Analysis 1

Lecture Notes 2013/2014

The original version of these Notes was written by

Vitali Liskevich

followed by minor adjustments by many Successors, and presently taught by

Misha Rudnev

University of Bristol
Bristol BS8 1TW, UK

Contents

Recommended Texts

[1] C.W. Clark, *Elementary Mathematical Analysis.* Wadsworth Publishers of Canada, 1982.

[2] G. H. Hardy, *A course of Pure Mathematics.* Cambridge University Press, 1908.

[3] J. M. Howie, *Real analysis.* Springer-Verlag, 2001.

[4] S. Krantz, *Real Analysis and Foundations.* Second Edition. Chapman and Hall/CRC Press, 2005.

[5] W. Rudin, *Principles of Mathematical Analysis.* McGraw Hill, 2006.

[6] I. Stewart and D. Tall, *The Foundations of Mathematics.* Oxford University Press, 1977.

Notation

$\mathbb{N}$ *set of (positive) natural numbers* $\{1, 2, 3, \dots\}$

$\mathbb{Z}$ *set of integer numbers* $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$

$\mathbb{Z}_+$ *set of non-negative integers* $\{0, 1, 2, 3, \dots\}$

$\mathbb{Q}$ *set of rational numbers* $\{r = \frac{p}{q} \mid p \in \mathbb{Z},\ q \in \mathbb{N},\ \mathrm{hcf}(p, q) = 1\}$

$\mathbb{R}$ *set of real numbers*

$\mathbb{R}_+$ *set of non-negative reals*

Note: You may find a small inconsistency in the literature as to whether or not zero is regarded as a natural number; in these notes it is not.

Part I

Introduction to Analysis

Chapter 1

Elements of Logic and Set Theory

In mathematics we always assume that our propositions are definite and unambiguous, so that such propositions are always true or false (there is no intermediate option). In this respect they differ from propositions in ordinary life, which are often ambiguous or indeterminate. We also assume that our mathematical propositions are objective, so that they are determinately true or determinately false independently of our knowing which of these alternatives holds. We will denote propositions by capital Roman letters.

Examples of propositions

1. $A \equiv$ London is the capital of the UK.
2. $B \equiv$ Paris is the capital of the UK.
3. $C \equiv 3 > 5$.
4. $D \equiv 3 < 5$.

(We will use the sign $\equiv$ in order to define propositions.)
A and D are true, whereas B and C are false, nonetheless they are propositions. Thus with respect to our agreement a proposition may take one of the two values : "*true*" or "*false*" (never simultaneously). We use the symbol T for an arbitrary *true* proposition and F for an arbitrary *false* one.

Not every sentence is a mathematical proposition. For instance, the following sentences are *not* mathematical propositions.

(i) It is easy to study Analysis.

(ii) The number 1000000 is very large.

(iii) Is there a number whose square is 2 ?

(iv) $x > 2$.

It is impossible to judge whether (i) is true or false (it depends on many factors ...). (ii) does not have any precise sense. (iii) is a question, so does not state anything. (iv) contains a letter x and whether it is true or false depends on the value of x.

At the same time, there are propositions for which it is not immediately easy to establish whether they are true or false. For example,

$$E \equiv (126^{3728} + 15^{15876}) + 8 \text{ is a prime number.}$$

This is obviously a definite proposition, but it would take a lot of computation actually to determine whether it is true or false.

There are some propositions in mathematics for which their truth value has not been determined yet. For instance, "in the decimal representation of π there are infinitely many digits 7". It is not known whether this mathematical proposition is true. The truth value of this proposition constitutes an open question. There are plenty of open questions in mathematics!

1.1 Propositional connectives

Propositional connectives are used to combine simple propositions into complex ones. They can be regarded as operations with propositions.

1.1.1 Negation

One can build a new proposition from an old one by negating it. Take A above as an example. The negation of A (not A) will mean

$$\neg A \equiv \text{ London is not the capital of the UK.}$$

We will use the symbol $\neg A$ to denote not A. For the proposition $D \equiv \{8 \text{ is a prime number}\}$, its negation is $\neg D \equiv \{8 \text{ is not a prime number}\}$. Since we agreed that we have one of the two truth values, true or false, we can *define* negation of a proposition A by saying that if A is true then $\neg A$ is false, and if A is false then $\neg A$ is true. This definition is reflected in the following table.

A	$\neg A$
T	F
F	T

This is called the *truth table* for negation.

1.1.2 Conjunction

Conjunction is a binary operation on propositions which corresponds to the word "and" in English. We stipulate *by definition* that "A and B" is true if A is true and B is true, and "A and B" is false if A is false or B is false. This definition is expressed in the following truth table. We use the notation $A \wedge B$ for the conjunction "A and B".

A	B	$A \wedge B$
T	T	T
T	F	F
F	T	F
F	F	F

The four rows of this table correspond to the four possible truth combinations of the proposition A and the proposition B. The last entry in each row stipulates the truth or falsity of the complex proposition in question. Conjunction is sometimes called the *logical product*.

1.1.3 Disjunction

Disjunction is a binary operation on propositions which corresponds to the word "or" in English. We stipulate *by definition* that "A or B" is true if A is true or B is true, and "A or B" is false if A is false and B is false. This definition is expressed in the following truth table. We use the notation $A \vee B$ for the disjunction "A or B".

A	B	$A \vee B$
T	T	T
T	F	T
F	T	T
F	F	F

Disjunction is sometimes called the *logical sum.*

1.1.4 Implication

Implication is a binary operation on propositions which corresponds to the phrase "if... then..." in English. We will denote this operation using the symbol "$\Rightarrow$". So $A \Rightarrow B$ can be read "if A then B", or "A implies B". A is called the antecedent (or premise), and B is called the consequent (or conclusion). The truth table for implication is the following.

A	B	$A \Rightarrow B$
T	T	T
T	F	F
F	T	T
F	F	T

So the implication is false only in the case where the antecedent is true and the consequent is false. It is true in all remaining cases.

The last two lines in this truth table may appear mysterious. But consider the implication

$$x > 2 \Rightarrow x^2 > 4$$

We would like this proposition to be true for all values of x. So let's substitute a few values for x and work out the truth values of the antecedent and consequent.

x	$x > 2$	$x^2 > 4$	$x > 2 \Rightarrow x^2 > 4$
3	T	T	T
-3	F	T	T
0	F	F	T

This accounts for the choice of truth values in the last two lines of the truth table.

Notice that our notion of "implies" differs from that used in ordinary speech. It entails, for example, that the proposition

"Snow is black implies grass is red"

is true (since it corresponds to the last line of the truth table).

The proposition $A \Rightarrow B$ can be also read as "A is sufficient for B", "B if A", or "B is necessary for A" (the meaning of the latter will be clarified later on).

1.1.5 Equivalence

The last binary operation on propositions we introduce is *equivalence*. Saying that A is equivalent to B means that A is true whenever B is true, and vice versa. We denote this by $A \Leftrightarrow B$. So we stipulate that $A \Leftrightarrow B$ is true in the cases in which the truth values of A and B are the same. In the remaining cases it is false. This is shown by the following truth table.

A	B	$A \Leftrightarrow B$
T	T	T
T	F	F
F	T	F
F	F	T

$A \Leftrightarrow B$ can be read as "A is equivalent to B", or "A if and only if B" (this is usually shortened to "A iff B"), or "A is necessary and sufficient for B". The equivalence $A \Leftrightarrow B$ can be also defined by means of conjunction and implication as follows

$$(A \Rightarrow B) \wedge (B \Rightarrow A).$$

Now let A be a proposition. Then $\neg A$ is also a proposition, so that we can construct its negation $\neg\neg A$. Now it is easy to understand that $\neg\neg A$ is the same proposition as A as we have only two truth values T and F. Thus

$$\boxed{\neg\neg A \Leftrightarrow A}$$

The last equivalence is called the *double negation law.*

1.2 Logical laws

In our definitions in the previous section A, B, etc. stand for arbitrary propositions. They themselves may be built from simple propositions by means of the introduced operations.

Logical laws or, in other words, *logical tautologies* are composite propositions built from simple propositions $A, B, \ldots$ (operands) by means of the introduced operations, that are *true*, no matter what are the truth values of the operands $A, B, \ldots$

The truth value of a composite proposition constructed from operands $A, B, \ldots$ does not depend on the operands themselves, but only on their truth values. Hence in order to check whether a composite proposition is a law or not, one can plug in T or F instead of $A, B, \ldots$ using all possible combinations, then determine the corresponding truth values of the proposition in question. If the composite proposition turns out to be true whatever the truth values of $A, B, \ldots$ are, then the composite proposition is a law. If there is a substitution which gives the value F, then it is *not* a law.

Example 1.2.1. The proposition

$$(A \wedge B) \Rightarrow (A \vee B)$$

is a law.

The shortest way to justify this is to build the truth table for $(A \wedge B) \Rightarrow (A \vee B)$.

A	B	$A \wedge B$	$A \vee B$	$(A \wedge B) \Rightarrow (A \vee B)$
T	T	T	T	T
T	F	F	T	T
F	T	F	T	T
F	F	F	F	T

The last column consists entirely of T's, which means that this is a law.

Below we list without proof some of the most important logical laws. We recommend that you verify them by constructing their truth tables.

- Commutative law of disjunction

$$(A \vee B) \Leftrightarrow (B \vee A). \tag{1.2.1}$$

- Associative law of disjunction

$$[(A \vee B) \vee C] \Leftrightarrow [A \vee (B \vee C)]. \tag{1.2.2}$$

- Commutative law of conjunction

$$(A \wedge B) \Leftrightarrow (B \wedge A). \tag{1.2.3}$$

- Associative law of conjunction

$$[(A \wedge B) \wedge C] \Leftrightarrow [A \wedge (B \wedge C)]. \tag{1.2.4}$$

- First distributive law

$$[A \wedge (B \vee C)] \Leftrightarrow [(A \wedge B) \vee (A \wedge C)]. \tag{1.2.5}$$

- Second distributive law

$$[A \vee (B \wedge C)] \Leftrightarrow [(A \vee B) \wedge (A \vee C)]. \tag{1.2.6}$$

- Idempotent laws

$$(A \wedge A) \Leftrightarrow A, \ (A \vee A) \Leftrightarrow A. \tag{1.2.7}$$

- Absorption laws

$$(A \wedge T) \Leftrightarrow A, \quad (A \wedge F) \Leftrightarrow F,$$
$$(A \vee T) \Leftrightarrow T, \quad (A \vee F) \Leftrightarrow A. \tag{1.2.8}$$

- Syllogistic law

$$[(A \Rightarrow B) \wedge (B \Rightarrow C)] \Rightarrow (A \Rightarrow C). \tag{1.2.9}$$

-

$$(A \vee \neg A) \Leftrightarrow T. \tag{1.2.10}$$

-

$$(A \wedge \neg A) \Leftrightarrow F. \tag{1.2.11}$$

- De Morgan's laws

$$\neg(A \vee B) \Leftrightarrow (\neg A \wedge \neg B), \tag{1.2.12}$$
$$\neg(A \wedge B) \Leftrightarrow (\neg A \vee \neg B). \tag{1.2.13}$$

- Contrapositive law

$$(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A). \tag{1.2.14}$$

-

$$(A \Rightarrow B) \Leftrightarrow (\neg A \vee B). \tag{1.2.15}$$

1.3 Sets

The notion of a set is one of the basic notions. We cannot, therefore, give it a precise mathematical definition. Roughly speaking:

A set is a collection of objects, to which one can assign a "size".

This is obviously not a definition (for what is a "collection", or "size"?). A rigorous set theory would be constructed axiomatically. But we confine ourselves to naïve set theory. We shall take for granted the notion of a set, as well as the relation "to be an element of a set".

If A is a set we write $x \in A$ to mean "x is an element of the set A", or "x belongs to A". If x is not an element of A we write $x \notin A$. So the following is true for any x and A:

$$(x \notin A) \Leftrightarrow \neg(x \in A).$$

One way to define a set is just by listing its elements.

Example 1.3.1. (i) $A = \{0, 1\}$. This means that the set A consists of two elements, 0 and 1.

(ii) $B = \{0, \{1\}, \{0, 1\}\}$. The set B contains three elements: the number 0; the set $\{1\}$ containing one element, namely the number 1; and the set containing two elements, the numbers 0 and 1.

The order in which the elements are listed is irrelevant. Thus $\{0, 1\} = \{1, 0\}$. An element may occur more than once. So $\{1, 2, 1\} = \{1, 2\}$. But $\{1, 2, \{1\}\} \neq \{1, 2\}$!

A set can be also specified by an *elementhood test.*

Example 1.3.2. (i) $C = \{x \in \mathbb{N} \mid x \text{ is a prime number}\}$. The set C contains all primes. We cannot list them for the reason that there are infinitely many primes. (Here $\mathbb{N}$ is the set of positive natural numbers $1, 2, 3 \ldots$)

(ii) $D = \{x \in \mathbb{Z}_+ \mid x^2 - x = 0\}$. The set D contains the roots of the equation $x^2 - x = 0$. But these are 0 and 1, so the set D contains the same elements as the set A in the previous example. In this case we say that the sets A and D coincide, and write $A = D$.

Sets satisfy the following fundamental law (axiom).

If sets A and B contain the same elements, then they coincide (or are equal).

1.3.1 Subsets. The empty set

Definition 1.3.1. If each element of the set A is also an element of the set B we say that A is a subset of B and write $A \subseteq B$.

This definition can be expressed as a logical proposition as follows.

$$\text{For every } x, \ \ [(x \in A) \Rightarrow (x \in B)].$$

Note that by the definition

$$(A = B) \Leftrightarrow [(A \subseteq B) \wedge (B \subseteq A)].$$

Thus, one way to show that two sets, A and B, coincide is to show that each element in A is contained in B and vice-versa.
Since a set may contain no elements, the following definition is useful.

Definition 1.3.2. The empty set is the set which contains no elements, and is denoted by $\varnothing$.

There is only one empty set. This means, in particular, that the set of elephants taking this Analysis course coincides with the set of natural numbers solving the equation $x^2 - 2 = 0$!

Let us illustrate the notion of a subset by a simple example.

Example 1.3.3. Let $S = \{0, 1, 2\}$. Then the subsets of S are:

$$\varnothing, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}.$$

Subsets of a set A which do not coincide with A are called *proper* subsets. In this example all subsets except for the last one are proper.

Definition 1.3.3. The set of all subsets of a set A is called the power set of A, and is denoted by $P(A)$ (or 2^A in some books). So $P(A) = \{S | S \subseteq A\}$.

Note that in the last example S contains 3 elements, whereas $P(S)$ has 8 elements ($2^3 = 8$). In general, if a set A has n elements then $P(A)$ has 2^n elements (provided n is finite). Prove it !

1.3.2 Operations on sets

Now we introduce operations on sets. The main operations are: union, intersection, difference and symmetric difference.

Definition 1.3.4. The union of sets A and B is the set containing the elements of A and the elements of B, and no other elements.

We denote the union of A and B by $A \cup B$.

Note: existence of the union for arbitrary A and B is accepted as an *axiom*.

For arbitrary x and arbitrary A and B the following proposition is *true*.

$$(x \in A \cup B) \Leftrightarrow (x \in A) \vee (x \in B).$$

Definition 1.3.5. The intersection of sets A and B is the set containing the elements which are elements of both A and B, and no other elements.

We denote the intersection of A and B by $A \cap B$. Thus for arbitrary x and arbitrary A and B the following proposition is *true*.

$$(x \in A \cap B) \Leftrightarrow (x \in A) \wedge (x \in B).$$

Definition 1.3.6. The difference of sets A and B is the set containing the elements of A which do *not* belong to B.

