \in \mathcal{D}(\mathbb{R}^{n+m})$ there exist positive integers $r_k \in \mathbb{N}$ and test functions $\varphi_k^{(m)}$, $\psi_k^{(m)}$ such that

$$\sum_{l=1}^{r_m} \varphi_k^{(m)} \otimes \psi_k^{(m)} \underset{\mathcal{D}}{\longrightarrow} \eta \quad \text{as } m \to \infty.$$

(b) Tensor Product of Distributions

Definition 16.11 Let $T \in \mathcal{D}'(\mathbb{R}^n)$ and $S \in \mathcal{D}'(\mathbb{R}^m)$ be two distributions. Then there exists a unique distribution $F \in \mathcal{D}'(\mathbb{R}^{n+m})$ such that for all $\varphi \in \mathcal{D}(\mathbb{R}^n)$ and $\psi \in \mathcal{D}(\mathbb{R}^m)$

$$F(\varphi \otimes \psi) = T(\varphi)S(\psi).$$

This distribution F is denoted by $T \otimes S$.

Indeed, $T \otimes S$ is linear on $\mathcal{D}(\mathbb{R}^n) \otimes \mathcal{D}(\mathbb{R}^n)$ such that $(T \otimes S)(\sum_{k=1}^r \varphi_k \otimes \psi_k) = \sum_{k=1}^r T(\varphi_k)S(\psi_k)$. By continuity it is extended from $\mathcal{D}(\mathbb{R}^n) \otimes \mathcal{D}(\mathbb{R}^n)$ to $\mathcal{D}(\mathbb{R}^{n+m})$. For example, if $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ then $\delta_a \otimes \delta_b = \delta_{(a,b)}$. Indeed, for $\varphi \in \mathcal{D}(\mathbb{R}^n)$ and $\psi \in \mathcal{D}(\mathbb{R}^m)$ we have

$$(\delta_a \otimes \delta_b)(\varphi \otimes \psi) = \varphi(a)\psi(b) = (\varphi \otimes \psi)(a,b) = \delta_{(a,b)}(\varphi \otimes \psi).$$

Lemma 16.6 *Let $F = T \otimes S$ be the unique distribution in $\mathcal{D}'(\mathbb{R}^{n+m})$ where $T \in \mathcal{D}(\mathbb{R}^n)$ and $S \in \mathcal{D}(\mathbb{R}^m)$ and $\eta(x,y) \in \mathcal{D}(\mathbb{R}^{n+m})$.*
Then $\varphi(x) = \langle S(y)\,,\, \eta(x,y)\rangle$ is in $\mathcal{D}(\mathbb{R}^n)$, $\psi(y) = \langle T(x)\,,\, \eta(x,y)\rangle$ is in $\mathcal{D}(\mathbb{R}^m)$ and we have

$$\langle (T \otimes S)\,,\, \eta\rangle = \langle S\,,\, \langle T\,,\, \eta\rangle\rangle = \langle T\,,\, \langle S\,,\, \eta\rangle\rangle\,.$$

For the proof, see [Wla72, II.7].

Example 16.8 (a) *Regular Distributions.* Let $f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$ and $g \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^m)$. Then $f \otimes g \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^{n+m})$ and $T_f \otimes T_g = T_{f\otimes g}$. Indeed, by Fubini's theorem, for test functions φ and ψ one has

$$\begin{aligned}\langle (T_f \otimes T_g)\,,\, \varphi\otimes\psi\rangle = \langle T_f\,,\,\varphi\rangle\ \langle T_g\,,\,\psi\rangle &= \int\limits_{\mathbb{R}^n} f(x)\varphi(x)\,\mathrm{d}x \int\limits_{\mathbb{R}^m} g(y)\psi(y)\,\mathrm{d}y \\ &= \int\limits_{\mathbb{R}^{n+m}} f(x)g(y)\,\varphi(x)\psi(y)\,\mathrm{d}x\mathrm{d}y = \langle T_{f\otimes g}\,,\,\varphi\otimes\psi\rangle\,.\end{aligned}$$

(b) $\langle \delta_{x_0} \otimes T\,,\, \eta\rangle = \langle T\,,\, \eta(x_0,y)\rangle$. Indeed,

$$\langle \delta_{x_0}\otimes T\,,\, \varphi(x)\psi(y)\rangle = \langle \delta_{x_0}\,,\,\varphi(x)\rangle\ \langle T\,,\,\psi\rangle = \varphi(x_0)\,\langle T\,,\,\psi(y)\rangle = \langle T\,,\,\varphi(x_0)\psi(y)\rangle\,.$$

In particular,

$$(\delta_a \otimes T_g)(\eta) = \int_{\mathbb{R}^m} g(y)\eta(a,y)\,\mathrm{d}y.$$

(c) For any $\alpha \in \mathbb{N}_0^n$, $\beta \in \mathbb{N}_0^m$,

$$D^{\alpha+\beta}(T\otimes S) = (D_x^\alpha T)\otimes(D_y^\beta S) = D^\beta((D^\alpha T)\otimes S) = D^\alpha(T\otimes D^\beta S).$$

Idea of proof in case $n = m = 1$. Let $\varphi, \psi \in \mathcal{D}(\mathbb{R})$. Then

$$\begin{aligned}\frac{\partial}{\partial x}\,(T\otimes S)\,(\varphi\otimes\psi) &= -\,(T\otimes S)\left(\frac{\partial}{\partial x}(\varphi\otimes\psi)\right) \\ &= -(T\otimes S)(\varphi'\otimes\psi) = -T(\varphi')S(\psi) = T'(\varphi)S(\psi) \\ &= (T'\otimes S)(\varphi\otimes\psi).\end{aligned}$$

16.3.3 Convolution Product

Motivation: Knowing the fundamental solution $\mathcal{E}$ of a linear differential operator L, that is $L(\mathcal{E}) = \delta$, one can immediately has the solution of the equation $L[u] = f$ for an arbitrary f, namely $u = \mathcal{E} * f$ where the $*$ is the convolution product already defined for functions in Definition 16.2.

(a) Convolution Product of Functions

The main problem with convolutions is: we run into trouble with the support. Even in case that f and g are locally integrable, $f * g$ need not to be a locally integrable function. However, there are three cases where all is fine:

1. One of the two functions f or g has compact support.
2. Both functions have support in $[0, +\infty)$.
3. Both functions are in $\mathrm{L}^1(\mathbb{R})$.

In the last case $(f * g)(x) = \int f(y)g(x-y)\,\mathrm{d}y$ is again integrable. The convolution product is a commutative and associative operation on $\mathrm{L}^1(\mathbb{R}^n)$.

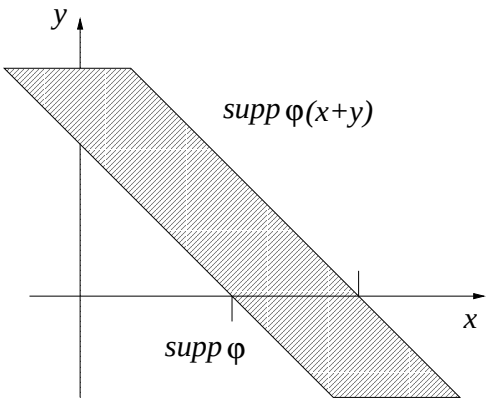
(b) Convolution Product of Distributions

Let us consider the case of regular distributions. Suppose that If $f, g, f * g \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$. As usual we want to have $T_f * T_g = T_{f*g}$. Let $\varphi \in \mathcal{D}(\mathbb{R}^n)$,

$$\langle T_{f*g}\,,\,\varphi\rangle = \int_{\mathbb{R}} (f*g)(x)\varphi(x)\,\mathrm{d}x = \iint\limits_{\mathbb{R}^2} f(y)g(x-y)\varphi(x)\,\mathrm{d}x\mathrm{d}y$$

$$\underset{t=x-y}{=} \iint\limits_{\mathbb{R}^2} f(y)g(t)\varphi(y+t)\,\mathrm{d}y\,\mathrm{d}t = T_{f\otimes g}(\tilde{\varphi}), \tag{16.7}$$

where $\tilde{\varphi}(y,t) = \varphi(y+t)$.



There are two problems: (a) in general $\tilde{\varphi}$ is *not* a test function since it has unbounded support in $\mathbb{R}^{2n}$. Indeed, $(y,t) \in \operatorname{supp}\tilde{\varphi}$ if $y + t = c \in \operatorname{supp}\varphi$, which is a family of parallel lines forming an unbounded strip. (b) the integral does not exist. We overcome the second problem if we impose the condition that the set

$$K_\varphi = \{(y,t) \in \mathbb{R}^{2n} \mid y \in \operatorname{supp} T_f,\ t \in \operatorname{supp} T_g,\ y + t \in \operatorname{supp}\varphi\}$$

is bounded for any $\varphi \in \mathcal{D}(\mathbb{R}^n)$; then the integral (16.7) makes sense.

We want to solve the problem (a) by "cutting" $\tilde{\varphi}$. Define

$$T_{f*g}(\varphi) = \lim_{k\to\infty} (T_f \otimes T_g)(\varphi(y+t)\eta_k(y,t)),$$

where $\eta_k \underset{n\to\infty}{\longrightarrow} 1$ as $k \to \infty$ and $\eta_k \in \mathcal{D}(\mathbb{R}^{2n})$. Such a sequence exists; let $\eta(y,t) \in \mathcal{D}(\mathbb{R}^{2n})$ with $\eta(y,t) = 1$ for $\|y\|^2 + \|t\|^2 \le 1$. Put $\eta_k(y,t) = \eta\left(\frac{y}{k}, \frac{t}{k}\right)$, $k \in \mathbb{N}$. Then $\lim_{k\to\infty} \eta_k(y,t) = 1$ for all $(y,t) \in \mathbb{R}^{2n}$.

Definition 16.12 Let $T, S \in \mathcal{D}'(\mathbb{R}^n)$ be distributions and assume that for every $\varphi \in \mathcal{D}(\mathbb{R}^n)$ the set

$$K_\varphi := \{(x,y) \in \mathbb{R}^{2n} \mid x+y \in \operatorname{supp}\varphi,\ x \in \operatorname{supp} T,\ y \in \operatorname{supp} S\}$$

is bounded. Define

$$\langle T * S\,,\,\varphi\rangle = \lim_{k\to\infty} \langle T\otimes S\,,\,\varphi(x+y)\eta_k(x,y)\rangle\,. \tag{16.8}$$

$T * S$ is called the *convolution product* of the distributions S and T.

Remark 16.7 (a) The sequence (16.8) becomes stationary for large k such that the limit exists. Indeed, for fixed φ, the set K_φ is bounded, hence there exists $k_0 \in \mathbb{N}$ such that $\eta_k(x,y) = 1$ for all $x, y \in K_\varphi$ and all $k \geq k_0$. That is $\varphi(x+y)\eta_k(x,y)$ does not change for $k \geq k_0$.
(b) The limit is a distribution in $\mathcal{D}'(\mathbb{R}^n)$
(c) The limit does not depend on the special choice of the sequence η_k.

Remarks 16.8 (Properties) (a) If S or T has compact support, then $T * S$ exists. Indeed, suppose that $\operatorname{supp} T$ is compact. Then $x+y \in \operatorname{supp}\varphi$ and $x \in \operatorname{supp} T$ imply $y \in \operatorname{supp}\varphi - \operatorname{supp} T = \{y_1 - y_2 \mid y_1 \in \operatorname{supp}\varphi, y_2 \in \operatorname{supp} T\}$. Hence, $(x,y) \in K_\varphi$ implies

$$\|(x,y)\| \leq \|x\| + \|y\| \leq \|x\| + \|y_1\| + \|y_2\| \leq 2C + D$$

if $\operatorname{supp} T \subset U_C(0)$ and $\operatorname{supp}\varphi \subset U_D(0)$. That is, K_φ is bounded.
(b) If $S * T$ exists, so does $T * S$ and $S * T = T * S$.
(c) If $T * S$ exists, so do $D^\alpha T * S$, $T * D^\alpha S$, $D^\alpha(T * S)$ and they coincide:

$$D^\alpha(T * S) = D^\alpha T * S = T * D^\alpha S.$$

Proof. For simplicity let $n = 1$ and $D^\alpha = \frac{\mathrm{d}}{\mathrm{d}x}$. Suppose that $\varphi \in \mathcal{D}(\mathbb{R})$ then

$$\begin{aligned}
\langle (S*T)'\,,\,\varphi\rangle &= -\langle S*T\,,\,\varphi'\rangle = -\lim_{k\to\infty}\langle S\otimes T\,,\,\varphi'(x+y)\,\eta_k(x,y)\rangle \\
&= -\lim_{k\to\infty}\left\langle S\otimes T\,,\,\frac{\partial}{\partial x}\left(\varphi(x+y)\eta_k(x,y)\right) - \varphi(x+y)\frac{\partial\eta_k}{\partial x}\right\rangle \\
&= \lim_{k\to\infty}\langle S'\otimes T\,,\,\varphi(x+y)\eta_k(x,y)\rangle - \lim_{k\to\infty}\left\langle S\otimes T\,,\,\varphi(x+y)\underbrace{\frac{\partial\eta_k}{\partial x}}_{=0 \text{ for large } k}\right\rangle \\
&= \langle S'*T\,,\,\varphi\rangle
\end{aligned}$$

The proof of the second equality uses commutativity of the convolution product. ∎

(d) If $\operatorname{supp} S$ is compact and $\psi \in \mathcal{D}(\mathbb{R}^n)$ such that $\psi(y) = 1$ in a neighborhood of $\operatorname{supp} S$. Then

$$(T*S)(\varphi) = \langle T\otimes S\,,\,\varphi(x+y)\psi(y)\rangle\,, \qquad \forall\,\varphi \in \mathcal{D}(\mathbb{R}^n).$$

(e) If $T_1, T_2, T_3 \in \mathcal{D}'(\mathbb{R}^n)$ all have compact support, then $T_1 * (T_2 * T_3)$ and $(T_1 * T_2) * T_3$ exist and $T_1 * (T_2 * T_3) = (T_1 * T_2) * T_3$.

16.3.4 Linear Change of Variables

Suppose that $y = Ax + b$ is a regular, linear change of variables; that is, A is a regular $n \times n$ matrix. As usual, consider first the case of a regular distribution $f(x)$. Let $\tilde{f}(x) = f(Ax + b)$ with $y = Ax + b$, $x = A^{-1}(y - b)$, $\mathrm{d}y = \det A \, \mathrm{d}x$. Then

$$\begin{aligned}\left\langle \tilde{f}(x)\,,\,\varphi(x)\right\rangle &= \int f(Ax+b)\varphi(x)\,\mathrm{d}x \\ &= \int f(y)\varphi(A^{-1}(y-b))\frac{1}{|\det A|}\,\mathrm{d}y \\ &= \frac{1}{|\det A|}\left\langle f(y)\,,\,\varphi(A^{-1}(y-b))\right\rangle .\end{aligned}$$

Definition 16.13 Let $T \in \mathcal{D}'(\mathbb{R}^n)$, A a regular $n \times n$-matrix and $b \in \mathbb{R}^n$. Then $T(Ax + b)$ denotes the distribution

$$\langle T(Ax+b)\,,\,\varphi(x)\rangle := \frac{1}{|\det A|}\left\langle T(y)\,,\,\varphi(A^{-1}(y-b))\right\rangle .$$

For example, $T(x) = T$, $\langle T(x-a)\,,\,\varphi(x)\rangle = \langle T\,,\,\varphi(x+a)\rangle$, in particular, $\delta(x-a) = \delta_a$.

$$\langle \delta(x-b)\,,\,\varphi(x)\rangle = \langle \delta(x)\,,\,\varphi(x+b)\rangle = \varphi(0+b) = \varphi(b) = \langle \delta_b\,,\,\varphi\rangle .$$

Example 16.9 (a) $\delta * S = S * \delta = S$ for all $S \in \mathcal{D}'$. The existence is clear since δ has compact support.

$$\begin{aligned}\langle (\delta * S)\,,\,\varphi\rangle &= \lim_{k\to\infty}\langle \delta(x)\otimes S(y)\,,\,\varphi(x+y)\eta_k(x,y)\rangle \\ &= \lim_{k\to\infty}\langle S(y)\,,\,\varphi(y)\eta_k(0,y)\rangle = \langle S\,,\,\varphi\rangle\end{aligned}$$

(b) $\delta_a * S = S(x-a)$. Indeed,

$$\begin{aligned}(\delta_a * S)(\varphi) &= \lim_{k\to\infty}\langle \delta_a\otimes S\,,\,\varphi(x+y)\eta_k(x,y)\rangle \\ &= \lim_{k\to\infty}\langle S(y)\,,\,\varphi(a+y)\eta_k(a,y)\rangle = \langle S(y)\,,\,\varphi(a+y)\rangle = \langle S(y-a)\,,\,\varphi\rangle .\end{aligned}$$

Inparticular $\delta_a * \delta_b = \delta_{a+b}$.

(c) Let $\varrho \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$ and $\operatorname{supp} T_f$ is compact.

Case $n = 2$ $f(x) = \log\frac{1}{\|x\|} \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^2)$. We call

$$V(x) = (\varrho * f)(x) = \iint_{\mathbb{R}^2}\varrho(y)\log\frac{1}{\|x-y\|}\,\mathrm{d}y$$

surface potential with density ϱ.

Case $n \geq 3$ $f(x) = \frac{1}{\|x\|^{n-2}} \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$. We call

$$V(x) = (\varrho * f)(x) = \int_{\mathbb{R}^n}\varrho(y)\frac{1}{\|x-y\|^{n-2}}\,\mathrm{d}y$$

vector potential with density ϱ.

(d) For $\alpha > 0$ and $x \in \mathbb{R}$ put $f_\alpha(x) = \frac{1}{\alpha\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2\alpha^2}}$. Then $f_\alpha * f_\beta = f_{\sqrt{\alpha^2+\beta^2}}$.

16.3.5 Fundamental Solutions

Suppose that $L[u]$ is a linear differential operator on $\mathbb{R}^n$,

$$L[u] = \sum_{|\,\alpha\,|\le k} c_\alpha(x) D^\alpha u,$$

where $c_\alpha \in \mathrm{C}^\infty(\mathbb{R}^n)$.

Definition 16.14 A distribution $\mathcal{E} \in \mathcal{D}'(\mathbb{R}^n)$ is said to be a *fundamental solution* of the differential operator L if

$$L(\mathcal{E}) = \delta.$$

Note that $\mathcal{E} \in \mathcal{D}'(\mathbb{R}^n)$ need not to be unique. It is a general result due to Malgrange and Ehrenpreis (1952) that any linear partial differential operator *with constant coefficients* possesses a fundamental solution.

(a) ODE

We start with an example from the theory of ordinary differential equations. Recall that $H = \chi_{(0,+\infty)}$ denotes the Heaviside function.

Lemma 16.7 *Suppose that $u(t)$ is a solution of the following initial value problem for the ODE*

$$\begin{aligned} L[u] = u^{(m)} + a_1(t)u^{(m-1)} + \cdots + a_m(t)u = 0,\\ u(0) = u'(0) = \cdots = u^{(m-2)}(0) = 0,\\ u^{(m-1)}(0) = 1. \end{aligned}$$

Then $\mathcal{E} = T_{H(t)\,u(t)}$ is a fundamental solution of L, that is, it satisfies $L(\mathcal{E}) = \delta$.

Proof. Using Leibniz' rule, Example 16.5 (b), and $u(0) = 0$ we find

$$\mathcal{E}' = \delta\, T_u + T_{Hu'} = u(0)\delta + T_{Hu'} = T_{Hu'}.$$

Similarly, on has

$$\mathcal{E}'' = T_{Hu''}, \ldots, \mathcal{E}^{(m-1)} = T_{Hu^{(m-1)}}, \quad \mathcal{E}^{(m)} = T_{Hu^{(m)}} + \delta(t).$$

This yields

$$L(\mathcal{E}) = \mathcal{E}^{(m)} + a_1(t)\mathcal{E}^{(m-1)} + \cdots + a_m(t)\mathcal{E}(t) = T_{H(t)\,L(u(t))} + \delta = T_0 + \delta = \delta.$$

■

Example 16.10 We have the following example of fundamental solutions:

$$\begin{aligned} y' + ay = 0, &\qquad \mathcal{E} = T_{H(x)\mathrm{e}^{-ax}},\\ y'' + a^2 y = 0, &\qquad \mathcal{E} = T_{H(x)\frac{\sin ax}{a}}. \end{aligned}$$

(b) PDE

Here is the main application of the convolution product: knowing the fundamental solution of a partial differential operator L one immediately knows a weak solution of the inhomogeneous equations $L(u) = f$ for $f \in \mathcal{D}'(\mathbb{R}^n)$.

Theorem 16.8 *Suppose that* $L[u] = \sum_{|\alpha|\le k} c_\alpha D^\alpha u$ *is a linear differential operator in* $\mathbb{R}^n$ *with constant coefficients* c_α*. Suppose further that* $\mathcal{E} \in \mathcal{D}'(\mathbb{R}^n)$ *is a fundamental solution of* L*. Let* $f \in \mathcal{D}'(\mathbb{R}^n)$ *be a distribution such that the convolution product* $S = \mathcal{E} * f$ *exists.*
Then $L(S) = f$ *in* $\mathcal{D}'$*.*
In the set of distributions of $\mathcal{D}'$ *which possess a convolution with* $\mathcal{E}$*,* S *is the unique solution of* $L(S) = f$

Proof. By Remark 16.8 (b) we have

$$L(S) = \sum_{|\alpha|\le k} c_a D^\alpha(\mathcal{E} * f) = \sum_{|\alpha|\le k} c_a D^\alpha(\mathcal{E}) * f = L(\mathcal{E}) * f = \delta * f = f.$$

Suppose that S_1 and S_2 are both solutions of $L(S) = f$, i. e. $L(S_1) = L(S_2) = f$. Then

$$S_1 - S_2 = (S_1 - S_2) * \delta = (S_1 - S_2) * \sum_{|\alpha|\le k} c_a D^\alpha \mathcal{E} =$$

$$= \left(\sum_{|\alpha|\le k} c_a D^\alpha (S_1 - S_2) \right) * \mathcal{E} = (f - f) * \mathcal{E} = 0. \quad (16.9)$$

■

16.4 Fourier Transformation in $\mathscr{S}(\mathbb{R}^n)$ and $\mathscr{S}'(\mathbb{R}^n)$

We want to define the Fourier transformation for test functions φ as well as for distributions. The problem with $\mathcal{D}(\mathbb{R}^n)$ is that its Fourier transformation

$$\mathcal{F}\varphi(\xi) = \alpha_n \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}x\xi} \varphi(x)\,\mathrm{d}x$$

of φ is an entire (analytic) function with real support $\mathbb{R}$. That is, $\mathcal{F}\varphi$ does not have compact support. The only test function in $\mathcal{D}$ which is analytic is 0. To overcome this problem, we enlarge the space of test function $\mathcal{D} \subset \mathscr{S}$ in such a way that $\mathscr{S}$ becomes invariant under the Fourier transformation $\mathcal{F}(\mathscr{S}) \subset \mathscr{S}$.

Lemma 16.9 *Let* $\varphi \in \mathcal{D}(\mathbb{R})$*. Then the Fourier transform* $g(z) = \alpha_n \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}tz} \varphi(t)\,\mathrm{d}t$ *is holomorphic in the whole complex plane and bounded in any half-plane* $H_a = \{z \in \mathbb{C} \mid \mathrm{Im}\,(z) \le a\}$.

Proof. (a) We show that the complex limit $\lim_{h\to 0}(g(z+h)-g(z))/h$ exists for all $z \in \mathbb{C}$. Indeed,

$$\frac{g(z+h)-g(z)}{h} = \alpha_n \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}zt} \frac{\mathrm{e}^{-\mathrm{i}ht}-1}{h} \varphi(t)\ \mathrm{d}t.$$

Since $\left| \mathrm{e}^{-\mathrm{i}zt} \frac{\mathrm{e}^{-\mathrm{i}ht}-1}{h} \varphi(t) \right| \le C$ for all $x \in \operatorname{supp}(\varphi)$, $h \in \mathbb{C}$, $|\,h\,| \le 1$, we can apply Lebesgue's Dominated Convergence theorem:

$$\lim_{h\to 0} \frac{g(z+h)-g(z)}{h} = \alpha_n \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}zt} \lim_{h\to 0} \frac{\mathrm{e}^{-\mathrm{i}ht}-1}{h} \varphi(t)\ \mathrm{d}t = \alpha_n \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}zt}(-\mathrm{i}t)\varphi(t)\,\mathrm{d}t = \mathcal{F}(-\mathrm{i}t\varphi(t)).$$

(b) Suppose that $\operatorname{Im}(z) \le a$. Then

$$|\,g(z)\,| \le \alpha_n \int_{\mathbb{R}} \left| \mathrm{e}^{-\mathrm{i}t\operatorname{Re}(z)} \right| \mathrm{e}^{t\operatorname{Im}(z)} \varphi(t)\,\mathrm{d}t \le \alpha_n \sup_{t\in K} |\,\varphi(t)\,| \int_K \mathrm{e}^{ta}\,\mathrm{d}t,$$

where K is a compact set which contains $\operatorname{supp}\varphi$.

■

16.4.1 The Space $\mathscr{S}(\mathbb{R}^n)$

Definition 16.15 $\mathscr{S}(\mathbb{R}^n)$ is the set of all functions $f \in \mathrm{C}^\infty(\mathbb{R}^n)$ such that for all multi-indices α and β

$$p_{\alpha,\beta}(f) = \sup_{x\in\mathbb{R}^n} \left| x^\beta D^\alpha f(x) \right| < \infty.$$

$\mathscr{S}$ is called the *Schwartz space* or the *space of rapidly decreasing functions*.
$\mathscr{S}(\mathbb{R}^n) = \{f \in \mathrm{C}^\infty(\mathbb{R}^n) \mid \ \forall\,\alpha,\beta : P_{\alpha,\beta}(f) < \infty\}$.

Roughly speaking, a Schwartz space function is a function decreasing to 0 (together with all its partial derivatives) faster than any rational function $\dfrac{1}{P(x)}$ as $x \to \infty$. In place of $p_{\alpha,\beta}$ one can also use the norms

$$p_{k,l}(\varphi) = \sum_{|\,\alpha\,|\le k,\ |\,\beta\,|\le l,} p_{\alpha,\beta}(\varphi), \quad k,l \in \mathbb{Z}_+$$

to describe $\mathscr{S}(\mathbb{R}^n)$.
The set $\mathscr{S}(\mathbb{R}^n)$ is a linear space and $p_{\alpha,\beta}$ are norms on $\mathscr{S}$.
For example, $P(x) \notin \mathscr{S}$ for any non-zero polynomial $P(x)$; however $\mathrm{e}^{-\|x\|^2} \in \mathscr{S}(\mathbb{R}^n)$.
$\mathscr{S}(\mathbb{R}^n)$ is an algebra. Indeed, the generalized Leibniz rule ensures $p_{kl}(\varphi\cdot\psi) < \infty$. For example, $f(x) = p(x)\mathrm{e}^{-ax^2+bx+c}$, $a > 0$, belongs to $\mathscr{S}(\mathbb{R})$ for p is a polynomial; $g(x) = \mathrm{e}^{-|\,x\,|}$ is not differentiable at 0 and hence not in $\mathscr{S}(\mathbb{R})$.

Convergence in $\mathscr{S}$

Definition 16.16 Let $\varphi_n, \varphi \in \mathscr{S}$. We say that the sequence (φ_n) converges in $\mathscr{S}$ to φ, abbreviated by $\varphi_n \underset{\mathscr{S}}{\longrightarrow} \varphi$, if one of the following equivalent conditions is satisfied for all multi-indices

α and β:

$$\begin{aligned} p_{\alpha,\beta}(\varphi-\varphi_n) &\underset{n\to\infty}{\longrightarrow} 0;\\ x^\beta D^\alpha(\varphi-\varphi_n) &\rightrightarrows 0, \quad \text{uniformly on } \mathbb{R}^n;\\ x^\beta D^\alpha\varphi_n &\rightrightarrows x^\beta D^\alpha\varphi, \quad \text{uniformly on } \mathbb{R}^n. \end{aligned}$$

Remarks 16.9 (a) In quantum mechanics one defines the *position* and *momentum operators* Q_k and P_k, $k=1,\dots,n$, by

$$(Q_k\varphi)(x) = x_k\varphi(x), \quad (P_k\varphi)(x) = -\mathrm{i}\frac{\partial\varphi}{\partial x_k},$$

respectively. The space $\mathscr{S}$ is invariant under both operators Q_k and P_k; that is $x^\beta D^\alpha\varphi(x) \in \mathscr{S}(\mathbb{R}^n)$ for all $\varphi\in\mathscr{S}(\mathbb{R}^n)$.
(b) $\mathscr{S}(\mathbb{R}^n)\subset \mathrm{L}^1(\mathbb{R}^n)$.
Recall that a rational function $P(x)/Q(x)$ is integrable over $[1,+\infty)$ if and only if $Q(x)\neq 0$ for $x\geq 1$ and $\deg Q \geq \deg P+2$. Indeed, C/x^2 is then an integrable upper bound. We want to find a condition on m such that

$$\int_{\mathbb{R}^n}\frac{\mathrm{d}x}{(1+\|x\|^2)^m} < \infty.$$

For, we use that any non-zero $x\in\mathbb{R}^n$ can uniquely be written as $x=r\,y$ where $r=\|x\|$ and y is on the unit sphere S^{n-1}. One can show that $\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_m = r^{n-1}\,\mathrm{d}r\,\mathrm{d}S$ where $\mathrm{d}S$ is the surface element of the unit sphere S^{n-1}. Using this and Fubini's theorem,

$$\int_{\mathbb{R}^n}\frac{\mathrm{d}x}{(1+\|x\|^2)^m} = \int_0^\infty\int_{\mathrm{S}^{n-1}}\frac{r^{n-1}\,\mathrm{d}r\,\mathrm{d}S}{(1+r^2)^m} = \omega_{n-1}\int_0^\infty\frac{r^{n-1}\,\mathrm{d}r}{(1+r^2)^m},$$

where ω_{n-1} is the $(n-1)$-dimensional measure of the unit sphere S^{n-1}. By the above criterion, the integral is finite if and only if $2m-n+1>1$ if and only if $m>n/2$. In particular,

$$\int_{\mathbb{R}^n}\frac{\mathrm{d}x}{1+\|x\|^{n+1}} < \infty.$$

In case $n=1$ the integral is π.
By the above argument

$$\begin{aligned} \int_{\mathbb{R}^n}|\,\varphi(x)\,|\,\mathrm{d}x &= \int_{\mathbb{R}^n}\Big|\,(1+\|x\|^{2n})\varphi(x)\,\Big|\,\frac{\mathrm{d}x}{1+\|x\|^{2n}}\\ &\leq C\,p_{0,2n}(\varphi)\int_{\mathbb{R}^n}\frac{\mathrm{d}x}{1+\|x\|^{2n}} < \infty. \end{aligned}$$

(c) $\mathcal{D}(\mathbb{R}^n)\subset\mathscr{S}(\mathbb{R}^n)$; indeed, the supremum $p_{\alpha,\beta}(\varphi)$ of any test function $\varphi\in\mathcal{D}(\mathbb{R}^n)$ is finite since the supremum of a continuous function over a compact set is finite. On the other hand, $\mathcal{D}\subsetneq\mathcal{S}$ since $\mathrm{e}^{-\|x\|^2}$ is in $\mathscr{S}$ but not in $\mathcal{D}$.
(d) In contrast to $\mathcal{D}(\mathbb{R}^n)$, the rapidly decreasing functions $\mathscr{S}(\mathbb{R}^n)$ form a metric space. Indeed, $\mathscr{S}(\mathbb{R}^n)$ is a locally convex space, that is a linear space V such that the topology is given by a

set of semi-norms p_α separating the elements of V, i. e. $p_\alpha(x) = 0$ for all α implies $x = 0$. Any locally convex linear space where the topology is given by a *countable* set of semi-norms is metrizable. Let $(p_n)_{n\in\mathbb{N}}$ be the defining family of semi-norms. Then

$$d(\varphi,\psi) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{p_n(\varphi-\psi)}{1+p_n(\varphi-\psi)}, \quad \varphi,\psi \in V$$

defines a metric on V describing the same topology. (In our case, use Cantor's first diagonal method to write the norms p_{kl}, $k,l \in \mathbb{N}$, from th array into a sequence p_n.)

Definition 16.17 Let $f(x) \in \mathrm{L}^1(\mathbb{R}^n)$, then the *Fourier transform* $\mathcal{F}f$ of the function $f(x)$ is given by

$$\mathcal{F}f(\xi) = \hat{f}(\xi) = \frac{1}{\sqrt{2\pi}^n} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}x\cdot\xi} f(x)\,\mathrm{d}x,$$

where $x = (x_1,\dots,x_n)$, $\xi = (\xi_1,\dots,\xi_n)$, and $x\cdot\xi = \sum_{k=1}^n x_k\xi_k$.

Let us abbreviate the normalization factor, $\alpha_n = \frac{1}{\sqrt{2\pi}^n}$. Caution, Wladimirow, [Wla72] uses another convention with $\mathrm{e}^{+\mathrm{i}\xi\cdot x}$ under the integral and normalization factor 1 in place of α_n. Note that $\mathcal{F}f(0) = \alpha_n \int_{\mathbb{R}^n} f(x)\,\mathrm{d}x$.

Example 16.11 We calculate the Fourier transform $\mathcal{F}\varphi$ of the function $\varphi(x) = \mathrm{e}^{-\|x\|^2/2} = \mathrm{e}^{-\frac{1}{2}x\cdot x}$, $x \in \mathbb{R}^n$.
(a) $n = 1$. From complex analysis, Lemma 14.30, we know

$$\mathcal{F}\left(\mathrm{e}^{-\frac{x^2}{2}}\right)(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{e}^{-\mathrm{i}x\xi}\,\mathrm{d}x = \mathrm{e}^{-\frac{\xi^2}{2}}. \tag{16.10}$$

(b) *Arbitrary* n. Thus,

$$\begin{aligned}
\mathcal{F}\varphi(\xi) = \hat{\varphi}(\xi) &= \alpha_n \int_{\mathbb{R}^n} \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^n x_k^2}\ \mathrm{e}^{-\mathrm{i}\sum_{k=1}^n x_k\xi_k}\ \mathrm{d}x \\
&= \alpha_n \int_{\mathbb{R}^n} \prod_{k=1}^n \mathrm{e}^{-\frac{1}{2}x_k^2 - \mathrm{i}x_k\xi_k}\,\mathrm{d}x_1\cdots\mathrm{d}x_n \\
&= \prod_{k=1}^n \alpha_n \int \mathrm{e}^{-\frac{1}{2}x_k^2 - \mathrm{i}x_k\xi_k}\,\mathrm{d}x_k \\
\mathcal{F}\varphi(\xi) &\underset{(16.10)}{=} \prod_{k=1}^n \mathrm{e}^{-\frac{1}{2}\xi_k^2} = \mathrm{e}^{-\frac{1}{2}\xi^2}.
\end{aligned}$$

Hence, the Fourier transform of $\mathrm{e}^{-\frac{1}{2}x^2}$ is the function itself. It follows via scaling $x \mapsto cx$ that

$$\mathcal{F}\left(\mathrm{e}^{-\frac{c^2x^2}{2}}\right)(\xi) = \frac{1}{c^n}\mathrm{e}^{-\frac{\xi^2}{2c^2}}.$$

Theorem 16.10 *Let $\varphi,\psi \in \mathscr{S}(\mathbb{R}^n)$. Then we have*

(i) $\mathcal{F}(x^\alpha\,\varphi(x)) = \mathrm{i}^{|\alpha|}\,D^\alpha(\mathcal{F}\varphi)$, *that is* $\mathcal{F}\circ Q_k = -P_k\circ\mathcal{F}$, $k = 1,\dots,n$.

(ii) $\mathcal{F}(D^\alpha\varphi(x))(\xi) = \mathrm{i}^{|\,\alpha\,|}\xi^\alpha\,(\mathcal{F}\varphi)(\xi)$, *that is* $\mathcal{F}\circ P_k = Q_k\circ\mathcal{F}$, $k = 1,\dots,n$.

(iii) $\mathcal{F}(\varphi)\in\mathscr{S}(\mathbb{R}^n)$, *moreover* $\varphi_n \underset{\mathscr{S}}{\longrightarrow} \varphi$ *implies* $\mathcal{F}\varphi_n \underset{\mathscr{S}}{\longrightarrow} \mathcal{F}\varphi$, *that is, the Fourier transform* $\mathcal{F}$ *is a continuous linear operator on* $\mathscr{S}$.

(iv) $\mathcal{F}(\varphi * \psi) = \alpha_n^{-1}\,\mathcal{F}(\varphi)\,\mathcal{F}(\psi)$.

(v) $\mathcal{F}(\varphi\cdot\psi) = \alpha_n\,\mathcal{F}(\varphi) * \mathcal{F}(\psi)$

(vi)
$$\mathcal{F}(\varphi(Ax+b))(\xi) = \frac{1}{|\det A|}\mathrm{e}^{\mathrm{i}A^{-1}b\cdot\xi}\mathcal{F}\varphi\left(A^{-\top}\xi\right),$$
where A *is a regular* $n\times n$ *matrix and* $A^{-\top}$ *denotes the transpose of* A^{-1}. *In particular,*
$$\mathcal{F}(\varphi(\lambda x))(\xi) = \frac{1}{|\,\lambda\,|^n}(\mathcal{F}\varphi)\left(\frac{\xi}{\lambda}\right),$$
$$\mathcal{F}(\varphi(x+b))(\xi) = \mathrm{e}^{\mathrm{i}b\cdot\xi}\,(\mathcal{F}\varphi)(\xi).$$

Proof. (i) We carry out the proof in case $\alpha = (1,0,\dots,0)$. The general case simply follows.
$$\frac{\partial}{\partial\xi_1}(\mathcal{F}\varphi)(\xi) = \alpha_n\frac{\partial}{\partial\xi_1}\int_{\mathbb{R}^n}\mathrm{e}^{-\mathrm{i}\xi\cdot x}\varphi(x)\,\mathrm{d}x.$$
Since $\frac{\partial}{\partial\xi_1}\left(\mathrm{e}^{-\mathrm{i}\xi\cdot x}\right)\varphi(x) = -\mathrm{i}x_1\mathrm{e}^{-\mathrm{i}\xi\cdot x}\varphi(x)$ tends to 0 as $x\to\infty$, we can exchange partial differentiation and integration, see Proposition 12.23 Hence,
$$\frac{\partial}{\partial\xi_1}(\mathcal{F}\varphi)(\xi) = -\alpha_n\int_{\mathbb{R}^n}\mathrm{e}^{-\mathrm{i}\xi\cdot x}\,\mathrm{i}x_1\varphi(x)\,\mathrm{d}x = \mathcal{F}(-\mathrm{i}x_1\varphi(x))(\xi).$$

(ii) Without loss of generality we again assume $\alpha = (1,0,\dots,0)$. Using integration by parts, we obtain
$$\begin{aligned}\mathcal{F}\left(\frac{\partial}{\partial x_1}\varphi(x)\right)(\xi) &= \alpha_n\int_{\mathbb{R}^n}\mathrm{e}^{-\mathrm{i}\xi\cdot x}\varphi_{x_1}(x)\,\mathrm{d}x = -\alpha_n\int_{\mathbb{R}^n}\frac{\partial}{\partial x_1}\left(\mathrm{e}^{-\mathrm{i}\xi\cdot x}\right)\,\varphi(x)\,\mathrm{d}x\\ &= \mathrm{i}\xi_1\alpha_n\int_{\mathbb{R}^n}\left(\mathrm{e}^{-\mathrm{i}\xi\cdot x}\right)\,\varphi(x)\,\mathrm{d}x = \mathrm{i}\xi_1\,(\mathcal{F}\varphi)(\xi).\end{aligned}$$

(iii) By (i) and (ii) we have for $|\,\alpha\,|\le k$ and $|\,\beta\,|\le l$
$$\begin{aligned}\left|\,\xi^\alpha D^\beta\mathcal{F}\varphi\,\right| &\le \alpha_n\int_{\mathbb{R}^n}\left|\,D^\alpha(x^\beta\varphi(x))\,\right|\,\mathrm{d}x \le c_1\int_{\mathbb{R}^n}(1+\|x\|^l)\sum_{|\,\gamma\,|\le k}|\,D^\gamma\varphi(x)\,|\,\mathrm{d}x\\ &\le c_2\int_{\mathbb{R}^n}\frac{(1+\|x\|^{l+n+1})}{(1+\|x\|^{n+1})}\sum_{|\,\gamma\,|\le k}|\,D^\gamma\varphi(x)\,|\,\mathrm{d}x\\ &\le c_3\sup_{x\in\mathbb{R}^n}\left((1+\|x\|^{l+n+1})\sum_{|\,\gamma\,|\le k}|\,D^\gamma\varphi(x)\,|\right)\\ &\le c_4\,p_{k,l+n+1}(\varphi).\end{aligned}$$

This implies $\mathcal{F}\varphi \in \mathscr{S}(\mathbb{R}^n)$ and, moreover $\mathcal{F}\colon \mathscr{S} \to \mathscr{S}$ being continuous.
(iv) First note that $\mathrm{L}^1(\mathbb{R}^n)$ is an algebra with respect to the convolution product where $\|f * g\|_{\mathrm{L}^1} \le \|f\|_{\mathrm{L}^1} \|g\|_{\mathrm{L}^1}$. Indeed,

$$\begin{aligned}\|f * g\|_{\mathrm{L}^1} &= \int_{\mathbb{R}^n} |\, f * g\,| \,\mathrm{d}x = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |\, f(y) g(x-y)\,| \,\mathrm{d}y \right) \mathrm{d}x \\ &\le \int_{\mathbb{R}^n} |\, f(y)\,| \left(\int_{\mathbb{R}^n} |\, g(x-y)\,| \,\mathrm{d}x \right) \mathrm{d}y \\ &\le \|g\|_{\mathrm{L}^1} \int_{\mathbb{R}^n} |\, f(y)\,| \,\mathrm{d}y = \|f\|_{\mathrm{L}^1} \, \|g\|_{\mathrm{L}^1} \,.\end{aligned}$$

This in particular shows that $|\, f * g(x)\,| < \infty$ is finite a.e. on $\mathbb{R}^n$. By definition and Fubini's theorem we have

$$\begin{aligned}\mathcal{F}(\varphi * \psi)(\xi) &= \alpha_n \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}x\cdot\xi} \int_{\mathbb{R}^n} \varphi(y)\psi(x-y) \,\mathrm{d}y \,\mathrm{d}x \\ &= \alpha_n^{-1} \int_{\mathbb{R}^n} \left(\alpha_n \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}(x-y)\cdot\xi} \psi(x-y) \,\mathrm{d}x \right) \alpha_n \mathrm{e}^{-\mathrm{i}y\cdot\xi} \varphi(y) \,\mathrm{d}y \\ &\underset{z=x-y}{=} \alpha_n^{-1} \mathcal{F}\psi(\xi)\, \mathcal{F}\varphi(\xi).\end{aligned}$$

(v) will be done later, after Proposition 16.11.
(vi) is straightforward using $\langle A^{-1}(y)\,,\, \xi\rangle = \langle y\,,\, A^{-\top}(\xi)\rangle$. ∎

Remark 16.10 Similar properties as $\mathcal{F}$ has the operator $\mathcal{G}$ which is also defined on $\mathrm{L}^1(\mathbb{R}^n)$:

$$\mathcal{G}\varphi(\xi) = \check{\varphi}(\xi) = \alpha_n \int_{\mathbb{R}^n} \mathrm{e}^{+\mathrm{i}x\cdot\xi} \, \varphi(x) \,\mathrm{d}x.$$

Put $\varphi_-(x) := \varphi(-x)$. Then

$$\mathcal{G}\varphi = \mathcal{F}\varphi_- = \overline{\mathcal{F}(\overline{\varphi})} \quad \text{and} \quad \mathcal{F}\varphi = \mathcal{G}\varphi_- = \overline{\mathcal{G}(\overline{\varphi})}.$$

It is easy to see that (iv) holds for $\mathcal{G}$, too, that is

$$\mathcal{G}(\varphi * \psi) = \alpha_n^{-1} \mathcal{G}(\varphi)\mathcal{G}(\psi).$$

Proposition 16.11 (Fourier Inversion Formula) *The Fourier transformation is a one-to-one mapping of $\mathscr{S}(\mathbb{R}^n)$ onto $\mathscr{S}(\mathbb{R}^n)$. The inverse Fourier transformation is given by $\mathcal{G}\varphi$:*

$$\mathcal{F}(\mathcal{G}\varphi) = \mathcal{G}(\mathcal{F}\varphi) = \varphi, \quad \varphi \in \mathscr{S}(\mathbb{R}^n).$$

Proof. Let $\psi(x) = \mathrm{e}^{-\frac{x\cdot x}{2}}$ and $\Psi(x) = \psi(\varepsilon x) = \mathrm{e}^{-\frac{\varepsilon^2 x^2}{2}}$. Then $\Psi_\varepsilon(x) := \mathcal{F}\Psi(x) = \frac{1}{\varepsilon^n}\psi\left(\frac{x}{\varepsilon}\right)$. We have

$$\alpha_n \int_{\mathbb{R}^n} \Psi_\varepsilon(x) \,\mathrm{d}x = \alpha_n \int_{\mathbb{R}^n} \frac{1}{\varepsilon^n} \psi\left(\frac{x}{\varepsilon}\right) \,\mathrm{d}x = \alpha_n \int_{\mathbb{R}^n} \psi(x) \,\mathrm{d}x = \hat{\psi}(0) = 1.$$

Further,

$$\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}\Psi_\varepsilon(x)\varphi(x)\,\mathrm{d}x=\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}\psi(x)\varphi(\varepsilon x)\,\mathrm{d}x\underset{\varepsilon\to0}{\longrightarrow}\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}\psi(x)\varphi(0)\,\mathrm{d}x=\varphi(0). \tag{16.11}$$

In other words, $\alpha_n\Psi_\varepsilon(x)$ is a δ-sequence.
We compute $\mathcal{G}(\mathcal{F}\varphi\,\Psi)(x)$. Using Fubini's theorem we have

$$\begin{aligned}\mathcal{G}(\mathcal{F}\varphi\,\Psi)(x)&=\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}(\mathcal{F}\varphi)(\xi)\Psi(\xi)\mathrm{e}^{\mathrm{i}x\cdot\xi}\mathrm{d}\xi=\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\Psi(\xi)\mathrm{e}^{\mathrm{i}x\cdot\xi}\int_{\mathbb{R}^n}\mathrm{e}^{-\mathrm{i}\xi\cdot y}\varphi(y)\,\mathrm{d}y\mathrm{d}\xi\\&=\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}\varphi(y)\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}\mathrm{e}^{-\mathrm{i}(y-x)\cdot\xi}\Psi(\xi)\mathrm{d}\xi\,\mathrm{d}y\\&=\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}\varphi(y)\,(\mathcal{F}\Psi)(y-x)\,\mathrm{d}y\\&\underset{z:=y-x}{=}\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}\mathcal{F}\Psi(z)\,\varphi(z+x)\,\mathrm{d}z=\frac{1}{\sqrt{2\pi}^n}\int_{\mathbb{R}^n}\Psi_\varepsilon(z)\,\varphi(z+x)\,\mathrm{d}z\\&\underset{\text{as }\varepsilon\to0,\text{ see (16.11)}}{\longrightarrow}\varphi(x).\end{aligned}$$

On the other hand, by Lebesgue's dominated convergence theorem,

$$\lim_{\varepsilon\to0}\mathcal{G}(\mathcal{F}\varphi\cdot\Psi)(x)=\alpha_n\int_{\mathbb{R}^n}(\mathcal{F}\varphi)(\xi)\psi(0)\mathrm{e}^{\mathrm{i}x\cdot\xi}\mathrm{d}\xi=\alpha_n\int_{\mathbb{R}^n}(\mathcal{F}\varphi)(\xi)\mathrm{e}^{\mathrm{i}x\cdot\xi}\mathrm{d}\xi=\mathcal{G}(\mathcal{F}\varphi)(x).$$

This proves the first part. The second part $\mathcal{F}(\mathcal{G}\varphi)=\varphi$ follows from $\mathcal{G}(\varphi)=\overline{\mathcal{F}(\overline{\varphi})}$, $\mathcal{F}(\varphi)=\overline{\mathcal{G}(\overline{\varphi})}$, and the first part. ∎

We are now going to complete the proof of Theorem 16.10 (v). For, let $\varphi=\mathcal{G}\varphi_1$ and $\psi=\mathcal{G}\psi_1$ with $\varphi_1,\psi_1\in\mathscr{S}$. By (iv) we have

$$\mathcal{F}(\varphi\cdot\psi)=\mathcal{F}(\mathcal{G}(\varphi_1)\mathcal{G}(\psi_1))=\mathcal{F}(\alpha_n\mathcal{G}(\varphi_1*\psi_1))=\alpha_n\varphi_1*\psi_1=\alpha_n\mathcal{F}\varphi*\mathcal{F}\psi.$$

Proposition 16.12 (Fourier–Plancherel formula) *For $\varphi,\,\psi\in\mathscr{S}(\mathbb{R}^n)$ we have*

$$\int_{\mathbb{R}^n}\varphi\,\overline{\psi}\,\mathrm{d}x=\int_{\mathbb{R}^n}\mathcal{F}(\varphi)\,\overline{\mathcal{F}(\psi)}\,\mathrm{d}x,$$

In particular, $\|\varphi\|_{\mathrm{L}^2(\mathbb{R}^n)}=\|\mathcal{F}(\varphi)\|_{\mathrm{L}^2(\mathbb{R}^n)}$

Proof. First note that

$$\mathcal{F}(\overline{\psi})(-y)=\alpha_n\int_{\mathbb{R}^n}\mathrm{e}^{\mathrm{i}x\cdot y}\overline{\psi(x)}\,\mathrm{d}x=\alpha_n\overline{\int_{\mathbb{R}^n}\mathrm{e}^{-\mathrm{i}x\cdot y}\psi(x)\,\mathrm{d}x}=\overline{\mathcal{F}(\psi)(y)}. \tag{16.12}$$

By Theorem 16.10 (v),

$$\begin{aligned}\int_{\mathbb{R}^n}\varphi(x)\,\overline{\psi(x)}\,\mathrm{d}x&=\alpha_n^{-1}\mathcal{F}(\varphi\cdot\overline{\psi})(0)=(\mathcal{F}(\varphi)*\mathcal{F}(\overline{\psi}))(0)\\&=\int_{\mathbb{R}^n}\mathcal{F}(\varphi)(y)\,\mathcal{F}(\overline{\psi})(0-y)\,\mathrm{d}y\underset{(16.12)}{=}\int_{\mathbb{R}^n}\mathcal{F}(\varphi)(y)\,\overline{\mathcal{F}(\psi)(y)}\,\mathrm{d}y.\end{aligned}$$

∎

Remark 16.11 $\mathscr{S}(\mathbb{R}^n) \subset \mathrm{L}^2(\mathbb{R}^n)$ is dense. Thus, the Fourier transformation has a unique extension to a unitary operator $\mathcal{F}\colon \mathrm{L}^2(\mathbb{R}^n) \to \mathrm{L}^2(\mathbb{R}^n)$. (To a given $f \in \mathrm{L}^2$ choose a sequence $\varphi_n \in \mathscr{S}$ converging to f in the L^2-norm. Since $\mathcal{F}$ preserves the L^2-norm, $\|\mathcal{F}\varphi_n - \mathcal{F}\varphi_m\| = \|\varphi_n - \varphi_m\|$ and (φ_n) is a Cauchy sequence in L^2, $(\mathcal{F}\varphi_n)$ is a Cauchy sequence, too; hence it converges to some $g \in \mathrm{L}^2$. We define $\mathcal{F}(f) = g$.)

16.4.2 The Space $\mathscr{S}'(\mathbb{R}^n)$

Definition 16.18 A *tempered distribution* (or slowly increasing distribution) is a continuous linear functional T on the space $\mathscr{S}(\mathbb{R}^n)$. The set of all tempered distributions is denoted by $\mathscr{S}'(\mathbb{R}^n)$.
A linear functional T on $\mathscr{S}(\mathbb{R}^n)$ is continuous if for all sequences $\varphi_n \in \mathscr{S}$ with $\varphi_n \underset{\mathscr{S}}{\longrightarrow} 0$,

$$\langle T\,,\,\varphi_n\rangle \to 0.$$

For $\varphi_n \in \mathcal{D}$ with $\varphi_n \underset{\mathcal{D}}{\longrightarrow} 0$ it follows that $\varphi_n \underset{\mathscr{S}}{\longrightarrow} 0$. So, every continuous linear functional on $\mathscr{S}$ (restricted to $\mathcal{D}$) is continuous on $\mathcal{D}$. Moreover, the mapping $\iota\colon \mathscr{S}' \to \mathcal{D}'$, $\iota(T) = T\restriction_{\mathcal{D}}$ is injective since $T(\varphi) = 0$ for all $\varphi \in \mathcal{D}$ implies $T(\psi) = 0$ for all $\psi \in \mathscr{S}$ which follows from $\mathcal{D} \subset \mathscr{S}$ is dense and continuity of T. Using the injection ι, we can identify $\mathscr{S}'$ with as a subspace of $\mathcal{D}'$, $\mathscr{S}' \subseteq \mathcal{D}'$. That is, every tempered distribution "is" a distribution.

Lemma 16.13 (Characterization of $\mathscr{S}'$) *A linear functional T defined on $\mathscr{S}(\mathbb{R}^n)$ belongs to $\mathscr{S}'(\mathbb{R}^n)$ if and only if there exist non-negative integers k and l and a positive number C such that for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$*

$$|\,T(\varphi)\,| \le C\,p_{kl}(\varphi),$$

where $p_{kl}(\varphi) = \sum_{|\,\alpha\,|\le k\ |\,\beta\,|\le l,} p_{\alpha\beta}(\varphi)$.

Remarks 16.12 (a) With the usual identification $f \leftrightarrow T_f$ of functions and regular distributions, $\mathrm{L}^1(\mathbb{R}^n) \subseteq \mathscr{S}'(\mathbb{R}^n)$, $\mathrm{L}^2(\mathbb{R}^n) \subseteq \mathscr{S}'(\mathbb{R}^n)$.
(b) $\mathrm{L}^1_{\mathrm{loc}} \not\subset \mathscr{S}'$, for example $T_f \notin \mathscr{S}'(\mathbb{R})$, $f(x) = \mathrm{e}^{x^2}$, since $T_f(\varphi)$ is not well-defined for all Schwartz function φ, for example $T_{\mathrm{e}^{x^2}}\left(\mathrm{e}^{-x^2}\right) = +\infty$.
(c) If $T \in \mathcal{D}'(\mathbb{R}^n)$ and $\operatorname{supp} T$ is compact then $T \in \mathscr{S}'(\mathbb{R}^n)$.
(d) Let $f(x)$ be measurable. If there exist $C > 0$ and $m \in \mathbb{N}$ such that

$$|\,f(x)\,| \le C(1 + \|x\|^2)^m \quad \text{a. e. } x \in \mathbb{R}^n.$$

Then $T_f \in \mathscr{S}'$. Indeed, the above estimate and Remark 16.9 imply

$$|\,\langle T_f\,,\,\varphi\rangle\,| = \left|\int_{\mathbb{R}^n} f(x)\varphi(x)\,\mathrm{d}x\right| \le C\int_{\mathbb{R}^n} (1+\|x\|^2)^m\,(1+\|x\|^2)^n\,\frac{1}{(1+\|x\|^2)^n}\,|\,\varphi(x)\,|\,\mathrm{d}x$$
$$\le C p_{0,2m+2n}(\varphi)\int_{\mathbb{R}^n}\frac{\mathrm{d}x}{(1+\|x\|^2)^n}.$$

By Lemma 16.13, f is a tempered regular distribution, $f(x) \in \mathscr{S}'$.

Operations on $\mathscr{S}'$

The operations are defined in the same way as in case of $\mathcal{D}'$. One has to show that the result is again in the (smaller) space $\mathscr{S}'$. If $T \in \mathscr{S}'$ then
(a) $D^\alpha T \in \mathscr{S}'$ for all multi-indices α.
(b) $f \cdot T \in \mathscr{S}'$ for all $f \in \mathrm{C}^\infty(\mathbb{R}^n)$ such that $D^\alpha f$ growth at most polynomially at infinity for all multi-indices α, (i. e. for all multi-indices α there exist $C_\alpha > 0$ and k_α such that $| D^\alpha f(x) | \leq C_\alpha(1 + \|x\|)^{k_\alpha}$.)
(c) $T(Ax + b) \in \mathscr{S}'$ for any regular real $n \times n$- matrix and $b \in \mathbb{R}^n$.
(d) $T \in \mathscr{S}'(\mathbb{R}^n)$ and $S \in \mathscr{S}'(\mathbb{R}^m)$ implies $T \otimes S \in \mathscr{S}'(\mathbb{R}^{n+m})$.
(e) Let $T \in \mathscr{S}'(\mathbb{R}^n)$, $\psi \in \mathscr{S}(\mathbb{R}^n)$. Define the convolution product

$$\begin{aligned}\langle \psi * T \,,\, \varphi \rangle &= \langle (1(x) \otimes T(y)) \,,\, \psi(x)\varphi(x+y) \rangle \,, \quad \varphi \in \mathscr{S}(\mathbb{R}^n) \\ &= \left\langle T \,,\, \int_{\mathbb{R}^n} \psi(x)\varphi(x+y)\,\mathrm{d}x \right\rangle\end{aligned}$$

Note that this definition coincides with the more general Definition 16.12 since

$$\lim_{k\to\infty} \psi(x)\varphi(x+y)\eta_k(x,y) = \psi(x)\varphi(x+y) \in \mathscr{S}(\mathbb{R}^{2n}).$$

16.4.3 Fourier Transformation in $\mathscr{S}'(\mathbb{R}^n)$

We are following our guiding principle to define the Fourier transform of a distribution $T \in \mathscr{S}'$: First consider the case of a regular tempered distribution. We want to define $\mathcal{F}(T_f) := T_{\mathcal{F}f}$. Suppose that $f(x) \in \mathrm{L}^1(\mathbb{R}^n)$ is integrable. Then its Fourier transformation $\mathcal{F}f$ exists and is a bounded continuous function:

$$| \mathcal{F}f(\xi) | \leq \alpha_n \int_{\mathbb{R}^n} \left| \mathrm{e}^{\mathrm{i}\xi\cdot x} f(x) \right| \,\mathrm{d}x = \alpha_n \int_{\mathbb{R}^n} | f(x) | \,\mathrm{d}x = \alpha_n \|T_f\|_{\mathrm{L}^1} < \infty.$$

By Remark 16.12 (d), $\mathcal{F}f$ defines a distribution in $\mathscr{S}'$. By Fubini's theorem

$$\begin{aligned}\langle T_{\mathcal{F}f} \,,\, \varphi \rangle &= \int_{\mathbb{R}^n} \mathcal{F}f(\xi)\varphi(\xi)\mathrm{d}\xi = \alpha_n \iint_{\mathbb{R}^{2n}} f(x)\mathrm{e}^{-\mathrm{i}\xi\cdot x}\varphi(\xi)\mathrm{d}\xi\,\mathrm{d}x \\ &= \int_{\mathbb{R}^n} f(x)\,\mathcal{F}\varphi(x)\,\mathrm{d}x = \langle T_f \,,\, \mathcal{F}\varphi \rangle \,.\end{aligned}$$

Hence, $\langle T_{\mathcal{F}f} \,,\, \varphi \rangle = \langle T_f \,,\, \mathcal{F}\varphi \rangle$. We take this equation as the definition of the Fourier transformation of a distribution $T \in \mathscr{S}'$.