We use the notation $A - B$ for the difference. The following is *true* for arbitrary x and arbitrary A and B :

$$(x \in A - B) \Leftrightarrow [(x \in A) \wedge (x \notin B)].$$

By the law $(P \Leftrightarrow Q) \Leftrightarrow (\neg P \Leftrightarrow \neg Q)$, de Morgan's law and the double negation law,

$$\neg(x \in A - B) \Leftrightarrow [\neg(x \in A) \vee (x \in B)],$$

and so

$$x \notin (A - B) \Leftrightarrow ((x \notin A) \vee (x \in B)).$$

Definition 1.3.7. The symmetric difference of the sets A and B is defined by

$$A \triangle B = (A - B) \cup (B - A).$$

Let us illustrate these operations with a simple example.

Example 1.3.4. Let $A = \{0, 1, 2, 3, 4, 5\}$ and $B = \{1, 3, 5, 7, 9\}$. Then

$$A \cup B = \{0, 1, 2, 3, 4, 5, 7, 9\}.$$

$$A \cap B = \{1, 3, 5\}.$$

$$A - B = \{0, 2, 4\}, B - A = \{7, 9\}.$$

$$A \triangle B = \{0, 2, 4, 7, 9\}.$$

Note that

$$A \cup B = (A \cap B) \cup (A \triangle B).$$

1.3.3 Laws for operations on sets

- Commutative laws

$$A \cup B = B \cup A, \tag{1.3.16}$$
$$A \cap B = B \cap A. \tag{1.3.17}$$

- Associative laws

$$A \cup (B \cup C) = (A \cup B) \cup C, \tag{1.3.18}$$
$$A \cap (B \cap C) = (A \cap B) \cap C. \tag{1.3.19}$$

- Distributive laws

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \tag{1.3.20}$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{1.3.21}$$

- Idempotent laws

$$A \cup A = A, \quad A \cap A = A. \tag{1.3.22}$$

The proofs of the above laws are based on the corresponding laws for conjunction and disjunction.

Now let us formulate and prove some laws involving the set difference operation.

$$A \cup (B - A) = A \cup B. \tag{1.3.23}$$

Proof.

$$\begin{aligned}[x \in A \cup (B - A)] &\Leftrightarrow \{(x \in A) \vee (x \in (B - A))\} \\ &\Leftrightarrow \{(x \in A) \vee [(x \in B) \wedge (x \notin A)]\} \\ &\Leftrightarrow \{(x \in A) \vee [(x \in B) \wedge \neg(x \in A)]\} \\ &\Leftrightarrow [(x \in A) \vee (x \in B)] \wedge [(x \in A) \vee \neg(x \in A)] \\ &\Leftrightarrow (x \in A) \vee (x \in B) \\ &\Leftrightarrow (x \in A \vee B)\end{aligned}$$

as $[(x \in A) \vee \neg(x \in A)]$ is true. □

From the last formula it follows that the difference is not an inverse operation to the union (in other words, that in general $A \cup (B - A) \neq B$).

$$A - B = A - (A \cap B). \tag{1.3.24}$$

Proof.

$$\begin{aligned}&[x \in A - (A \cap B)] \Leftrightarrow \{(x \in A) \wedge \neg(x \in A \cap B)\} \\ &\Leftrightarrow \{(x \in A) \wedge \neg[(x \in A) \wedge (x \in B)] \\ &\Leftrightarrow \{(x \in A) \wedge [\neg(x \in A) \vee \neg(x \in B)]\} \\ &\Leftrightarrow \{[(x \in A) \wedge \neg(x \in A)] \vee [(x \in A) \wedge \neg(x \in B)]\} \\ &\Leftrightarrow [(x \in A) \wedge \neg(x \in B)] \\ &\Leftrightarrow (x \in A - B)\end{aligned}$$

since $[(x \in A) \wedge \neg(x \in A)]$ is false. □

De Morgan's laws (for sets)

$$A - (B \cap C) = (A - B) \cup (A - C), \tag{1.3.25}$$
$$A - (B \cup C) = (A - B) \cap (A - C). \tag{1.3.26}$$

The proof is based on De Morgan's laws for propositions.

1.3.4 Universe. Complement

In many applications of set theory one considers only sets which are contained in some fixed set. (For example, in plane geometry we study only sets containing points in the plane.) This containing set is called the *universe* or *universal set*. We will denote it by $\mathcal{U}$.

Definition 1.3.8. Let $\mathcal{U}$ be the universe. The set $\mathcal{U} - \mathcal{A}$ is called the complement of A. It is denoted by A^c.

It is easy to see that the following properties hold

$$(A^c)^c = A, \tag{1.3.27}$$

$$(A \cap B)^c = A^c \cup B^c, \tag{1.3.28}$$

$$(A \cup B)^c = A^c \cap B^c. \tag{1.3.29}$$

Prove these properties.

1.4 Predicates and quantifiers

In mathematics, along with propositions, one deals with statements that depend on one or more *variables*, letters denoting elements of sets. In this case we speak of predicates. For instance, "n is a prime number" is a predicate. As we see, it may be true or false depending on the value n. A predicate becomes a proposition after substituting the variable by a fixed element from the set of definition (the set to which the variable belongs). Generally, a predicate can be written as

$$A(x)(x \in S),$$

where S is the set of definition (which we often omit when it is clear what S is). A subset of S containing all the elements of S which make $A(x)$ true is called the *truth set* for $A(x)$. For the truth set of $A(x)$ we write

$$\text{Truth set of } A = \{x \in S \,|\, A(x)\}.$$

Example 1.4.1. Let $A \equiv \{x \in \mathbb{Z} \,|\, x^2 - x = 0\}$. Then

$$\{x \in \mathbb{Z} \,|\, A(x)\} = \{0, 1\}.$$

We often want to say that some property holds for every element from S. In this case we use the universal quantifier $\forall$. So for "for all $x \in S$, $A(x)$" we write $(\forall x \in S)\ A(x)$. After applying the universal quantifier to a predicate we obtain a proposition (which may be true or false, as usual). The universal quantifier $\forall$ substitutes for the words "every", "for every", "any", "for all".

Example 1.4.2. (i) The proposition "Every integer has a non-negative square" can be written as

$$(\forall x \in \mathbb{Z})\ [x^2 \geq 0].$$

(ii) The proposition "Every integer is non-negative" can be written as

$$(\forall x \in \mathbb{Z})\ [x \geq 0].$$

Evidently, (i) is true and (ii) is false.

Note that the proposition $[(\forall x \in S)\ A(x)]$ means that the truth set of $A(x)$ is the whole set S. Thus if for an element x of S, $A(x)$ is false, the proposition $[(\forall x \in S)\ A(x)]$ is false, hence in order to state that it is false it is enough to find one element in S for which $A(x)$ is false.

To state that a property holds for some element of S, in other words, *"there exists an element in S such that $A(x)$ holds"*, we use the *existential quantifier* $\exists$ and write

$$(\exists x \in S)\ A(x).$$

$\exists$ substitutes for the words "for some", "there exists".

Note that in order to establish that $[(\exists x \in S)\ A(x)]$ is true it is enough to find one element in S for which $A(x)$ is true.

Example 1.4.3. The proposition "Some integers are smaller than their squares" can be written as

$$(\exists x \in \mathbb{Z})\ [x < x^2].$$

It is true, of course.

The propositions "$(\forall x)\ P(x)$" and "$(\exists x)\ P(x)$", $P(x)$ may themselves contain quantifiers.

Example 1.4.4. (i)

$$(\forall x \in \mathbb{N})(\exists y \in \mathbb{N})\ [y = x + x].$$

(ii)

$$(\forall x \in \mathbb{Z})(\exists y \in \mathbb{Z})\ [y < x].$$

We have the following laws.

$$\neg[(\forall x \in S)\, P(x)] \Leftrightarrow [(\exists x \in S)\, \neg P(x)],$$

$$\neg[(\exists x \in S)\, P(x)] \Leftrightarrow [(\forall x \in S)\, \neg P(x)].$$

This means that we can negate a predicate with an initial quantifier by first changing $\forall$ to $\exists$ and $\exists$ to $\forall$ then negating the predicate.

Example 1.4.5. Consider the following proposition

$$(\forall x \in \mathbb{Z}_+)\, (\exists y \in \mathbb{Z}_+)\, (2y > x).$$

The negation of this is

$$(\exists x \in \mathbb{Z}_+)\, (\forall y \in \mathbb{Z}_+)\, (2y \leq x).$$

Example 1.4.6.

$$\begin{aligned} &\neg[(\forall x \in X)(\exists y \in Y)(\forall z \in Z)\, P(x, y, z)] \\ \Leftrightarrow\ &[(\exists x \in X)(\forall y \in Y)(\exists z \in Z)\, \neg P(x, y, z)]. \end{aligned}$$

1.5 Ordered pairs. Cartesian products

Let us talk about sets from a universe $\mathcal{U}$. Recall that $\{a\}$ denotes the set containing one element a. The set $\{a, b\}$ contains two elements if $a \neq b$ and one element otherwise. Obviously, $\{a, b\} = \{b, a\}$. In many problems in mathematics we need an object in which the order in a pair is important. So we want to define an *ordered pair*.

You have already seen ordered pairs studying points in the xy plane. The use of x and y coordinates to identify points in the plane works by assigning to each point in the plane an ordered pair of real numbers x and y. The pair must be ordered because, for example, $(2, 5)$ and $(5, 2)$ correspond to different points.

Definition 1.5.1.

$$(a, b) = \{\{a\}, \{a, b\}\}.$$

Theorem 1.5.1.

$$[(a, b) = (x, y)] \Leftrightarrow [(a = x) \wedge (b = y)].$$

Proof. $(\Leftarrow)$ Suppose $a = x$ and $b = y$. Then

$$(a, b) = \{\{a\}, \{a, b\}\} = \{\{x\}, \{x, y\}\} = (x, y).$$

$(\Rightarrow)$ Suppose that $(a, b) = (x, y)$, i.e.

$$\{\{a\}, \{a, b\}\} = \{\{x\}, \{x, y\}\}.$$

There are two cases to consider.

Case 1. $a = b$. Then $(a, b) = \{\{a\}\}$. Since $(a, b) = (x, y)$ both sets (a, b) and (x, y) are both sets with one element. (A set with one element is called a singleton.) So $x = y$. Since $\{x\} \in (a, b) = \{\{a\}\}$ we have that $\{x\} = \{a\}$ and so $a = b = x = y$.

Case 2. $a \neq b$. Then both (a, b) and (x, y) contain exactly one singleton namely $\{a\}$ and $\{x\}$ respectively. So $a = x$. Since in this case it is also true that both (a, b) and (x, y) contain exactly one unordered pair namely $\{a, b\}$ and $\{x, y\}$ we conclude that $\{a, b\} = \{x, y\}$. Since b cannot be equal to x (for then we should have $a = x$ and $b = x$ and therefore $a = b$) we must have $b = y$. $\square$

Now we define the *Cartesian product*, which is an essential tool for further development.

Definition 1.5.2. Let A and B be sets. The *Cartesian product of A and B, denoted by $A \times B$, is the set of all ordered pairs (a, b) in which $a \in A$ and $b \in B$, i.e.*

$$A \times B = \{(a, b) \mid (a \in A) \wedge (b \in B)\}.$$

Thus

$$(p \in A \times B) \Leftrightarrow \{(\exists a \in A)(\exists b \in B)\ [p = (a, b)]\}.$$

Example 1.5.1. (i) If $A =$\{red, green\} and $B = \{1, 2, 3\}$ then

$$A \times B = \{(\text{red}, 1), (\text{red}, 2), (\text{red}, 3), (\text{green}, 1), (\text{green}, 2), (\text{green}, 3)\}.$$

(ii) $\mathbb{Z} \times \mathbb{Z} = \{(x, y) \mid x \text{ and } y \text{ are integers}\}$. This is the set of integer coordinates points in the x, y-plane. The notation $\mathbb{Z}^2$ is usually used for this set.

Remark 1.5.1. $X \times X$ is sometimes called the Cartesian square of X.

The following theorem provides some basic properties of the Cartesian product.

Theorem 1.5.2. *Let A, B, C, D be sets. Then*

$$A \times (B \cap C) = (A \times B) \cap (A \times C), \tag{1.5.30}$$
$$A \times (B \cup C) = (A \times B) \cup (A \times C), \tag{1.5.31}$$
$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D), \tag{1.5.32}$$
$$(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D), \tag{1.5.33}$$
$$A \times \varnothing = \varnothing \times A = \varnothing. \tag{1.5.34}$$

Proof of (1.5.30). ($\Rightarrow$) Let $p \in A \times (B \cap C)$. Then

$$(\exists a \in A)(\exists x \in B \cap C)\ [p = (a, x)]$$

In particular,

$$(\exists a \in A)(\exists x \in B)\ [p = (a, x)] \text{ and } (\exists a \in A)(\exists x \in C)\ [p = (a, x)]$$

So $p \in (A \times B) \cap (A \times C)$.
($\Leftarrow$) Let $p \in (A \times B) \cap (A \times C)$. Then $p \in (A \times B)$ and $p \in (A \times C)$. So

$$(\exists a \in A)(\exists b \in B)\ [p = (a, b)] \text{ and } (\exists a' \in A)(\exists c \in C)\ [p = (a', c)]$$

But then $(a, b) = p = (a', c)$ and hence $a = a'$ and $b = c$. Thus $p = (a, x)$ for some $a \in A$ and $x \in B \cap C$, i.e. $p \in A \times (B \cap C)$.
This proves (1.5.30). □

The proof of (1.5.31), (1.5.32), (1.5.33) and (1.5.34) are left as exercises.

1.6 Relations

Definition 1.6.1. Let X, Y be sets. A set $R \subseteq X \times Y$ is called a relation from X to Y.

If $(x, y) \in R$, we say that x is in the relation R to y. We will also write in this case xRy.

Example 1.6.1. (i) Let $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$. The set $R = \{(1, 3), (1, 5), (3, 3)\}$ is a relation from A to B since $R \subseteq A \times B$.

(ii) $G = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid x > y\}$ is a relation from $\mathbb{Z}$ to $\mathbb{Z}$.

Definition 1.6.2. Let R be a relation from X to Y. The *domain of R is the set*

$$D(R) = \{x \in X \mid \exists y \in Y\ [(x, y) \in R]\}.$$

The range of R is the set

$$Ran(R) = \{y \in Y \mid \exists x \in X\ [(x, y) \in R]\}.$$

The inverse of R is the relation R^{-1} from Y to X defined as follows

$$R^{-1} = \{(y, x) \in Y \times X \mid (x, y) \in R\}.$$

Definition 1.6.3. Let R be a relation from X to Y, S be a relation from Y to Z. The *composition* of S and R is a relation from X to Z defined as follows

$$S \circ R = \{(x, z) \in X \times Z \,|\, \exists y \in Y \,[(x, y) \in R] \wedge [(y, z) \in S]\}.$$

Theorem 1.6.1. *Let R be a relation from X to Y, S be a relation from Y to Z, T be a relation from Z to V. Then*

1. $(R^{-1})^{-1} = R$.
2. $D(R^{-1}) = Ran(R)$.
3. $Ran(R^{-1}) = D(R)$.
4. $T \circ (S \circ R) = (T \circ S) \circ R$.
5. $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$.

For the proof see Stewart and Tall (1977).