Definition 16.19 For $T \in \mathscr{S}'$ and $\varphi \in \mathscr{S}$ define

$$\langle \mathcal{F}T \,,\, \varphi \rangle = \langle T \,,\, \mathcal{F}\varphi \rangle \,. \tag{16.13}$$

We call $\mathcal{F}T$ the *Fourier transform* of the distribution T.

Since $\mathcal{F}\varphi \in \mathscr{S}(\mathbb{R}^n)$, $\mathcal{F}T$ it is well-defined on $\mathscr{S}$. Further, $\mathcal{F}$ is a linear operator and T a linear functional, hence $\mathcal{F}T$ is again a linear functional. We show that $\mathcal{F}T$ is a continuous linear functional on $\mathscr{S}$. For, let $\varphi_n \underset{\mathscr{S}}{\longrightarrow} \varphi$ in $\mathscr{S}$. By Theorem 16.10, $\mathcal{F}\varphi_n \underset{\mathscr{S}}{\longrightarrow} \mathcal{F}\varphi$. Since T is continuous,

$$\langle \mathcal{F}T\,,\,\varphi_n\rangle = \langle T\,,\,\mathcal{F}\varphi_n\rangle \longrightarrow \langle T\,,\,\mathcal{F}\varphi\rangle = \langle \mathcal{F}T\,,\,\varphi\rangle$$

which proves continuity.

Lemma 16.14 *The Fourier transformation $\mathcal{F}\colon \mathscr{S}'(\mathbb{R}^n) \to \mathscr{S}'(\mathbb{R}^n)$ is a continuous bijection.*

Proof. (a) We show continuity (see also homework 53.3). Suppose that $T_n \to T$ in $\mathscr{S}'$; that is for all $\varphi \in \mathscr{S}$, $\langle T_n\,,\,\varphi\rangle \to \langle T\,,\,\varphi\rangle$. Hence,

$$\langle \mathcal{F}T_n\,,\,\varphi\rangle = \langle T_n\,,\,\mathcal{F}\varphi\rangle \underset{n\to\infty}{\longrightarrow} \langle T\,,\,\mathcal{F}\varphi\rangle = \langle \mathcal{F}T\,,\,\varphi\rangle$$

which proves the assertion.
(b) We define a second transformation $\mathcal{G}\colon \mathscr{S}' \to \mathscr{S}'$ via

$$\langle \mathcal{G}T\,,\,\varphi\rangle := \langle T\,,\,\mathcal{G}\varphi\rangle$$

and show that $\mathcal{F}\circ\mathcal{G} = \mathcal{G}\circ\mathcal{F} = \mathrm{id}$ on $\mathscr{S}'$. Taking into account Proposition 16.11 we have

$$\langle \mathcal{G}(\mathcal{F}T)\,,\,\varphi\rangle = \langle \mathcal{F}T\,,\,\mathcal{G}\varphi\rangle = \langle T\,,\,\mathcal{F}(\mathcal{G}\varphi)\rangle = \langle T\,,\,\varphi\rangle\,;$$

thus, $\mathcal{G}\circ\mathcal{F} = \mathrm{id}$. The proof of the direction $\mathcal{F}\circ\mathcal{G} = \mathrm{id}$ is similar; hence, $\mathcal{F}$ is a bijection. ∎

Remark 16.13 All the properties of the Fourier transformation as stated in Theorem 16.10 (i), (ii), (iii), (iv), and (v) remain valid in case of $\mathscr{S}'$. In particular, $\mathcal{F}(x^\alpha T) = \mathrm{i}^{|\alpha|} D^\alpha(\mathcal{F}T)$. Indeed, for $\varphi \in \mathscr{S}(\mathbb{R}^n)$, by Theorem 16.10 (ii)

$$\begin{aligned}\mathcal{F}(x^\alpha T)(\varphi) &= \langle x^\alpha T\,,\,\mathcal{F}\varphi\rangle = \langle T\,,\,x^\alpha\mathcal{F}\varphi\rangle = \big\langle T\,,\,(-\mathrm{i})^{|\alpha|}\mathcal{F}(D^\alpha\varphi)\big\rangle \\ &= (-1)^{|\alpha|}(-\mathrm{i})^{|\alpha|}\,\langle D^\alpha(\mathcal{F}T)\,,\,\varphi\rangle = \big\langle \mathrm{i}^{|\alpha|}D^\alpha T\,,\,\varphi\big\rangle\,.\end{aligned}$$

Example 16.12 (a) Let $a \in \mathbb{R}^n$. We compute $\mathcal{F}\,\delta_a$. For $\varphi \in \mathscr{S}(\mathbb{R}^n)$,

$$\mathcal{F}\delta_a(\varphi) = \delta_a(\mathcal{F}\varphi) = (\mathcal{F}\varphi)(a) = \alpha_n \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}x\cdot a}\varphi(x)\,\mathrm{d}x = T_{\alpha_n \mathrm{e}^{-\mathrm{i}x\cdot a}}(\varphi).$$

Hence, $\mathcal{F}\delta_a$ is the regular distribution corresponding to $f(x) = \alpha_n \mathrm{e}^{-\mathrm{i}x\cdot a}$. In particular, $\mathcal{F}(\delta) = T_{\alpha_n 1}$ is the constant function. Note that $\mathcal{F}(\delta) = \mathcal{G}(\delta) = \frac{1}{\sqrt{2\pi}^n} T_1$. Moreover, $\mathcal{F}(T_1) = \mathcal{G}(T_1) = \alpha_n^{-1}\,\delta$.
(b) $n = 1$, $b > 0$.

$$\begin{aligned}\mathcal{F}(H(b - |\,x\,|)) &= \alpha_1 \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}x\xi} H(b - |\,x\,|)\,\mathrm{d}x \\ &= \alpha_1 \int_{-b}^{b} \mathrm{e}^{-\mathrm{i}x\xi}\,\mathrm{d}x = \frac{2}{\sqrt{2\pi}}\,\frac{\sin(b\xi)}{\xi}.\end{aligned}$$

(c) *The single-layer distribution.* Suppose that S is a compact, regular, piecewise differentiable, non self-intersecting surface in $\mathbb{R}^3$ and $\varrho(x) \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^3)$ is a function on S (a density function or distribution—in the physical sense). We define the distribution $\varrho\delta_S$ by the scalar surface integral

$$\langle \varrho\delta_S \, , \, \varphi \rangle = \iint\limits_S \varrho(x)\varphi(x) \, \mathrm{d}S.$$

The support of $\varrho\delta_S$ is S, a set of measure zero with respect to the 3-dimensional Lebesgue measure. Hence, $\varrho\delta_S$ is a singular distribution.
Similarly, one defines the *double-layer distribution* (which comes from dipoles) by

$$\left\langle -\frac{\partial}{\partial \vec{n}} \left(\varrho\delta_S\right) \, , \, \varphi \right\rangle = \iint\limits_S \varrho(x) \, \frac{\partial \varphi(x)}{\partial \vec{n}} \, \mathrm{d}S,$$

where $\vec{n}$ denotes the unit normal vector to the surface.
We compute the Fourier transformation of the single layer $\mathcal{F}(\varrho\delta_S)$ in case of a sphere of radius r, $\mathrm{S}_r = \{x \in \mathbb{R}^3 \mid \|x\| = r\}$ and density $\varrho = 1$. By Fubini's theorem,

$$\begin{aligned}
\langle \mathcal{F}\delta_{\mathrm{S}_r} \, , \, \varphi \rangle &= \left\langle \delta_{\mathrm{S}_r(0)} \, , \, \mathcal{F}\varphi \right\rangle \frac{1}{\sqrt{2\pi}^3} \iint_{\mathrm{S}_r} \left(\int_{\mathbb{R}^3} \mathrm{e}^{-\mathrm{i}x\cdot\xi} \varphi(x) \, \mathrm{d}x \right) \mathrm{d}S_\xi \\
&= \frac{1}{\sqrt{2\pi}^3} \int_{\mathbb{R}^3} \left(\iint_{\mathrm{S}_r} (\cos(x\cdot\xi) - \mathrm{i} \underbrace{\sin(x\cdot\xi)}_{\text{is } 0}) \, \mathrm{d}S_\xi \right) \varphi(x) \, \mathrm{d}x
\end{aligned}$$

Using spherical coordinates on S_r, where x is fixed to be the z-axis and ϑ is the angle between x and $\xi \in \mathrm{S}_r$, we have $\mathrm{d}S = r^2 \sin\vartheta \, \mathrm{d}\varphi \, \mathrm{d}\vartheta$ and $x \cdot \xi = r \, \|x\| \cos\vartheta$. Hence, the inner (surface) integral reads

$$\begin{aligned}
&= \int_0^{2\pi} \int_0^{\pi} \cos(\|x\| \, r \cos\vartheta) r^2 \sin\vartheta \mathrm{d}\vartheta \mathrm{d}\varphi, \quad s = \|x\| \, r \cos\vartheta, \quad \mathrm{d}s = -\|x\| \, r \sin\vartheta \mathrm{d}\vartheta \\
&= 2\pi \int_{\|x\| r}^{-\|x\| r} -\cos s \, \frac{r}{\|x\|} \, \mathrm{d}s = 4\pi \frac{r}{\|x\|} \sin(\|x\| \, r).
\end{aligned}$$

Hence,

$$\langle \mathcal{F}\delta_{\mathrm{S}_r} \, , \, \varphi \rangle = \frac{2r}{\sqrt{2\pi}} \int_{\mathbb{R}^3} \varphi(x) \, \frac{\sin(r\,\|x\|)}{\|x\|} \, \mathrm{d}x;$$

the Fourier transformation of δ_{S_r} is the regular distribution

$$\mathcal{F}\delta_{\mathrm{S}_r}(x) = \frac{2r}{\sqrt{2\pi}} \, \frac{\sin(r\,\|x\|)}{\|x\|}.$$

(d) *The Resolvent of the Laplacian* $-\Delta$. Consider the Hilbert space $H = \mathrm{L}^2(\mathbb{R}^n)$ and its dense subspace $\mathscr{S}(\mathbb{R}^n)$. For $\varphi \in \mathscr{S}$ there is defined the Laplacian $-\Delta\varphi$. Recall that the resolvent of a linear operator A at λ is the bounded linear operator on H, given by $R_\lambda(A) = (A - \lambda I)^{-1}$. Given $f \in H$ we are looking for $u \in H$ with $R_\lambda(A) \, f = u$. This is equivalent to solve

$f = (A - \lambda I)(u)$ for u. In case of $A = -\Delta$ we can apply the Fourier transformation to solve this equation. By Theorem 16.10 (ii)

$$-\Delta u - \lambda u = f, \quad -\mathcal{F}\left(\sum_{k=1}^{n} \frac{\partial^2}{\partial x_k^2} u\right) - \lambda \mathcal{F}u = \mathcal{F}f,$$

$$\sum_{k=1}^{n} \xi_k^2 (\mathcal{F}u)(\xi) - \lambda \mathcal{F}u(\xi) = (\mathcal{F}u)(\xi)(\xi^2 - \lambda) = \mathcal{F}f(\xi)$$

$$\mathcal{F}u(\xi) = \frac{\mathcal{F}f(\xi)}{\xi^2 - \lambda}$$

$$u(x) = \mathcal{G}\left(\frac{1}{\xi^2 - \lambda}\, \mathcal{F}f(\xi)\right)(x)$$

Hence,

$$R_\lambda(-\Delta) = \mathcal{F}^{-1} \circ \frac{1}{\xi^2 - \lambda} \circ \mathcal{F},$$

where in the middle is the multiplication operator by the function $1/(\xi^2 - \lambda)$. One can see that this operator is bounded in H if and only if $\lambda \in \mathbb{C} \setminus \mathbb{R}_+$ such that the spectrum of $-\Delta$ satisfies $\sigma(-\Delta) \subset \mathbb{R}_+$.

16.5 Appendix—More about Convolutions

Since the following proposition is used in several places, we make the statement explicit.

Proposition 16.15 *Let $T(x,t)$ and $S(x,t)$ be distributions in $\mathcal{D}'(\mathbb{R}^{n+1})$ with $\operatorname{supp} T \subset \mathbb{R}^n \times \mathbb{R}_+$ and $\operatorname{supp} S \subset \Gamma_+(0,0)$. Here $\Gamma_+(0,0) = \{(x,t) \in \mathbb{R}^{n+1} \mid \|x\| \le at\}$ denotes the forward light cone at the origin.*
Then the convolution $T * S$ exists in $\mathcal{D}'(\mathbb{R}^{n+1})$ and can be written as

$$\langle T * S\,,\, \varphi \rangle = \langle T(x,t) \otimes S(y,s)\,,\, \eta(t)\eta(s)\eta(as - \|y\|)\, \varphi(x+y, t+s) \rangle\,, \tag{16.14}$$

$\varphi \in \mathcal{D}(\mathbb{R}^{n+1})$, where $\eta \in \mathcal{D}(\mathbb{R})$ with $\eta(t) = 1$ for $t > -\varepsilon$ and $\varepsilon > 0$ is any fixed positive number. The convolution $(T*S)(x,t)$ vanishes for $t < 0$ and is continuous in both components, that is
(a) *If $T_k \to T$ in $\mathcal{D}'(\mathbb{R}^{n+1})$ and $\operatorname{supp} f_k$, $f \subseteq \mathbb{R}^n \times \mathbb{R}$, then $T_k * S \to T * S$ in $\mathcal{D}'(\mathbb{R}^{n+1})$.*
(b) *If $S_k \to S$ in $\mathcal{D}'(\mathbb{R}^{n+1})$ and $\operatorname{supp} S_k$, $S \subseteq \Gamma_+(0,0)$, then $T * S_k \to T * S$ in $\mathcal{D}'(\mathbb{R}^{n+1})$.*

Proof. Since $\eta \in \mathcal{D}(\mathbb{R})$, there exists $\delta > 0$ with $\eta(x) = 0$ for $x < -\delta$. Let $\varphi(x,t) \in \mathcal{D}(\mathbb{R}^{n+1})$ with $\operatorname{supp}\varphi \in U_R(0)$ for some $R > 0$. Let $\eta_K(x,t,y,s)$, $K \to \mathbb{R}^{2n+2}$, be a sequence in $\mathcal{D}(\mathbb{R}^{2n+2})$ converging to 1 in $\mathbb{R}^{2n+2}$, see before Definition 16.12. For sufficiently large K we then have

$$\begin{aligned} \psi_K &:= \eta(s)\eta(t)\eta(as - \|y\|)\eta_K(x,t,y,s)\varphi(x+y, t+s) \\ &= \eta(s)\eta(t)\eta(at - \|y\|)\varphi(x+y, t+s) =: \psi. \end{aligned} \tag{16.15}$$

To prove this it suffices to show that $\psi \in \mathcal{D}(\mathbb{R}^{2n+2})$. Indeed, ψ is arbitrarily often differentiable and its support is contained in

$$\{(x,t,y,s) \mid s,t \geq -\delta, as - \|y\| \geq -\delta, \quad \|x+y\|^2 + |\, r+s\,|^2 \leq R^2\},$$

which is a bounded set.
Since $\eta(t) = 1$ in a neighborhood of $\operatorname{supp} T$ and $\eta(s)\eta(as - \|y\|) = 1$ in a neighborhood of $\operatorname{supp} S$, $T(x,t) = \eta(x)T(x,t)$ and $S(y,s) = \eta(s)\eta(as - \|y\|)S(y,s)$. Using (16.15) we have

$$\begin{aligned}\langle T * S\,,\, \varphi\rangle &= \lim_{K\to\mathbb{R}^{2n+2}} \langle T(x,t)\otimes S(y,s)\,,\, \eta_K(x,t,y,s)\varphi(x+y,t+s)\rangle \\ &= \lim_{K\to\mathbb{R}^{2n+2}} \langle T(x,t)\otimes S(y,s)\,,\, \psi_K\rangle\,, \quad \varphi \in \mathcal{D}(\mathbb{R}^{2n+2}).\end{aligned}$$

This proves the first assertion.
We now prove that the right hand side of (16.14) defines a continuous linear functional on $\mathcal{D}(\mathbb{R}^{n+1})$. Let $\varphi_k \underset{\mathcal{D}}{\longrightarrow} \varphi$ as $k \to \infty$. Then

$$\psi_k := \eta(t)\eta(s)\eta(as - \|y\|)\,\varphi_k(x+y,t+s) \underset{\mathcal{D}}{\longrightarrow} \psi$$

as $k \to \infty$. Hence,

$$\langle T * S\,,\, \varphi_k\rangle = \langle T(x,t)\otimes S(y,s)\,,\, \psi_k\rangle \to \langle T(x,t)\otimes S(y,s)\,,\, \psi\rangle = \langle T * S\,,\, \varphi\rangle\,, \quad k \to \infty,$$

and $T * S$ is continuous.
We show that $T * S$ vanishes for $t < 0$. For, let $\varphi \in \mathcal{D}(\mathbb{R}^{n+1})$ with $\operatorname{supp}\varphi \subseteq \mathbb{R}^n \times (-\infty, -\delta_1]$. Choosing $\delta > \delta_1/2$ one has

$$\eta(t)\eta(s)\eta(as - \|y\|)\varphi(x+y,t+s) = 0,$$

such that $\langle T * S\,,\, \varphi\rangle = 0$. Continuity of the convolution product follows from the continuity of the tensor product. ∎

Chapter 17

PDE II — The Equations of Mathematical Physics

In this chapter we study in detail the Laplace equation, wave equation as well as the heat equation. Firstly, for all space dimensions n we determine the fundamental solutions to the corresponding differential operators; then we consider initial value problems and initial boundary value problems. We study eigenvalue problems for the Laplace equation.
Recall Green's identities, see Proposition 10.2,

$$\iiint_G (u\Delta(v) - v\Delta(u))\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint_{\partial G} \left(u\frac{\partial v}{\partial \vec{n}} - v\frac{\partial u}{\partial \vec{n}}\right)\,\mathrm{d}S,$$
$$\iiint_G \Delta(u)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint_{\partial G} \frac{\partial u}{\partial \vec{n}}\,\mathrm{d}S. \tag{17.1}$$

We also need that for $x \in \mathbb{R}^n \setminus \{0\}$,

$$\Delta\left(\frac{1}{\|x\|^{n-2}}\right) = 0, \quad n \geq 3, \quad \Delta(\log\|x\|) = 0, \quad n = 2,$$

see Example 7.5.

17.1 Fundamental Solutions

17.1.1 The Laplace Equation

Let us denote by ω_n the measure of the unit sphere S^{n-1} in $\mathbb{R}^n$, that is, $\omega_2 = 2\pi$, $\omega_3 = 4\pi$.

Theorem 17.1 *The function*

$$\mathcal{E}_n(x) = \begin{cases} \frac{1}{2\pi}\log\|x\|\,, & n = 2, \\ -\frac{1}{(n-2)\omega_n}\,\frac{1}{\|x\|^{n-2}}, & n \geq 3 \end{cases}$$

is locally integrable; the corresponding regular distribution $\mathcal{E}_n$ satisfies the equation $\Delta\mathcal{E}_n = \delta$, and hence is a fundamental solution for the Laplacian in $\mathbb{R}^n$.

Proof. Step 1. Example 7.5 shows that $\Delta(\mathcal{E}(x)) = 0$ if $x \neq 0$.
Step 2. By homework 50.4, $\log \|x\|$ is in $\mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^2)$ and $1/\|x\|^\alpha \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$ if and only if $\alpha < n$. Hence, $\mathcal{E}_n$, $n \geq 2$, define regular distributions in $\mathbb{R}^n$.
Let $n \geq 3$ and $\varphi \in \mathcal{D}(\mathbb{R}^n)$. Using that $1/\|x\|^{n-2}$ is locally integrable and Example 12.6 (a), $\psi \in \mathcal{D}$ implies

$$\int_{\mathbb{R}^n} \frac{\psi(x)}{r^{n-2}}\,\mathrm{d}x = \lim_{\varepsilon\to 0} \int_{\|x\|\geq\varepsilon} \frac{\psi(x)}{r^{n-2}}\,\mathrm{d}x.$$

Abbreviating $\beta_n = -1/((n-2)\omega_n)$ we have

$$\begin{aligned}\langle \Delta\mathcal{E}_n\,,\,\varphi\rangle &= \beta_n \int_{\mathbb{R}^n} \frac{\Delta\varphi(x)\,\mathrm{d}x}{\|x\|^{n-2}} \\ &= \beta_n \lim_{\varepsilon\to 0} \int_{\|x\|\geq\varepsilon} \frac{\Delta\varphi(x)\,\mathrm{d}x}{\|x\|^{n-2}}.\end{aligned}$$

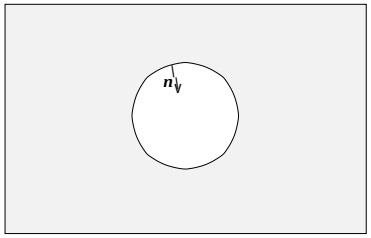
We compute the integral on the right using $v(x) = \frac{1}{r^{n-2}}$, which is harmonic for $\|x\| \geq \varepsilon$, $\Delta v = 0$. Applying Green's identity, we have

$$\beta_n \int_{\|x\|\geq\varepsilon} \frac{\Delta\varphi(x)\,\mathrm{d}x}{\|x\|^{n-2}} = \beta_n \int_{\|x\|=\varepsilon} \left(\frac{1}{r^{n-2}}\frac{\partial}{\partial r}\varphi(x) - \varphi(x)\frac{\partial}{\partial r}\left(\frac{1}{r^{n-2}}\right)\right)\mathrm{d}S$$

Let us consider the first integral as $\varepsilon \to 0$. Note that φ and $\operatorname{grad}\varphi$ are both bounded by a constant C since hφ is a test function. We make use of the estimate

$$\left|\int_{\|x\|=\varepsilon} \frac{\partial}{\partial r}\varphi(x)\frac{\mathrm{d}S}{r^{n-2}}\right| \leq \frac{1}{\varepsilon^{n-2}}\int_{\|x\|=\varepsilon}\left|\frac{\partial\varphi(x)}{\partial r}\,\mathrm{d}S\right| \leq \frac{c}{\varepsilon^{n-2}}\int_{\|x\|=\varepsilon}\mathrm{d}S = \frac{c'\varepsilon^{n-1}}{\varepsilon^{n-2}} = c'\varepsilon$$

which tends to 0 as $\varepsilon \to 0$.



Hence we are left with computing the second integral. Note that the outer unit normal vector to the sphere is $\vec{n} = -\dfrac{x}{\|x\|}$ such that $\frac{\partial}{\partial\vec{n}}\left(\frac{1}{\|x\|^{n-2}}\right) = (n-2)\frac{1}{r^{n-1}}$ and we have

$$\text{second integral} = \beta_n \int_{\|x\|=\varepsilon} \varphi(x)\frac{-(n-2)}{r^{n-1}}\,\mathrm{d}S = \frac{1}{\omega_n}\frac{1}{\varepsilon^{n-1}}\int_{\|x\|=\varepsilon}\varphi(x)\,\mathrm{d}S.$$

Note that $\omega_n\varepsilon^{n-1}$ is exactly the $(n-1)$-dimensional measure of the sphere of radius ε. So, the integral is the mean value of φ over the sphere of radius ε. Since φ is continuous at 0, the mean value tends to $\varphi(0)$. This proves the assertion in case $n \geq 3$.
The proof in case $n = 2$ is quite analogous. ∎

Corollary 17.2 *Suppose that $f(x)$ is a continuous function with compact support. Then $S = \mathcal{E} * f$ is a regular distribution and we have $\Delta S = f$ in $\mathcal{D}'$. In particular,*

$$\begin{aligned} S(x) &= \frac{1}{2\pi} \iint\limits_{\mathbb{R}^2} \log \|x-y\| \; f(y)\,\mathrm{d}y, \quad n=2; \\ S(x) &= -\frac{1}{4\pi} \iiint\limits_{\mathbb{R}^3} \frac{f(y)}{\|x-y\|}\,\mathrm{d}y, \quad n=3. \end{aligned} \tag{17.2}$$

Proof. By Theorem 16.8, $S = \mathcal{E} * f$ is a solution of $Lu = f$ if $\mathcal{E}$ is a fundamental solution of the differential operator L. Inserting the fundamental solution of the Laplacian for $n = 2$ and $n = 3$ and using that f has compact support, the assertion follows. ∎

Remarks 17.1 (a) The given solution (17.2) is even a *classical* solutions of the Poisson equation. Indeed, we can differentiate the parameter integral as usual.
(b) The function $G(x, y) = \mathcal{E}_n(x - y)$ is called the *Green's function* of the Laplace equation.

17.1.2 The Heat Equation

Proposition 17.3 *The function*

$$F(x,t) = \frac{1}{(4\pi a^2 t)^{\frac{n}{2}}}\, H(t)\, \mathrm{e}^{-\frac{\|x\|^2}{4a^2 t}}$$

defines a regular distribution $\mathcal{E} = T_F$ and a fundamental solution of the heat equation $u_t - a^2\Delta u = 0$, that is

$$\mathcal{E}_t - a^2\, \Delta_x \mathcal{E} = \delta(x) \otimes \delta(t). \tag{17.3}$$

Proof. Step 1. The function $F(x,t)$ is locally integrable since $F = 0$ for $t \le 0$ and $F \ge 0$ for $t > 0$ and

$$\int_{\mathbb{R}^n} F(x,t)\,\mathrm{d}x = \frac{1}{(4\pi a^2 t)^{n/2}} \int_{\mathbb{R}^n} \mathrm{e}^{-\frac{r^2}{4a^2 t}}\,\mathrm{d}x = \prod_{k=1}^{n} \left(\frac{1}{\sqrt{\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\xi_k^2} \mathrm{d}\xi_k \right) = 1. \tag{17.4}$$

Step 2. For $t > 0$, $F \in \mathrm{C}^\infty$ and therefore

$$\begin{aligned} \frac{\partial F}{\partial t} &= \left(\frac{x^2}{4a^2 t^2} - \frac{n}{2t} \right) F, \\ \frac{\partial F}{\partial x_i} &= -\frac{x_i}{2a^2 t} F; \quad \frac{\partial^2 F}{\partial x_i^2} = \left(\frac{x_i^2}{4a^4 t^2} - \frac{1}{2a^2 t} \right), \\ \frac{\partial F}{\partial t} - a^2 \Delta F &= 0. \end{aligned} \tag{17.5}$$

See also homework 59.2.

We give a proof using the Fourier transformation with respect to the spatial variables. Let $E(\xi,t) = (\mathcal{F}_x F)(\xi,t)$. We apply the Fourier transformation to (17.3) and obtain a first order ODE with respect to the time variable t:

$$\frac{\partial}{\partial t}E(\xi,t) + a^2\xi^2 E(\xi,t) = \alpha_n 1(\xi)\delta(t).$$

Recall from Example 16.10, that $u' + bu = \delta$ has the fundamental solution $u(t) = H(t)\,\mathrm{e}^{-bt}$; hence

$$E(\xi,t) = H(t)\,\alpha_n \mathrm{e}^{-a^2\xi^2 t}$$

We want to apply the inverse Fourier transformation with respect to the spatial variables. For, note that by Example 16.11,

$$\mathcal{F}^{-1}\left(\mathrm{e}^{-\frac{\xi^2}{2c^2}}\right) = c^n \mathrm{e}^{-\frac{c^2x^2}{2}},$$

where, in our case, $\frac{1}{2c^2} = a^2 t$ or $c = \frac{1}{\sqrt{2a^2t}}$. Hence,

$$\mathcal{E}(x,t) = H(t)\,\alpha_n \mathcal{F}^{-1}\left(\mathrm{e}^{-a^2\xi^2 t}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}}\cdot\frac{1}{(2a^2t)^{\frac{n}{2}}}\mathrm{e}^{-\frac{x^2}{2\cdot 2a^2 t}} = \frac{1}{(4\pi a^2 t)^{\frac{n}{2}}}\,\mathrm{e}^{-\frac{x^2}{4a^2t}}.$$

■

Corollary 17.4 *Suppose that $f(x,t)$ is a continuous function on $\mathbb{R}^n \times \mathbb{R}_+$ with compact support. Let*

$$V(x,t) = H(t)\,\frac{1}{(4a^2\pi)^{\frac{n}{2}}}\int_0^t \int_{\mathbb{R}^n} \frac{\mathrm{e}^{-\frac{\|x-y\|^2}{4a^2(t-s)}}}{(t-s)^{\frac{n}{2}}}\, f(y,s)\,\mathrm{d}y\,\mathrm{d}s$$

Then $V(x,t)$ is a regular distribution in $\mathcal{D}'(\mathbb{R}^n \times \mathbb{R}_+)$ and a solution of $u_t - a^2\Delta u = f$ in $\mathcal{D}'(\mathbb{R}^n \times \mathbb{R}_+)$.