Next we take a look at some particular types of relations. Let us consider relations from X to X, i.e. subsets of $X \times X$. In this case we talk about relations on X. A simple example of such a relation is the *identity relation* on X which is defined as follows

$$i_X = \{(x, y) \in X \times X \,|\, x = y\}.$$

Definition 1.6.4. 1. A relation R on X is said to be *reflexive* if

$$(\forall x \in X)\,(x, x) \in R.$$

2. R is said to be *symmetric* if

$$(\forall x \in X)(\forall y \in X)\,\{[(x, y) \in R] \Rightarrow [(y, x) \in R]\}.$$

3. R is said to be *transitive* if

$$(\forall x \in X)(\forall y \in X)(\forall z \in X)\{[((x, y) \in R) \wedge ((y, z) \in R)] \Rightarrow [(x, z) \in R]\}.$$

A particularly important class of relations are *equivalence relations.*

Definition 1.6.5. A relation R on X is called *equivalence relation* if it is reflexive, symmetric and transitive.

Example 1.6.2. (i) Let X be a set of students. A relation on $X \times X$: "to be friends".

It is reflexive (I presume that everyone is a friend to himself/herself). It is symmetric. But it's not transitive.

(ii) Let $X = \mathbb{Z}$, $a \in \mathbb{N}$. Define $R \subseteq X \times X$ as

$$R = \{(x, y) \,|\, |x - y| \leq a\}.$$

R is reflexive, symmetric, but not transitive.

(iii) Let $X = \mathbb{Z}$, $m \in \mathbb{N}$. Define the *congruence* mod m on $X \times X$ as follows:

$$x \equiv y \text{ if } (\exists k \in \mathbb{Z} \mid x - y = km).$$

This is an equivalence relation on X.

Definition 1.6.6. Let R be an equivalence relation on X. Let $x \in X$. The equivalence class of x with respect to R is the set

$$[x]_R = \{y \in X \mid (y, x) \in R\}.$$

Let us take a look at several properties of equivalence classes.

Proposition 1.6.1. *Let R be an equivalence relation on X. Then*

1. $(\forall x \in X)\ x \in [x]_R$.
2. $(\forall x \in X)(\forall y \in X)\ [(y \in [x]_R) \Leftrightarrow ([y]_R = [x]_R)]$.

Proof. 1. Since R is reflexive, $(x, x) \in R$, hence $x \in [x]_R$.
2. First, let $y \in [x]_R$. (a) Suppose that $z \in [y]_R$ (an arbitrary element of $[y]_R$. Then $(x, y) \in R$, $(y, z) \in R$ and by transitivity $(x, z) \in R$. Therefore $z \in [x]_R$, which shows that $[y]_R \subset [x]_R$. (b) Suppose that $z \in [x]_R$. Similarly show that $z \in [y]_R$.
Therefore $[x]_R = [y]_R$.
The implication $([y]_R = [x]_R) \Rightarrow (y \in [x]_R)$ is obvious. □

From the above proposition it follows that the classes of equivalence are disjoint and every element of the set X belongs to a class of equivalence (the union of the classes equals to the set X).

Remark 1.6.1. See more on this in Stewart and Tall (1977).

1.7 Functions

1.7.1 On the definition of functions

The notion of a function is of fundamental importance in all branches of mathematics. For the purposes of this course, by a function F from a set X to a set Y, we mean a "rule", which assigns one and only one value $y \in Y$ to each value $x \in X$. For this y we write $y = F(x)$. This y is called the image of x under F. Of course, this is not a proper definition because we have not defined what is meant by a "rule", for instance. A proper definition follows below in order for this set of notes to be complete, but it will be dealt with in the course Further Topics in Analysis.

Definition 1.7.1. Let X and Y be sets. Let F be a relation from X to Y. Then F is called a *function if the following properties are satisfied*

(i) $(\forall x \in X)(\exists y \in Y)\,[(x, y) \in F]$.

(ii) $(\forall x \in X)(\forall y \in Y)(\forall z \in Y)\{([(x,y) \in F] \wedge [(x,z) \in F]) \Rightarrow (y = z)\}$.
(In other words, for every $x \in X$ there is only one *$y \in Y$ such that $(x,y) \in F$).*

X is called the domain of F and Y is called codomain. It is customary to write

$$F : X \to Y, \; y = F(x).$$

Let us consider several examples.

Example 1.7.1. (i) Let $X = \{1,2,3\}$, $Y = \{4,5,6\}$. Define $F \subseteq X \times Y$ as

$$F = \{(1,4),(2,5),(3,5)\}.$$

Then F is a function.
(In contrast to that define $G \subseteq X \times Y$ as

$$G = \{(1,4),(1,5),(2,6),(3,6)\}.$$

Then G is *not* a function.)
(ii) Let $X = \mathbb{Z}$ and $Y = \mathbb{Z}$. Define $F \subseteq X \times Y$ as

$$F = \{(x,y) \in \mathbb{Z} \times \mathbb{Z} \,|\, y = x^2\}.$$

Then F is a function from $\mathbb{Z}$ to $\mathbb{Z}$.
(In contrast to that define $G \subseteq X \times Y$ as

$$G = \{(x,y) \in \mathbb{Z} \times \mathbb{Z} \,|\, x^2 + y^2 = 25\}.$$

Then G is *not* a function.)

Theorem 1.7.1. *Let X, Y be sets, F, G be functions from X to Y. Then*

$$[(\forall x \in X)(F(x) = G(x))] \Leftrightarrow (F = G).$$

Proof. $(\Rightarrow)$. Let $x \in X$. Then $(\exists y \in Y)(y = F(x))$. But $G(x) = F(x)$, so $(x,y) \in G$. Therefore $F \subseteq G$. Analogously one sees that $G \subseteq F$.
2. $(\Leftarrow)$. Obvious since the sets F and G are equal. □

The above theorem says that in order to establish that two functions are equal one has to establish that they have the same domain and codomain and for every element of the domain they have equal images. Note that equal function may be defined by different "rules".

Example 1.7.2. Let $f : \mathbb{Z} \to \mathbb{Z}$, $g : \mathbb{Z} \to \mathbb{Z}$, $h : \mathbb{Z} \to \mathbb{Z}_+$. Let $\forall x \in \mathbb{Z}$

$$f(x) = (x+1)^2, \quad g(x) = x^2 + 2x + 1, \quad h(x) = (x+1)^2.$$

Then f and g are equal, but f and h are not since they have different codomains.

Let $f : X \to Y$ and $g : Y \to Z$. Then we can define the composition of g with f, which we denote as $g \circ f : X \to Z$, by

$$(\forall x \in X)\ [(g \circ f)(x) = g(f(x))].$$

The definition of composition of relations can be also applied to functions. If $f : X \to Y$ and $g : Y \to Z$ then

$$g \circ f = \{(x, z) \in X \times Z \mid (\exists y \in Y)[(x, y) \in f] \wedge [(y, z) \in g]\}.$$

Theorem 1.7.2. *Let $f : X \to Y$ and $g : Y \to Z$. Then $g \circ f : X \to Z$ and*

$$(\forall x \in X)\ [(g \circ f)(x) = g(f(x))].$$

Proof. We know that $g \circ f$ is a relation. So we must prove that for every $x \in X$ there exists a unique element $z \in Z$ such that $(x, z) \in g \circ f$.
Existence: Let $x \in X$ be arbitrary. Then $\exists y \in Y$ such that $y = f(x)$, or in other words, $(x, y) \in f$. Also $\exists z \in Z$ such that $z = g(y)$, or in other words, $(y, z) \in g$. By the definition it means that $(x, z) \in g \circ f$. Moreover, we see that

$$(g \circ f)(x) = g(f(x)).$$

Uniqueness: Suppose that $(x, z_1) \in g \circ f$ and $(x, z_2) \in g \circ f$. Then by the definition of composition $(\exists y_1 \in Y)\,[(x, y_1) \in f] \wedge [(y_1, z_1) \in g]$ and $(\exists y_2 \in Y)\,[(x, y_2) \in f] \wedge [(y_2, z_2) \in g]$. But f is a function. Therefore $y_1 = y_2$. g is also a function, hence $z_1 = z_2$. □

Example 1.7.3. Let $f : \mathbb{Z} \to \mathbb{Z}$, $g : \mathbb{Z} \to \mathbb{Z}$,

$$f(x) = x^2 + 2, \quad g(x) = 2x - 1.$$

Find $(f \circ g)(x)$ and $(g \circ f)(x)$.

Solution.

$$\begin{aligned}(f \circ g)(x) = f(g(x)) = g(x)^2 + 2 = 4x^2 - 4x + 3,\\ (g \circ f)(x) = g(f(x)) = 2f(x) - 1 = 2x^2 + 3.\end{aligned}$$

As you clearly see from the above, $f \circ g \neq g \circ f$ in general.

Definition 1.7.2. Let $f : X \to Y$ and $A \subseteq X$. The *image of A under f is the set*

$$f(A) = \{f(x) \mid x \in A\}.$$

Example 1.7.4. Let $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = x^2$. Let $A = \{x \in \mathbb{Z} \mid 0 \leq x \leq 2\}$. Then $f(A) = \{0, 1, 4\}$.

The following theorem (which we give without proof) establishes some properties of images of sets.

Theorem 1.7.3. *Let $f : X \to Y$ and $A \subset X$, $B \subset X$. Then*

(i) $f(A \cup B) = f(A) \cup f(B)$,

(ii) $f(A \cap B) \subset f(A) \cap f(B)$,

(iii) $(A \subset B) \Rightarrow [f(A) \subset f(B)]$.

Remark 1.7.1. Note that in (ii) there is no equality in general. Consider the following example. $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = x^2$. Let $A = \{-2, -1, 0\}$ and $B = \{0, 1, 2\}$. Then $f(A) = \{0, 1, 4\}$, $f(B) = \{0, 1, 4\}$, so that $f(A) \cap f(B) = \{0, 1, 4\}$. At the same time $A \cap B = \{0\}$, and hence $f(A \cap B) = \{0\}$.

Definition 1.7.3. Let $f : X \to Y$. The set $Ran(f) = f(X)$ is called the *range of* f. *We also have that*

$$Ran(f) = \{y \in Y \mid (\exists x \in X)(f(x) = y)\}.$$

Example 1.7.5. Let $f : \mathbb{Z} \to \mathbb{Z}$, $f(x) = x^2 - 4x$. Then $Ran(f) = \{x^2 - 4x | x \in \mathbb{Z}\}$.

Definition 1.7.4. Let $f : X \to Y$ and $B \subset Y$. The *inverse image of* B *under* f *is the set*

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}.$$

Example 1.7.6. Let $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = x^2$. Let $B = \{y \in \mathbb{Z} | y \leq 10\}$. Then $f^{-1}(B) = \{-3, -2, -1, 0, 1, 2, 3\}$.

The following theorem (which we give without proof) establishes some properties of inverse images of sets.

Theorem 1.7.4. *Let $f : X \to Y$ and $A \subset Y$, $B \subset Y$. Then*

(i) $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$,

(ii) $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$,

(iii) $(A \subset B) \Rightarrow [f^{-1}(A) \subset f^{-1}(B)]$.

Remark 1.7.2. Note the difference with the previous theorem.

1.7.2 Injections and surjections. Bijections. Inverse functions

In the last section we saw that the composition of two functions is again a function. If $f : X \to Y$ is a relation from X to Y, one can define the inverse relation f^{-1}. Then the question arises: Is f^{-1} a function? In general the answer is "no". In this section we will study particular cases of functions and find out when the answer is "yes".

Definition 1.7.5. Let $f : X \to Y$. Then f is called an *injection if*

$$(\forall x_1 \in X)(\forall x_2 \in X)[(f(x_1) = f(x_2)) \Rightarrow (x_1 = x_2)].$$

The above definition means that f sets up a one-to-one correspondence between X and $Ran(f)$. Using the contrapositive law one can rewrite the above definition as follows.

$$(\forall x_1 \in X)(\forall x_2 \in X)[(x_1 \neq x_2) \Rightarrow (f(x_1) \neq f(x_2))].$$

Example 1.7.7. (i) Let $X = \{1, 2, 3\}$, $Y = \{4, 5, 6, 7\}$. Define $f : X \to Y$ as

$$f = \{(1, 5), (2, 6), (3, 7)\}.$$

Then f is an injection.
In contrast define $g : X \to Y$ as

$$g = \{(1, 5), (2, 6), (3, 5)\}.$$

Then g is *not* an injection since $g(1) = g(3)$.
(ii) Let $f : \mathbb{N} \to \mathbb{N}$ defined by $f(n) = n^2$. Then f is an injection.
In contrast let $g : \mathbb{Z} \to \mathbb{Z}$ defined by $g(n) = n^2$. Then g is *not* an injection since, for instance, $g(1) = g(-1)$.

Definition 1.7.6. Let $f : X \to Y$. Then f is called a *surjection if*

$$(\forall y \in Y)(\exists x \in X)[f(x) = y].$$

The above definition means that $Ran(f) = Y$. For this reason surjections are sometimes said to be *onto*.

Example 1.7.8. (i) Let $X = \{1, 2, 3, 4\}$, $Y = \{5, 6, 7\}$. Define $f : X \to Y$ as

$$f = \{(1, 5), (2, 6), (3, 7), (4, 6)\}.$$

Then f is an surjection.
In contrast define $g : X \to Y$ as

$$g = \{(1, 5), (2, 6), (3, 5), (4, 6)\}.$$

Then g is *not* a surjection since 7 is not in its range.
(ii) Let $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = n + 2$. Then f is a surjection.
In contrast let $g : \mathbb{Z} \to \mathbb{Z}$ defined by $g(n) = n^2$. Then g is *not* a surjection since, for instance, there is no integer such that its square is 2.

Definition 1.7.7. Let $f : X \to Y$. Then f is called a *bijection if it is both an injection and a surjection.*

Example 1.7.9. (i) Let $X = \{1, 2, 3\}$ and $Y = \{4, 5, 6\}$. Define $f : X \to Y$ by

$$f = \{(1, 4), (2, 5), (3, 6)\}.$$

Then f is a bijection.
(ii) Let $X = Y$ be a non-empty set. Then i_X is a bijection.

Definition 1.7.8. Let $f : X \to Y$. If a function $f^{-1} : Y \to X$ exists, satisfying

$$f^{-1} \circ f = i_X \text{ and } f \circ f^{-1} = i_Y,$$

then this function f^{-1} is called the *inverse function and we say that f is invertible.*

Theorem 1.7.5. *Let $f : X \to Y$. Suppose functions $g : Y \to X$ and $h : Y \to X$ exist, satisfying*

$$g \circ f = i_X \textit{ and } f \circ h = i_Y,$$

Then $g = h = f^{-1}$.

Proof. For all $y \in Y$, $g(y) = g(f(h(y))) = h(y)$. Since g and h also have the same domain and codomain, they are equal and also equal to the inverse function by Definition 1.7.8. □

Corollary 1.7.1. *If it exists, the inverse function is unique.*

Proof. Suppose f_1 and f_2 are two inverse functions of a function f. Then by Definition 1.7.8, we have

$$f_1 \circ f = i_X \text{ and } f \circ f_2 = i_Y.$$

But then Theorem 1.7.5 implies $f_1 = f_2$, proving the uniqueness of the inverse function. □

Theorem 1.7.6. *Let $f : X \to Y$. Then*

$$(\textit{the inverse function}\, f^{-1} : Y \to X\, \textit{exists}) \Leftrightarrow (f \textit{ is a bijection}).$$

Proof. $(\Rightarrow)$. Since $f \circ f^{-1} = i_Y$,

$$Y = Ran(i_Y) = Ran(f \circ f^{-1}) \subseteq Ran(f).$$

Also $Ran(f) \subseteq Y$. Hence, $Ran(f) = Y$, i.e. f is a surjection. To prove that f is an injection, let $y_1 = f(x_1)$, $y_2 = f(x_2)$ and $y_1 = y_2$. Then

$$x_1 = f^{-1} \circ f(x_1) = f^{-1}(y_1) = f^{-1}(y_2) = f^{-1} \circ f(x_2) = x_2.$$

Hence f is also an injection and, therefore, it is a bijection.
$(\Leftarrow)$. If f is a bijection then we can define a function $f^{-1} : Y \to X$ by

$$f^{-1}(y) = x \Leftrightarrow f(x) = y.$$

That is f^{-1} is the function that assigns to every $y \in Y$, the unique $x \in X$ satisfying $f(x) = y$. This x exists because f is a surjection and is unique because f is also an injection. With this definition of f^{-1},

$$f^{-1} \circ f(x) = f^{-1}(y) = x,$$

i.e. $f^{-1} \circ f = i_X$, and

$$f \circ f^{-1}(y) = f(x) = y,$$

i.e. $f \circ f^{-1} = i_Y$, proving that this f^{-1} is indeed the (unique) inverse function. □

Example 1.7.10. Let $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = ax + b$, where $b \in \mathbb{Z}$ and $a \in \mathbb{Z} - \{0\}$. Then f is an injection, and $g : \mathbb{Z} \to Ran(f)$ is a bijection.