Proof. This follows from Theorem 16.8. ■

17.1.3 The Wave Equation

We shall determine the fundamental solutions for the wave equation in dimensions $n = 3$, $n = 2$, and $n = 1$. In case $n = 3$ we again apply the Fourier transformation. For the other dimensions we use the *method of descent*.

(a) Case $n = 3$

Proposition 17.5

$$\mathcal{E}(x,t) = \delta_{\mathrm{S}_{at}} \otimes \frac{H(t)}{4\pi a^2 t} \in \mathcal{D}'(\mathbb{R}^4)$$

is a fundamental solution for the wave operator $\Box_a u = u_{tt} - a^2(u_{x_1x_1} + u_{x_2x_2} + u_{x_3x_3})$ where $\delta_{\mathrm{S}_{at}}$ denotes the single-layer distribution of the sphere of radius $a{\cdot}t$ around 0.

Proof. As in case of the heat equation let $E(\xi,t) = \mathcal{F}\mathcal{E}(\xi,t)$ be the Fourier transform of the fundamental solution $\mathcal{E}(x,t)$. Then $E(\xi,t)$ satisfies

$$\frac{\partial^2}{\partial t^2}E + a^2\xi^2 E = \alpha_3 1(\xi)\delta(t)$$

Again, this is an ODE of order 2 in t. Recall from Example 16.10 that $u'' + a^2u = \delta$, $a \neq 0$, has a solution $u(t) = H(t)\frac{\sin at}{a}$. Thus,

$$E(\xi,t) = \alpha_3\, H(t)\, \frac{\sin(a\,\|\xi\|\,t)}{a\,\|\xi\|},$$

where ξ is thought to be a parameter. Apply the inverse Fourier transformation $\mathcal{F}_x^{-1}$ to this function. Recall from Example 16.12 (b), the Fourier transform of the single layer of the sphere of radius at around 0 is

$$\mathcal{F}\delta_{\mathrm{S}_{at}}(\xi) = \frac{2at}{\sqrt{2\pi}}\,\frac{\sin(at\,\|\xi\|)}{\|\xi\|}.$$

This shows

$$\mathcal{E}(x,t) = \frac{1}{2\pi}\,\frac{1}{2at}\,H(t)\delta_{\mathrm{S}_{at}}(x)\frac{1}{a} = \frac{1}{4\pi a^2 t}\,H(t)\delta_{\mathrm{S}_{at}}(x)$$

■

Let's evaluate $\langle \mathcal{E}_3\,,\,\varphi(x,t)\rangle$. Using $\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 = \mathrm{d}S_r\;\mathrm{d}r$ where $x = (x_1,x_2,x_3)$ and $r = \|x\|$ as well as the transformation $r = at$, $\mathrm{d}r = a\,\mathrm{d}t$ and $\mathrm{d}S$ is the surface element of the sphere $\mathrm{S}_r(0)$, we obtain

$$\begin{aligned}\langle \mathcal{E}_3\,,\,\varphi(x,t)\rangle &= \frac{1}{4\pi a^2}\int_0^\infty \frac{1}{t}\iint\limits_{\mathrm{S}_{at}} \varphi(x,t)\,\mathrm{d}S\,\mathrm{d}t \qquad (17.6)\\ &= \frac{1}{4\pi a^2}\int_0^\infty \frac{a}{r}\iint\limits_{\mathrm{S}_r} \varphi\left(x,\frac{r}{a}\right)\,\mathrm{d}S\frac{\mathrm{d}r}{a}\\ &= \frac{1}{4\pi a^2}\int_{\mathbb{R}^3}\frac{\varphi\left(x,\frac{\|x\|}{a}\right)}{\|x\|}\,\mathrm{d}x. \qquad (17.7)\end{aligned}$$

(b) The Dimensions $n = 2$ and $n = 1$

To construct the fundamental solution $\mathcal{E}_2(x,t)$, $x = (x_1,x_2)$, we use the so-called method of descent.

Lemma 17.6 *A fundamental solution $\mathcal{E}_2$ of the 2-dimensional wave operator $\Box_{a,2}$ is given by*

$$\langle \mathcal{E}_2\,,\,\varphi(x_1,x_2,t)\rangle = \lim_{k\to\infty}\langle \mathcal{E}_3(x_1,x_2,x_3,t)\,,\,\varphi(x_1,x_2,t)\eta_k(x_3)\rangle\,,$$

where $\mathcal{E}_3$ denotes a fundamental solution of the 3-dimensional wave operator $\Box_{a,3}$ and $\eta_k \in \mathcal{D}(\mathbb{R})$ is the function converging to 1 as $k\to\infty$.

Proof. Let $\varphi \in \mathcal{D}\mathbb{R}^2)$. Noting that $\eta_k'' \to 0$ uniformly on $\mathbb{R}$ as $k \to \infty$, we get

$$\begin{aligned}
\langle \Box_{a,2}\mathcal{E}_2\,,\, \varphi(x_1,x_2,t)\rangle &= \langle \mathcal{E}_2\,,\, \Box_{a,2}\varphi(x_1,x_2,t)\rangle \\
&= \lim_{k\to\infty} \langle \mathcal{E}_3\,,\, \Box_{a,2}(\varphi(x_1,x_2,t)\eta_k(x_3))\rangle \\
&= \lim_{k\to\infty} \langle \mathcal{E}_3\,,\, \Box_{a,2}\varphi(x_1,x_2,t)\cdot\eta_k(x_3)+\varphi\cdot\eta_k''(x_3)\rangle \\
&= \lim_{k\to\infty} \big\langle \mathcal{E}_3\,,\, \Box_{a,3}\big(\varphi(x_1,x_2,t)\,\eta_k(x_3)\big)\big\rangle \\
&= \lim_{k\to\infty} \langle \Box_{a,3}\mathcal{E}_3\,,\, \varphi(x_1,x_2,t)\eta_k(x_3)\rangle = \\
&\lim_{k\to\infty} \langle \delta(x_1,x_2,x_3)\delta(t)\,,\, \varphi(x_1,x_2,t)\eta_k(x_3)\rangle \\
&= \varphi(0,0,0) = \langle \delta(x_1,x_2,t)\,,\, \varphi(x_1,x_2,t)\rangle\,.
\end{aligned}$$

In the third line we used that $\Delta_{x_1,x_2,x_3}\big(\varphi(x_1,x_2,t)\cdot\eta(x_3)\big) = \Delta_{x_1,x_2}\varphi(x_1,x_2,t)\,\eta + \varphi\,\eta''(x_3)$. ∎

Proposition 17.7 (a) *For $x = (x_1,x_2) \in \mathbb{R}^2$ and $t \in \mathbb{R}$, the regular distribution*

$$\mathcal{E}_2(x,t) = \frac{1}{2\pi a}\,\frac{H(at-\|x\|)}{\sqrt{a^2t^2-x^2}} = \begin{cases} \frac{1}{2\pi a}\,\frac{1}{\sqrt{a^2t^2-x^2}}, & at > \|x\|\,, \\ 0, & at \le \|x\| \end{cases}$$

is a fundamental solution of the 2-dimensional wave operator.
(b) *The regular distribution*

$$\mathcal{E}_1(x,t) = \frac{1}{2a}H(at-|\,x\,|) = \begin{cases} \frac{1}{2a}, & |\,x\,| < at, \\ 0, & |\,x\,| \ge at \end{cases}$$

is a fundamental solution of the one-dimensional wave operator.

Proof. By the above lemma,

$$\langle \mathcal{E}_2\,,\, \varphi(x_1,x_2,t)\rangle = \langle \mathcal{E}_3\,,\, \varphi(x_1,x_2,t)1(x_3)\rangle = \frac{1}{4\pi a^2}\int_0^\infty \frac{1}{t}\iint\limits_{\mathrm{S}_{at}} \varphi(x_1,x_2,t)\,\mathrm{d}S\,\mathrm{d}t.$$

We compute the surface element of the sphere of radius at around 0 in terms of x_1, x_2. The surface is the graph of the function $x_3 = f(x_1,x_2) = \sqrt{a^2t^2-x_1^2-x_2^2}$. By the formula before Example 10.4, $\mathrm{d}S = \sqrt{1+f_{x_1}^2+f_{x_2}^2}\,\mathrm{d}x_1\,\mathrm{d}x_2$. In case of the sphere we have

$$\mathrm{d}S_{x_1,x_2} = \frac{at\,\mathrm{d}x_1\,\mathrm{d}x_2}{\sqrt{a^2t^2-x_1^2-x_2^2}}$$

Integration over both the upper and the lower half-sphere yields factor 2,

$$\begin{aligned}
&= 2\,\frac{1}{4\pi a^2}\int_0^\infty \frac{1}{t} \iint\limits_{x_1^2+x_2^2\le a^2t^2} \frac{at\varphi(x_1,x_2,t)}{\sqrt{a^2t^2-x_1^2-x_2^2}}\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}t \\
&= \frac{1}{2\pi a}\int_0^\infty \iint\limits_{\|x\|\le at} \frac{\varphi(x_1,x_2,t)}{\sqrt{a^2t^2-x_1^2-x_2^2}}\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}t.
\end{aligned}$$

This shows that $\mathcal{E}_2(x,t)$ is a regular distribution of the above form.
One can show directly that $\mathcal{E}_2 \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^3)$. Indeed, $\iiint\limits_{\mathbb{R}^2\times[-R,R]} \mathcal{E}_2(x_1,x_2,t)\,\mathrm{d}x\,\mathrm{d}t < \infty$ for all $R>0$.
(b) It was already shown in homework 57.2 that $\mathcal{E}_1$ is the fundamental solution to the one-dimensional wave operator. A short proof is to be found in [Wla72, II. §6.5 Example g)]. ∎

Use the method of descent to complete this proof. Let $\varphi \in \mathcal{D}(\mathbb{R}^2)$. Since $\int_{\mathbb{R}} |\,\mathcal{E}_2(x_1,x_2,t)\,|\,\mathrm{d}x_2 < \infty$ and $\int_{\mathbb{R}} \mathcal{E}_2(x_1,x_2,t)\,\mathrm{d}x_2$ again defines a locally integrable function, we have as in Lemma 17.6

$$\mathcal{E}_1(\varphi) = \lim_{k\to\infty} \langle \mathcal{E}_2(x_1,x_2,t)\,,\,\varphi(x_1,t)\eta_k(x_2)\rangle = \lim_{k\to\infty} \int_{\mathbb{R}^3} \mathcal{E}_2(x_1,x_2,t)\eta_k(x_2)\varphi(x_1,t)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}t.$$

Hence, the fundamental solution $\mathcal{E}_1$ is the regular distribution

$$\mathcal{E}_1(x_1,t) = \int_{-\infty}^{\infty} \mathcal{E}_2(x_1,x_2,t)\,\mathrm{d}x_2$$

17.2 The Cauchy Problem

In this section we formulate and study the classical and generalized Cauchy problems for the wave equation and for the heat equation.

17.2.1 Motivation of the Method

To explain the method, we first apply the theory of distribution to solve an initial value problem of a linear second order ODE.
Consider the Cauchy problem

$$u''(t) + a^2 u(t) = f(t), \quad u\,|_{t=0+} = u_0, \quad u'\,|_{t=0+} = u_1, \tag{17.8}$$

where $f \in \mathrm{C}(\mathbb{R}_+)$. We extend the solution $u(t)$ as well as $f(t)$ by 0 for negative values of t, $t<0$. We denote the new function by $\tilde{u}$ and $\tilde{f}$, respectively. Since $\tilde{u}$ has a jump of height u_0 at 0, by Example 16.6, $\tilde{u}'(t) = \{u'(t)\} + u_0\delta(t)$. Similarly, $u'(t)$ jumps at 0 by u_1 such that $\tilde{u}''(t) = \{u''(t)\} + u_0\delta'(t) + u_1\delta(t)$. Hence, $\tilde{u}$ satisfies on $\mathbb{R}$ the equation

$$\tilde{u}'' + a^2\tilde{u} = \tilde{f}(t) + u_0\delta'(t) + u_1\delta(t). \tag{17.9}$$

We construct the solution $\tilde{u}$. Since the fundamental solution $\mathcal{E}(t) = H(t)\sin at/a$ as well as the right hand side of (17.9) has positive support, the convolution product exists and equals

$$\tilde{u} = \mathcal{E} * (\tilde{f} + u_0\delta'(t) + u_1\delta(t)) = \mathcal{E} * \tilde{f} + u_0\mathcal{E}'(t) + u_1\mathcal{E}(t)$$
$$\tilde{u} = \frac{1}{a}\int_0^t f(\tau)\sin a(t-\tau)\mathrm{d}\tau + u_0\mathcal{E}'(t) + u_1\mathcal{E}(t).$$

Since in case $t>0$, $\tilde{u}$ satisfies (17.9) and the solution of the Cauchy problem is unique, the above formula gives the classical solution for $t>0$, that is

$$u(t) = \frac{1}{a}\int_0^t f(\tau)\sin a(t-\tau)\mathrm{d}\tau + u_0\cos at + u_1\frac{\sin at}{a}.$$

17.2.2 The Wave Equation

(a) The Classical and the Generalized Initial Value Problem—Existence, Uniqueness, and Continuity

Definition 17.1 (a) The problem

$$\Box_a u = f(x,t), \quad x \in \mathbb{R}^n, t > 0, \tag{17.10}$$

$$u|_{t=0+} = u_0(x), \tag{17.11}$$

$$\left.\frac{\partial u}{\partial t}\right|_{t=0+} = u_1(x), \tag{17.12}$$

where we assume that

$$f \in \mathrm{C}(\mathbb{R}^n \times \mathbb{R}_+), \quad u_0 \in \mathrm{C}^1(\mathbb{R}^n), \quad u_1 \in \mathrm{C}(\mathbb{R}^n).$$

is called the *classical initial value problem* (CIVP, for short) to the wave equation.
A function $u(x,t)$ is called *classical solution* of the CIVP if

$$u(x,t) \in \mathrm{C}^2(\mathbb{R}^n \times \mathbb{R}^+) \cap \mathrm{C}^1(\mathbb{R}^n \times \mathbb{R}_+),$$

$u(x,y)$ satisfies the wave equation (17.10) for $t > 0$ and the initial conditions (17.11) and (17.12) as $t \to 0+0$.
(b) The problem

$$\Box_a U = F(x,t) + U_0(x) \otimes \delta'(t) + U_1(x) \otimes \delta(t)$$

with $F \in \mathcal{D}'(\mathbb{R}^{n+1})$, $U_0, U_1 \in \mathcal{D}'(\mathbb{R}^n)$, and $\operatorname{supp} F \subset \mathbb{R}^n \times [0,+\infty)$ is called *generalized initial value problem* (GIVP). A generalized function $U \in \mathcal{D}'(\mathbb{R}^{n+1})$ with $\operatorname{supp} U \subset \mathbb{R}^n \times [0,+\infty)$ which satisfies the above equation is called a (generalized, weak) solution of the GIVP.

Proposition 17.8 (a) *Suppose that $u(x,t)$ is a solution of the CIVP with the given data f, u_0, and u_1. Then the regular distribution T_u is a solution of the GIVP with the right hand side $T_f + T_{u_0} \otimes \delta'(t) + T_{u_1} \otimes \delta(t)$ provided that $f(x,t)$ and $u(x,t)$ are extended by 0 into the domain $\{(x,t) \mid (x,t) \in \mathbb{R}^{n+1},\ t < 0\}$.*
(b) *Conversely, suppose that U is a solution of the GIVP. Let the distributions $F = T_f$, $U_0 = T_{u_0}$, $U_1 = T_{u_1}$ and $U = T_u$ be regular and they satisfy the regularity assumptions of the CIVP. Then, $u(x,t)$ is a solution of the CIVP.*

Proof. (b) Suppose that U is a solution of the GIVP; let $\varphi \in \mathcal{D}(\mathbb{R}^{n+1})$. By definition of the tensor product and the derivative,

$$\begin{aligned}\big\langle U_{tt} - a^2 \Delta U\,,\, \varphi \big\rangle &= \langle F\,,\,\varphi\rangle + \langle U_0 \otimes \delta'\,,\,\varphi\rangle + \langle U_1 \otimes \delta\,,\,\varphi\rangle \\ &= \int_0^\infty \int_{\mathbb{R}^n} f(x,t)\varphi(x,t)\,\mathrm{d}x\,\mathrm{d}t - \int_{\mathbb{R}^n} u_0(x)\frac{\partial \varphi}{\partial t}(x,0)\,\mathrm{d}x + \int_{\mathbb{R}^n} u_1(x)\varphi(x,0)\,\mathrm{d}x. \end{aligned} \tag{17.13}$$

Applying integration by parts with respect to t twice, we find

$$\begin{aligned}\int_0^\infty u\varphi_{tt}\,\mathrm{d}t &= u\varphi_t|_0^\infty - \int_0^\infty u_t\varphi_t\,\mathrm{d}t\\ &= -u(x,0)\varphi_t(x,0) - u_t\varphi|_0^\infty + \int_0^\infty u_t t\varphi\,\mathrm{d}t\\ &= -u(x,0)\varphi_t(x,0) + u_t(x,0)\varphi(x,0) + \int_0^\infty u_t t\varphi\,\mathrm{d}t.\end{aligned}$$

Since $\mathbb{R}^n$ has no boundary and φ has compact support, integration by parts with respect to the spatial variables x yields no boundary terms.
Hence, by the above formula and $\iint u\,\Delta\varphi\,\mathrm{d}t\,\mathrm{d}x = \iint \Delta u\,\varphi\,\mathrm{d}t\,\mathrm{d}x$, we obtain

$$\begin{aligned}&\left\langle U_{tt} - a^2\Delta U\,,\,\varphi\right\rangle = \left\langle U\,,\,\varphi_{tt} - a^2\Delta\varphi\right\rangle = \int_{\mathbb{R}^n}\int_0^\infty u(x,t)\left(\varphi_{tt} - a^2\Delta\varphi\right)\,\mathrm{d}t\,\mathrm{d}x\\ &= \int_{\mathbb{R}^n}\int_0^\infty \left(u_{tt} - a^2\Delta u\right)\varphi(x,t)\,\mathrm{d}x\,\mathrm{d}t - \int_{\mathbb{R}^n}\left(u(x,0)\varphi_t(x,0) - u_t(x,0)\varphi(x,0)\right)\,\mathrm{d}x.\end{aligned} \tag{17.14}$$

For any $\varphi \in \mathcal{D}(\mathbb{R}^n \times \mathbb{R}_+)$, $\operatorname{supp}\varphi$ is contained in $\mathbb{R}^n \times (0,+\infty)$ such that $\varphi(x,0) = \varphi_t(x,0) = 0$. From (17.13) and (17.14) it follows that

$$\int_{\mathbb{R}^n}\int_0^\infty \left(f(x,t) - u_{tt} + a^2\Delta u\right)\varphi(x,t)\,\mathrm{d}t\,\mathrm{d}x = 0.$$

By Lemma 16.2 (Du Bois Reymond) it follows that $u_{tt} - a^2\Delta u = f$ on $\mathbb{R}^n \times \mathbb{R}_+$. Inserting this into (17.13) and (17.14) we have

$$\int_{\mathbb{R}^n}(u_0(x) - u(x,0))\varphi_t(x,0)\,\mathrm{d}x - \int_{\mathbb{R}^n}(u_1(x) - u_t(x,0))\,\varphi(x,0)\,\mathrm{d}x = 0.$$

If we set $\varphi(x,t) = \psi(x)\eta(t)$ where $\eta \in \mathcal{D}(\mathbb{R})$ and $\eta(t) = 1$ is constant in a neighborhood of 0, $\varphi_t(x,0) = 0$ and therefore

$$\int_{\mathbb{R}^n}(u_1(x) - u_t(x,0))\psi(x) = 0, \quad \psi \in \mathcal{D}(\mathbb{R}^n).$$

Moreover,

$$\int_{\mathbb{R}^n}(u_0(x) - u(x,0))\psi(x)\,\mathrm{d}x = 0, \quad \psi \in \mathcal{D}(\mathbb{R}^n)$$

if we set $\varphi(x,t) = t\eta(t)\psi(x)$. Again, Lemma 16.2 yields

$$u_0(x) = u(x,0), \quad u_1(x) = u_t(x,0)$$

and $u(x,t)$ is a solution of the CIVP.
(a) Conversely, if $u(x,t)$ is a solution of the CIVP then (17.14) holds with

$$U(x,t) = H(t)\,u(x,t).$$

Comparing this with (17.13) it is seen that

$$U_{tt} - a^2\Delta U = F + U_0 \otimes \delta' + U_1 \otimes \delta$$

where $F(x,t) = H(t)f(x,t)$, $U_0(x) = u(x,0)$ and $U_1(x) = u_t(x,0)$. ∎

Corollary 17.9 *Suppose that F, U_0, and U_1 are data of the GIVP. Then there exists a unique solution U of the GIVP. It can be written as*

$$U = V + V^{(0)} + V^{(1)}$$

where

$$V = \mathcal{E}_n * F, \quad V^{(1)} = \mathcal{E}_n *_x U_1, \quad V^{(0)} = \frac{\partial \mathcal{E}_n}{\partial t} *_x U_0.$$

Here $\mathcal{E}_n *_x U_1 := \mathcal{E}_n * (U_1(x) \otimes \delta(t))$ denotes the convolution product with respect to the spatial variables x only. The solution U depends continuously in the sense of the convergence in $\mathcal{D}'$ on F, U_0, and U_1. Here $\mathcal{E}_n$ denote the fundamental solution of the n-dimensional wave operator $\Box_{a,n}$.

Proof. The supports of the distributions $U_0 \otimes \delta'$ and $U_1 \otimes \delta$ are contained in the hyperplane $\{(x,t) \in \mathbb{R}^{n+1} \mid t = 0\}$. Hence the support of the distribution $F + U_0 \otimes \delta' + U_1 \otimes \delta$ is contained in the half space $\mathbb{R}^n \times \mathbb{R}_+$.
It follows from Proposition 16.15 below that the convolution product

$$U = \mathcal{E}_n * (F + U_0 \otimes \delta' + U_1 \otimes \delta)$$

exists and has support in the positive half space $t \geq 0$. It follows from Theorem 16.8 that U is a solution of the GIVP. On the other and, any solution of the GIVP has support in $\mathbb{R}^n \times \mathbb{R}_+$ and therefore,
by Proposition 16.15, posses the convolution with $\mathcal{E}_n$. By Theorem 16.8, the solution U is unique.
Suppose that $U_k \longrightarrow U_1$ as $k \to \infty$ in $\mathcal{D}'(\mathbb{R}^{n+1})$ then $\mathcal{E}_n * U_k \longrightarrow \mathcal{E}_n * U_1$ by the continuity of the convolution product in $\mathcal{D}'$ (see Proposition 16.15). ∎

(b) Explicit Solutions for $n = 1, 2, 3$

We will make the above formulas from Corollary 17.9 explicit, that is, we compute the above convolutions to obtain the potentials V, $V^{(0)}$, and $V^{(1)}$.

Proposition 17.10 *Let $f \in \mathrm{C}^2(\mathbb{R}^n \times \mathbb{R}_+)$, $u_0 \in \mathrm{C}^3(\mathbb{R}^n)$, and $u_1 \in \mathrm{C}^2(\mathbb{R}^n)$ for $n = 2, 3$; let $f \in \mathrm{C}^1(\mathbb{R}_+)$, $u_0 \in \mathrm{C}^2(\mathbb{R})$, and $u_1 \in \mathrm{C}^1(\mathbb{R})$ in case $n = 1$.*
Then there exists a unique solution of the CIVP. It is given in case $n = 3$ by Kirchhoff's formula

$$u(x,t) = \frac{1}{4\pi a^2}\left(\iiint\limits_{U_{at}(x)} \frac{f\left(y, t - \left\|\frac{x-y}{a}\right\|\right)}{\|x-y\|}\,\mathrm{d}y + \frac{1}{t}\iint\limits_{\mathrm{S}_{at}(x)} u_1(y)\,\mathrm{d}S_y + \frac{\partial}{\partial t}\left(\frac{1}{t}\iint\limits_{\mathrm{S}_{at}(x)} u_0(y)\,\mathrm{d}S_y\right)\right). \tag{17.15}$$

The first term V is called retarded potential.
In case $n = 2$, $x = (x_1, x_2)$, $y = (y_1, y_2)$, it is given by Poisson's formula

$$u(x,t) = \frac{1}{2\pi a}\int_0^t \iint\limits_{U_{a(t-s)}(x)} \frac{f(y,s)\,\mathrm{d}y\,\mathrm{d}s}{\sqrt{a^2(t-s)^2 - \|x-y\|^2}} + \frac{1}{2\pi a}\iint\limits_{U_{at}(x)} \frac{u_1(y)\,\mathrm{d}y}{\sqrt{a^2t^2 - \|x-y\|^2}} + \frac{1}{2\pi a}\frac{\partial}{\partial t}\iint\limits_{U_{at}(x)} \frac{u_0(y)\,\mathrm{d}y}{\sqrt{a^2t^2 - \|x-y\|^2}}. \quad (17.16)$$

In case $n = 1$ it is given by d'Alembert's formula

$$u(x,t) = \frac{1}{2a}\int_0^t \int_{x-a(t-s)}^{x+a(t-s)} f(y,s)\,\mathrm{d}y\,\mathrm{d}s + \frac{1}{2a}\int_{x-at}^{x+at} u_1(y)\,\mathrm{d}y + \frac{1}{2}\left(u_0(x+at) + u_0(x-at)\right). \quad (17.17)$$

The solution $u(x,t)$ depends continuously on u_0, u_1, and f in the following sense: If

$$\left| f - \tilde{f} \right| < \varepsilon, \quad |\, u_0 - \tilde{u}_0 \,| < \varepsilon_0, \quad |\, u_1 - \tilde{u}_1 \,| < \varepsilon_1, \quad \|\operatorname{grad}(u_0 - \tilde{u}_0)\| < \varepsilon_0'$$

(where we impose the last inequality only in cases $n = 3$ and $n = 2$), then the corresponding solutions $u(x,t)$ and $\tilde{u}(x,t)$ satisfy in a strip $0 \le t \le T$

$$|\, u(x,t) - \tilde{u}(x,y) \,| < \frac{1}{2}T^2\varepsilon + T\varepsilon_1 + \varepsilon_0 + (aT\varepsilon_0'),$$

where the last term is omitted in case $n = 1$.