Proof. Let $x_1, x_2 \in \mathbb{Z}$ be arbitrary and $ax_1 + b = ax_2 + b$. Then, of course, $ax_1 = ax_2$, and $x_1 = x_2$ since $a \neq 0$. So f is an injection. Since $(\forall x \in \mathbb{Z})[f(x) = g(x)]$ $Ran(f) = Ran(g)$. Hence g is onto its own range. So g is surjective. As before g is injective. Hence g is bijective. □

1.8 Some methods of proof. Proof by induction

1. First we discuss a couple of widely used methods of proof: contrapositive proof and proof by contradiction.

The idea of *contrapositive proof* is the following equivalence

$$\boxed{(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A).}$$

So to prove that $A \Rightarrow B$ is true is the same as to prove that $\neg B \Rightarrow \neg A$ is true.

Example 1.8.1. For integers m and n, if mn is odd then so are m and n.

Proof. We have to prove that $(\forall m, n \in \mathbb{Z}_+)$

$$(mn \text{ is odd}) \Rightarrow [(m \text{ is odd}) \wedge (n \text{ is odd})],$$

which is the same as to prove that

$$[(m \text{ is even}) \vee (n \text{ is even})] \Rightarrow (mn \text{ is even}).$$

The latter is evident. □

The idea of *proof by contradiction* is the following equivalence

$$\boxed{(A \Rightarrow B) \Leftrightarrow (\neg A \vee B) \Leftrightarrow \neg(A \wedge \neg B).}$$

So to prove that $A \Rightarrow B$ is true is the same as to prove that $\neg A \vee B$ is true or else that $A \wedge \neg B$ is false.

Example 1.8.2. Let $x, y \in \mathbb{R}$ be positive. If $x^2 + y^2 = 25$ and $x \neq 3$, then $y \neq 4$.

Proof. In order to prove by contradiction we assume that

$$[(x^2 + y^2 = 25) \wedge (x \neq 3)] \wedge (y = 4).$$

Then $x^2 + y^2 = x^2 + 16 = 25$. Hence $(x = 3) \wedge (x \neq 3)$ which is a contradiction.

2. In the remaining part of this short section we will discuss an important property of natural numbers. The set of natural numbers

$$\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}$$

which we will always denote by $\mathbb{N}$, is taken for granted. We shall denote the set of positive natural numbers $\mathbb{N} = \{1, 2, 3, 4, \dots\}$.
The *Principle of Mathematical Induction* is often used when one needs to prove statements of the form

$$\boxed{(\forall n \in \mathbb{N})\ P(n)}$$

or similar types of statements.

Since there are infinitely many natural numbers we cannot check one by one that they all have property P. The idea of mathematical induction is that to list all natural numbers one has to start from 1 and then repeatedly add 1. Thus one can show that 1 has property P and that whenever one adds 1 to a number that has property P, the resulting number also has property P.

Principle of Mathematical Induction. *If for a statement $P(n)$*

(*i*) *$P(1)$ is true,*

(*ii*) *$[P(n) \Rightarrow P(n+1)]$ is true,*

then $(\forall n \in \mathbb{N})\, P(n)$ is true.

Part (i) is called the *base case*; (ii) is called the *induction step*.

Example 1.8.3. Prove that $\forall n \in \mathbb{N}$:

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Solution. Base case: $n = 1$. $1^2 = \frac{1 \cdot 2 \cdot 3}{6}$ is true.
Induction step: Suppose that the statement is true for $n = k\,(k \geqslant 1)$. We have to prove that it is true for $n = k + 1$. So our assumption is

$$1^2 + 2^2 + 3^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}.$$

Therefore we have

$$1^2 + 2^2 + 3^2 + \cdots + k^2 + (k+1)^2 = \frac{k(k+1)(2k+1)}{6} + (k+1)^2 = \frac{(k+1)(k+2)(2k+3)}{6},$$

which proves the statement for $n = k + 1$. By the principle of mathematical induction the statement is true for all $n \in \mathbb{N}$. $\square$

Example 1.8.4. $1 + 3 + 5 + \cdots + (2n - 1) =?$

Solution. First we have to work out a conjecture. For this let us try several particular cases:

$$n = 1;\quad 1 = 1;$$
$$n = 2;\quad 1 + 3 = 4 = 2^2;$$
$$n = 3;\quad 1 + 3 + 5 = 9 = 3^2.$$

So a reasonable conjecture is that $\forall n \in \mathbb{N}$,

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2.$$

We have already established the base case, so let's move on to the induction step.

Induction step: Let

$$1 + 3 + 5 + \cdots + (2k - 1) = k^2\ (k \geqslant 1).$$

Then

$$1 + 3 + 5 + \cdots + (2k - 1) + (2k + 1) = k^2 + 2k + 1 = (k + 1)^2.$$

This completes the proof by induction. □

The base case need not start with $n = 1$. It can start from any integer onwards.

Example 1.8.5. Prove that $(\forall n \in \mathbb{Z}_+)(3|(n^3 - n))$.

Solution. Base case: $n = 0$. $(3|0)$ is true.
Induction step: Assume that $(3|(k^3 - k))$ for $k \geqslant 0$. Then

$$(k + 1)^3 - (k + 1) = k^3 + 3k^2 + 3k + 1 - k - 1 = \underbrace{(k^3 - k)}_{\text{divisible by 3}} + \underbrace{3(k^2 + k)}_{\text{divisible by 3}},$$

which completes the proof. □

Example 1.8.6. Prove that $(\forall n \in \mathbb{Z}_+)[(n \geqslant 5) \Rightarrow (2^n > n^2)]$.

Solution. Base case: $n = 5$. True.
Induction step: Suppose that $2^k > k^2\ (k \geqslant 5)$. Then

$$2^{k+1} = 2 \cdot 2^k > 2k^2.$$

Now it is sufficient to prove that

$$2k^2 > (k + 1)^2.$$

Consider the difference:

$$2k^2 - (k + 1)^2 = k^2 - 2k - 1 = (k - 1)^2 - 2.$$

Since $k \geqslant 5$ we have $k-1 \geqslant 4$ and $(k-1)^2-2 \geqslant 14 > 0$, which proves the above inequality. □

Theorem 1.8.1. (The Binomial Theorem) *Let $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$. Then*

$$\begin{aligned}(a+b)^n &= a^n + \frac{n}{1}a^{n-1}b + \frac{n(n-1)}{2.1}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{3.2.1}a^{n-3}b^3 + \dots \\ &+ \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}a^{n-k}b^k + \dots + \frac{n(n-1)(n-2)\cdots 2}{(n-1)(n-2)\cdots 2.1}ab^{n-1} + b^n.\end{aligned}$$

Before proving this theorem it is convenient to introduce some notation. For $n \in \mathbb{N}$ we define $n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$, and $0! = 1$. If $a_0, a_1, a_2, ..., a_n$ are real numbers then

$$\sum_{k=0}^{n} a_k = a_1 + a_2 + ... + a_n.$$

The k-th *binomial coefficient* is denoted by the symbol $\left(\begin{array}{c} n \\ k \end{array}\right)$, and is defined by

$$\left(\begin{array}{c} n \\ k \end{array}\right) = \frac{n!}{(n-k)!\; k!}.$$

Note that $\left(\begin{array}{c} n \\ 0 \end{array}\right) = \left(\begin{array}{c} n \\ n \end{array}\right) = 1$, and for all $k \in \mathbb{Z}_+$ with $k \le n$, $\left(\begin{array}{c} n \\ k \end{array}\right) = \left(\begin{array}{c} n \\ n-k \end{array}\right)$. The Binomial Formula can now be rewritten as

$$(a+b)^n = \sum_{k=0}^{n} a^k b^{n-k} \left(\begin{array}{c} n \\ k \end{array}\right),$$

The following identity will be needed in the proof of the Binomial Theorem.

Lemma 1.8.1. $(\forall n \in \mathbb{N})(\forall k \in \mathbb{Z}_+)\left[(k \le n-1) \Rightarrow \left(\begin{array}{c} n \\ k+1 \end{array}\right) + \left(\begin{array}{c} n \\ k \end{array}\right) = \left(\begin{array}{c} n+1 \\ k+1 \end{array}\right)\right].$

Proof. Since k and n are natural numbers which satisfy $k \le n-1, n \ge 1$ the binomial coefficients below are well defined. Hence

$$\left(\begin{array}{c} n \\ k+1 \end{array}\right) + \left(\begin{array}{c} n \\ k \end{array}\right) = \left(\begin{array}{c} n \\ k \end{array}\right)\left(\frac{n-k}{k+1} + 1\right) = \left(\begin{array}{c} n \\ k \end{array}\right)\frac{n+1}{k+1} = \left(\begin{array}{c} n+1 \\ k+1 \end{array}\right).$$

□

Proof of the Binomial Theorem The Binomial Formula holds for $n = 1$ since $\left(\begin{array}{c} 1 \\ 0 \end{array}\right) =$

$\left(\begin{array}{c}1\\1\end{array}\right) = 1$ (Base case). Next suppose that the formula holds for n. Then (Induction step)

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)\sum_{k=0}^{n} a^k b^{n-k}\left(\begin{array}{c}n\\k\end{array}\right)\\
&= \sum_{k=0}^{n} a^{k+1} b^{n-k}\left(\begin{array}{c}n\\k\end{array}\right) + \sum_{k=0}^{n} a^k b^{n-k+1}\left(\begin{array}{c}n\\k\end{array}\right)\\
&= a^{n+1} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k}\left(\begin{array}{c}n\\k\end{array}\right) + \sum_{k=1}^{n} a^k b^{n-k+1}\left(\begin{array}{c}n\\k\end{array}\right) + b^{n+1}\\
&= a^{n+1} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k}\left(\begin{array}{c}n\\k\end{array}\right) + \sum_{k=0}^{n-1} a^{k+1} b^{n-k}\left(\begin{array}{c}n\\k+1\end{array}\right) + b^{n+1}.
\end{aligned}$$

By Lemma 1.8.1

$$\begin{aligned}
(a+b)^{n+1} &= a^{n+1} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k}\left(\left(\begin{array}{c}n\\k\end{array}\right) + \left(\begin{array}{c}n\\k+1\end{array}\right)\right) + b^{n+1}\\
&= a^{n+1} + \sum_{k=0}^{n-1} a^{k+1} b^{n-k}\left(\begin{array}{c}n+1\\k+1\end{array}\right) + b^{n+1}\\
&= a^{n+1} + \sum_{k=1}^{n} a^{k} b^{n+1-k}\left(\begin{array}{c}n+1\\k\end{array}\right) + b^{n+1}\\
&= \sum_{k=0}^{n+1} a^{k} b^{n+1-k}\left(\begin{array}{c}n+1\\k\end{array}\right).
\end{aligned}$$

□

Finally note that you can use Mathematical Induction to prove statements, depending on the parameter $n \in \mathbb{Z}_+$, rather than $n \in \mathbb{N}$ by making the base case $n = 0$, as has already been done in Examples 1.8.5, 1.8.6 above. If $n \in \mathbb{Z}$, then the induction step shall involve not only the transition from n to $n+1$, but to $n-1$ as well.

Chapter 2

Numbers

The aim of this chapter is to introduce axiomatically the set of Real numbers.

2.1 Various Sorts of Numbers

2.1.1 Integers

We take for granted the system $\mathbb{N}$ of natural numbers

$$\mathbb{N} = \{1, 2, 3, 4 \ldots\}$$

stressing only that for $\mathbb{N}$ the following properties hold.

$$(\forall\, a \in \mathbb{N})(\forall\, b \in \mathbb{N})(\exists\, c \in \mathbb{N})(\exists\, d \in \mathbb{N})[(a + b = c) \wedge (ab = d)].$$

$$(\textit{closure under addition and multiplication})$$

$$(\forall\, a \in \mathbb{N})[a \cdot 1 = 1 \cdot a = a].$$

$$(\textit{existence of a multiplicative identity})$$

$$(\forall\, a \in \mathbb{N})(\forall\, b \in \mathbb{N})[(a = b) \vee (a < b) \vee (a > b)].$$

$$(\textit{order})$$

The first difficulty occurs when we try to come up with the additive analogue of $a \cdot 1 = 1 \cdot a = a$ for $a \in \mathbb{N}$. Namely, there is no $x \in \mathbb{N} : a + x = a$, for all $a \in \mathbb{N}$. This necessitates adding $\{0\}$ to $\mathbb{N}$, declaring $0 < 1$, thereby obtaining the set of non-negative integers $\mathbb{Z}_+$. Still, we cannot solve the equation

$$a + x = b$$

for $x \in \mathbb{Z}_+$ with $b < a$. In order to make this equation soluble we have to enlarge the set $\mathbb{Z}_+$ by introducing negative integers as unique solutions of the equations

$$a + x = 0 \text{ (existence of additive inverse)}$$

for each $a \in \mathbb{N}$. Our extended system, which is denoted by $\mathbb{Z}$, now contains all integers and can be arranged in order

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\} = \mathbb{N} \cup \{0\} \cup \{-a \,|\, a \in \mathbb{N}\}.$$

We also mention at this point the *Fundamental theorem of arithmetic.*

Theorem 2.1.1. (Fundamental theorem of arithmetic) *Every positive integer except* 1 *can be expressed uniquely as a product of primes.*

This is proved in the Unit *Number Theory and Group Theory.*

2.1.2 Rational Numbers

Let $a \in \mathbb{Z}$, $b \in \mathbb{Z}$. The equation

$$ax = b \tag{2.1.1}$$

need not have a solution $x \in \mathbb{Z}$. In order to solve (2.1.1) (for $a \neq 0$) we have to enlarge our system of numbers again so that it includes fractions b/a (existence of multiplicative inverse in $\mathbb{Z} - \{0\}$). This motivates the following definition.

Definition 2.1.1. The set of rational numbers (or rationals) $\mathbb{Q}$ is the set

$$\mathbb{Q} = \left\{ r = \frac{p}{q} : \, p \in \mathbb{Z}, \, q \in \mathbb{N}, \, \mathrm{hcf(p, q)} = 1 \right\}.$$

Here $\mathrm{hcf}(p, q)$ stands for the highest common factor of p and q, so when writing p/q for a rational we often assume that the numbers p and q have no common factor greater than 1. All the arithmetical operations in $\mathbb{Q}$ are straightforward. Let us introduce a relation of order for rationals.

Definition 2.1.2. Let $b \in \mathbb{N}$, $d \in \mathbb{N}$. Then

$$\left(\frac{a}{b} > \frac{c}{d}\right) \Longleftrightarrow \Big(ad > bc\Big).$$

The following theorem provides a very important property of rationals.

Theorem 2.1.2. *Between any two rational numbers there is another (and, hence, infinitely many others).*

Proof. Let $b \in \mathbb{N}$, $d \in \mathbb{N}$, and

$$\frac{a}{b} > \frac{c}{d}.$$

Notice that

$$(\forall m \in \mathbb{N}) \, [\, \frac{a}{b} > \frac{a + mc}{b + md} > \frac{c}{d} \,].$$

Indeed, since b, d and m are positive we have

$$[a(b + md) > b(a + mc)] \Longleftrightarrow [mad > mbc] \Leftrightarrow (ad > bc),$$

and

$$[d(a + mc) > c(b + md)] \Leftrightarrow (ad > bc). \quad \square$$

2.1.3 Irrational Numbers

Suppose that $a \in \mathbb{Q}$ and consider the equation

$$x^2 = a. \tag{2.1.2}$$

In general (2.1.2) does not have rational solutions. For example, the following theorem holds.

Theorem 2.1.3. *No rational number has square 2.*

Proof. Suppose for a contradiction that the rational number $\frac{p}{q}$ ($p \in \mathbb{Z}, q \in \mathbb{N}$, in lowest terms) is such that $(\frac{p}{q})^2 = 2$. Then $p^2 = 2q^2$. Hence, appealing to the Fundamental Theorem of Arithmetic, p^2 is even, and hence p is even. Thus $(\exists k \in \mathbb{Z})\,[\,p = 2k\,]$. This implies that

$$2k^2 = q^2$$

and therefore q is also even. The last statement contradicts our assumption that p and q have no common factor. $\square$

The last theorem provides an example of a number which is not rational. We call such numbers *irrational*. Here are some other examples of irrational numbers.

Theorem 2.1.4. *No rational x satisfies the equation $x^3 = x + 7$.*

Proof. First we show that there are no integers satisfying the equation $x^3 = x + 7$. For a contradiction suppose that there is. Then $x(x+1)(x-1) = 7$ from which it follows that x divides 7. Hence x can be only $\pm 1, \pm 7$. Direct verification shows that these numbers do not satisfy the equation.

Second, show that there are no fractions satisfying the equation $x^3 = x+7$. For a contradiction suppose that there is. Let $\frac{m}{n}$ with $m \in \mathbb{Z}$, $n \in \mathbb{Z}_+$, $n \neq 0, 1$, and m, n have no common factors greater than 1, is such that $(\frac{m}{n})^3 = \frac{m}{n} + 7$. Multiplying this equality by n^2 we obtain $\frac{m^3}{n} = mn + 7n^2$, which is impossible since the right-hand side is an integer and the left-hand side is not. $\square$

We leave the following as an exercise.

Example 2.1.1. No rational satisfies the equation $x^5 = x + 4$.

2.2 The Field of Real Numbers

In the previous sections we discussed the need to extend $\mathbb{N}$ to $\mathbb{Z}$, and $\mathbb{Z}$ to $\mathbb{Q}$. The rigorous construction of $\mathbb{N}$ can be found in a standard course on Set Theory. The rigorous construction of the set of real numbers from the rationals is rather complicated and lengthy. An excellent exposition can be found in [5]. W. Rudin *Principles of Mathematical Analysis.* McGraw Hill, 2006. In this course we postulate the existence of the set of real numbers $\mathbb{R}$ as well as basic properties summarized in a collection of axioms.

Notation. We use the symbol $\exists!$ to say that there exists a *unique...* So $(\exists!\, x \in \mathbb{R})$ is read: 'there exists a unique real number x'.

Those of you familiar with basic concepts of algebra will find that axioms A.1 - A.11 characterize $\mathbb{R}$ as an *algebraic field.*

A.1. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})\,[(a+b) \in \mathbb{R}]$ (*closed under addition*).

A.2. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})\,[a+b = b+a]$ (*commutativity of addition*).

A.3. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})(\forall\, c \in \mathbb{R})\,[(a+b)+c = a+(b+c)]$ (*associativity of addition*).

A.4. $(\exists\, 0 \in \mathbb{R})(\forall\, a \in \mathbb{R})\,[0+a = a]$ (*existence of additive identity*).

A.5. $(\forall\, a \in \mathbb{R})(\exists\,!\, x \in \mathbb{R})\,[a+x = 0]$ (*existence of additive inverse*). We write $x = -a$.

Axioms A.6-A.10 are analogues of A.1-A.5 for the operation of multiplication.

A.6. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})\,[ab \in \mathbb{R}]$ (*closed under multiplication*).

A.7. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})\,[ab = ba]$ (*commutativity of multiplication*).

A.8. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})(\forall\, c \in \mathbb{R})\,[(ab)c = a(bc)]$ (*associativity of multiplication*).

A.9. $(\exists\, 1 \in \mathbb{R})(\forall\, a \in \mathbb{R})\,[1 \cdot a = a]$ (*existence of multiplicative identity*).

A.10. $(\forall\, a \in \mathbb{R} - \{0\})(\exists\,!\, y \in \mathbb{R})\,[ay = 1]$ (*existence of multiplicative inverse*). We write $y = 1/a$.

The last axiom links the operations of summation and multiplication.

A.11. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})(\forall\, c \in \mathbb{R})\,[(a+b)c = ac+bc]$ (*distributive law*).

Familiar rules for the manipulation of real numbers can be deduced from the axioms above. Here is an example.

Example 2.2.1. $(\forall\, a \in \mathbb{R})[0a = 0]$.

Indeed, we have

$$\begin{array}{rcll} a+0a & = & 1a+0a & \text{(by A.9)} \\ & = & (1+0)a & \text{(by A.11)} \\ & = & 1a & \text{(by A.2 and A.4)} \\ & = & a & \text{(by A.9)} \end{array}$$

Now add $-a$ to both sides.

$$\begin{array}{rlcll} & -a+(a+0a) & = & -a+a & \\ \Rightarrow & (-a+a)+0a & = & 0 & \text{(by A.3 and A.5)} \\ \Rightarrow & 0+0a & = & 0 & \text{(by A.5)} \\ \Rightarrow & 0a & = & 0 & \text{(by A.4).} \end{array}$$

Remark 2.2.1. The set of rationals $\mathbb{Q}$ also forms an algebraic field (that is, the rational numbers satisfy axioms A.1 - A.11). But the integers $\mathbb{Z}$ do not form a field, as axiom A.10 does not hold.

Axioms A.1 - A.11 represent algebraic properties of real numbers. Now we add axioms of order.

O.1. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})[(a = b) \vee (a < b) \vee (a > b)]$
$\equiv (\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})[(a \geq b) \wedge (b \geq a) \Rightarrow (a = b)]$ (*trichotomy law*).

O.2. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})(\forall\, c \in \mathbb{R})[(a > b) \wedge (b > c) \Rightarrow (a > c)]$ (*transitive law*).

O.3. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})(\forall\, c \in \mathbb{R})[(a > b) \Rightarrow (a + c > b + c)]$ (*compatibility with addition*).

O.4. $(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})(\forall\, c \in \mathbb{R})[(a > b) \wedge (c > 0) \Rightarrow (ac > bc)]$
(*compatibility with multiplication*).

Remark 2.2.2. Note that

$$(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})\{[a > b] \Leftrightarrow [a - b > 0]\}.$$

This follows from (O.3) upon adding $-b$.

Axioms A.1 - A.11 and O.1 - O.4 define $\mathbb{R}$ to be an *ordered field*. Observe that the rational numbers also satisfy axioms A.1 - A.11 and O.1 - O.4, so $\mathbb{Q}$ is also an ordered field. It is the *completeness axiom* that distinguishes the reals from the rationals. Roughly speaking, the completeness axiom states that there are no gaps in the real numbers. We formulate and discuss this in further detail later on. Before doing so, let us consider various techniques for proving inequalities using the order axioms.

The order axioms express properties of the order relation (inequality) on the set of real numbers. Inequalities play an extremely important role in analysis. Here we explore several ideas around how to prove inequalities. First, we give the definition of the absolute value of a real number and derive a simple consequence from it.

Definition 2.2.1. Let $a \in \mathbb{R}$. The *absolute value* $|a|$ of a is defined by

$$|a| = \begin{cases} \;\; a & if \quad a \geq 0, \\ -a & if \quad a < 0. \end{cases}$$

Theorem 2.2.1. (triangle inequality)

$$(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R}) \;\; [\; |a + b| \leq |a| + |b| \;] \,.$$

Proof. We split the proof into two cases. We use the fact that $a \leqslant |a|$ for all $a \in \mathbb{R}$.

Case $a + b \geqslant 0$. Then

$$|a + b| = a + b \leqslant |a| + |b|.$$

Case $a + b < 0$. Then

$$|a + b| = -(a + b) = (-a) + (-b) \leqslant |a| + |b|.$$

The first idea for proving inequalities is very simple and is based on the equivalence

$$(\forall\, x \in \mathbb{R})(\forall\, y \in \mathbb{R})[(x \geq y) \Leftrightarrow (x - y \geq 0)].$$

Example 2.2.2. Prove that

$$(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})\ [a^2 + b^2 \geq 2ab].$$

Proof. The result is equivalent to $a^2 + b^2 - 2ab \geq 0$. But,

$$a^2 + b^2 - 2ab = (a - b)^2 \geq 0.$$

Note that equality holds if and only if $a = b$. $\square$

Example 2.2.3. Prove that

$$(\forall\, a \in \mathbb{R}_+)(\forall\, b \in \mathbb{R}_+)\ \left[\frac{a+b}{2} \geq \sqrt{ab}\right].$$

Recall that $\mathbb{R}_+ = \{x \in \mathbb{R} \,|\, x \geq 0\}$.

Proof. As above, let us prove that the difference between the left-hand side (LHS) and the right-hand side (RHS) is non-negative:

$$\frac{a+b}{2} - \sqrt{ab} = \frac{(\sqrt{a} - \sqrt{b})^2}{2} \geq 0.$$

Equality holds if and only if $a = b$. $\square$

The second idea is to use already proved inequalities and axioms to derive new ones.

Example 2.2.4. Prove that

$$(\forall\, a \in \mathbb{R})(\forall\, b \in \mathbb{R})(\forall\, c \in \mathbb{R})\ [a^2 + b^2 + c^2 \geq ab + ac + bc].$$

Proof. Adding three inequalities derived from Example 2.2.2,

$$\begin{aligned} a^2 + b^2 &\geq 2ab, \\ a^2 + c^2 &\geq 2ac, \\ b^2 + c^2 &\geq 2bc, \end{aligned}$$

we obtain $2(a^2 + b^2 + c^2) \geq 2ab + 2ac + 2bc$, which proves the desired result. Equality holds if and only if $a = b = c$. $\square$

Example 2.2.5. Prove that

$$(\forall\, a \in \mathbb{R}_+)(\forall\, b \in \mathbb{R}_+)(\forall\, c \in \mathbb{R}_+)(\forall\, d \in \mathbb{R}_+)\ \left[\frac{a+b+c+d}{4} \geq \sqrt[4]{abcd}\right].$$

Proof. By Example 2.2.3 and by (O.2) we have

$$\frac{a+b+c+d}{4} \geq \frac{2\sqrt{ab} + 2\sqrt{cd}}{4} \geq \sqrt[4]{abcd}.$$

Equality holds if and only if $a = b = c = d$. $\square$

The third idea is to use the transitivity property

$$(\forall\, a \in \mathbb{R}_+)(\forall\, b \in \mathbb{R}_+)(\forall\, c \in \mathbb{R}_+)[(a \geq b) \wedge (b \geq c) \Rightarrow (a \geq c)].$$

In general this means that when proving that $a \geq c$ we have to find b such that $a \geq b$ and $b \geq c$.

Example 2.2.6. Let $n \geq 2$ be a natural number. Prove that

$$\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} > \frac{1}{2}.$$

Proof.

$$\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} > \underbrace{\frac{1}{2n} + \frac{1}{2n} + \cdots + \frac{1}{2n}}_{n} = n\frac{1}{2n} = \frac{1}{2}. \quad \square$$

Inequalities involving integers may be proved by induction as discussed before. Here we give an inequality which is of particular importance in analysis.

Theorem 2.2.2. (Bernoulli's inequality)

$$(\forall n \in \mathbb{Z}_+)(\forall \alpha > -1)[(1+\alpha)^n \geq 1 + n\alpha].$$

Proof. Base case. The inequality holds for $n = 0, 1$.

Induction step. Suppose that the inequality is true for $n = k$ with $k \geq 1$; that is,

$$(1+\alpha)^k \geq 1 + k\alpha.$$

We have to prove that it is true for $n = k+1$; in other words,

$$(1+\alpha)^{k+1} \geq 1 + (k+1)\alpha.$$

Now,

$$\begin{aligned}(1+\alpha)^{k+1} &= (1+\alpha)^k(1+\alpha) \geq (1+k\alpha)(1+\alpha)\\ &= 1 + (k+1)\alpha + k\alpha^2 \geq 1 + (k+1)\alpha.\end{aligned}$$

This concludes the induction step. By the principle of mathematical induction, the result is true for all $n \in \mathbb{N}$. $\square$

2.3 Bounded sets of numbers

The simplest examples of bounded sets of real numbers are finite sets. Such sets always have a maximum and minimum. The maximum $\max S$ and minimum $\min S$ of a set S of real numbers is defined as below.

Definition 2.3.1. Let $S \subseteq \mathbb{R}$. Then

$$(x = \max S) \Leftrightarrow \{(x \in S) \wedge [(\forall\, y \in S)(y \leq x)]\},$$
$$(x = \min S) \Leftrightarrow \{(x \in S) \wedge [(\forall\, y \in S)(y \geq x)]\}.$$

Note that the definition says nothing about existence of $\max S$ and $\min S$. In fact, these numbers may or may not exist depending on the set S (consider $S = (0,\, 1)$ and $S = [0,\, 1]$ for example).

Definition 2.3.2. (i) A set $S \subseteq \mathbb{R}$ is said to be *bounded above* if

$$(\exists K \in \mathbb{R})(\forall\, x \in S)(x \leq K).$$

The number K is called an *upper bound* of S.

(ii) A set $S \subseteq \mathbb{R}$ is said to be *bounded below* if

$$(\exists\, k \in \mathbb{R})(\forall x \in S)(x \geq k).$$

The number k is called a *lower bound* of S.

(iii) A set $S \subseteq \mathbb{R}$ is said to be *bounded* if it is bounded above and below.

Note that if S is bounded above with upper bound K then any real number greater than K is also an upper bound of S. Similarly, if S is bounded below with lower bound k then any real number smaller than k is also a lower bound of S.

Example 2.3.1. Let $a, b \in \mathbb{R}$ with $a < b$.

(i) let $A = \{x \in \mathbb{R} \,|\, a \leq x \leq b\} = [a, b]$ be a closed interval. Then A is bounded above since $(\forall\, x \in A)(x \leq b)$ and is bounded below since $(\forall\, x \in A)(x \geq a)$. As $a \in A$ and $b \in A$, we have that $\max A = b$ and $\min A = a$.

(ii) Let $B = \{x \in \mathbb{R} \,|\, a < x < b\} = (a, b)$ be an open interval. It is bounded above and below for the same reasons as A. However, the maximum and minimum of B do not exist.

Note that if $A \subseteq \mathbb{R}$ is a bounded set and $B \subseteq A$ then B is also bounded (exercise).

Example 2.3.2. Let

$$A = \left\{ \frac{n}{n+1} \,:\, n \in \mathbb{N} \right\}.$$

Then A is bounded. Indeed, $(\forall\, n \in \mathbb{N}) \left(\frac{n}{n+1} \geqslant 0\right)$ so A is bounded below by 0. Moreover,

$$(\forall\, n \in \mathbb{N}) \left(\frac{n}{n+1} < 1 \right)$$

so A is bounded above, and therefore bounded. Also, $\min A = \frac{1}{2}$ but $\max A$ does not exist.