Proof. (idea of proof) We show Kirchhoff's formula.
(a) *The potential term with f.*
By Proposition 16.15 below, the convolution product $\mathcal{E}_3 * f$ exists. It is shown in [Wla72, p. 153] that for a locally integrable function $f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^{n+1})$ with $\operatorname{supp} f \subset \mathbb{R}^n \times \mathbb{R}_+$, $\mathcal{E}_n * T_f$ is again a locally integrable function.
Formally, the convolution product is given by

$$(\mathcal{E}_3 * f)(x,t) = \int_{\mathbb{R}^4} \mathcal{E}_3(y,s) f(x-y, t-s)\,\mathrm{d}y\,\mathrm{d}s = \int_{\mathbb{R}^4} \mathcal{E}_3(x-y, t-s)\, f(y,s)\,\mathrm{d}y\,\mathrm{d}s,$$

where the integral is to be understood the evaluation of $\mathcal{E}_3(y,s)$ on the shifted function $f(x-y, t-s)$. Since f has support on the positive time axis, one can restrict oneselves to $s > 0$ and $t - s > 0$, that is to $0 < s < t$. That is formula (17.6) gives

$$\mathcal{E}_3 * f(x,t) = \frac{1}{4\pi a^2}\int_0^t \frac{1}{s}\iint\limits_{\mathrm{S}_{as}} f(x-y, t-s)\,\mathrm{d}S(y)\,\mathrm{d}s$$

Using $r = as$, $\mathrm{d}r = a\,\mathrm{d}s$, we obtain

$$V(x,t) = \frac{1}{4\pi a^2}\int_0^{at} \iint\limits_{\mathrm{S}_r} \frac{1}{r} f\left(x-y, t-\frac{r}{a}\right)\mathrm{d}S(y)\,\mathrm{d}r.$$

Using $\mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}y_3 = \mathrm{d}r\,\mathrm{d}S$ as well as $\|y\| = r = as$ we can proceed

$$V(x,t) = \frac{1}{4\pi a^2} \iiint\limits_{U_{at(0)}} \frac{f\left(x-y, t-\frac{\|y\|}{a}\right)}{\|y\|}\,\mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}y_3.$$

The shift $z = x - y$, $\mathrm{d}z_1\,\mathrm{d}z_2\,\mathrm{d}z_3 = \mathrm{d}y_1\,\mathrm{d}y_2\,\mathrm{d}y_3$ finally yields

$$V(x,t) = \frac{1}{4\pi a^2} \iiint\limits_{U_{at(x)}} \frac{f\left(z, t-\frac{\|x-z\|}{a}\right)}{\|x-z\|}\,\mathrm{d}z.$$

This is the first potential term of Kirchhoff's formula.
(b) We compute $V^{(1)}(x,t)$. By definition,

$$V^{(1)} = \mathcal{E}_3 * (u_1 \otimes \delta) = \mathcal{E}_3 *_x u_1.$$

Formally, this is given by,

$$\begin{aligned} V^{(1)}(x,t) &= \frac{1}{4\pi a^2 t} \iiint\limits_{\mathbb{R}^3} \delta_{\mathrm{S}_{at}}(y)\, u_1(x-y)\,\mathrm{d}y = \frac{1}{4\pi a^2 t} \iint\limits_{\mathrm{S}_{at}} u_1(x-y)\,\mathrm{d}S(y) \\ &= \frac{1}{4\pi a^2 t} \iint\limits_{\mathrm{S}_{at}(x)} u_1(y)\,\mathrm{d}S(y). \end{aligned}$$

(c) Recall that $(D^\alpha S) * T = D^\alpha (S * T)$, by Remark 16.8 (b). In particular

$$\mathcal{E}_3 * (u_0 \otimes \delta') = \frac{\partial}{\partial t}\left(\mathcal{E}_3 *_x u_0\right);$$

which immediately gives (c) in view of (b). ∎

Remark 17.2 (a) The stronger regularity (differentiability) conditions on f, u_0, u_1 are necessary to prove $u \in \mathrm{C}^2(\mathbb{R}^n \times \mathbb{R}^+)$ and to show stability.
(b) Proposition 17.10 and Corollary 17.9 show that the GIVP for the wave wave equation is a well-posed problem (existence, uniqueness, stability).

17.2.3 The Heat Equation

Definition 17.2 (a) The problem

$$u_t - a^2 \Delta u = f(x,t), \quad x \in \mathbb{R}^n, t > 0 \tag{17.18}$$

$$u(x,0) = u_0(x), \tag{17.19}$$

where we assume that

$$f \in \mathrm{C}(\mathbb{R}^n \times \mathbb{R}_+), \quad u_0 \in \mathrm{C}(\mathbb{R}^n)$$

is called the *classical initial value problem* (CIVP, for short) to the heat equation.
A function $u(x,t)$ is called *classical solution* of the CIVP if

$$u(x,t) \in \mathrm{C}^2(\mathbb{R}^n \times (0,+\infty)) \cap \mathrm{C}(\mathbb{R}^n \times [0,+\infty)),$$

and $u(x,t)$ satisfies the heat equation (17.18) and the initial condition (17.19).
(b) The problem

$$U_t - a^2 \Delta U = F + U_0 \otimes \delta$$

with $F \in \mathcal{D}'(\mathbb{R}^{n+1})$, $U_0 \in \mathcal{D}'(\mathbb{R}^n)$, and $\operatorname{supp} F \subset \mathbb{R}^n \times \mathbb{R}_+$ is called *generalized initial value problem* (GIVP). A generalized function $U \in \mathcal{D}'(\mathbb{R}^{n+1})$ with $\operatorname{supp} U \subset \mathbb{R}^n \times \mathbb{R}_+$ which satisfies the above equation is called a *generalized solution* of the GIVP.

The fundamental solution of the heat operator has the following properties:

$$\int_{\mathbb{R}^n} \mathcal{E}(x,t)\, \mathrm{d}x = 1,$$
$$\mathcal{E}(x,t) \longrightarrow \delta(x), \quad \text{as} \quad t \to 0+ .$$

The fundamental solution describes the heat distribution of a point-source at the origin $(0,0)$. Since $\mathcal{E}(x,t) > 0$ for all $t > 0$ and all $x \in \mathbb{R}^n$, the heat propagates with infinite speed. This is in contrast to our experiences. However, for short distances, the heat equation is gives sufficiently good results. For long distances one uses the transport equation. We summarize the results which are similar to that of the wave equation.

Proposition 17.11 (a) *Suppose that $u(x,t)$ is a solution of the CIVP with the given data f and u_0. Then the regular distribution $T_{\tilde{u}}$ is a solution of the GIVP with the right hand side $T_{\tilde{f}} + T_{u-0} \otimes \delta$ provided that $f(x,t)$ and $u(x,t)$ are extended to $\tilde{f}(x,t)$ and $\tilde{u}(x,t)$ by 0 into the left half-space $\{(x,t) \mid (x,t) \in \mathbb{R}^{n+1},\ t < 0\}$.*
(b) *Conversely, suppose that U is a solution of the GIVP. Let the distributions $F = T_f$, $U_0 = T_{u_0}$, and $U = T_u$ be regular and they satisfy the regularity assumptions of the CIVP.*
Then, $u(x,t)$ is a solution of the CIVP.

Proposition 17.12 *Suppose that F and U_0 are data of the GIVP. Suppose further that F and U_0 both have compact support. Then there exists a solution U of the GIVP which can be written as*

$$U = V + V^{(0)}$$

where

$$V = \mathcal{E} * F, \quad V^{(0)} = \mathcal{E} *_x U_0.$$

The solution U varies continuously with F and U_0.

Remark 17.3 The theorem differs from the corresponding result for the wave equation in that there is no proposition on uniqueness. It turns out that the GIVP cannot be solved uniquely. A. Friedman, Partial differential equations of parabolic type, gave an example of a non-vanishing

distribution which solves the GIVP with $F = 0$ and $U_0 = 0$.
However, if all distributions are regular and we place additional requirements on the growth for $t, \|x\| \to \infty$ of the regular distribution, uniqueness can be achieved.

For existence and uniqueness we introduce the following class of functions

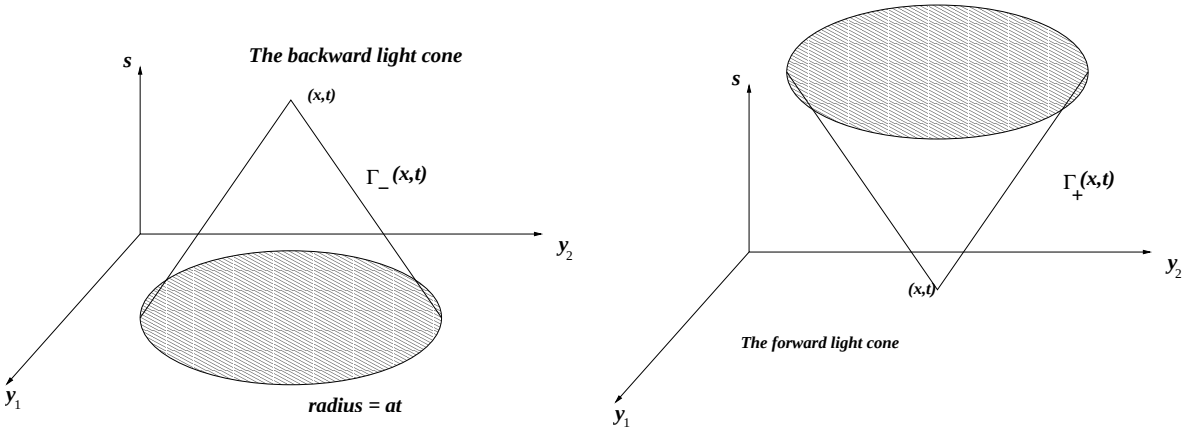
$$\mathcal{M} = \{f \in \mathrm{C}(\mathbb{R}^n \times \mathbb{R}_+) \mid f \quad \text{is bounded on the strip} \quad \mathbb{R}^n \times [0,T] \quad \text{for all } T > 0\},$$
$$\mathrm{C}_b(\mathbb{R}^n) = \{f \in \mathrm{C}(\mathbb{R}^n) \mid f \quad \text{is bounded on} \quad \mathbb{R}^n\}$$

Corollary 17.13 (a) *Let $f \in \mathcal{M}$ and $u_0 \in \mathrm{C}_b(\mathbb{R}^n)$. Then the two potentials $V(x,t)$ as in Corollary 17.4 and*

$$V^{(0)}(x,t) = \mathcal{E} * T_{u_0} \otimes \delta = \frac{H(t)}{(4\pi a^2 t)^{\frac{n}{2}}} \int_{\mathbb{R}^n} u_0(y) \mathrm{e}^{-\frac{\|x-y\|^2}{4a^2 t}} \, \mathrm{d}y$$

are regular distributions and $u = V + V^{(0)}$ is a solution of the GIVP.
(b) *In case $f \in \mathrm{C}^2(\mathbb{R}^n \times \mathbb{R}_+)$ with $D^\alpha f \in \mathcal{M}$ for all α with $|\alpha| \leq 1$ (first order partial derivatives), the solution in* (a) *is a solution of the CIVP. In particular, $V^{(0)}(x,t) \longrightarrow u_0(x)$ as $t \to 0+$.*
(c) *The solution u of the GIVP is unique in the class $\mathcal{M}$.*

17.2.4 Physical Interpretation of the Results



Definition 17.3 We introduce the two cones in $\mathbb{R}^{n+1}$

$$\Gamma_-(x,t) = \{(y,s) \mid \|x-y\| < a(t-s)\}, \quad s < t,$$
$$\Gamma_+(x,t) = \{(y,s) \mid \|x-y\| < a(s-t)\}, \quad s > t,$$

which are called *domain of dependence* (backward light cone) and *domain of influence* (forward light cone), respectively.

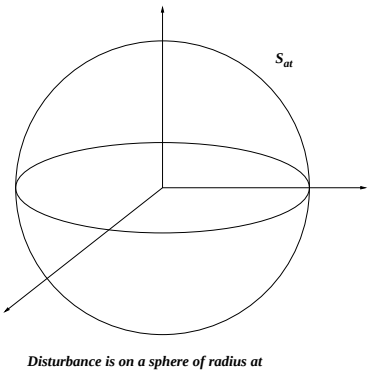
Recall that the boundaries $\partial\Gamma_+$ and $\partial\Gamma_-$ are characteristic surfaces of the wave equation.

(a) Propagation of Waves in Space

Consider the fundamental solution

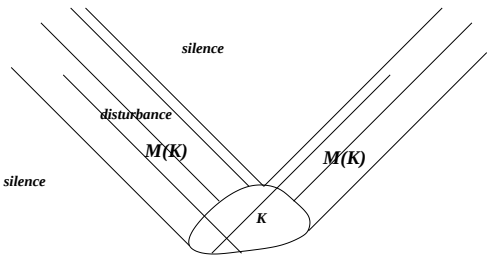
$$\mathcal{E}_3(x,t) = \frac{1}{4\pi a^2}\delta_{\mathrm{S}_{at}} \otimes T_{\frac{1}{t}}$$

of the 3-dimensional wave equation.



It shows that the disturbance at time $t > 0$ effected by a point source $\delta(x)\delta(t)$ in the origin is located on a sphere of radius at around 0. The disturbance moves like a spherical wave, $\|x\| = at$, with velocity a. In the beginning there is silence, then disturbance (on the sphere), and afterwards, again silence. This is called *Huygens' principle*.

It follows by the superposition principle that the solution $u(x_0, t_0)$ of an initial disturbance $u_0(x)\delta'(t) + u_1(x)\delta(t)$ is completely determined by the values of u_0 and u_1 on the sphere of the backwards light-cone at $t = 0$; that is by the values $u_0(x)$ and $u_1(x)$ at all values x with $\|x - x_0\| = at_0$.



Now, let the disturbance be situated in a compact set K rather than in a single point. Suppose that d and D are the minimal and maximal distances of x from K. Then the disturbance starts to act in x at time $t_0 = d/a$ it lasts for $(D - d)/a$; and again, for $t > D/a = t_1$ there is silence at x. Therefore, we can observe a forward wave front at time t_0 and a backward wave front at time t_1.

This shows that the domain of influence $M(K)$ of compact set K is the union of all boundaries of forward light-cones $\Gamma_+(y, 0)$ with $y \in K$ at time $t = 0$.

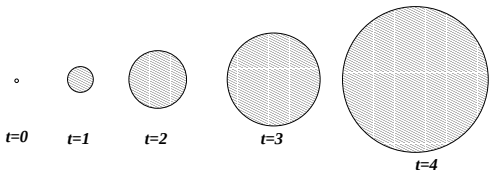
$$M(K) = \{(y,s) \mid \ \exists\, x \in K : \|x - y\| = as\}.$$

(b) Propagation of Plane Waves

Consider the fundamental solution

$$\mathcal{E}_2(x,t) = \frac{H(at - \|x\|)}{2\pi a\sqrt{a^2t^2 - \|x\|^2}}, \quad x = (x_1, x_2)$$

of the 2-dimensional wave equation.



It shows that the disturbance effected by a point source $\delta(x)\delta(t)$ in the origin at time 0 is a disc U_{at} of radius at around 0. One observes a forward wave front moving with speed a. In contrast to the 3-dimensional picture, there exists *no* back front. The disturbance is permanently present from t_0 on. We speak of wave *diffusion*; Huygens' principle does not hold.

Diffusion can also be observed in case of arbitrary initial disturbance $u_0(x)\delta'(t) + u_1(x)\delta(t)$. Indeed, the superposition principle shows that the domain of dependence of a compact initial disturbance K is the union of all discs $U_{at}(y)$ with $y \in K$.

(c) Propagation on a Line

Recall that $\mathcal{E}_1(x,t) = \frac{1}{2a}H(at - |\,x\,|)$. The disturbance at time $t > 0$ which is effected by a point source $\delta(x)\delta(t)$ is the whole closed interval $[-at, at]$. We have two forward wave "fronts" one at the point $x = at$ and one at $x = -at$; one moving to the right and one moving to the left. As in the plane case, there does not exist a back wave font; we observe diffusion.
For more details, see the discussion in Wladimirow, [Wla72, p. 155 – 159].

17.3 Fourier Method for Boundary Value Problems

A good, easy accessable introduction to the Fourier method is to be found in [KK71].
In this section we use Fourier series to solve BEVP to the Laplace equation as well as initial boundary value problems to the wave and heat equations.
Recall that the following sets are CNOS in the Hilbert space H

$$\left\{\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\mathrm{i}nt} \mid n \in \mathbb{Z},\right\}, \quad H = \mathrm{L}^2(a, a+2\pi),$$
$$\left\{\frac{1}{\sqrt{b-a}}\mathrm{e}^{\frac{2\pi\mathrm{i}}{b-a}nt} \mid n \in \mathbb{Z}\right\}, \quad H = \mathrm{L}^2(a,b),$$
$$\left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\sin(nt), \frac{1}{\sqrt{\pi}}\cos(nt) \mid n \in \mathbb{N}\right\}, \quad H = \mathrm{L}^2(a, a+2\pi),$$
$$\left\{\frac{1}{\sqrt{b-a}}, \sqrt{\frac{2}{b-a}}\sin\left(\frac{2\pi}{b-a}nt\right), \sqrt{\frac{2}{b-a}}\cos\left(\frac{2\pi}{b-a}nt\right) \mid n \in \mathbb{N}\right\}, \quad H = \mathrm{L}^2(a,b),$$

For any function $f \in \mathrm{L}^1(0, 2\pi)$ one has an associated Fourier series

$$f \sim \sum_{n\in\mathbb{Z}} c_n\,\mathrm{e}^{\mathrm{i}nt}, \quad c_n = \frac{1}{\sqrt{2\pi}}\int_0^{2\pi} f(t)\mathrm{e}^{-\mathrm{i}nt}\,\mathrm{d}t.$$

Lemma 17.14 *Each of the following two sets forms a CNOS in* $H = \mathrm{L}^2(0,\pi)$ *(on the half interval).*

$$\left\{\sqrt{\frac{2}{\pi}}\,\sin(nt) \;\mid\; n\in\mathbb{N}\right\},\quad \left\{\sqrt{\frac{2}{\pi}}\,\cos(nt) \;\mid\; n\in\mathbb{N}_0\right\}.$$

Proof. To check that they form an NOS is left to the reader. We show completeness of the first set. Let $f\in\mathrm{L}^2(0,\pi)$. Extend f to an odd function $\tilde f\in\mathrm{L}^2(-\pi,\pi)$, that is $\tilde f(x)=f(x)$ and $\tilde f(-x)=-f(x)$ for $x\in(0,\pi)$. Since $\tilde f$ is an odd function, in its Fourier series

$$\frac{a_0}{2}+\sum_{n=1}^{\infty} a_n\cos(nt)+b_n\sin(nt)$$

we have $a_n=0$ for all n. Since the Fourier series $\sum_{n=1}^{\infty} b_n\sin(nt)$ converges to $\tilde f$ in $\mathrm{L}^2(-\pi,\pi)$, it converges to f in $\mathrm{L}^2(0,\pi)$. Thus, the sine system is complete. The proof for the cosine system is analogous. ∎

17.3.1 Initial Boundary Value Problems

(a) The Homogeneous Heat Equation, Periodic Boundary Conditions

We consider heat conduction in a closed wire loop of length 2π. Let $u(x,t)$ be the temperature of the wire at position x and time t. Since the wire is closed (a loop), $u(x,t)=u(x+2\pi,t)$; u is thought to be a 2π periodic function on $\mathbb{R}$ for every fixed t. Thus, we have the following periodic boundary conditions

$$u(0,t)=u(2\pi,t),\quad u_x(0,t)=u_x(2\pi,t),\quad t\in\mathbb{R}_+. \tag{PBC}$$

The initial temperature distribution at time $t=0$ is given such that the BIVP reads

$$\begin{aligned} &u_t-a^2u_{xx}=0,\quad x\in\mathbb{R},\ t>0,\\ &u(x,0)=u_0(x),\quad x\in\mathbb{R},\\ &\text{(PBC)}. \end{aligned} \tag{17.20}$$

Separation of variables. We are seeking solutions of the form

$$u(x,t)=f(x)\cdot g(t)$$

ignoring the initial conditions for a while. The heat equation then takes the form

$$f(x)g'(t)=a^2f''(x)g(t)\iff\frac{1}{a^2}\frac{g'(t)}{g(t)}=\frac{f''(x)}{f(x)}=\kappa=\text{const.}$$

We obtain the system of two independent ODE only coupled by κ:

$$f''(x)-\kappa f(x)=0,\quad g'(t)-a^2\kappa g(t)=0.$$

The periodic boundary conditions imply:

$$f(0)g(t) = f(2\pi)g(t), \quad f'(x)g(t) = f''(2\pi)g(t)$$

which for nontrivial g gives $f(0) = f(2\pi)$ and $f'(0) = f'(2\pi)$. In case $\kappa = 0$, $f''(x) = 0$ has the general solution $f(x) = ax + b$. The only periodic solution is $f(x) = b = \text{const}$. Suppose now that $\kappa = -\nu^2 < 0$. Then the general solution of the second order ODE is

$$f(x) = c_1 \cos(\nu x) + c_2 \sin(\nu x).$$

Since f is periodic with period 2π, only a *discrete* set of values ν are possible, namely $\nu_n = n$, $n \in \mathbb{Z}$. This implies $\kappa_n = -n^2$, $n \in \mathbb{N}$.
Finally, in case $\kappa = \nu^2 > 0$, the general solution

$$f(x) = c_1 \mathrm{e}^{\nu x} + c_2 \mathrm{e}^{-\nu x}$$

provides no periodic solutions f. So far, we obtained a set of solutions

$$f_n(x) = a_n \cos(nx) + b_n \sin(nx), \quad n \in \mathbb{N}_0,$$

corresponding to $\kappa_n = -n^2$. The ODE for $g_n(t)$ now reads

$$g_n'(t) + a^2 n^2 g_n(t) = 0.$$

Its solution is $g_n(t) = c\mathrm{e}^{-a^2 n^2 t}$, $n \in \mathbb{N}$. Hence, the solutions of the BVP are given by

$$u_n(x,t) = \mathrm{e}^{-a^2 n^2 t}(a_n \cos(nx) + b_n \sin(nx)),\ n \in \mathbb{N}, \quad u_0(x,t) = \frac{a_0}{2}$$

and finite or "infinite" linear combinations:

$$u(x,t) = \frac{a_0}{2} + \sum_{n=0}^{\infty} \mathrm{e}^{-a^2 n^2 t}(a_n \cos(nx) + b_n \sin(nx)), \tag{17.21}$$

Consider now the initial value, that is $t = 0$. The corresponding series is

$$u(x,0) = \frac{a_0}{2} + \sum_{n=0}^{\infty} a_n \cos(nx) + b_n \sin(nx), \tag{17.22}$$

which gives the ordinary Fourier series of $u_0(x)$. That is, the Fourier coefficient a_n and b_n of the initial function $u_0(x)$ *formally* give a solution $u(x,t)$.

1. If the Fourier series of u_0 pointwise converges to u_0, the initial conditions are satisfied by the function $u(x,t)$ given in (17.21)

2. If the Fourier series of u_0 is twice differentiable (with respect to x), so is the function $u(x,t)$ given by (17.21).

Lemma 17.15 *Consider the BIVP* (17.20). (a) Existence. *Suppose that* $u_0 \in \mathrm{C}^4(\mathbb{R})$ *is periodic. Then the function* $u(x,t)$ *given by* (17.21) *is in* $\mathrm{C}^{2,1}_{x,t}([0,2\pi]\times\mathbb{R}_+)$ *and solves the classical BIVP* (17.20).
(b) Uniqueness and Stability. *In the class of functions* $\mathrm{C}^{2,1}_{x,t}([0,2\pi]\times\mathbb{R}_+)$ *the solution of the above BIVP is unique.*

Proof. (a) The Fourier coefficients of $u_0^{(4)}$ are bounded such that the Fourier coefficients of u_0 have growth $1/n^4$ (integrate the Fourier series of $u_0^{(4)}$ four times). Then the series for $u_{xx}(x,t)$ and $u_t(x,t)$ both are dominated by the series $\displaystyle\sum_{n=0}^{\infty}\frac{1}{n^2}$; hence they converge uniformly. This shows that the series $u(x,t)$ can be differentiated term by term twice w.r.t. x and once w.r.t. t.
(b) For any fixed $t\geq 0$, $u(x,t)$ is continuous in x. Consider $v(t) := \|u(x,t)\|^2_{\mathrm{L}^2(0,2\pi)}$. Then

$$\begin{aligned} v'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_0^{2\pi} u(x,t)^2\,\mathrm{d}x\right) = 2\int_0^{2\pi} u(x,t)u_t(x,t)\,\mathrm{d}x = 2\int_0^{2\pi} u(x,t)a^2u_{xx}(x,t)\,\mathrm{d}x \\ &= 2\left(a^2u_x\,u\Big|_0^{2\pi} - a^2\int_0^{2\pi}(u_x(x,t))^2\,\mathrm{d}x\right) = -2a^2\int_0^{2\pi} u_x^2\,\mathrm{d}x \leq 0. \end{aligned}$$

This shows that $v(t)$ is monotonically decreasing in t.
Suppose now u_1 and u_2 both solve the BIVP in the given class. Then $u = u_1 - u_2$ solves the BIVP in this class with homogeneous initial values, that is $u(x,0) = 0$, hence, $v(0) = \|u(x,0)\|^2_{\mathrm{L}^2} = 0$. Since $v(t)$ is decreasing for $t\geq 0$ and non-negative, $v(t) = 0$ for all $t\geq 0$; hence $u(x,t) = 0$ in $\mathrm{L}^2(0,2\pi)$ for all $t\geq 0$. Since $u(x,t)$ is continuous in x, this implies $u(x,t) = 0$ for all x and t. Thus, $u_1(x,t) = u_2(x,t)$—the solution is unique.
Stability. Since $v(t)$ is decreasing

$$\sup_{t\in\mathbb{R}_+}\|u(x,t)\|_{\mathrm{L}^2(0,2\pi)} \leq \|u_0\|_{\mathrm{L}^2(0,2\pi)}\,.$$

This shows that small changes in the initial conditions u_0 imply small changes in the solution $u(x,t)$. The problem is well-posed. ∎

(b) The Inhomogeneous Heat Equation, Periodic Boundary Conditions

We study the IBVP

$$\begin{aligned} u_t - a^2u_{xx} &= f(x,t), \\ u(x,0) &= 0, \\ &\text{(PBC)}. \end{aligned} \tag{17.23}$$

Solution. Let $\mathrm{e}_n(x) = \mathrm{e}^{\mathrm{i}nx}/\sqrt{2\pi}$, $n\in\mathbb{Z}$, be the CNOS in $\mathrm{L}^2(0,2\pi)$. These functions are all eigen functions with respect to the differential operator $\frac{\mathrm{d}^2}{\mathrm{d}x^2}$, $\mathrm{e}_n''(x) = -n^2\,\mathrm{e}_n(x)$. Let $t>0$ be fixed and

$$f(x,t) \sim \sum_{n\in\mathbb{Z}} c_n(t)\,\mathrm{e}_n(x)$$

be the Fourier series of $f(x,t)$ with coefficients $c_n(t)$. For u, we try the following ansatz

$$u(x,t) \sim \sum_{n\in\mathbb{Z}} d_n(t)\,\mathrm{e}_n(x) \tag{17.24}$$

If $f(x,t)$ is continuous in x and piecewise continuously differentiable with respect to x, its Fourier series converges pointwise and we have

$$u_t - a^2 u_{xx} = \sum_{n\in\mathbb{Z}} \big(\mathrm{e}_n\, d'(t) + a^2 n^2 \mathrm{e}_n d(t)\big) = f(x,t) = \sum_{n\in\mathbb{N}} c_n(t)\mathrm{e}_n.$$

For each n this is an ODE in t

$$d_n'(t) + a^2 n^2 d_n(t) = c_n(t), \quad d_n(0) = 0.$$

From ODE the solution is well-known

$$d_n(t) = \mathrm{e}^{-a^2n^2t} \int_0^t \mathrm{e}^{a^2n^2s} c_n(s)\,\mathrm{d}s.$$

Under certain regularity and growth conditions on f, (17.24) solves the inhomogeneous IBVP.

(c) The Homogeneous Wave Equation with Dirichlet Conditions

Consider the initial boundary value problem of the vibrating string of length π.

$$\begin{aligned}
&\text{(E)} && u_{tt} - a^2 u_{xx} = 0, && 0 < x < \pi, \quad t > 0;\\
&\text{(BC)} && u(0,t) = u(\pi,t) = 0, &&\\
&\text{(IC)} && u(x,0) = \varphi(x), && u_t(x,0) = \psi(x), \quad 0 < x < \pi.
\end{aligned}$$

The ansatz $u(x,t) = f(x)g(t)$ yields

$$\frac{f''}{f} = \kappa = \frac{g''}{a^2 g}, \quad f''(x) = \kappa f(x), \quad g'' = \kappa a^2 g.$$

The boundary conditions imply $f(0) = f(\pi) = 0$. Hence, the first ODE has the only solutions

$$f_n(x) = c_n \sin(nx), \quad \kappa_n = -n^2, \quad n \in \mathbb{N}.$$

The corresponding ODEs for g then read

$$g_n'' + n^2 a^2 g_n = 0,$$

which has the general solution $a_n \cos(nat) + b_n \sin(nat)$. Hence,

$$u(x,t) = \sum_{n=1}^{\infty} (a_n \cos(nat) + b_n \sin(nat)) \sin(nx)$$

solves the boundary value problem in the sense of $\mathcal{D}'(\mathbb{R}^2)$ (choose any a_n, b_n of polynomial growth). Now, insert the initial conditions, $t = 0$:

$$u(x,0) = \sum_{n=1}^{\infty} a_n \sin(nx) \stackrel{!}{=} \varphi(x), \quad u_t(x,0) = \sum_{n=1}^{\infty} nab_n \sin(nx) \stackrel{!}{=} \psi(x).$$

Since $\{\sqrt{\frac{2}{\pi}} \sin(nx) \mid n \in \mathbb{N}\}$ is a CNOS in $\mathrm{L}^2(0,\pi)$, we can determine the Fourier coefficients of φ and ψ with respect to this CNOS and obtain a_n and b_n, respectively.
Regularity. Suppose that $\varphi \in \mathrm{C}^4([0,\pi])$, $\psi \in \mathrm{C}^3([0,\pi])$. Then the Fourier-Sine coefficients a_n and anb_n of φ and ψ have growth $1/n^4$ and $1/n^3$, respectively. Hence, the series

$$u(x,t) = \sum_{n=1}^{\infty} (a_n \cos(nat) + b_n \sin(nat)) \sin(nx) \tag{17.25}$$

can be differentiated twice with respect to x or t since the differentiated series have a summable upper bound $\sum c/n^2$. Hence, (17.25) solves the IBVP.