2.4 Supremum and infimum. Completeness axiom

If a set $S \subseteq \mathbb{R}$ is bounded above one often needs to find the smallest upper bound. The least upper bound of a set is called the supremum. The precise definition is as follows.

Definition 2.4.1. Let $S \subseteq \mathbb{R}$ be bounded above. Then $a \in \mathbb{R}$ is called the *supremum* (or *least upper bound*; *l.u.b.* for short) of S if

(i) a is an upper bound for S; that is,

$$(\forall\, x \in S)(x \leq a);$$

(ii) a is the least upper bound for S; that is,

$$(\forall\, \varepsilon > 0)(\exists\, b \in S)(b > a - \varepsilon).$$

We write $a = \sup\, S$.

Now we are in a position to formulate the completeness axiom for real numbers. This property distinguishes the reals from the rationals.

The Completeness Axiom. Every nonempty subset of $\mathbb{R}$ that is bounded above has a least upper bound.

The set of real numbers may then be characterized via:

The set of real numbers $\mathbb{R}$ is an ordered field satisfying the Completeness Axiom.

Remark 2.4.1. The set of rational numbers $\mathbb{Q}$ does not satisfy the Completeness Axiom. To see this, consider the set $\{x \in \mathbb{Q} | x^2 \leq 2\}$.

The greatest lower bound of a set bounded below can be defined analogously.

Definition 2.4.2. Let $S \subseteq \mathbb{R}$ be bounded below. Then $a \in \mathbb{R}$ is called the *infimum* (or *greatest lower bound*; *g.l.b.* for short) of S if

(i) a is a lower bound for S; that is,

$$(\forall\, x \in S)(x \geq a);$$

(ii) a is the greatest lower bound for S; that is,

$$(\forall\, \varepsilon > 0)(\exists\, b \in S)(b < a + \varepsilon).$$

We write $b = \inf\, S$.

We use the following notation

$$-S = \{x \in \mathbb{R} \,:\, -x \in S\}.$$

Theorem 2.4.1. *Let $S \subseteq \mathbb{R}$ be non-empty and bounded below. Then* $\inf S$ *exists and is equal to* $-\sup(-S)$.

Proof. Since S is bounded below, it follows that $\exists\, k \in \mathbb{R}$ such that $(\forall\, x \in S)(x \geq k)$. Hence $(\forall\, x \in S)(-x \leq -k)$ which means that $-S$ is bounded above. By the Completeness Axiom there exists $\xi = \sup(-S)$. We will show that $-\xi = \inf S$. First, we have that $(\forall\, x \in S)(-x \leq \xi)$, so that $x \geq -\xi$. Hence $-\xi$ is a lower bound of S. Next, let η be another lower bound for S, i.e. $(\forall\, x \in S)(x \geq \eta)$. Then $(\forall\, x \in S)(-x \leq -\eta)$. Hence $-\eta$ is an upper bound of $-S$, and by the definition of supremum $-\eta \geq \sup(-S) = \xi$. Hence $\eta \leq -\xi$ which proves the required result. □

Theorem 2.4.2. (The Archimedean Principle) *Let $x \in \mathbb{R}$. Then there exists $n \in \mathbb{Z}$ such that $x < n$.*

Proof. Suppose for a contradiction that

$$(\forall\, n \in \mathbb{Z})\,(n \leq x).$$

Then the set $A := \{n \,:\, n \in \mathbb{Z}\}$ is bounded above and x is an upper bound. By the Completeness Axiom, the supremum $\sup A$ exists. The number $\sup A - 1 < \sup A$ is *not* an upper bound for A, so there exists $m \in \mathbb{Z}$ such that $m > \sup A - 1$, in other words,

$$m + 1 > \sup A. \tag{2.4.3}$$

But $m + 1 \in \mathbb{Z}$ so $m + 1 \in A$, and (2.4.3) contradicts the fact that $\sup A$ is an upper bound for A. □

Corollary 2.4.1. $\mathbb{N}$ *is unbounded.*

Corollary 2.4.2.

$$(\forall\, x > 0)(\exists\, n \in \mathbb{N})\ (\frac{1}{n} < x).$$

Proof. By the Archimedean Principle (Theorem 2.4.2),

$$(\exists\, n \in \mathbb{Z})\ (n > \frac{1}{x}).$$

Since $\frac{1}{x} > 0$ it follows that $n \in \mathbb{N}$. Therefore $\frac{1}{n} < x$. □

Example 2.4.1.

$$\inf \left\{ \frac{1}{n} \,:\, n \in \mathbb{N} \right\} = 0.$$

Example 2.4.2. Let

$$A := \left\{ \frac{n-1}{2n} \,:\, n \in \mathbb{N} \right\}.$$

Then $\sup A = \frac{1}{2}$ and $\inf A = 0$.

Proof. All the elements of A are positive except for the first one which is 0. Therefore $\min A = 0$, and hence $\inf A = 0$. Notice that

$$(\forall\, n \in \mathbb{N})\ \left(\frac{n-1}{2n} = \frac{1}{2} - \frac{1}{2n} < \frac{1}{2} \right).$$

So $\frac{1}{2}$ is an upper bound for A.

We have to prove that $\frac{1}{2}$ is the least upper bound. For this we have to prove that

$$(\forall\,\varepsilon > 0)(\exists\,n \in \mathbb{N})\ \left(\frac{n-1}{2n} > \frac{1}{2} - \varepsilon\right).$$

Now

$$\left(\frac{n-1}{2n} > \frac{1}{2} - \varepsilon\right) \Leftrightarrow \left(\frac{1}{2} - \frac{1}{2n} > \frac{1}{2} - \varepsilon\right)$$
$$\Leftrightarrow \left(\frac{1}{2n} < \varepsilon\right) \Leftrightarrow [n > \frac{1}{2\varepsilon}].$$

By the Archimedean Principle there exists such an $n \in \mathbb{N}$. $\square$

Note that for any set $A \subseteq \mathbb{R}$, if $\max A$ exists then $\sup A = \max A$ and if $\min A$ exists then $\inf A = \min A$.

Theorem 2.4.3. (well-ordering principle) *Any non-empty subset S of the integers $\mathbb{Z}$ which is bounded below has a minimum.*

Proof. As S is non-empty and bounded below the infimum $s = \inf S \in \mathbb{R}$ exists by the Completeness Axiom. Now $s + 1$ is not a lower bound for S so we can find $m \in S$ such that $m < s + 1$. So $m - 1 < s$. But then $m \geqslant s > m - 1$. We claim that m is the minimum element of S (that is, $(\forall\,n \in S)[n \leqslant m]$ and $m \in S$). Certainly, $m \in S$. Suppose we can find $n \in S$ with $n < m$. Then $m > n \geqslant s > m - 1$. So using the fact that

$$(\forall\,k \in \mathbb{Z})(\forall\,l \in \mathbb{Z})[(k > l) \Rightarrow (k \geqslant l + 1)].$$

, $m > n \geqslant m$, which is a contradiction. $\square$

The counterpart to this result for subsets of $\mathbb{Z}$ that are bounded above reads as follows (prove it).

Theorem 2.4.4. *Any non-empty subset S of the integers $\mathbb{Z}$ which is bounded above has a maximum.*

Theorem 2.4.5. *For any interval (a, b) there exists a rational $r \in (a, b)$. In other words,*

$$(\forall\,a \in \mathbb{R})(\forall\,b \in \mathbb{R})[(b > a) \Rightarrow (\exists\,r \in \mathbb{Q})(a < r < b)].$$

Proof. Let $h = b - a > 0$. Then by the Archimedean Principle,

$$(\exists\,n \in \mathbb{N})\ \left(\frac{1}{n} < h\right).$$

Note that $1 < nh = n(b - a)$ or $na < nb - 1$. Put $S := \{m \in \mathbb{Z} : m < nb\}$. Then S is non-empty and bounded above so has a maximum element $m = \max S$. Now $m < nb \leqslant m+1$ where the latter inequality follows from the fact that m is the maximum element of S. We then have

$$na < nb - 1 \leqslant (m + 1) - 1 = m < nb.$$

Dividing through by $n \in \mathbb{N}$ we obtain finally,

$$a < r := \frac{m}{n} < b.$$

□

The next theorem is a multiplicative analogue of the Archimedean Principle.

Theorem 2.4.6. (multiplicative Archimedean Principle) *Let* $x > 1,\ y > 0$. *Then*

$$(\exists\, n \in \mathbb{N})\ (x^n > y).$$

Prove this theorem repeating the argument from the proof of Theorem 2.4.2.

Theorem 2.4.7. $(\exists!\, x > 0)\ (x^2 = 2)$.

Proof. Define a set

$$A := \{y > 0\ :\ y^2 < 2\}.$$

Then $A \neq \varnothing$ since $1 \in A$. A is bounded above since $(\forall\, y \in A)\,(y < 2)$. By the Completeness Axiom, $\sup A = x$ exists. We prove that $x^2 = 2$.

a) First suppose that $x^2 > 2$. Let $\varepsilon = \frac{x^2-2}{2x}$. Then $\varepsilon > 0$ and

$$(x-\varepsilon)^2 = x^2 - 2x\varepsilon + \varepsilon^2 > x^2 - 2x\varepsilon = x^2 - 2x\frac{x^2-2}{2x} = 2.$$

Hence $x - \varepsilon$ is another upper bound for A, so that x is not the least upper bound for A. This is a contradiction.

b) Now suppose that $x^2 < 2$. Let $\varepsilon = \frac{2-x^2}{2x+1}$. Then by assumption $0 < \varepsilon < 1$ so that

$$(x+\varepsilon)^2 = x^2 + 2x\varepsilon + \varepsilon^2 < x^2 + 2x\varepsilon + \varepsilon =$$
$$= x^2 + \varepsilon(2x+1) = x^2 + \frac{2-x^2}{2x+1}(2x+1) = 2.$$

Hence $x + \varepsilon$ is also in A, in which case x cannot be an upper bound for A, which is a contradiction. □

Exercise 2.4.1. Prove that $(\forall\, a > 0)(\forall\, n \in \mathbb{N})(\exists!\, x > 0)[x^n = a]$.

Part II

Real Analysis

Chapter 3

Sequences and Limits

The notion of a *sequence* is central in Analysis. Among sequences, *convergent sequences* form a special important class. They are introduced and studied in this chapter. Further, based on this, the notion of a *convergent series* is introduced.

3.1 Sequences

In general, any function $f : \mathbb{N} \to X$ with domain $\mathbb{N}$ and arbitrary codomain X is called a sequence. This means that a sequence $(a_n)_{n\in\mathbb{N}}$ is a subset of X each element of which has an index, $a_n = f(n)$. In this course we confine ourselves to sequences of real numbers. In particular, in this chapter we always assume that $X = \mathbb{R}$. So we adopt the following definition.

Definition 3.1.1. A **sequence** of real numbers is a function $f : \mathbb{N} \to \mathbb{R}$. We use the notation $a_n = f(n)$ for the general term and $(a_n)_{n\in\mathbb{N}}$ for the whole sequence, so

$$(a_n)_{n\in\mathbb{N}} = (a_1, a_2, a_3, \cdots, a_n, \ldots).$$

Example 3.1.1. *(i)* $a_n = \frac{1}{n}$. The sequence is $\left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right)$.

(ii) $a_n = n^2$. The sequence is $(1, 4, 9, 16, 25, \ldots)$.

(iii) $a_n = \frac{(-1)^n}{n}$. The sequence is $\left(-1, \frac{1}{2}, -\frac{1}{3}, \frac{1}{4}, \ldots\right)$.

3.1.1 Null sequences

In Example *(i)* above, the n-th element of the sequence becomes smaller as n becomes larger; in other words, a_n tends to zero. The same applies to Example *(iii)*. Such sequences are called null sequences. We now give a precise mathematical definition for the notion of a null sequence.

Definition 3.1.2. $(a_n)_{n\in\mathbb{N}}$ is a **null sequence** if

$$(\forall\, \varepsilon > 0)(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})\ [(n > N) \Rightarrow (|a_n| < \varepsilon)].$$

We illustrate this definition with some examples.

Example 3.1.2. $(a_n)_{n\in\mathbb{N}}$ with $a_n = \dfrac{1}{n}$ is a null sequence.

Proof. We have to prove that

$$(\forall\, \varepsilon > 0)(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N}) \left[(n > N) \Rightarrow \left(\frac{1}{n} < \varepsilon\right)\right].$$

In other words, we need n to satisfy the inequality $n > \frac{1}{\varepsilon}$. By the AP (Archimedian Principle),

$$(\exists\, N \in \mathbb{N}) \left(N > \frac{1}{\varepsilon}\right).$$

(In particular, one can take $N = [1/\varepsilon]+1$; here, $[x]$ denotes the integer part of the real number x). Then

$$\left[(n > N) \wedge \left(N > \frac{1}{\varepsilon}\right)\right] \Rightarrow \left(n > \frac{1}{\varepsilon}\right),$$

which proves the statement. □

The next theorem follows straight from the definition of a null sequence and the fact that $(\forall\, a \in \mathbb{R})\ (|a| = ||a||)$.

Theorem 3.1.1. *A sequence $(a_n)_{n\in\mathbb{N}}$ is a null sequence if and only if the sequence $(|a_n|)_{n\in\mathbb{N}}$ is a null sequence.*

Example 3.1.3. $(a_n)_{n\in\mathbb{N}}$ with $a_n = \dfrac{(-1)^n}{n}$ is a null sequence.

Example 3.1.4. $(a_n)_{n\in\mathbb{N}}$ with $a_n = \dfrac{1}{n^2}$ is a null sequence.

Definition 3.1.3. Let $x \in \mathbb{R}$ and $\varepsilon > 0$. The set

$$B(x, \varepsilon) := (x - \varepsilon, x + \varepsilon) = \{y \in \mathbb{R} :\ |y - x| < \varepsilon\}$$

is called the **ε-neighbourhood** of the point x.

Using this notion one can say that a sequence $(a_n)_{n\in\mathbb{N}}$ is a null sequence if and only if for any $\varepsilon > 0$ all the elements of the sequence belong to the ε-neighbourhood of zero, apart from only finitely many.

3.1.2 Sequences converging to a limit

A null sequence is one whose terms approach zero. This a particular case of a sequence tending to a limit.

Example 3.1.5. Let $a_n = \frac{n}{n+1}$. Then $(a_n)_{n\in\mathbb{N}^+}$ tends to 1 as $n \to \infty$.

Definition 3.1.4. A sequence $(a_n)_{n\in\mathbb{N}}$ **converges** to a limit $a \in \mathbb{R}$ if

$$(\forall\, \varepsilon > 0)(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})\ [(n > N) \Rightarrow (|a_n - a| < \varepsilon)].$$

We write in this case that

$$\lim_{n\to\infty} a_n = a$$

or abbreviating,

$$\lim a_n = a$$

or, again

$$a_n \to a \text{ as } n \to \infty.$$

Remark 3.1.1. *(i)* From the above definition it is clear that

$$\left(\lim_{n\to\infty} a_n = a\right) \Leftrightarrow ((a_n - a)_n \text{ is a null sequence}).$$

(ii) $\lim_{n\to\infty} a_n = a$ if and only if for any $\varepsilon > 0$ starting from some number N all the elements of the sequence $(a_n)_{n=N}^{\infty}$ belong to the ε-neighbourhood of a.

(iii) Note that in Definition 3.1.4, N depends on ε.

Let us now prove the statement in Example 3.1.5.