(d) The Wave Equation with Inhomogeneous Boundary Conditions

Consider the following problem in $\Omega \subset \mathbb{R}^n$

$$\begin{aligned} u_{tt} - a^2 \Delta u &= 0, \\ u(x,0) = u_t(x,0) &= 0 \\ u\,|_{\partial\Omega} &= w(x,t). \end{aligned}$$

Idea. Find an extension $v(x,t)$ of $w(x,t)$, $v \in \mathrm{C}^2(\overline{\Omega} \times \mathbb{R}_+)$, and look for functions $\tilde{u} = u - v$. Then $\tilde{u}$ has homogeneous boundary conditions and satisfies the IBVP

$$\begin{aligned} \tilde{u}_{tt} - a^2 \Delta \tilde{u} &= -v_{tt} + a^2 \Delta v, \\ \tilde{u}(x,0) = -v(x,0), &\quad \tilde{u}_t(x,0) = -v_t(x,0) \\ \tilde{u}\,|_{\partial\Omega} &= 0. \end{aligned}$$

This problem can be split into two problems, one with zero initial conditions and one with homogeneous wave equation.

17.3.2 Eigenvalue Problems for the Laplace Equation

In the previous subsection we have seen that BIVPs using Fourier's method often lead to boundary eigenvalue problems (BEVP) for the Laplace equation.
We formulate the problems. Let $n = 1$ and $\Omega = (0,l)$. One considers the following types of BEVPs to the Laplace equation; $f'' = \lambda f$:

- Dirichlet boundary conditions: $f(0) = f(l) = 0$.
- Neumann boundary conditions: $f'(0) = f'(l) = 0$.

- Periodic boundary conditions : $f(0) = f(l)$, $f'(0) = f'(l)$.
- Mixed boundary conditions: $\alpha_1 f(0) + \alpha_2 f'(0) = 0, \quad \beta_1 f(l) + \beta_2 f'(l) = 0.$
- Symmetric boundary conditions: If u and v are function satisfying these boundary conditions, then $(u'v - uv')|_0^l = 0$. In this case integration by parts gives

$$\int_0^l (u''v - uv'')\,\mathrm{d}x = u'v - v'u|_0^l - \int_0^l (u'v' - v'u')\,\mathrm{d}x = 0.$$

  That is $u''\cdot v = u\cdot v''$ and the Laplace operator becomes symmetric.

Proposition 17.16 *Let $\Omega \subset \mathbb{R}^n$. The BEVP with Dirichlet conditions*

$$\Delta u = \lambda u, \quad u\,|_{\partial\Omega} = 0, \quad u \in \mathrm{C}^2(\Omega) \cap \mathrm{C}^1(\overline{\Omega}) \tag{17.26}$$

has countably many eigenvalues λ_k. All eigenvalues are negative and of finite multiplicity. Let $0 > \lambda_1 > \lambda_2 > \cdots$ then sequence $(\frac{1}{\lambda_k})$ tends to 0. The eigenfunctions u_k corresponding to λ_k form a CNOS in $\mathrm{L}^2(\Omega)$.

Sketch of proof. (a) Let $H = \mathrm{L}^2(\Omega)$. We use Green's 1^{st} formula with $u = v$, $u\,|_{\partial\Omega} = 0$,

$$\int_\Omega u\,\Delta u\,\mathrm{d}x + \int_\Omega (\nabla u)^2\,\mathrm{d}x = 0.$$

to show that all eigenvalues of Δ are negative. Let $\Delta u = \lambda u$. First note, that $\lambda = 0$ is not an eigenvalue of Δ. Suppose to the contrary $\Delta u = 0$, that is, u is harmonic. Since $u\,|_{\partial\Omega} = 0$, by the uniqueness theorem for the Dirichlet problem, $u = 0$ in Ω. Then

$$\lambda\,\|u\|^2 = \lambda\,\langle u\,,\,u\rangle = \langle \lambda u\,,\,u\rangle = \langle \Delta u\,,\,u\rangle = \int_\Omega u\,\Delta u\,\mathrm{d}x = -\int_\Omega (\nabla u)^2\,\mathrm{d}x < 0.$$

Hence, λ is negative.
(b) Assume that a Green's function G for Ω exists. By (17.34), that is

$$u(y) = \int_\Omega G(x,y)\,\Delta u(x)\,\mathrm{d}x + \int_{\partial\Omega} u(x)\,\frac{\partial G(x,y)}{\partial \vec{n}_x}\,\mathrm{d}S(x),$$

$u\,|_{\partial\Omega} = 0$ implies

$$u(y) = \int_\Omega G(x,y)\,\Delta u(x)\,\mathrm{d}x.$$

This shows that the integral operator $A\colon \mathrm{L}^2(\Omega) \to \mathrm{L}^2(\Omega)$ defined by

$$(Av)(y) := \int_\Omega G(x,y) v(x)\,\mathrm{d}x$$

is inverse to the Laplacian. Since $G(x,y) = G(y,x)$ is real, A is self-adjoint. By (a), its eigenvalues, $1/\lambda_k$ are all negative. If

$$\iint_{\Omega\times\Omega} |\,G(x,y)\,|^2\,\mathrm{d}x\mathrm{d}y < \infty,$$

A is a *compact* operator.
We want to justify the last statement. Let $(Kf)(x) = \int_\Omega k(x,y)f(y)\,\mathrm{d}y$ be an integral operator on $H = \mathrm{L}^2(\Omega)$, with kernel $k(x,y) \in \mathcal{H} = \mathrm{L}^2(\Omega \times \Omega)$. Let $\{u_n \mid n \in \mathbb{N}\}$ be a CNOS in H; then $\{u_n(x)u_m(y) \mid n,m \in \mathbb{N}\}$ is a CNOS in $\mathcal{H}$. Let k_{nm} be the Fourier coefficients of k with respect to the basis $\{u_n(x)u_m(y)\}$ in $\mathcal{H}$. Then

$$\begin{aligned}(Kf)(x) &= \int_\Omega f(y) \sum_{n,m} k_{nm}u_n(x)u_m(y)\,\mathrm{d}y \\ &= \sum_n u_n(x) \left(\sum_m k_{nm} \int_\Omega u_m(y)f(y)\,\mathrm{d}y \right) \\ &= \sum_n u_n(x) \sum_m k_{nm} \langle f\,,\, u_m\rangle = \sum_{n,m} k_{nm} \langle f\,,\, u_m\rangle\, u_n.\end{aligned}$$

This in particular shows that

$$\|Kf\|^2 = \sum_{m,n} k_{mn}^2 \langle f\,,\, u_m\rangle^2 \le \sum_{m,n} k_{mn}^2 \|f\|^2 = \|f\|^2 \int_{\Omega\times\Omega} k(x,y)^2\,\mathrm{d}x\mathrm{d}y = \|f\|^2 \sum_n \|Ku_n\|^2$$

$$\|K\|^2 \le \sum_n \|Ku_n\|^2$$

We show that K is approximated by the sequence (K_n) defined by

$$K_n f = \sum_m \sum_{r=1}^n k_{rm} \langle f\,,\, u_m\rangle\, u_r$$

of finite rank operators. Indeed,

$$\|(K-K_n)f\|^2 = \sum_m \sum_{r=n+1}^\infty k_{rm}^2 \,|\langle f\,,\, u_m\rangle|^2 \le \sup_m \sum_{r=n+1}^\infty k_{rm}^2 \|f\|^2$$

such that

$$\|K-K_n\|^2 = \sup_m \sum_{r=n+1}^\infty k_{rm}^2 \longrightarrow 0$$

as $n \to \infty$. Hence, K is compact.
(c) By (a) and (b), A is a negative, compact, self-adjoint operator. By the spectral theorem for compact self-adjoint operators, Theorem 13.33, there exists an NOS (u_k) of eigenfunctions to $1/\lambda_k$ of A. The NOS (u_k) is complete since 0 is not an eigenvalue of A.

Example 17.1 Dirichlet Conditions on the Square. Let $Q = (0,\pi) \times (0,\pi) \subset \mathbb{R}^2$. The Laplace operator with Dirichlet boundary conditions on Ω has eigenfunctions

$$u_{mn}(x,y) = \frac{2}{\pi} \sin(mx)\, \sin(ny),$$

corresponding to the eigenvalues $\lambda_{mn} = -(m^2+n^2)$. The eigenfunctions $\{u_{mn} \mid m,n \in \mathbb{N}\}$ form a CNOS in the Hilbert space $\mathrm{L}^2(\Omega)$.

Example 17.2 Dirichlet Conditions on the ball $U_1(0)$ in $\mathbb{R}^2$. We consider the BEVP with Dirichlet boundary conditions on the ball.

$$-\Delta u = \lambda u, \quad u\,|_{\mathrm{S}_1(0)} = 0.$$

In polar coordinates $u(x,y) = \tilde{u}(r,\varphi)$ this reads,

$$\Delta\tilde{u}(r,\varphi) = \frac{1}{r}\frac{\partial}{\partial r}\,(r\,\tilde{u}_r) + \frac{1}{r^2}\tilde{u}_{\varphi\varphi} = -\lambda\tilde{u}, \quad 0 < r < 1, \quad 0 \le \varphi < 2\pi.$$

Separation of variables. We try the ansatz $\tilde{u}(r,\varphi) = R(r)\Phi(\varphi)$. We have the boundary condition $R(1) = 0$ and in $R(r)$ is bounded in a neighborhood of $r = 0$. Also, Φ is periodic. Then $\frac{\partial}{\partial r}\tilde{u} = R'\Phi$ and

$$\frac{\partial}{\partial r}\,(r\tilde{u}_r) = \frac{\partial}{\partial r}\,(rR'\Phi) = (R' + rR'')\Phi, \quad \tilde{u}_{\varphi\varphi} = R\Phi''.$$

Hence, $\Delta u = -\lambda u$ now reads

$$\begin{aligned}
\left(\frac{R'}{r} + R''\right)\Phi + \frac{R}{r^2}\Phi'' &= -\lambda R\Phi \\
\frac{\frac{R'}{r} + R''}{R} + \frac{1}{r^2}\frac{\Phi''}{\Phi} &= -\lambda \\
\frac{rR' + r^2R''}{R} + \lambda r^2 &= -\frac{\Phi''}{\Phi} = \mu.
\end{aligned}$$

In this way, we obtain the two one-dimensional problems

$$\begin{aligned}
\Phi'' + \mu\Phi = 0, \quad \Phi(0) = \Phi(2\pi); \\
r^2R'' + rR' + (\lambda r^2 - \mu)R = 0, \quad |\,R(0)\,| < \infty, \quad R(1) = 0.
\end{aligned} \tag{17.27}$$

The eigenvalues and eigenfunctions to the first problem are

$$\mu_k = k^2, \quad \Phi_k(\varphi) = \mathrm{e}^{\mathrm{i}k\varphi}, \quad k \in \mathbb{Z}.$$

Equation (17.27) is the Bessel ODE. For $\mu = k^2$ the solution of (17.27) bounded in $r = 0$ is given by the Bessel function $J_k(r\sqrt{\lambda})$. Recall from homework 21.2 that

$$J_k(x) = \sum_{n=0}^{\infty} \frac{(-1)^n \left(\frac{x}{2}\right)^{2n+k}}{n!(n+k)!}, \quad k \in \mathbb{N}_0.$$

To determine the eigenvalues λ we use the boundary condition $R(1) = 0$ in (17.27), namely $J_k(\sqrt{\lambda}) = 0$. Hence, $\sqrt{\lambda} = \mu_{kj}$, where μ_{kj}, $j = 1, 2, \ldots$, denote the positive zeros of J_k. We obtain

$$\lambda_{kj} = \mu_{kj}^2, \quad R_{kj}(r) = J_k(\mu_{kj}\,r), \quad j = 1, 2, \cdots.$$

The solution of the BEVP is

$$\lambda_{kj} = \mu_{kj}^2, \quad u_{kj}(x) = J_{|\,k\,|}(\mu_{|\,k\,|j}r)\mathrm{e}^{\mathrm{i}k\varphi}, \quad k \in \mathbb{Z}, \quad j = 1, 2, \cdots.$$

Note that the Bessel functions $\{J_k \mid k \in \mathbb{Z}_+\}$ and the system $\{\mathrm{e}^{\mathrm{i}kt} \mid k \in \mathbb{Z}\}$ form a complete OS in $\mathrm{L}^2((0,1), r\,\mathrm{d}r)$ and in in $\mathrm{L}^2(0, 2\pi)$, respectively. Hence, the OS $\{u_{kl} \mid k \in \mathbb{Z},\, l \in \mathbb{Z}_+\}$ is a complete OS in $\mathrm{L}^2(U_1(0))$. Thus, there are no further solutions to the given BEVP. For more details on Bessel functions, see [FK98, p. 383].

17.4 Boundary Value Problems for the Laplace and the Poisson Equations

Throughout this section (if nothing is stated otherwise) we will assume that Ω is a bounded region in $\mathbb{R}^n$, $n \geq 2$. We suppose further that Ω belongs to the class C^2, that is, the boundary $\partial\Omega$ consists of finitely many twice continuously differentiable hypersurfaces; $\Omega' := \mathbb{R}^n \setminus \overline{\Omega}$ is assumed to be connected (i. e. it is a region, too). All functions are assumed to be real valued.

17.4.1 Formulation of Boundary Value Problems

(a) The Inner Dirichlet Problem:

Given $\varphi \in \mathrm{C}(\partial\Omega)$ and $f \in \mathrm{C}(\overline{\Omega})$, find $u \in \mathrm{C}(\overline{\Omega}) \cap \mathrm{C}^2(\Omega)$ such that

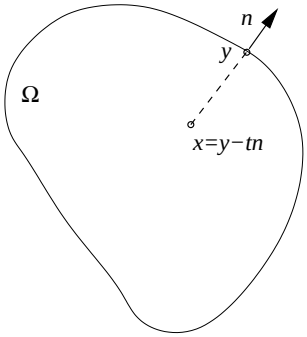
$$\begin{aligned} \Delta u(x) &= f(x) \quad \forall\, x \in \Omega, \quad \text{and} \\ u(y) &= \varphi(y), \ \forall\, y \in \partial\Omega. \end{aligned}$$

(b) The Exterior Dirichlet Problem:

Given $\varphi \in \mathrm{C}(\partial\Omega)$ and $f \in \mathrm{C}(\overline{\Omega'})$, find $u \in \mathrm{C}(\overline{\Omega'}) \cap \mathrm{C}^2(\Omega')$ such that

$$\begin{aligned} \Delta u(x) &= f(x), \ \forall\, x \in \Omega', \quad \text{and} \\ u(y) &= \varphi(y), \ \forall\, y \in \partial\Omega, \\ \lim_{|x|\to\infty} u(x) &= 0. \end{aligned}$$

(c) The Inner Neumann Problem:



Given $\varphi \in \mathrm{C}(\partial\Omega)$ and $f \in \mathrm{C}(\overline{\Omega})$, find $u \in \mathrm{C}^1(\overline{\Omega}) \cap \mathrm{C}^2(\Omega)$ such that

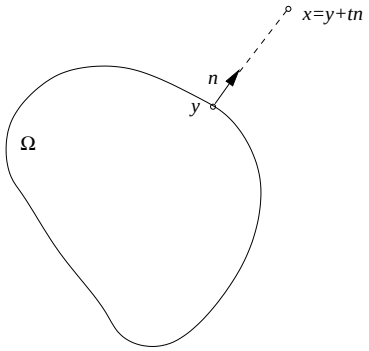
$$\begin{aligned} \Delta u(x) &= f(x) \quad \forall\, x \in \Omega, \quad \text{and} \\ \frac{\partial u}{\partial \vec{n}_-}(y) &= \varphi(y), \ \forall\, y \in \partial\Omega. \end{aligned}$$

Here $\frac{\partial u}{\partial \vec{n}_-}(y)$ denotes the limit of directional derivative

$$\frac{\partial u}{\partial \vec{n}_-}(y) = \lim_{t\to 0+0} \vec{n}(y) \cdot \operatorname{grad} u(y - t\vec{n}(y))$$

and $\vec{n}(y)$ is the outer normal to Ω at $y \in \partial\Omega$. That is, $x \in \Omega$ approaches $y \in \partial\Omega$ in the direction of the normal vector $\vec{n}(y)$. We assume that this limit exists for all boundary points $y \in \partial\Omega$.

(d) The Exterior Neumann Problem:



Given $\varphi \in \mathrm{C}(\partial\Omega)$ and $f \in \mathrm{C}(\overline{\Omega'})$, find $u \in \mathrm{C}^1(\overline{\Omega'}) \cap \mathrm{C}^2(\Omega')$ such that

$$\begin{aligned} \Delta u(x) &= f(x) \quad \forall\, x \in \Omega', \quad \text{and} \\ \frac{\partial u}{\partial \vec{n}_+}(y) &= \varphi(y), \ \forall\, y \in \partial\Omega \\ \lim_{|x|\to\infty} u(x) &= 0. \end{aligned}$$

Here $\frac{\partial u}{\partial \vec{n}_+}(y)$ denotes the limit of directional derivative

$$\frac{\partial u}{\partial \vec{n}_+}(y) = \lim_{t\to 0+0} \vec{n}(y) \cdot \operatorname{grad} u(y + t\vec{n}(y))$$

and $\vec{n}(y)$ is the outer normal to Ω at $y \in \partial\Omega$. We assume that this limit exists for all boundary points $y \in \partial\Omega$. In both Neumann problems one can also look for a function $u \in \mathrm{C}^2(\Omega) \cap \mathrm{C}(\overline{\Omega})$ or $u \in \mathrm{C}^2(\Omega') \cap \mathrm{C}(\overline{\Omega}')$, respectively, provided the above limits exist and define continuous functions on the boundary.
These four problems are intimately connected with each other, and we will obtain solutions to all of them simultaneously.

17.4.2 Basic Properties of Harmonic Functions

Recall that a function $u \in \mathrm{C}^2(\Omega)$, is said to be *harmonic* if $\Delta u = 0$ in Ω.
We say that an operator L on a function space V over $\mathbb{R}^n$ is *invariant* under a affine transformation T, $T(x) = A(x) + b$, where $A \in \mathscr{L}(\mathbb{R}^n)$, $b \in \mathbb{R}^n$, if

$$L \circ T^* = T^* \circ L,$$

where $T^* \colon V \to V$ is given by $(T^* f)(x) = f(T(x))$, $f \in V$. It follows that the Laplacian is invariant under translations ($T(x) = x + b$) and rotations T (i. e. $T^\top T = T T^\top = I$). Indeed, for translations, the matrix B with $\tilde{A} = BAB^\top$ is the identity and in case of the rotation, $B = T^{-1}$. Since $A = I$, in both cases, $\tilde{A} = A = I$; the Laplacian is invariant.
In this section we assume that $\Omega \subset \mathbb{R}^n$ is a region where Gauß' divergence theorem is valid for all vector fields $f \in \mathrm{C}^1(\Omega) \cap \mathrm{C}(\overline{\Omega})$ for :

$$\int_\Omega \operatorname{div} f(x)\, \mathrm{d}x = \int_{\partial\Omega} f(y) \cdot \vec{\mathrm{d}S}(y),$$

where the dot $\cdot$ denotes the inner product in $\mathbb{R}^n$. The term under the integral can be written as

$$\begin{aligned} \omega(y) = f(y) \cdot \vec{\mathrm{d}S} &= (f_1(y), \cdots, f_n(y)) \cdot (\, \mathrm{d}y_2 \wedge \mathrm{d}y_3 \wedge \cdots \wedge \mathrm{d}y_n, -\mathrm{d}y_1 \wedge \mathrm{d}y_3 \wedge \cdots \wedge \mathrm{d}y_n, \\ &\qquad \cdots, (-1)^{n-1}\, \mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_{n-1}\big) \\ &= \sum_{k=1}^{n} (-1)^{k-1} f_k(y)\ \mathrm{d}y_1 \wedge \cdots \wedge \widehat{\mathrm{d}y_k} \wedge \cdots \wedge \mathrm{d}y_n, \end{aligned}$$

where the hat means ommission of this factor. In this way $\omega(y)$ becomes a differential $(n-1)$-form. Using differentiation of forms, see Definition 11.7, we obtain

$$\mathrm{d}\omega = \operatorname{div} f(y)\,\mathrm{d}y_1 \wedge \mathrm{d}y_2 \wedge \cdots \wedge \mathrm{d}y_n.$$

This establishes the above generalized form of Gauß' divergence theorem. Let $U\colon \partial\Omega \to \mathbb{R}$ be a continuous scalar function on $\partial\Omega$, one can define $U(y)\,\mathrm{d}S(y) := U(y)\vec{n}(y)\cdot \mathrm{d}\vec{S}(y)$, where $\vec{n}$ is the outer unit normal vector to the surface $\partial\Omega$.
Recall that we obtain Green's first formula inserting $f(x) = v(x)\nabla u(x)$, $u, v \in \mathrm{C}^2(\overline{\Omega})$:

$$\int_\Omega v(x)\Delta u(x)\,\mathrm{d}x + \int_\Omega \nabla u(x)\cdot\nabla v(x)\,\mathrm{d}x = \int_{\partial\Omega} v(y)\frac{\partial u}{\partial\vec{n}}(y)\,\mathrm{d}S(y).$$

Interchanging the role of u and v and taking the difference, we obtain Green's second formula

$$\int_\Omega \left(v(x)\Delta u(x) - u(x)\Delta v(x)\right)\,\mathrm{d}x = \int_{\partial\Omega}\left(v(y)\,\frac{\partial u}{\partial\vec{n}}(y) - u(y)\frac{\partial v}{\partial\vec{n}}(y)\right)\,\mathrm{d}S(y). \tag{17.28}$$

Recall that

$$\begin{aligned}\mathcal{E}_2(x) &= \frac{1}{2\pi}\log\|x\|, \quad n=2,\\ \mathcal{E}_n(x) &= -\frac{1}{(n-2)\omega_n}\,\|x\|^{-n+2}, \quad n\geq 3\end{aligned}$$

are the fundamental solutions of the Laplacian in $\mathbb{R}^n$.

Theorem 17.17 (Green's representation formula) *Let $u \in \mathrm{C}^2(\overline{\Omega})$.*
Then for $x \in \Omega$ we have

$$u(x) = \int_\Omega \mathcal{E}_n(x-y)\,\Delta u(y)\,\mathrm{d}y + \int_{\partial\Omega}\left(u(y)\,\frac{\partial\mathcal{E}_n}{\partial\vec{n}_y}(x-y) - \mathcal{E}_n(x-y)\,\frac{\partial u}{\partial\vec{n}}(y)\right)\,\mathrm{d}S(y) \tag{17.29}$$

Here $\frac{\partial}{\partial\vec{n}_y}$ denotes the derivative in the direction of the outer normal with respect to the variable y.

Note that the distributions $\{\Delta u\}$, $\left(\frac{\partial u}{\partial\vec{n}}\,\delta_{\partial\Omega}\right)$, and $\frac{\partial}{\partial\vec{n}}\,(u\,\delta_{\partial\Omega})$ have compact support such that the convolution products with $\mathcal{E}_n$ exist.
Proof. Idea of proof For sufficiently small $\varepsilon > 0$, $U_\varepsilon(x) \subset \Omega$, since Ω is open. We apply Green's second formula with $v(y) = \mathcal{E}_n(x-y)$ and $\Omega\setminus U_\varepsilon(x)$ in place of Ω. Since $\mathcal{E}_n(x-y)$ is harmonic with respect to the variable y in $\Omega\setminus\{x\}$ (recall from Example 7.5, that $\mathcal{E}_n(x)$ is harmonic in $\mathbb{R}^n\setminus\{0\}$), we obtain

$$\begin{aligned}\int\limits_{\Omega\setminus U_\varepsilon(x)} \mathcal{E}_n(x-y)\Delta(y)\,\mathrm{d}y &= \int_{\partial\Omega}\left(\mathcal{E}_n(x-y)\frac{\partial u}{\partial\vec{n}}(y) - u(y)\frac{\partial\mathcal{E}_n(x-y)}{\partial\vec{n}_y}\right)\,\mathrm{d}S(y)\\ &\quad + \int\limits_{\partial U_\varepsilon(x)}\left(\mathcal{E}_n(x-y)\frac{\partial u}{\partial\vec{n}}(y) - u(y)\frac{\partial\mathcal{E}_n(x-y)}{\partial\vec{n}_y}\right)\,\mathrm{d}S(y).\end{aligned} \tag{17.30}$$

In the second integral $\vec{n}$ denotes the outer normal to $\Omega \setminus U_\varepsilon(x)$ hence the inner normal of $U_\varepsilon(x)$. We wish to evaluate the limits of the individual integrals in this formula as $\varepsilon \to 0$. Consider the left-hand side of (17.30). Since $u \in \mathrm{C}^2(\overline{\Omega})$, Δu is bounded; since $\mathcal{E}_n(x-y)$ is locally integrable, the lhs converges to

$$\int_\Omega \mathcal{E}_n(x-y)\,\Delta u(y)\,\mathrm{d}y.$$

On $\partial U_\varepsilon(x)$, we have $\mathcal{E}_n(x-y) = \beta_n \varepsilon^{-n+2}$, $\beta_n = -1/(\omega_n(n-2))$. Thus as $\varepsilon \to 0$,

$$\begin{aligned}\left| \int\limits_{\partial U_\varepsilon(x)} \mathcal{E}_n(x-y)\frac{\partial u}{\partial \vec{n}}(y)\,\mathrm{d}S \right| &\le \frac{|\,\beta_n\,|}{\varepsilon^{n-2}} \int\limits_{\partial U_\varepsilon(x)} \left| \frac{\partial u}{\partial \vec{n}}(y) \right| \mathrm{d}S \le \frac{|\,\beta_n\,|}{\varepsilon^{n-2}} \sup_{U_\varepsilon(x)} \left| \frac{\partial u(y)}{\partial \vec{n}} \right| \int\limits_{\mathrm{S}_\varepsilon(x)} \mathrm{d}S \\ &\le \frac{1}{(n-2)\omega_n \varepsilon^{n-2}} \omega_n \varepsilon^{n-1} \sup_{U_\varepsilon(x)} \left| \frac{\partial u(y)}{\partial \vec{n}} \right| = C\varepsilon \longrightarrow 0.\end{aligned}$$

Furthermore, since $\vec{n}$ is the interior normal of the ball $U_\varepsilon(y)$, the same calculations as in the proof of Theorem 17.1 show that $\frac{\partial \mathcal{E}_n(x-y)}{\partial \vec{n}_y} = -\beta_n \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\varepsilon^{-n+2}\right) = -\varepsilon^{-n+1}/\omega_n$. We obtain,

$$-\int\limits_{\partial U_\varepsilon(x)} u(y)\frac{\partial \mathcal{E}_n(x-y)}{\partial \vec{n}_y}\,\mathrm{d}S(y) = \underbrace{\frac{1}{\omega_n \varepsilon^{n-1}} \int\limits_{\mathrm{S}_\varepsilon(x)} u(y)\,\mathrm{d}S(y)}_{\text{spherical mean}} \longrightarrow u(x).$$

In the last line we used that the integral is the mean value of u over the sphere $\mathrm{S}_\varepsilon(x)$, and u is continuous at x. ∎

Remarks 17.4 (a) Green's representation formula is also true for functions $u \in \mathrm{C}^2(\Omega) \cap \mathrm{C}^1(\overline{\Omega})$. To prove this, consider Green's representation theorem on smaller regions $\Omega_\varepsilon \subset \Omega$ such that $\overline{\Omega_\varepsilon} \subset \Omega$.
(b) Applying Green's representation formula to a test function $\varphi \in \mathcal{D}(\Omega)$, see Definition 16.1, $\varphi(y) = \frac{\partial \varphi}{\partial \vec{n}}(y) = 0$, $y \in \partial\Omega$, we obtain

$$\varphi(x) = \int_\Omega \mathcal{E}_n(x-y)\Delta\varphi(x)\,\mathrm{d}x$$

(c) We may now draw the following consequence from Green's representation formula: If one knows Δu, then u is completely determined by its values and those of its normal derivative on $\partial\Omega$. In particular, a harmonic function on Ω can be reconstructed from its boundary data. One may ask conversely whether one can construct a harmonic function for arbitrary given values of u and $\frac{\partial u}{\partial \vec{n}}$ on $\partial\Omega$. Ignoring regularity conditions, we will find out that this is not possible in general. Roughly speaking, only one of these data is sufficient to describe u completely.
(d) In case of a harmonic function $u \in \mathrm{C}^2(\Omega) \cap \mathrm{C}^1(\overline{\Omega})$, $\Delta u = 0$, Green's representation formula reads ($n = 3$):

$$u(x) = \frac{1}{4\pi} \int_{\partial\Omega} \left(\frac{1}{\|x-y\|} \frac{\partial u(y)}{\partial \vec{n}} - u(y) \frac{\partial}{\partial \vec{n}_y} \frac{1}{\|x-y\|} \right) \mathrm{d}S(y). \tag{17.31}$$

In particular, the surface potentials $V^{(0)}(x)$ and $V^{(1)}(x)$ can be differentiated arbitrarily often for $x \in \Omega$. Outside $\partial\Omega$, $V^{(0)}$ and $V^{(1)}$ are harmonic. It follows from (17.31) that any harmonic function is a C^∞-function.

Spherical Means and Ball Means

First of all note that $u \in \mathrm{C}^1(\overline{\Omega})$ and u harmonic in Ω implies

$$\int_{\partial\Omega} \frac{\partial u(y)}{\partial \vec{n}} \, \mathrm{d}S = 0. \tag{17.32}$$

Indeed, this follows from Green's first formula inserting $v = 1$ and u harmonic, $\Delta u = 0$.

Proposition 17.18 (Mean Value Property) *Suppose that u is harmonic in $U_R(x_0)$ and continuous in $\overline{U_R(x_0)}$.*
(a) *Then $u(x_0)$ coincides with its spherical mean over the sphere $\mathrm{S}_R(x_0)$.*

$$u(x_0) = \frac{1}{\omega_n R^{n-1}} \int\limits_{\mathrm{S}_R(x_0)} u(y) \, \mathrm{d}S(y) \quad \textit{(spherical mean)}. \tag{17.33}$$

(b) *Further,*

$$u(x_0) = \frac{n}{\omega_n R^n} \int\limits_{U_R(x_0)} u(x) \, \mathrm{d}x \quad \textit{(ball mean)}.$$

Proof. (a) For simplicity, we consider only the case $n = 3$ and $x_0 = 0$. Apply Green's representation formula (17.31) to any ball $\Omega = U_\rho(0)$ with $\rho < R$. Noting (17.32) from (17.31) it follows that

$$\begin{aligned} u(0) &= \frac{1}{4\pi} \left(\frac{1}{\rho} \int_{\mathrm{S}_\rho(0)} \frac{\partial u(y)}{\partial \vec{n}} \, \mathrm{d}S - \int_{\mathrm{S}_\rho(0)} u(y) \, \frac{\partial}{\partial \vec{n}_y} \frac{1}{\|y\|} \, \mathrm{d}S \right) \\ &= -\frac{1}{4\pi} \int_{\mathrm{S}_\rho(0)} u(y) \, \frac{\partial}{\partial \vec{n}_y} \frac{1}{\|y\|} \, \mathrm{d}S = \frac{1}{4\pi} \int_{\mathrm{S}_\rho(0)} u(y) \, \frac{1}{\rho^2} \, \mathrm{d}S \\ &= \frac{1}{4\pi\rho^2} \int_{\mathrm{S}_\rho(0)} u(y) \, \mathrm{d}S, \end{aligned}$$

Since u is continuous on the closed ball of radius R, the formula remains valid as $\rho \to R$.
(b) Use $\mathrm{d}x = \mathrm{d}x_1 \cdots \mathrm{d}x_n = \mathrm{d}r \, \mathrm{d}S_r$ where $\|x\| = r$. Multiply both sides of (17.33) by $r^{n-1} \, \mathrm{d}r$ and integrate with respect to r from 0 to R:

$$\begin{aligned} \int_0^R r^{n-1} u(x_0) \, \mathrm{d}r &= \int_0^R r^{n-1} \left(\frac{1}{\omega_n r^{n-1}} \int_{\mathrm{S}_R(x_0)} u(y) \, \mathrm{d}S \right) \mathrm{d}r \\ \frac{1}{n} R^n u(x_0) &= \frac{1}{\omega_n} \int_{U_R(x_0)} u(x) \, \mathrm{d}x. \end{aligned}$$

The assertion follows. Note that $R^n \omega_n / n$ is exactly the n-dimensional volume of $U_R(x_0)$. The proof in case $n = 2$ is similar. ∎

Proposition 17.19 (Minimum-Maximum Principle) *Let u be harmonic in Ω and continuous in $\overline{\Omega}$. Then*

$$\max_{x\in\overline{\Omega}} u(x) = \max_{x\in\partial\Omega} u(x);$$

i. e. u attains its maximum on the boundary $\partial\Omega$. The same is true for the minimum.

Proof. Suppose to the contrary that $M = u(x_0) = \max\limits_{x\in\overline{\Omega}} u(x)$ is attained at an inner point $x_0 \in \Omega$ and $M > m = \max\limits_{x\in\partial\Omega} u(x) = u(y_0)$, $y_0 \in \partial\Omega$.
(a) We show that $u(x) = M$ is constant in any ball $U_\varepsilon(x_0) \subset \Omega$ around x_0.
Suppose to the contrary that $u(x_1) < M$ for some $x_1 \in U_\varepsilon(x_0)$. By continuity of u, $u(x) < M$ for all $x \in U_\varepsilon(x_0) \cap U_\eta(x_1)$. In particular

$$M = u(x_0) = \frac{n}{\omega_n \varepsilon^n} \int_{B_\varepsilon(x_0)} u(x)\,\mathrm{d}x < \frac{n}{\omega_n \varepsilon^n} \int_{B_\varepsilon(x_0)} M\,\mathrm{d}y = M;$$

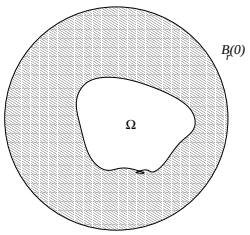
this is a contradiction; u is constant in $U_\varepsilon(x_0)$.
(b) $u(x) = M$ is constant in Ω. Let $x_1 \in \Omega$; we will show that $u(x_1) = M$. Since Ω is connected and bounded, there exists a path from x_0 to x_1 which can be covered by a chain of finitely many balls in Ω. In all balls, starting with the ball around x_0 from (a), $u(x) = M$ is constant. Hence, u is constant in Ω. Since u is continuous, u is constant in $\overline{\Omega}$. This contradicts the assumption; hence, the maximum is assumed on the boundary $\partial\Omega$.
Passing from u to $-u$, the statement about the minimum follows. ∎

Remarks 17.5 (a) A stronger proposition holds with "local maximum" in place of "maximum"
(b) Another stricter version of the maximum principle is:

> Let $u \in \mathrm{C}^2(\Omega) \cap \mathrm{C}(\overline{\Omega})$ and $\Delta u \geq 0$ in Ω. Then either u is constant or
>
> $$u(y) < \max_{x\in\partial\Omega} u(x)$$
>
> for all $y \in \Omega$.

Corollary 17.20 (Uniqueness) *The inner and the outer Dirichlet problem has at most one solution, respectively.*

Proof. Suppose that u_1 and u_2 both are solutions of the Dirichlet problem, $\Delta u_1 = \Delta u_2 = f$. Put $u = u_1 - u_2$. Then $\Delta u(x) = 0$ for all $x \in \Omega$ and $u(y) = 0$ on the boundary $y \in \partial\Omega$.
(a) Inner problem. By the maximum principle, $u(x) = 0$ for all $x \in \overline{\Omega}$; that is $u_1 = u_2$.



(b) Suppose that $u \not\equiv 0$. Without loss of generality we may assume that $u(x_1) = \alpha > 0$ for some $x_1 \in \Omega'$. By assumption, $|\,u(x)\,| \to 0$ as $x \to \infty$. Hence, there exists $r > 0$ such that $|\,u(x)\,| < \alpha/2$ for all $x \geq r$. Since u is harmonic in $B_r(0) \setminus \overline{\Omega}$, the maximum principle yields

$$\alpha = u(x_1) \leq \max_{x\in \mathrm{S}_R(0)\cup\partial\Omega} u(x) \leq \alpha/2;$$

a contradiction.

■

Corollary 17.21 (Stability) *Suppose that u_1 and u_2 are solutions of the inner Dirichlet problem $\Delta u_1 = \Delta u_2 = f$ with boundary values $\varphi_1(y)$ and $\varphi_2(y)$ on $\partial\Omega$, respectively. Suppose further that*

$$|\,\varphi_1(y) - \varphi_2(y)\,| \le \varepsilon \qquad \forall\, y \in \partial\Omega.$$

Then $|\,u_1(x) - u_2(x)\,| \le \varepsilon$ for all $x \in \overline{\Omega}$.

A similar statement is true for the exterior Dirichlet problem.
Proof. Put $u = u_1 - u_2$. Then $\Delta u = 0$ and $|\,u(y)\,| \le \varepsilon$ for all $y \in \partial\Omega$. By the Maximum Principle, $|\,u(x)\,| \le \varepsilon$ for all $x \in \overline{\Omega}$. ■

Lemma 17.22 *Suppose that u is a* non-constant *harmonic function on Ω and the maximum of $u(x)$ is attained at $y \in \partial\Omega$.*
Then $\frac{\partial u}{\partial \vec{n}}(y) > 0$.

For the proof see [Tri92, 3.4.2. Theorem, p. 174].

Proposition 17.23 (Uniqueness) (a) *The exterior Neumann problem has at most one solution.*
(b) *A necessary condition for solvability of the inner Neumann problem is*

$$\int_{\partial\Omega} \varphi \,\mathrm{d}S = \int_{\Omega} f(x)\,\mathrm{d}x.$$

Two solutions of the inner Neumann problem differ by a constant.

Proof. (a) Suppose that u_1 and u_2 are solutions of the exterior Neumann problem, then $u = u_1 - u_2$ satisfies $\Delta u = 0$ and $\dfrac{\partial u}{\partial \vec{n}}(y) = 0$. The above lemma shows that $u(x) = c$ is constant in Ω. Since $\lim_{|\,x\,|\to\infty} u(x) = 0$, the constant c is 0; hence $u_1 = u_2$.
(b) Inner problem. The uniqueness follows as in (a). The necessity of the formula follows from (17.28) with $v \equiv 1$, $\Delta u = f$, $\frac{\partial v}{\partial \vec{n}} = 0$. ■

Proposition 17.24 (Converse Mean Value Theorem) *Suppose that $u \in \mathrm{C}(\Omega)$ and that whenever $x_0 \in \Omega$ such that $\overline{U_r(x_0)} \subset \Omega$ we have the mean value property*

$$u(x_0) = \frac{1}{\omega_n r^{n-1}} \int_{\mathrm{S}_r(x_0)} u(y)\,\mathrm{d}S(y) = \frac{1}{\omega_n} \int_{\mathrm{S}_1(0)} u(x_0 + ry)\,\mathrm{d}S(y).$$

Then $u \in \mathrm{C}^\infty(\Omega)$ and u is harmonic in Ω.

Proof. (a) We show that $u \in \mathrm{C}^\infty(\Omega)$. The Mean Value Property ensures that the mollification $h_\varepsilon * u$ equals u as long as $U_{1/\varepsilon}(x_0) \subset \Omega$; that is, the mollification does not change u. By

homework 49.4, $u \in \mathrm{C}^\infty$ since h_ε is. We prove $h_\varepsilon * u = u$. Here g denotes the 1-dimensional bump function, see page 419.

$$\begin{aligned} u_\varepsilon(x) = (u * h_\varepsilon)(x) &= \int_{\mathbb{R}^n} u(y) h_\varepsilon(x-y)\,\mathrm{d}y = \int_{\mathbb{R}^n} u(x-y) h_\varepsilon(y)\,\mathrm{d}y \\ &= \int_{U_\varepsilon(0)} u(x-y) h(y/\varepsilon)\varepsilon^{-n}\,\mathrm{d}y \underset{z_i=\frac{y_i}{\varepsilon}}{=} \int_{U_1(0)} u(x-\varepsilon z) h(z)\,\mathrm{d}z \\ &= \int_0^1 \int_{\mathrm{S}_1(0)} u(x-r\varepsilon y) g(r) r^{n-1}\,\mathrm{d}S(y)\,\mathrm{d}r \\ &= \omega_n u(x) \int_0^1 g(r) r^{n-1}\,\mathrm{d}r = u(x) \int_{\mathbb{R}^n} h(y)\,\mathrm{d}y = u(x). \end{aligned}$$

Second Part. Differentiating the above equation with respect to r yields $\int_{U_r(x)} \Delta u(y)\,\mathrm{d}y = 0$ for any ball in Ω, since the left-hand side $u(x)$ does not depend on r.

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}r} \int_{\mathrm{S}_1(0)} u(x+ry)\,\mathrm{d}S(y) = \int_{\mathrm{S}_1(0)} y \cdot \nabla u(x+ry)\,\mathrm{d}S(y) \\ &= \int_{\mathrm{S}_r(0)} (r^{-1}z) \nabla u(x+z) r^{1-n}\,\mathrm{d}S(z) \\ &= r^{-n} \int_{\mathrm{S}_r(0)} \vec{n}(z) \cdot \nabla u(x+z)\,\mathrm{d}S(z) \\ &= r^{-n} \int_{\mathrm{S}_r(0)} \frac{\partial u}{\partial \vec{n}}(x+z)\,\mathrm{d}S(z) \\ &= r^{-n} \int_{\mathrm{S}_r(x_0)} \frac{\partial u(y)}{\partial \vec{n}}\,\mathrm{d}S(y) = r^{-n} \int_{U_r(x_0)} \Delta u(x)\,\mathrm{d}x. \end{aligned}$$

In the last line we used Green's 2nd formula with $v = 1$. Thus $\Delta u = 0$. Suppose to the contrary that $\Delta u(x_0) \neq 0$, say $\Delta u(x_0) > 0$. By continuity of $\Delta u(x)$, $\Delta u(x) > 0$ for $x \in U_\varepsilon(x_0)$. Hence $\int_{U_\varepsilon(x_0)} \Delta u(x)\,\mathrm{d}x > 0$ which contradicts the above equation. We conclude that u is harmonic in $U_r(x_0)$. ∎

Remark 17.6 A regular distribution $u \in \mathcal{D}'(\Omega)$ is called *harmonic* if $\Delta u = 0$, that is,

$$\langle \Delta u\,,\, \varphi \rangle = \int_\Omega u(x)\,\Delta\varphi(x)\,\mathrm{d}x = 0, \quad \varphi \in \mathcal{D}(\Omega).$$

Weyl's Lemma: Any harmonic regular distribution is a harmonic function, in particular, $u \in \mathrm{C}^\infty(\Omega)$.

Example 17.3 Solve

$$\begin{aligned} \Delta u &= -2, \quad (x,y) \in \Omega = (0,a) \times (-b/2, b/2), \\ u\,|_{\partial\Omega} &= 0. \end{aligned}$$

Ansatz: $u = w + v$ with $\Delta w = 0$. For example, choose $w = -x^2 + ax$. Then $\Delta v = 0$ with boundary conditions

$$v(0, y) = v(a, y) = 0, \quad v(x, -b/2) = v(x, b/2) = x^2 - ax.$$

Use separation of variables, $u(x, y)) = X(x)Y(y)$ to solve the problem.

17.5 Appendix

17.5.1 Existence of Solutions to the Boundary Value Problems

(a) Green's Function

Let $u \in \mathrm{C}^2(\Omega) \cap \mathrm{C}^1(\overline{\Omega})$. Let us combine Green's representation formula and Green's 2^{nd} formula with a harmonic function $v(x) = v_y(x)$, $x \in \Omega$, where $y \in \Omega$ is thought to be a parameter.

$$\begin{aligned} u(y) &= \int_\Omega \mathcal{E}_n(x-y)\,\Delta u(x)\,\mathrm{d}x + \int_{\partial\Omega} \left(u(x)\,\frac{\partial \mathcal{E}_n}{\partial \vec{n}_x}(x-y) - \mathcal{E}_n(x-y)\,\frac{\partial u}{\partial \vec{n}}(x) \right) \mathrm{d}S(x) \\ 0 &= \int_\Omega v_y(x)\Delta u(x)\,\mathrm{d}x + \int_{\partial\Omega} \left(u(x)\frac{\partial v_y}{\partial \vec{n}}(x) - v_y(x)\,\frac{\partial u}{\partial \vec{n}}(x) \right) \mathrm{d}S(x) \end{aligned}$$

Adding up these two lines and denoting $G(x, y) = \mathcal{E}_n(x - y) + v_y(x)$ we get

$$u(y) = \int_\Omega G(x,y)\,\Delta u(x)\,\mathrm{d}x + \int_{\partial\Omega} \left(u(x)\,\frac{\partial G(x,y)}{\partial \vec{n}_x} - G(x,y)\,\frac{\partial u}{\partial \vec{n}}(x) \right) \mathrm{d}S(x).$$

Suppose now that $G(x, y)$ vanishes for all $x \in \partial\Omega$ then the last surface integral is 0 and

$$u(y) = \int_\Omega G(x,y)\,\Delta u(x)\,\mathrm{d}x + \int_{\partial\Omega} u(x)\,\frac{\partial G(x,y)}{\partial \vec{n}_x}\,\mathrm{d}S(x). \tag{17.34}$$

In the above formula, u is completely determined by its boundary values and Δu in Ω. This motivates the following definition.

Definition 17.4 A function $G\colon \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}$ satisfying

(a) $G(x, y) = 0$ for all $x \in \partial\Omega$, $y \in \overline{\Omega}$, $x \neq y$.
(b) $v_y(x) = G(x, y) - \mathcal{E}_n(x - y)$ is harmonic in $x \in \Omega$ for all $y \in \Omega$.

is called a *Green's function* of Ω. More precisely, $G(x, y)$ is a Green's function to the inner Dirichlet problem on Ω.

Remarks 17.7 (a) The function $v_y(x)$ is in particular harmonic in $x = y$. Since $\mathcal{E}_n(x - y)$ has a pole at $x = y$, $G(x, y)$ has a pole of the same order at $x = y$ such that $G(x, y) - \mathcal{E}_n(x - y)$ has no singularity.
If such a function $G(x, y)$ exists, for all $u \in \mathrm{C}^2(\overline{\Omega})$ we have (17.34).

In particular, if in addition, u is harmonic in Ω,

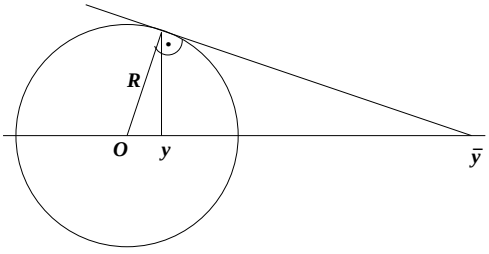
$$u(y) = \int_{\partial\Omega} u(x) \frac{\partial G(x,y)}{\partial \vec{n}_x} \, \mathrm{d}S(x). \tag{17.35}$$

This is the so called *Poisson's formula for* Ω. In general, it is difficult to find Green's function. For most regions Ω it is even impossible to give $G(x,y)$ explicitely. However, if Ω has kind of symmetry, one can use the reflection principle to construct $G(x,y)$ explicitely. Nevertheless, $G(x,y)$ exists for all "well-behaved" Ω (the boundary is a C^2-set and Gauß' divergence theorem holds for Ω).

(c) The Reflection Principle

This is a method to calculate Green's function explicitely in case of domains Ω with the following property: Using repeated reflections on spheres and on hyperplanes occurring as boundaries of Ω and its reflections, the whole $\mathbb{R}^n$ can be filled up without overlapping.

Example 17.4 Green's function on a ball $U_R(0)$. For, we use the reflection on the sphere $\mathrm{S}_R(0)$. For $y \in \mathbb{R}^n$ put



$$\overline{y} := \begin{cases} y \frac{R^2}{\|y\|^2}, & y \neq 0, \\ \infty, & y = 0. \end{cases}$$

Note that this map has the the property $y \cdot \overline{y} = R^2$ and $\|y\|^2 \overline{y} = R^2 y$. Points on the sphere $\mathrm{S}_R(0)$ are fix under this map, $\overline{y} = y$. Let $E_n \colon \mathbb{R}_+ \to \mathbb{R}$ denote the corresponding to $\mathcal{E}_n$ radial scalar function with $\mathcal{E}(x) = E_n(\|x\|)$, that is $E_n(r) = -1/((n-2)\omega_n r^{n-2})$, $n \geq 2$. Then we put

$$G(x,y) = \begin{cases} E_n(\|x-y\|) - E_n\left(\frac{\|y\|}{R} \, \|x - \overline{y}\|\right), & y \neq 0, \\ E_n(\|x\|) - E_n(R), & y = 0. \end{cases} \tag{17.36}$$

For $x \neq y$, $G(x,y)$ is harmonic in x, since for $\|y\| < R$, $\|\overline{y}\| > R$ and therefore $x - \overline{y} \neq 0$. The function $G(x,y)$ has only one singularity in $U_R(0)$ namely at $x = y$ and this is the same as that of $\mathcal{E}_n(x-y)$. Therefore,

$$v_y(x) = G(x,y) - \mathcal{E}_n(x-y) = \begin{cases} -E_n\left(\frac{\|y\|}{R} \, \|x - \overline{y}\|\right), & y \neq 0, \\ E_n(R), & y = 0. \end{cases}$$

is harmonic for all $x \in \Omega$. For $x \in \partial\Omega = \mathrm{S}_R(0)$ we have for $y \neq 0$

$$\begin{aligned} G(x,y) &= E_n\left(\left(\|x\|^2 + \|y\|^2 - 2x\cdot y\right)^{\frac{1}{2}}\right) - E_n\left(\frac{\|y\|}{R}\left(\|x\|^2 + \|\overline{y}\|^2 - 2x\cdot\overline{y}\right)^{\frac{1}{2}}\right) \\ &= E_n\left(\left(R^2 + \|y\|^2 - 2x\cdot y\right)^{\frac{1}{2}}\right) - E_n\left(\left(\|y\|^2 + \frac{\|y\|^2\|\overline{y}\|^2}{R^2} - 2\|y\|^2 x\cdot\frac{\overline{y}}{R^2}\right)^{\frac{1}{2}}\right) \\ &= E_n\left(\left(R^2 + \|y\|^2 - 2x\cdot y\right)^{\frac{1}{2}}\right) - E_n\left(\left(\|y\|^2 + R^2 - 2x\cdot y\right)^{\frac{1}{2}}\right) = 0. \end{aligned}$$

For $y = 0$ we have

$$G(x,0) = E_n(\|x\|) - E_n(R) = E_n(R) - E_n(R) = 0.$$

This proves that $G(x,y)$ is a Green's function for $U_R(0)$. In particular, the above calculation shows that $r = \|x-y\|$ and $\overline{r} = \frac{\|y\|}{R}\|x-\overline{y}\|$ are equal if $x \in \partial\Omega$.
One can show that Green's function is symmetric, that is $G(x,y) = G(y,x)$. This is a general property of Green's function.
To apply formula (17.34) we have to compute the normal derivative $\dfrac{\partial}{\partial\vec{n}_x}G(x,y)$. Note first that for any constant $z \in \mathbb{R}^n$ and $x \in \mathrm{S}_R(0)$

$$\frac{\partial}{\partial\vec{n}_x}f(\|x-z\|) = \vec{n}\cdot\nabla f(\|x-z\|) = \frac{x}{\|x\|}\cdot f'(\|x-z\|)\frac{x-z}{\|x-z\|}.$$

Note further that for $\|x\| = R$ we have

$$r = \|x-y\| = \frac{\|y\|}{R}\|x-\overline{y}\|, \tag{17.37}$$

$$(x-y)\cdot x - \frac{\|y\|^2}{R^2}(x-\overline{y})\cdot x = R^2 - \|y\|^2. \tag{17.38}$$

Hence, for $y \neq 0$,

$$\begin{aligned} \frac{\partial}{\partial\vec{n}_x}G(x,y) &= -\frac{1}{(n-2)\omega_n}\left(\frac{\partial}{\partial\vec{n}_x}\|x-y\|^{-n+2} - \frac{\partial}{\partial\vec{n}_x}\left(\frac{\|y\|^{-n+2}}{R^{-n+2}}\|x-\overline{y}\|^{-n+2}\right)\right) \\ &= \frac{1}{\omega_n}\left(\|x-y\|^{-n+1}\frac{x-y}{\|x-y\|}\cdot\frac{x}{\|x\|} - \frac{\|y\|^{-n+2}}{R^{-n+2}}\|x-\overline{y}\|^{-n+1}\frac{x-\overline{y}}{\|x-\overline{y}\|}\cdot\frac{x}{\|x\|}\right) \\ &= \frac{1}{\omega_n r^n R}\left((x-y)\cdot x - (x-\overline{y})\frac{\|y\|^2}{R^2}\cdot x\right) \end{aligned}$$

By (17.38), the expression in the brackets is $R^2 - \|y\|^2$. Hence,

$$\frac{\partial}{\partial\vec{n}_x}G(x,y) = \frac{R^2-\|y\|^2}{\omega_n R}\frac{1}{\|x-y\|^n}.$$

This formula holds true in case $y = 0$. Inserting this into (17.34) we have for any harmonic function $u \in \mathrm{C}^2(U_R(0)) \cap \mathrm{C}(\overline{U_R(0)})$ we have

$$u(y) = \frac{R^2-\|y\|^2}{\omega_n R}\int\limits_{\mathrm{S}_R(0)}\frac{u(x)}{\|x-y\|^n}\,\mathrm{d}S(x). \tag{17.39}$$

This is the so called *Poisson's formula for the ball* $U_R(0)$.

Proposition 17.25 *Let $n \geq 2$. Consider the inner Dirichlet problem in $\Omega = U_R(0)$ and $f = 0$. The function*

$$u(y) = \begin{cases} \frac{R^2 - \|y\|^2}{\omega_n R} \int\limits_{\mathrm{S}_R(0)} \frac{\varphi(x)}{\|x-y\|^n} \, \mathrm{d}S(x), & \|y\| < R, \\ \varphi(y), & \|y\| = R \end{cases}$$

is continuous on the closed ball $\overline{U_R}(0)$ and harmonic in $U_R(0)$.
In case $n = 2$ the function $u(y)$, can be written in the following form

$$u(y) = \operatorname{Re}\left(\frac{1}{2\pi \mathrm{i}} \int_{\mathrm{S}_R(0)} \varphi(z) \, \frac{z+y}{z-y} \, \frac{\mathrm{d}z}{z}\right), \quad y \in U_R(0) \subset \mathbb{C}.$$

For the proof of the general statement with $n \geq 2$, see [Jos02, Theorem 1.1.2] or [Joh82, p. 107]. We show the last statement for $n = 2$. Since $y\overline{z} - \overline{y}z$ is purely imaginary,

$$\operatorname{Re} \frac{z+y}{z-y} = \operatorname{Re} \frac{(z+y)(\overline{z}-\overline{y})}{(z-y)(\overline{z}-\overline{y})} = \operatorname{Re} \frac{|\, z \,|^2 - |\, y \,|^2 + y\overline{z} - \overline{y}z}{|\, z-y \,|^2} = \frac{R^2 - |\, y \,|^2}{|\, z-y \,|^2}.$$

Using the parametrization $z = R\mathrm{e}^{\mathrm{i}t}$, $\mathrm{d}t = \dfrac{\mathrm{d}z}{\mathrm{i}z}$ we obtain

$$\operatorname{Re}\left(\frac{1}{2\pi} \int_{\mathrm{S}_R(0)} \varphi(z) \, \frac{z+y}{z-y} \, \frac{\mathrm{d}z}{\mathrm{i}z}\right) = \frac{1}{2\pi} \int_0^{2\pi} \frac{R^2 - |\, y \,|^2}{|\, z-y \,|^2} \, \varphi(z) \, \mathrm{d}t = \frac{R^2 - |\, y \,|^2}{2\pi R} \int_{\mathrm{S}_R(0)} \frac{\varphi(x)}{|\, x-y \,|^2} \, |\, \mathrm{d}x \,|.$$

In the last line we have a (real) line integral of the first kind, using $x = (x_1, x_2) = x_1 + \mathrm{i}x_2 = z$, $x \in \mathrm{S}_R(0)$ and $|\, \mathrm{d}x \,| = R \, \mathrm{d}t$ on the circle.

Other Examples. (a) $n = 3$. The half-space $\Omega = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_3 > 0\}$. We use the ordinary reflection map with respect to the plane $x_3 = 0$ which is given by $y = (y_1, y_2, y_3) \mapsto y' = (y_1, y_2, -y_3)$. Then Green's function to Ω is

$$G(x,y) = \mathcal{E}_3(x,y) - \mathcal{E}_3(x,y') = \frac{1}{4\pi} \left(\frac{1}{\|x-y'\|} - \frac{1}{\|x-y\|}\right).$$

(see homework 57.1)
(b) $n = 3$. The half ball $\Omega = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid \|x\| < R, \quad x_3 > 0\}$. We use the reflections $y \to y'$ and $y \to \overline{y}$ (reflection with respect to the sphere $\mathrm{S}_R(0)$). Then

$$G(x,y) = \mathcal{E}_3(x-y) - \frac{R}{\|y\|}\mathcal{E}_3(x-\overline{y}) - \mathcal{E}_3(x-y') + \frac{R}{\|y\|}\mathcal{E}_3(x-\overline{y}')$$

is Green's function to Ω.
(c) $n = 3$, $\Omega = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_2 > 0, \quad x_3 > 0\}$. We introduce the reflection $y = (y_1, y_2, y_3) \mapsto y^* = (y_1, -y_2, y_3)$. Then Green's function to Ω is

$$G(x,y) = \mathcal{E}_3(x-y) - \mathcal{E}_3(x-y') - \mathcal{E}_3(x-y^*) + \mathcal{E}_3(x-(y^*)').$$

Consider the Neumann problem and the ansatz for Green's function in case of the Dirichlet problem:

$$u(y) = \int_\Omega H(x,y)\,\Delta u(x)\,\mathrm{d}x + \int_{\partial\Omega}\left(u(x)\,\frac{\partial H(x,y)}{\partial \vec{n}_x} - H(x,y)\,\frac{\partial u}{\partial \vec{n}}(x)\right)\,\mathrm{d}S(x). \quad (17.40)$$

We want to choose a Green's function of the second kind $H(x,y)$ in such a way that only the last surface integral remains present.
Inserting $u = 1$ in the above formula, we have

$$1 = \int_{\partial\Omega} \frac{\partial G(x,y)}{\partial \vec{n}_x}\,\mathrm{d}S(x).$$

Imposing, $\frac{\partial G(x,y)}{\partial \vec{n}_x} = \alpha = \text{ const.}$, this constant must be $\alpha = 1/\mathrm{vol}(\partial\Omega)$.
Note, that one defines a Green's function to the Neumann problem replacing $G(x,y) = 0$ on $\partial\Omega$ by the condition $\frac{\partial}{\partial \vec{n}_y} G(x,y) = \text{ const.}$
Green's function of second kind (Neumann problem) to the ball of radius R in $\mathbb{R}^3$, $U_R(0)$.

$$H(x,y) = -\frac{1}{4\pi}\left(\frac{1}{\|x-y\|} + \frac{R}{\|y\|\,\|x-\overline{y}\|} + \frac{1}{R}\log\frac{2R^2}{R^2 - x\cdot y + \|y\|\,\|x-\overline{y}\|}\right)$$

(c) Existence Theorems

The aim of this considerations is to sketch the method of proving *existence* of solutions of the four BVPs.
Definition. Suppose that $\Omega \subset \mathbb{R}^n$, $n \geq 3$ and let $\mu(y)$ be a continuous function on the boundary $\partial\Omega$, that is $\mu \in \mathrm{C}(\partial\Omega)$. We call

$$w(x) = \int_{\partial\Omega} \mu(y)\,\mathcal{E}_n(x-y)\,\mathrm{d}S(y), \quad x \in \mathbb{R}^n, \quad (17.41)$$

a *single-layer potential* and

$$v(x) = \int_{\partial\Omega} \mu(y)\,\frac{\partial \mathcal{E}_n}{\partial \vec{n}}(x-y)\,\mathrm{d}S(y), \quad x \in \mathbb{R}^n \quad (17.42)$$

a *double-layer potential*

Remarks 17.8 (a) For $x \notin \partial\Omega$ the integrals (17.41) and (17.42) exist.
(b) The single layer potential $u(x)$ is continuous on $\mathbb{R}^n$. The double-layer potential jumps at $y_0 \in \partial\Omega$ by $\mu(y_0)$ as x approaches y_0, see (17.43) below.