Proof. Let $\varepsilon > 0$ be given. We have to find $N \in \mathbb{N}$ such that for all $n > N$ we have $|a_n - a| < \varepsilon$, with $a = 1$. The last inequality for our example reads

$$\left|\frac{n}{n+1} - 1\right| < \varepsilon.$$

An equivalent form of this inequality is

$$\left(\frac{1}{n+1} < \varepsilon\right) \Leftrightarrow \left(n > \frac{1}{\varepsilon} - 1\right).$$

Let N be any integer greater than or equal to $\frac{1}{\varepsilon} - 1$ (which exists by the AP). Then, if $n > N$, we have

$$n > N \geq \frac{1}{\varepsilon} - 1.$$

Therefore

$$\left(\frac{1}{n+1} < \varepsilon\right) \Leftrightarrow \left(\left|\frac{n}{n+1} - 1\right| < \varepsilon\right). \quad \square$$

If a sequence $(a_n)_n$ does not converge, we say that it **diverges**.

Definition 3.1.5. 1. The sequence $(a_n)_{n\in\mathbb{N}}$ **diverges to** ∞ if

$$(\forall\, M \in \mathbb{R})(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})\ [(n > N) \Rightarrow (a_n > M)].$$

2. The sequence $(a_n)_{n\in\mathbb{N}}$ **diverges to** $-\infty$ if

$$(\forall\, M \in \mathbb{R})(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})\ [(n > N) \Rightarrow (a_n < M)].$$

Example 3.1.6. 1. The sequence $a_n = n^2$ diverges to ∞.

2. The sequence $a_n = -n^2$ diverges to $-\infty$.

3. The sequence $a_n = (-1)^n$ diverges.

3.1.3 Properties of convergent sequences

Theorem 3.1.2. *A sequence $(a_n)_{n\in\mathbb{N}}$ can have at most one limit.*

Proof. We have to prove that the following implication is true for all $a, b \in \mathbb{R}$

$$\left[\left(\lim_{n\to\infty} a_n = a\right) \wedge \left(\lim_{n\to\infty} a_n = b\right)\right] \Rightarrow (a = b).$$

By definition,

$$(\forall\, \varepsilon > 0)(\exists\, N_1 \in \mathbb{N})(\forall\, n \in \mathbb{N})\ [(n > N_1) \Rightarrow (|a_n - a| < \varepsilon)],$$

$$(\forall\, \varepsilon > 0)(\exists\, N_2 \in \mathbb{N})(\forall\, n \in \mathbb{N})\ [(n > N_2) \Rightarrow (|a_n - b| < \varepsilon)].$$

Suppose for a contradiction that the statement is not true. Then its negation

$$\left[\left(\lim_{n\to\infty} a_n = a\right) \wedge \left(\lim_{n\to\infty} a_n = b\right)\right] \wedge (a \neq b)$$

is true. Fix $\varepsilon = \frac{1}{3}|a - b| > 0$ and take $N = \max\{N_1, N_2\}$. Then, if $n > N$, we have (by the triangle inequality),

$$|a - b| = |(a - a_n) + (a_n - b)| \leq |a_n - a| + |a_n - b| < 2\varepsilon = \frac{2}{3}|a - b|,$$

which is a contradiction. □

Theorem 3.1.3. *Any convergent sequence is bounded.*

Proof. We have to prove that the following implication is true

$$\left[\left(\lim_{n\to\infty} a_n = a\right)\right] \Rightarrow [(\exists\, m,\ M \in \mathbb{R})(\forall\, n \in \mathbb{N})\ (m \leq a_n \leq M)].$$

Take $\varepsilon = 1$ in the definition of the limit. Then $\exists\, N \in \mathbb{N}$ such that for $n > N$ we have

$$(|a_n - a| < 1) \Leftrightarrow (a - 1 < a_n < a + 1).$$

Set $M = \max\{a + 1, a_1, \ldots, a_N\}$ and $m = \min\{a - 1, a_1, \ldots, a_N\}$. Then

$$(\forall\, n \in \mathbb{N})\ (m \leq a_n \leq M). \quad \square$$

Note that Theorem 3.1.3 can be expressed by the contrapositive law in the following way: if a sequence is unbounded then it diverges. So boundedness of a sequence is a necessary condition for its convergence.

Theorem 3.1.4. *If $\lim_{n\to\infty} a_n = a \neq 0$ then*

$$(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})\ \left[(n > N) \Rightarrow (|a_n| > \frac{|a|}{2})\right].$$

Moreover, if $a > 0$ then for the above n, $a_n > \frac{a}{2}$; and if $a < 0$, then for the above n, $a_n < \frac{a}{2}$.

Proof. Fix $\varepsilon = \frac{|a|}{2} > 0$. Then $\exists\, N \in \mathbb{N}$ such that for $n > N$

$$\frac{|a|}{2} > |a - a_n| \geq |a| - |a_n|,$$

which implies $|a_n| > |a| - \frac{|a|}{2} = \frac{|a|}{2}$, and the first assertion is proved. On the other hand,

$$(\frac{|a|}{2} > |a - a_n|) \Leftrightarrow \left(a - \frac{|a|}{2} < a_n < a + \frac{|a|}{2}\right).$$

Thus, if $a > 0$ then for $n > N$,

$$a_n > a - \frac{|a|}{2} = \frac{a}{2}.$$

Also, if $a < 0$ then for $n > N$,

$$a_n < a + \frac{|a|}{2} = \frac{a}{2}. \qquad \square$$

Theorem 3.1.5. *Let $a, b \in \mathbb{R}$, $(a_n)_n$, $(b_n)_n$ be real sequences. Then*

$$\left\{\left(\lim_{n\to\infty} a_n = a\right) \wedge \left(\lim_{n\to\infty} b_n = b\right) \wedge [(\forall\, n \in \mathbb{N})\,(a_n \leq b_n)]\right\} \Rightarrow (a \leq b).$$

Proof. For a contradiction assume that $b < a$. Fix $\varepsilon < \frac{a-b}{2}$ so that $b + \varepsilon < a - \varepsilon$ and choose N_1 and N_2 such that the following is true

$$[(\forall\, n > N_1)(a_n > a - \varepsilon)] \wedge [(\forall\, n > N_2)(b_n < b + \varepsilon)].$$

If $N \geq \max\{N_1, N_2\}$ then

$$b_n < b + \varepsilon < a - \varepsilon < a_n,$$

which contradicts to the condition $(\forall\, n \in \mathbb{N})\,(a_n \leq b_n)$. $\square$

Theorem 3.1.6. (Sandwich rule). *Let $a \in \mathbb{R}$ and $(a_n)_n, (b_n)_n, (c_n)_n$ be real sequences. Then*

$$\{[(\forall\, n \in \mathbb{N})\,(a_n \leq b_n \leq c_n)] \wedge (\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = a)\} \Rightarrow (\lim_{n\to\infty} b_n = a).$$

Proof. Let $\varepsilon > 0$. Then one can find $N_1, N_2 \in \mathbb{N}$ such that

$$[(n > N_1) \Rightarrow (a - \varepsilon < a_n)] \wedge [(n > N_2) \Rightarrow (c_n < a + \varepsilon)].$$

Then choosing $N = \max\{N_1, N_2\}$ we have

$$(a - \varepsilon < a_n \leq b_n \leq c_n < a + \varepsilon) \Rightarrow (|b_n - a| < \varepsilon) \qquad \square.$$

The next theorem is a useful tool in computing limits.

Theorem 3.1.7. *Let $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$. Then*

(i) $\lim_{n\to\infty} (a_n + b_n) = a + b$;

(ii) $\lim_{n\to\infty}(a_n b_n) = ab$.

(iii) *If in addition* $b \neq 0$ *and* $(\forall\, n \in \mathbb{N})\,(b_n \neq 0)$ *then* $\lim_{n\to\infty}\left(\frac{a_n}{b_n}\right) = \frac{a}{b}$.

Proof. *(i)* Let $\varepsilon > 0$. Choose N_1 such that for all $n > N_1$ $|a_n - a| < \varepsilon/2$. Choose N_2 such that for all $n > N_2$ $|b_n - b| < \varepsilon/2$. Then for all $n > \max\{N_1, N_2\}$,

$$|a_n + b_n - (a + b)| \leq |a_n - a| + |b_n - b| < \varepsilon.$$

(ii) Since $(b_n)_n$ is convergent, it is bounded by Theorem 3.1.3. Let K be such that $(\forall\, n \in \mathbb{N})\,(|b_n| \leq K)$ and $|a| \leq K$. We have

$$\begin{aligned}|a_n b_n - ab| &= |a_n b_n - ab_n + ab_n - a_n b_n|\\ &\leq |a_n - a||b_n| + |a||b_n - b| \leq K|a_n - a| + K|b_n - b|.\end{aligned}$$

Choose N_1 such that for $n > N_1$, $K|a_n - a| \leq \varepsilon/2$. Choose N_2 such that for $n > N_2$, $K|b_n - b| \leq \varepsilon/2$. Then for $n > N = \max\{N_1, N_2\}$,

$$|a_n b_n - ab| < \varepsilon.$$

(iii) We have

$$\left|\frac{a_n}{b_n} - \frac{a}{b}\right| = \left|\frac{a_n b - ab_n}{b_n b}\right| \leq \frac{|a_n - a||b| + |a||b_n - b|}{|b_n b|}.$$

Choose N_1 such that for $n > N_1$,

$$|b_n| > \frac{1}{2}|b| \quad \text{(possible by Theorem 3.1.4)}.$$

Choose N_2 such that for $n > N_2$,

$$|a_n - a| < \frac{1}{4}|b|\varepsilon.$$

Choose N_3 such that for $n > N_3$,

$$|a||b_n - b| < \frac{1}{4}b^2\varepsilon.$$

Then for $n > N := \max\{N_1, N_2, N_3\}$,

$$\left|\frac{a_n}{b_n} - \frac{a}{b}\right| < \varepsilon. \qquad \square$$

3.2 Monotone sequences

In this section we consider a particular class of sequences which often occur in applications.

Definition 3.2.1. *(i)* A sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ is called **increasing** if

$$(\forall\, n \in \mathbb{N})\ (a_n \leq a_{n+1}).$$

(ii) A sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ is called **decreasing** if

$$(\forall\, n \in \mathbb{N})\ (a_n \geq a_{n+1}).$$

(iii) A sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ is called **monotone** if it is either increasing or decreasing.

Example 3.2.1. *(i)* $a_n = n^2$ is increasing.

(ii) $a_n = \dfrac{1}{n}$ is decreasing.

(iii) $a_n = (-1)^n \dfrac{1}{n}$ is not monotone.

The next theorem is one of the main theorems in the theory of converging sequences.

Theorem 3.2.1. *If a sequence of real numbers is bounded above and increasing then it is convergent.*

Proof. Put $A = \{a_n : n \in \mathbb{N}^+\}$. As A is bounded above, we may define

$$\alpha = \sup\{a_n : n \in \mathbb{N}^+\} \in \mathbb{R}.$$

Let $\varepsilon > 0$. Now $\alpha - \varepsilon$ is not an upper bound for A, so there exists $N \in \mathbb{N}^+$ such that $a_N > \alpha - \varepsilon$. By monotonicity, $\alpha - \varepsilon < a_N \leqslant a_n \leqslant \alpha$ for each $n \geqslant N$. In particular, $|a_n - \alpha| < \varepsilon$ for $n \geqslant N$. □

The following theorem is the counterpart to the one above.

Theorem 3.2.2. *If a sequence of real numbers is bounded below and decreasing then it is convergent.*

In many cases, before computing the limit of a sequence, one has to prove that the limit exists, i.e. that the sequence converges. In particular, this is the case when a sequence is defined by a recurrence formula: in other words, the nth element of the sequence is determined by preceding elements. Let us illustrate this point with an example.

Example 3.2.2. Let $(a_n)_n$ be a sequence defined by

$$a_n = \sqrt{a_{n-1} + 2},\ a_1 = \sqrt{2}. \tag{3.2.1}$$

(i) $(a_n)_n$ is bounded.

Indeed, we prove that $(\forall\, n \in \mathbb{N})\ (0 < a_n \leq 2)$.

Proof. Positivity is obvious. For the upper bound we use induction. For $n = 1$, $a_1 = \sqrt{2} \leq 2$, and the statement is true. Suppose now that it is true for $n = k$ $(k \geq 1)$; that is $a_k \leq 2$. For $n = k + 1$ we have

$$a_{k+1} = \sqrt{a_k + 2} \leq \sqrt{2 + 2} = 2,$$

and by the principle of induction the statement is proved. □

(ii) $(a_n)_n$ is increasing.

We have to prove that

$$(\forall\, n \in \mathbb{N})\ (a_{n+1} \geq a_n).$$

Proof. This is equivalent to proving that

$$[(\forall\, n \in \mathbb{N})\ (\sqrt{a_n + 2} \geq a_n)] \Leftrightarrow [(\forall\, n \in \mathbb{N})\ (a_n + 2 \geq a_n^2)].$$

But,

$$a_n + 2 - a_n^2 = (2 - a_n)(a_n + 1) \geq 0.$$

□

By Theorem 3.2.1, we conclude that (a_n) is convergent. Let $\lim_n a_n = x$; then, of course, $\lim_n a_{n-1} = x$ also. Write

$$a_n^2 = a_{n-1} + 2$$

and take limits in both sides to obtain

$$x^2 = x + 2 \text{ or } x^2 - x - 2 = 0.$$

As $x \geqslant 0$, we conclude that $x = 2$. □

Now we are ready to prove a very important property of real numbers. This Theorem will appear in the proof of the Bolzano-Weierstrass Theorem (in *Further Topics in Analysis*).

Theorem 3.2.3. *Let* $([a_n, b_n])_{n\in\mathbb{N}}$ *be a sequence of closed intervals on the real axis such that:*

(i) $(\forall\, n \in N)\ ([a_{n+1}, b_{n+1}] \subseteq [a_n, b_n])$*;*

(ii) $\lim\limits_{n\to\infty}(b_n - a_n) = 0$ *(the length of the intervals tends to zero).*

Then there is a unique point which belongs to all the intervals.

Proof. The sequence $(a_n)_n$ is increasing and bounded above as $a_n \leqslant b_n \leqslant b_1$ because $[a_n, b_n] \subseteq [a_1, b_1]$; similarly, the sequence $(b_n)_n$ is decreasing and bounded below. Therefore the sequences $(a_n)_n$ and $(b_n)_n$ are both convergent in virtue of the above Theorems. Set $a := \lim_n a_n$ and $b := \lim_n b_n$. Then

$$a \leq b$$

by Theorem 3.1.5. Moreover,

$$b - a = \lim_n (b_n - a_n) = 0.$$

Now $a \in \bigcap_{n\in\mathbb{N}^+}[a_n, b_n]$ as, for any $n \in \mathbb{N}^+$, $(a_k)_{k\geqslant n}$ belongs to $[a_n, b_n]$ so that $a = \lim_k a_k \in [a_n, b_n]$. Hence

$$b = a \in \bigcap_{n\in\mathbb{N}} [a_n, b_n].$$

□

3.3 Series

In this section, we discuss infinite series. Let us start with an example likely to be familiar from secondary school. Consider the geometric progression

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}\dots$$

Its sum is understood as follows. Define the (partial) sum of the first n terms by

$$s_n=\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\dots+\frac{1}{2^n}=1-\frac{1}{2^n};$$

then define the infinite sum s as $\lim_{n\to\infty} s_n$, so that $s=1$. This idea is used to define the sum of an arbitrary series formally.

Definition 3.3.1. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real (complex) numbers. Let

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

We say that the series

$$\sum_{k=1}^{\infty} a_k=a_1+a_2+\dots$$

is **convergent** (or **converges**) if the sequence of the partial sums $(s_n)_{n\in\mathbb{N}}$ is convergent. The limit of this sequence is called the sum of the series. If the series is not convergent we say that it is **divergent** (or **diverges**).