Theorem 17.26 *Let Ω be a connected bounded region in $\mathbb{R}^n$ of the class C^2 and $\Omega' = \mathbb{R}^n \setminus \overline{\Omega}$ also be connected.*
Then the interior Dirichlet problem to the Laplace equation has a unique solution. It can be represented in form of a double-layer potential. The exterior Neumann problem likewise has a unique solution which can be represented in form of a single-layer potential.

Theorem 17.27 *Under the same assumptions as in the previous theorem, the inner Neumann problem to the Laplace equation has a solution if and only if* $\int_{\partial\Omega} \varphi(y)\,\mathrm{d}S(y) = 0$. *If this condition is satisfies, the solution is unique up to a constant.*
The exterior Dirichlet problem has a unique solution.

Remark. Let μ_{ID} denote the continuous functions which produces the solution $v(x)$ of the interior Dirichlet problem, i. e. $v(x) = \int_{\partial\Omega} \mu_{\mathrm{ID}}(y)\, K(x,y)\,\mathrm{d}S(y)$, where $K(x,y) = \dfrac{\partial}{\partial \vec{n}_y}\mathcal{E}_n(x-y)$. Because of the jump relation for $v(x)$ at $x_0 \in \partial\Omega$:

$$\lim_{x\to x_0, x\in\Omega} v(x) - \frac{1}{2}\mu(x_0) = v(x_0) = \lim_{x\to x_0, x\in\Omega'} v(x) + \frac{1}{2}\mu(x_0), \tag{17.43}$$

μ_{ID} satisfies the integral equation

$$\varphi(x) = \frac{1}{2}\mu_{\mathrm{ID}}(x) + \int_{\partial\Omega} \mu_{\mathrm{ID}}(y) K(x,y)\,\mathrm{d}S(y), \quad x \in \partial\Omega.$$

The above equation can be written as $\varphi = (A + \frac{1}{2}I)\mu_{\mathrm{ID}}$, where A is the above integral operator in $\mathrm{L}^2(\partial\Omega)$. One can prove the following facts: A is compact, $A + \frac{1}{2}I$ is injective and surjective, φ continuous implies μ_{ID} continuous. For details, see [Tri92, 3.4].

Application to the Poisson Equation

Consider the inner Dirichlet problem $\Delta u = f$, and $u = \varphi$ on $\partial\Omega$. We suppose that $f \in \mathrm{C}(\overline{\Omega}) \cap \mathrm{C}^1(\Omega)$. We already know that

$$w(x) = (\mathcal{E}_n * f)(x) = -\frac{1}{(n-2)\omega_n} \int_{\mathbb{R}^n} \frac{f(y)}{\|x-y\|^{n-2}}\,\mathrm{d}y$$

is a distributive solution of the Poisson equation, $\Delta w = f$. By the assumptions on f, $w \in \mathrm{C}^2(\Omega)$ and therefore is a classical solution. To solve the problem we try the ansatz $u = w + v$. Then $\Delta u = \Delta w + \Delta v = f + \Delta v$. Hence, $\Delta u = f$ if and only if $\Delta v = 0$. Thus, the inner Dirichlet problem for the Poisson equation reduces to the inner Dirichlet problem for the Laplace equation $\Delta v = 0$ with boundary values

$$v(y) = u(y) - w(y) = \varphi(y) - w(y) =: \tilde{\varphi}(y), \quad y \in \partial\Omega.$$

Since φ and w are continuous on $\partial\Omega$, so is $\tilde{\varphi}$.

17.5.2 Extremal Properties of Harmonic Functions and the Dirichlet Principle

(a) The Dirichlet Principle

Consider the inner Dirichlet problem to the Poisson equation with given data $f \in \mathrm{C}(\overline{\Omega})$ and $\varphi \in \mathrm{C}(\partial\Omega)$.
Put

$$\mathrm{C}^1_\varphi(\overline{\Omega}) := \{v \in \mathrm{C}^1(\overline{\Omega}) \mid v = \varphi \quad \text{on} \quad \partial\Omega\}.$$

On this space define the *Dirichlet integral* by

$$E(v) = \frac{1}{2}\int_\Omega \|\nabla v\|^2 \,\mathrm{d}x + \int_\Omega f\cdot v\,\mathrm{d}x, \qquad v \in \mathrm{C}^1_\varphi(\overline{\Omega}). \tag{17.44}$$

This integral is also called *energy integral*. The Dirichlet principle says that among all functions v with given boundary values φ, the function u with $\Delta u = f$ minimizes the energy integral E.

Proposition 17.28 *A function $u \in \mathrm{C}^1_\varphi(\overline{\Omega}) \cap \mathrm{C}^2(\Omega)$ is a solution of the inner Dirichlet problem if and only if the energy integral E attains its minimum on $\mathrm{C}^1_\varphi(\overline{\Omega})$ at u.*

Proof. (a) Suppose first that $u \in \mathrm{C}^1_\varphi(\overline{\Omega}) \cap \mathrm{C}^2(\Omega)$ is a solution of the inner Dirichlet problem, $\Delta u = f$. For $v \in \mathrm{C}^1_\varphi(\overline{\Omega})$ let $w = v - u \in \mathrm{C}^1_\varphi(\overline{\Omega})$. Then

$$\begin{aligned} E(v) = E(u+w) &= \frac{1}{2}\int_\Omega (\nabla u + \nabla w)\cdot(\nabla u + \nabla w)\,\mathrm{d}x + \int_\Omega (u+w) f\,\mathrm{d}x \\ &= \frac{1}{2}\int_\Omega \|\nabla u\|^2 + \frac{1}{2}\int_\Omega \|\nabla w\|^2 + \int_\Omega \nabla u\cdot\nabla w\,\mathrm{d}x + \int_\Omega (u+w) f\,\mathrm{d}x \end{aligned}$$

Since u and v satisfy the same boundary conditions, $w\,|_{\partial\Omega} = 0$. Further, $\Delta u = f$. By Green's 1^{st} formula,

$$\int_\Omega \nabla u\cdot\nabla w\,\mathrm{d}x = -\int_\Omega (\Delta u)\, w\,\mathrm{d}x + \int_{\partial\Omega} \frac{\partial u}{\partial \vec{n}}\, w\,\mathrm{d}S = -\int_\Omega f\, w\,\mathrm{d}x.$$

Inserting this into the above equation, we have

$$\begin{aligned} E(v) &= \frac{1}{2}\int_\Omega \|\nabla u\|^2 + \frac{1}{2}\int_\Omega \|\nabla w\|^2 - \int_\Omega f w\,\mathrm{d}x + \int_\Omega (u+w) f\,\mathrm{d}x \\ &= E(u) + \frac{1}{2}\int_\Omega \|\nabla w\|^2 \,\mathrm{d}x \geq E(u). \end{aligned}$$

This shows that $E(u)$ is minimal.
(b) Conversely, let $u \in \mathrm{C}^1_\varphi(\overline{\Omega}) \cap \mathrm{C}^2(\Omega)$ minimize the energy integral. In particular, for any test function $\psi \in \mathcal{D}(\Omega)$, ψ has zero boundary values, the function

$$g(t) = E(u + t\psi) = E(u) + t\int_\Omega (\nabla u\cdot\nabla\psi + f\psi)\,\mathrm{d}x + \frac{1}{2}t^2\int_\Omega \|\nabla\psi\|^2\,\mathrm{d}x$$

has a local minimum at $t = 0$. Hence, $g'(0) = 0$ which is, again by Green's 1^{st} formula and $\psi\,|_{\partial\Omega} = 0$, equivalent to

$$0 = \int_\Omega (\nabla u\cdot\nabla\psi + f\psi)\,\mathrm{d}x = \int_\Omega (-\Delta u + f)\,\psi\,\mathrm{d}x.$$

By the fundamental Lemma of calculus of variations, $\Delta u = f$ almost everywhere on Ω. Since both Δu and f are continuous, this equation holds pointwise for all $x \in \Omega$.

■

(b) Hilbert Space Methods

We want to give another reformulation of the Dirichlet problem. Consider the problem

$$\Delta u = -f, \quad u\,|_{\partial\Omega} = 0.$$

On $\mathrm{C}^1(\overline{\Omega})$ define a bilinear map

$$u \cdot v_{\mathrm{E}} = \int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x.$$

$\mathrm{C}^1(\overline{\Omega})$ is not yet an inner product space since for any non-vanishing constant function u, $u \cdot u_{\mathrm{E}} = 0$. Denote by $\mathrm{C}^1_0(\overline{\Omega})$ the subspace of functions in $\mathrm{C}^1(\overline{\Omega})$ vanishing on the boundary $\partial\Omega$. Now, $u \cdot v_{\mathrm{E}}$ is an inner product on $\mathrm{C}^1_0(\overline{\Omega})$. The positive definiteness is a consequence of the Poincaré inequality below. Its corresponding norm is $\|u\|^2_{\mathrm{E}} = \int_\Omega \|\nabla u\|^2 \,\mathrm{d}x$. Let u be a solution of the above Dirichlet problem. Then for any $v \in \mathrm{C}^1_0(\overline{\Omega})$, by Green's 1^{st} formula

$$v \cdot u_{\mathrm{E}} = \int_\Omega \nabla v \cdot \nabla u \,\mathrm{d}x = -\int_\Omega v\, \Delta u \,\mathrm{d}x = \int_\Omega v\, f \,\mathrm{d}x = v \cdot f_{\mathrm{L}^2}.$$

This suggests that u can be found by representing the known linear functional in v

$$F(v) = \int_\Omega v\, f \,\mathrm{d}x$$

as an inner product $v \cdot u_{\mathrm{E}}$. To make use of Riesz's representations theorem, Theorem 13.8, we have to complete $\mathrm{C}^1_0(\overline{\Omega})$ into a Hilbert space W with respect to the energy norm $\|\cdot\|_{\mathrm{E}}$ and to prove that the above linear functional F is *bounded* with respect to the energy norm. This is a consequence of the next lemma. We make the same assumtions on Ω as in the beginning of Section 17.4.

Lemma 17.29 (Poincaré inequality) *Let $\Omega \subset \mathbb{R}^n$. Then there exists $C > 0$ such that for all $u \in \mathrm{C}^1_0(\overline{\Omega})$*

$$\|u\|_{\mathrm{L}^2(\Omega)} \le C \,\|u\|_{\mathrm{E}}\,.$$

Proof. Let Ω be contained in the cube $\Gamma = \{x \in \mathbb{R}^n \mid |\,x_i\,| \le a,\ i = 1, \ldots, n\}$. We extend u by zero outside Ω. For any $x = (x_1, \ldots, x_n)$, by the Fundamental Theorem of Calculus

$$\begin{aligned} u(x)^2 &= \left(\int_{-a}^{x_1} u_{x_1}(y_1, x_2, \ldots, x_n)\,\mathrm{d}y_1\right)^2 = \left(\int_{-a}^{x_1} 1 \cdot u_{x_1}(y_1, x_2, \ldots, x_n)\,\mathrm{d}y_1\right)^2 \\ &\underset{\text{CSI}}{\le} \int_{-a}^{x_1} \mathrm{d}y_1 \int_{-a}^{x_1} u_{x_1}^2 \,\mathrm{d}y_1 = (x_1 + a) \int_{-a}^{x_1} u_{x_1}^2 \,\mathrm{d}y_1 \le 2a \int_{-a}^{a} u_{x_1}^2 \,\mathrm{d}y_1. \end{aligned}$$

Since the last integral does not depend on x_1, integration with respect to x_1 gives

$$\int_{-a}^{a} u(x)^2 \,\mathrm{d}x_1 \le 4a^2 \int_{-a}^{a} u_{x_1}^2 \,\mathrm{d}y_1.$$

Integrating over $x_2, \ldots, x_n$ from $-a$ to a we find

$$\int_\Gamma u^2 \,\mathrm{d}x \le 4a^2 \int_\Gamma u_{x_1}^2 \,\mathrm{d}y.$$

The same inequality holds for x_i, $i = 2, \ldots, n$ in place of x_1 such that

$$\|u\|_{\mathrm{L}^2}^2 = \int_\Gamma u^2 \,\mathrm{d}x \le \frac{4a^2}{n} \int_\Gamma (\nabla u)^2 \,\mathrm{d}x = C^2 \|u\|_{\mathrm{E}}^2 ,$$

where $C = 2a/\sqrt{n}$. ∎

The Poincaré inequality is sometimes called Poincaré–Friedrich inequality. It remains true for functions u in the completion W. Let us discuss the elements of W in more detail. By definition, $f \in W$ if there is a Cauchy sequence (f_n) in $\mathrm{C}_0^1(\overline{\Omega})$ such that (f_n) "converges to f" in the energy norm. By the Poincaré inequality, (f_n) is also an L^2-Cauchy sequence. Since $\mathrm{L}^2(\Omega)$ is complete, (f_n) has an L^2-limit f. This shows $W \subseteq \mathrm{L}^2(\Omega)$. For simplicity, let $\Omega \subset \mathbb{R}^1$. By definition of the energy norm, $\int_\Omega |\, (f_n - f_m)' \,|^2 \,\mathrm{d}x \to 0$, as $m, n \to \infty$; that is (f_n') is an L^2-Cauchy sequence, too. Hence, (f_n') has also some L^2-limit, say $g \in \mathrm{L}^2(\Omega)$. So far,

$$\|f_n - f\|_{\mathrm{L}^2} \longrightarrow 0, \quad \|f_n' - g\|_{\mathrm{L}^2} \longrightarrow 0. \tag{17.45}$$

We will show that the above limits imply $f' = g$ in $\mathcal{D}'(\Omega)$. Indeed, by (17.45) and the Cauchy–Schwarz inequality, for all $\varphi \in \mathcal{D}(\Omega)$,

$$\int_\Omega (f_n - f)\varphi' \,\mathrm{d}x \le \left(\int_\Omega |\, f_n - f \,|^2 \,\mathrm{d}x \right)^{\frac{1}{2}} \left(\int_\Omega |\, \varphi' \,|^2 \,\mathrm{d}x \right)^{\frac{1}{2}} \longrightarrow 0,$$
$$\int_\Omega (f_n' - g)\varphi \,\mathrm{d}x \le \left(\int_\Omega |\, f_n' - g \,|^2 \,\mathrm{d}x \right)^{\frac{1}{2}} \left(\int_\Omega |\, \varphi \,|^2 \,\mathrm{d}x \right)^{\frac{1}{2}} \longrightarrow 0.$$

Hence,

$$\int_\Omega f' \,\varphi \,\mathrm{d}x = -\int_\Omega f \,\varphi' \,\mathrm{d}x = -\lim_{n\to\infty} \int_\Omega f_n \,\varphi' \,\mathrm{d}x = \lim_{n\to\infty} \int_\Omega f_n' \,\varphi \,\mathrm{d}x = \int_\Omega g \,\varphi \,\mathrm{d}x.$$

This shows $f' = g$ in $\mathcal{D}'(\Omega)$. One says that the elements of W provide *weak derivatives*, that is, its distributive derivative is an L^2-function (and hence a regular distribution).
Also, the inner product $\cdots_{\mathrm{E}}$ is positive definite since the L^2-inner product is. It turns out that W is a separable Hilbert space. W is the so called *Sobolev space* $\mathrm{W}_0^{1,2}(\Omega)$ sometimes also denoted by $\mathrm{H}_0^1(\Omega)$. The upper indices 1 and 2 in $\mathrm{W}_0^{1,2}(\Omega)$ refer to the highest order of partial derivatives ($|\,\alpha\,| = 1$) and the L^p-space ($p = 2$) in the definition of W, respectively. The lower index 0 refers to the so called *generalized boundary values* 0. For further readings on Sobolev spaces, see [Fol95, Chapter 6].

Corollary 17.30 *$F(v) = v \cdot f_{\mathrm{L}^2} = \int_\Omega f\, v \,\mathrm{d}x$ defines a bounded linear functional on W.*

Proof. By the Cauchy–Schwarz and Poincaré inequalities,

$$|\, F(v) \,| \le \int_\Omega |\, f\, v \,|\, \mathrm{d}x \le \|f\|_{\mathrm{L}^2} \,\|v\|_{\mathrm{L}^2} \le C \,\|f\|_{\mathrm{L}^2} \,\|v\|_{\mathrm{E}} .$$

Hence, F is bounded with $\|F\| \le C \,\|f\|_{\mathrm{L}^2}$. ∎

Corollary 17.31 *Let $f \in \mathrm{C}(\overline{\Omega})$. Then there exists a unique $u \in W$ such that*

$$v \cdot u_{\mathrm{E}} = v \cdot f_{\mathrm{L}^2}, \qquad \forall\, v \in W.$$

This u solves

$$\Delta u = -f, \quad \textit{in} \quad \mathcal{D}'(\Omega).$$

The first statement is a consequence of Riesz's representation theorem; note that F is a bounded linear functional on the Hilbert space W. The last statement follows from $\mathcal{D}(\Omega) \subset \mathrm{C}_0^1(\overline{\Omega})$ and $u \cdot \varphi_{\mathrm{E}} = -\langle \Delta u\,,\, \varphi \rangle = f \cdot \varphi_{\mathrm{L}^2}$. This is the so called *modified Dirichlet problem*. It remains open the task to identify the solution $u \in W$ with an *ordinary function* $u \in \mathrm{C}^2(\Omega)$.

17.5.3 Numerical Methods

(a) Difference Methods

Since most ODE and PDE are not solvable in a closed form there are a lot of methods to find approximative solutions to a given equation or a given problem. A general principle is *discretization*. One replaces the derivative $u'(x)$ by one of its *difference quotients*

$$\partial^+ u(x) = \frac{u(x+h) - u(x)}{h}, \quad \partial^- u(x) = \frac{u(x) - u(x-h)}{h},$$

where h is called the step size. One can also use a symmetric difference $\frac{u(x+h)-u(x-h)}{2h}$. The Five-Point formula for the Laplacian in $\mathbb{R}^2$ is then given by

$$\begin{aligned}\Delta_h u(x,y) &:= (\partial_x^- \partial_x^+ + \partial_y^- \partial_y^+) u(x,y) = \\ &= \frac{u(x-h,y) + u(x+h,y) + u(x,y-h) + u(x,y+h) - 4u(x,y)}{h^2}\end{aligned}$$

Besides the equation, the domain Ω as well as its boundary $\partial\Omega$ undergo a discretization: If $\Omega = (0,1) \times (0,1)$ then

$$\Omega_h = \{(nh, mh) \in \Omega \mid n, m \in \mathbb{N}\}, \quad \partial\Omega_h = \{(nh, mh) \in \partial\Omega \mid n, m \in \mathbb{Z}\}.$$

The discretization of the inner Dirichlet problem then reads

$$\begin{aligned}\Delta_h u &= f, \quad x \in \Omega_h, \\ u\,|_{\partial\Omega_h} &= \varphi.\end{aligned}$$

Also, Neumann problems have discretizations, [Hac92, Chapter 4].

(b) The Ritz–Galerkin Method

Suppose we have a boundary value problem in its variational formulation:

Find $u \in V$, so that $u \cdot v_{\mathrm{E}} = F(v)$ for all $v \in V$,

where we are thinking of the Sobolev space $V = W$ from the previous paragraph. Of course, F is assumed to be bounded.
Difference methods arise through discretising the differential operator. Now we wish to leave the differential operator which is hidden in $\cdots_{\mathrm{E}}$ unchanged. The Ritz–Galerkin method consists in replacing the infinite-dimensional space V by a finite-dimensional space V_N,

$$V_N \subset V, \quad \dim V_N = N < \infty.$$

V_N equipped with the norm $\|\cdot\|_{\mathrm{E}}$ is still a Banach space. Since $V_N \subseteq V$, both the inner product $\cdots_{\mathrm{E}}$ and F are defined for $u, v \in V_N$. Thus, we may pose the problem

Find $u_N \in V_N$, so that $u_N \cdot v_{\mathrm{E}} = F(v)$ for all $v \in V_N$,

The solution to the above problem, if it exists, is called *Ritz–Galerkin solution* (belonging to V_N).
An introductory example is to be found in [Hac92, 8.1.11, p. 164], see also [Bra01, Chapter 2].

Bibliography

[AF01] I. Agricola and T. Friedrich. *Globale Analysis (in German)*. Friedr. Vieweg & Sohn, Braunschweig, 2001.

[Ahl78] L. V. Ahlfors. *Complex analysis. An introduction to the theory of analytic functions of one complex variable*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York, 3 edition, 1978.

[Arn04] V. I. Arnold. *Lectures in Partial Differential Equations*. Universitext. Springer and Phasis, Berlin. Moscow, 2004.

[Bra01] D. Braess. *Finite elements. Theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, Cambridge, 2001.

[Bre97] Glen E. Bredon. *Topology and geometry*. Number 139 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1997.

[Brö92] Th. Bröcker. *Analysis II (German)*. B. I. Wissenschaftsverlag, Mannheim, 1992.

[Con78] J. B. Conway. *Functions of one complex variable*. Number 11 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1978.

[Con90] J. B. Conway. *A course in functional analysis*. Number 96 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1990.

[Cou88] R. Courant. *Differential and integral calculus I–II*. Wiley Classics Library. John Wiley & Sons, New York etc., 1988.

[Die93] J. Dieudonné. *Treatise on analysis. Volume I – IX*. Pure and Applied Mathematics. Academic Press, Boston, 1993.

[Els02] J. Elstrodt. *Maß- und Integrationstheorie*. Springer, Berlin, third edition, 2002.

[Eva98] L. C. Evans. *Partial differential equations*. Number 19 in Graduate Studies in Mathematics. AMS, Providence, 1998.

[FB93] E. Freitag and R. Busam. *Funktionentheorie*. Springer-Lehrbuch. Springer-Verlag, Berlin, 1993.

[FK98] H. Fischer and H. Kaul. *Mathematik für Physiker. Band 2*. Teubner Studienbücher: Mathematik. Teubner, Stuttgart, 1998.

[FL88] W. Fischer and I. Lieb. *Ausgewählte Kapitel der Funktionentheorie*. Number 48 in Vieweg Studium. Friedrich Vieweg & Sohn, Braunschweig, 1988.

[Fol95] G. B. Folland. *Introduction to partial differential equations*. Princeton University Press, Princeton, 1995.

[For81] O. Forster. *Analysis 3 (in German)*. Vieweg Studium: Aufbaukurs Mathematik. Vieweg, Braunschweig, 1981.

[For01] O. Forster. *Analysis 1 – 3 (in German)*. Vieweg Studium: Grundkurs Mathematik. Vieweg, Braunschweig, 2001.

[GS64] I. M. Gelfand and G. E. Schilow. *Verallgemeinerte Funktionen (Distributionen). III: Einige Fragen zur Theorie der Differentialgleichungen. (German)*. Number 49 in Hochschulbücher für Mathematik. VEB Deutscher Verlag der Wissenschaften, Berlin, 1964.

[GS69] I. M. Gelfand and G. E. Schilow. *Verallgemeinerte Funktionen (Distributionen). I: Verallgemeinerte Funktionen und das Rechnen mit ihnen. (German)*. Number 47 in Hochschulbücher für Mathematik. VEB Deutscher Verlag der Wissenschaften, Berlin, 1969.

[Hac92] W. Hackbusch. *Theory and numerical treatment*. Number 18 in Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 1992.

[Hen88] P. Henrici. *Applied and computational complex analysis. Vol. 1*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1988.

[How03] J. M. Howie. *Complex analysis*. Springer Undergraduate Mathematics Series. Springer-Verlag, London, 2003.

[HS91] F. Hirzebruch and W. Scharlau. *Einführung in die Funktionalanalysis*. Number 296 in BI-Hochschultaschenbücher. BI-Wissenschaftsverlag, Mannheim, 1991.

[HW96] E. Hairer and G. Wanner. *Analysis by its history*. Undergraduate texts in Mathematics. Readings in Mathematics. Springer-Verlag, New York, 1996.

[Jän93] K. Jänich. *Funktionentheorie*. Springer-Lehrbuch. Springer-Verlag, Berlin, third edition, 1993.

[Joh82] F. John. *Partial differential equations*. Number 1 in Applied Mathematical Sciences. Springer-Verlag, New York, 1982.

[Jos02] J. Jost. *Partial differential equations*. Number 214 in Graduate Texts in Mathematics. Springer-Verlag, New York, 2002.

[KK71] A. Kufner and J. Kadlec. *Fourier Series*. G. A. Toombs Iliffe Books, London, 1971.

[Kno78] K. Knopp. *Elemente der Funktionentheorie*. Number 2124 in Sammlung Göschen. Walter de Gruyter, Berlin, New York, 9 edition, 1978.

[Kön90] K. Königsberger. *Analysis 1 (English)*. Springer-Verlag, Berlin, Heidelberg, New York, 1990.

[Lan89] S. Lang. *Undergraduate Analysis*. Undergraduate texts in mathematics. Springer, New York-Heidelberg, second edition, 1989.

[MW85] J. Marsden and A. Weinstein. *Calculus. I, II, III*. Undergraduate Texts in Mathematics. Springer-Verlag, New York etc., 1985.

[Nee97] T. Needham. *Visual complex analysis*. Oxford University Press, New York, 1997.

[O'N75] P. V. O'Neil. *Advanced calculus*. Collier Macmillan Publishing Co., London, 1975.

[RS80] M. Reed and B. Simon. *Methods of modern mathematical physics. I. Functional Analysis*. Academic Press, Inc., New York, 1980.

[Rud66] W. Rudin. *Real and Complex Analysis*. International Student Edition. McGraw-Hill Book Co., New York-Toronto, 1966.

[Rud76] W. Rudin. *Principles of mathematical analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, third edition, 1976.

[Rüh83] F. Rühs. *Funktionentheorie*. Hochschulbücher für Mathematik. VEB Deutscher Verlag der Wissenschaften, Berlin, 4 edition, 1983.

[Spi65] M. Spivak. *Calculus on manifolds*. W. A. Benjamin, New York, Amsterdam, 1965.

[Spi80] M. Spivak. *Calculus*. Publish or Perish, Inc., Berkeley, California, 1980.

[Str92] W. A. Strauss. *Partial differential equations*. John Wiley & Sons, New York, 1992.

[Tri92] H. Triebel. *Higher analysis*. Hochschulbücher für Mathematik. Johann Ambrosius Barth Verlag GmbH, Leipzig, 1992.

[vW81] C. von Westenholz. *Differential forms in mathematical physics*. Number 3 in Studies in Mathematics and its Applications. North-Holland Publishing Co., Amsterdam-New York, second edition, 1981.

[Wal74] W. Walter. *Einführung in die Theorie der Distributionen (in German)*. Bibliographisches Institut, B.I.- Wissenschaftsverlag, Mannheim-Wien-Zürich, 1974.

[Wal02] W. Walter. *Analysis 1–2 (in German)*. Springer-Lehrbuch. Springer, Berlin, fifth edition, 2002.

[Wla72] V. S. Wladimirow. *Gleichungen der mathematischen Physik (in German)*. Number 74 in Hochschulbücher für Mathematik. VEB Deutscher Verlag der Wissenschaften, Berlin, 1972.

PD DR. A. SCHÜLER
MATHEMATISCHES INSTITUT
UNIVERSITÄT LEIPZIG
04009 LEIPZIG
Axel.Schueler@math.uni-leipzig.de