Theorem 3.3.1. *If the series*

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

is convergent then $\lim\limits_{n\to\infty} a_n=0$.

Proof. Let $s_n=\sum_{k=1}^{n} a_k$. Then by the definition the limit $\lim_n s_n$ exists. Denote it by s. Then of course $\lim_n s_{n-1}=s$. Note that $a_n=s_n-s_{n-1}$ for $n\geq 2$. Hence

$$\lim_{n\to\infty} a_n=\lim_{n\to\infty} s_n-\lim_{n\to\infty} s_{n-1}=s-s=0. \qquad \square$$

The same idea can be used to prove the following theorem.

Theorem 3.3.2. *If the series*

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

is convergent then

$$s_{2n}-s_n=a_{n+1}+a_{n+2}+\dots+a_{2n}\to 0 \text{ as } n\to\infty.$$

The above theorem expresses the simplest necessary condition for the convergence of a series. For example, each of the following series is divergent:

$$1 + 1 + 1 + 1 + \dots,$$

$$1 - 1 + 1 - 1 + \dots,$$

since the necessary condition is not satisfied.

Let us take a look at several important examples.

Example 3.3.1. The series

$$\sum_{k=0}^{\infty} x^k = 1 + x + x^2 + \dots$$

is convergent if and only if $-1 < x < 1$.

Indeed,

$$s_n = \begin{cases} \frac{1-x^n}{1-x} & \text{if} \quad x \neq 1, \\ n & \text{if} \quad x = 1. \end{cases}$$

Note that in the above example the first term of the series was labeled by $k = 0$, rather than $k = 1$. Basically, for notational conveniences, a series may start with any integer value of k, not necessarily $k = 1$.

Example 3.3.2. (harmonic series) The harmonic series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n} + \dots$$

diverges.

Indeed,

$$s_{2n} - s_n = \frac{1}{n+1} + \frac{1}{n+2} + \dots + \frac{1}{2n} > \frac{1}{2}.$$

Example 3.3.3. (generalised harmonic series) The series

$$1 + \frac{1}{2^p} + \frac{1}{3^p} + \frac{1}{4^p} + \dots + \frac{1}{n^p} + \dots$$

for $p \in \mathbb{R}$ converges if $p > 1$ and diverges if $p \leq 1$. The result for $p \leqslant 1$ follows by comparison with the harmonic series (see the comparison test below). The result for $p > 1$ follows by the **integral test** developed further in Section 8.7, see Example 8.7.3.

3.3.1 Series of positive terms

In this subsection, we confine ourselves to considering series whose terms are non-negative numbers; that is,

$$\sum_{k=1}^{\infty} a_k \text{ with } a_k \geq 0 \text{ for every } k \in \mathbb{N}.$$

The main feature of this case is that the sequence of partial sums $s_n = \sum_{k=1}^{n} a_k$ is increasing. Indeed, $s_n - s_{n-1} = a_n \geq 0$. This allows us to formulate a simple criterion of convergence for such a series.

Theorem 3.3.3. *A series of positive terms is convergent if and only if the sequence of partial sums is bounded above.*

Proof. If a series is convergent then the sequence of partial sums is convergent (by definition) and therefore the partial sums are bounded. Conversely, if the sequence of partial sums is bounded above then these are convergent by Theorem 3.2.1 (since the partial sums are increasing); therefore the series is convergent. □

3.3.2 Comparison tests

Here we establish various tests which make it possible to infer convergence or divergence of a series by comparison with another series whose convergence or divergence properties are known.

Theorem 3.3.4. *Let* $\sum_{k=1}^{\infty} a_k, \sum_{k=1}^{\infty} b_k$ *be two series of positive terms. Assume that*

$$(\forall\, n \in \mathbb{N})\ (a_n \leq b_n). \tag{3.3.2}$$

Then

(i) if $\sum_{k=1}^{\infty} b_k$ *is convergent then* $\sum_{k=1}^{\infty} a_k$ *is convergent;*

(ii) if $\sum_{k=1}^{\infty} a_k$ *is divergent then* $\sum_{k=1}^{\infty} b_k$ *is divergent.*

Proof. Let $s_n = \sum_{k=1}^{n} a_k$, $s'_n = \sum_{k=1}^{n} b_k$. Then

$$(\forall\, n \in \mathbb{N})\ (s_n \leq s'_n).$$

So the theorem follows from Theorem 3.3.3. □

Remark 3.3.1. Theorem 3.3.4 holds if instead of (3.3.2) one assumes

$$(\exists\, K > 0)(\forall\, n \in \mathbb{N})\ (a_n \leq K b_n).$$

Now, since for $p \leq 1$,

$$(\forall\, n \in \mathbb{N}) \ \left(\frac{1}{n^p} \geq \frac{1}{n}\right),$$

we obtain one of the assertions of Example 3.3.3 from the above theorem.

The following observation (which follows immediately from the definition of convergence of a series) is important for the next theorem.

Remark 3.3.2. The convergence or divergence of a series is unaffected if a *finite* number of terms are inserted, suppressed, or altered.

Theorem 3.3.5. *Let $(a_n)_n$, $(b_n)_n$ be two sequences of positive numbers. Assume that*

$$\lim_{n\to\infty} \frac{a_n}{b_n} = L \in (0,\, \infty).$$

Then the series $\displaystyle\sum_{k=1}^{\infty} a_k$ is convergent if and only if the series $\displaystyle\sum_{k=1}^{\infty} b_k$ is convergent.

Proof. By definition of the limit of a sequence,

$$(\exists\, N \in \mathbb{N})(\forall\, n > N) \ \left(\frac{1}{2}L < \frac{a_n}{b_n} < \frac{3}{2}L\right).$$

Now the assertion follows from Remarks 3.3.1 and 3.3.2. □

Example 3.3.4. *(i)* The series $\displaystyle\sum_{n=1}^{\infty} \frac{2n}{n^2+1}$ is divergent since

$$\lim_{n\to\infty} \frac{\frac{2n}{n^2+1}}{\frac{1}{n}} = 2$$

and the series

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

is divergent.

(ii) The series $\displaystyle\sum_{n=1}^{\infty} \frac{n}{2n^3+2}$ is convergent since $\displaystyle\lim_{n\to\infty} \frac{\frac{n}{2n^3+2}}{\frac{1}{n^2}} = \frac{1}{2}$ and the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$ is convergent.

3.3.3 Other tests of convergence

Theorem 3.3.6. *(***root test** *or* **Cauchy's test***) Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers. Suppose that*

$$\lim_{n\to\infty} \sqrt[n]{a_n} = l.$$

Then if $l < 1$, the series $\displaystyle\sum_{n=1}^{\infty} a_n$ converges; if $l > 1$, the series $\displaystyle\sum_{n=1}^{\infty} a_n$ diverges. If $l = 1$, no conclusion can be drawn.

Proof. Suppose that $l < 1$. Choose r such that $l < r < 1$. Then

$$(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})[(n > N) \Rightarrow (\sqrt[n]{a_n} < r)].$$

So $a_n < r^n$ for all $n > N$. Convergence follows from comparison with the convergent series $\sum_{n=1}^{\infty} r^n$.
Next suppose that $l > 1$. Then

$$(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})[(n > N) \Rightarrow (\sqrt[n]{a_n} > 1)].$$

So $a_n > 1$ for all $n > N$. Hence the series diverges. □

Theorem 3.3.7. *(***ratio test** *or* **d'Alembert's test***) Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers. Suppose that*

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = l.$$

Then if $l < 1$, the series $\sum\limits_{n=1}^{\infty} a_n$ converges; if $l > 1$, the series $\sum\limits_{n=1}^{\infty} a_n$ diverges. If $l = 1$, no conclusion can be drawn.

Proof. Suppose that $l < 1$. Choose r such that $l < r < 1$. Then

$$(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})[(n > N) \Rightarrow (\frac{a_{n+1}}{a_n} < r)].$$

Therefore

$$a_n = \frac{a_n}{a_{n-1}} \cdot \frac{a_{n-1}}{a_{n-2}} \cdot \cdots \cdot \frac{a_{N+2}}{a_{N+1}} \cdot a_{N+1} < r^{n-N-1} a_{N+1} = \frac{a_{N+1}}{r^{N+1}} \cdot r^n.$$

The convergence follows now from comparison with the convergent series $\sum_{n=1}^{\infty} r^n$.
Next suppose that $l > 1$. Then

$$(\exists\, N \in \mathbb{N})(\forall\, n \in \mathbb{N})[(n > N) \Rightarrow (\frac{a_{n+1}}{a_n} > 1)].$$

Therefore $a_{n+1} > a_n$ for all $n > N$. Hence the series diverges. □

Example 3.3.5. Investigate the convergence of the series

$$\sum_{n=1}^{\infty} \frac{2^n + n}{3^n - n}.$$

Let us use the ratio test. Compute the limit

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{2^{n+1} + n + 1}{2^n + n} \cdot \frac{3^n - n}{3^{n+1} - n - 1} = \frac{2}{3}.$$

Therefore the series converges.

Example 3.3.6. Investigate the convergence of the series

$$\sum_{n=1}^{\infty} 2^{(-1)^n} \left(\frac{1}{2}\right)^n .$$

Let us use the root test. Compute the limit

$$\lim_{n\to\infty} \sqrt[n]{a_n} = \lim_{n\to\infty} 2^{\frac{(-1)^n}{n}} \frac{1}{2} = \frac{1}{2}.$$

We used the fact that $\lim_{n\to\infty} \sqrt[n]{2} = 1$ (see Exercise sheets). Therefore the series converges.

Chapter 4

Limits of functions and continuity

In this chapter the key notion of a *continuous function* is introduced, followed by several important theorems about continuous functions.

4.1 Limits of function

In this chapter we deal exclusively with functions taking values in the set of real numbers (that is, **real-valued functions**).

We aim to define what is meant by

$$f(x) \to b \text{ as } x \to a.$$

Definition 4.1.1. (**Cauchy**) Let $f : D(f) \to \mathbb{R}$ and let $c < a < d$ be such that $(c, a) \cup (a, d) \subseteq D(f)$. We say that $f(x) \to b$ as $x \to a$ and write $\lim\limits_{x\to a} f(x) = b$ if

$$(\forall\, \varepsilon > 0)(\exists\, \delta > 0)(\forall\, x \in D(f))\ \Big[(0 < |x - a| < \delta) \Rightarrow \big(\ |f(x) - b| < \varepsilon\ \big)\Big].$$

Example 4.1.1. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x$. Then for any $a \in \mathbb{R}$,

$$\lim_{x\to a} x = a.$$

Proof. Let $a \in \mathbb{R}$ be arbitrary. Fix $\varepsilon > 0$. *(We need to show that* $0 < |x - a| < \delta$ *implies* $|x - a| < \varepsilon$*. Hence we may choose* $\delta = \varepsilon$*).* Choose $\delta = \varepsilon$. Then

$$(\forall x \in \mathbb{R})\ \Big[(0 < |x - a| < \delta) \Rightarrow (|x - a| < \varepsilon)\Big]. \quad \square$$

Example 4.1.2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Then for any $a \in \mathbb{R}$,

$$\lim_{x\to a} x^2 = a^2.$$

Proof. Let $a \in \mathbb{R}$ be arbitrary. Fix $\varepsilon > 0$. *(We need to show that* $0 < |x-a| < \delta$ *implies* $|x^2 - a^2| < \varepsilon$. *We may assume that* $|x-a| < 1$. *Then* $|x| \le |x-a| + |a| < 1 + |a|$, *and* $|x+a| \le |x| + |a| < 1 + 2|a|$. *Note that* $|x^2 - a^2| = |x-a| \cdot |x+a| < (1+2|a|)\,|x-a|$.*)* Choose $\delta := \min\{1, \frac{\varepsilon}{1+2|a|}\}$. Then

$$(\forall x \in \mathbb{R}) \left[(0 < |x-a| < \delta) \Rightarrow (|x^2 - a^2| < \varepsilon)\right]. \qquad \square$$

The limit of a function may be defined in another way, based on the definition of a limit of a sequence.

Definition 4.1.2. (**Heine**) Let f be be such that $(c,a) \cup (a,d) \subseteq D(f)$. We say that

$$f(x) \to b \text{ as } x \to a$$

if for any sequence $(x_n)_{n \in \mathbb{N}}$ such that

(i) $(\forall n \in \mathbb{N}) \left[(x_n \in (c,d)) \wedge (x_n \neq a)\right]$,

(ii) $x_n \to a$ as $n \to \infty$,

we have

$$\lim_{n \to \infty} f(x_n) = b.$$

In order to make use of the above two definitions of the limit of a function, we need first to make sure that they are compatible. This is the content of the next result.

Theorem 4.1.1. *The Cauchy and Heine definitions for the limit of a function are equivalent.*

Proof. (Cauchy $\Rightarrow$ Heine.)
Assume that $\lim_{x \to a} f(x) = b$ in the sense of Definition (4.1.1). Let $(x_n)_{n \in \mathbb{N}}$ be a sequence satisfying the conditions of Definition (4.1.2). We need to prove that $\lim_{n \to \infty} f(x_n) = b$. Fix $\varepsilon > 0$. Then

$$(\exists \delta > 0)(\forall x \in \mathbb{R}) \left[(0 < |x-a| < \delta) \Rightarrow \left(\, |f(x) - b| < \varepsilon \, \right)\right]. \tag{4.1.1}$$

Fix δ as found above. Then

$$(\exists N \in \mathbb{N})(\forall n \in \mathbb{N}) \left[(n > N) \Rightarrow \left(|x_n - a| < \delta\right)\right]. \tag{4.1.2}$$

Then by (4.1.1) and (4.1.2),

$$(n > N) \Rightarrow \left(\, |f(x_n) - b| < \varepsilon \, \right)$$

which proves that

$$\lim_{n \to \infty} f(x_n) = b.$$

(Heine $\Rightarrow$ Cauchy.)
We argue by contradiction. Suppose that $f(x) \to b$ as $x \to a$ in the sense of Definition (4.1.2) but **not** in the sense of Definition (4.1.1). This means that

$$(\exists \varepsilon > 0)(\forall \delta > 0)(\exists x \in D(f)) \ \Big[(0 < |x-a| < \delta) \wedge \big(\ |f(x) - b| \geq \varepsilon \ \big)\Big].$$

Take $\delta = \frac{1}{n}$. Find x_n such that

$$(0 < |x_n - a| < \frac{1}{n}) \wedge \big(\ |f(x_n) - b| \geq \varepsilon \ \big).$$

We have that $x_n \to a$ as $n \to \infty$. Therefore by Definition (4.1.2), $f(x_n) \to b$ as $n \to \infty$. This is a contradiction. $\square$

Definition 4.1.3. We say that A is the limit of the function $f : D(f) \to \mathbb{R}$ as $x \to +\infty$, and write $\lim\limits_{x \to +\infty} f(x) = A$, if $\exists K \in \mathbb{R}$ such that $(K, \infty) \subseteq D(f)$ and

$$(\forall \varepsilon > 0)(\exists M > K)(\forall x \in D(f)) \ \Big[(x > M) \Rightarrow \big(\ |f(x) - A| < \varepsilon \ \big)\Big].$$

The limit

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

is defined similarly.

Example 4.1.3.

$$\lim_{x \to \infty} \frac{1}{x} = 0.$$

Proof. We have to prove that

$$(\forall \varepsilon > 0)(\exists M > 0)(\forall x \in D(f)) \ \Big[(x > M) \Rightarrow \Big(\left|\frac{1}{x}\right| < \varepsilon \Big)\Big]$$

It is sufficient to choose M with $M\varepsilon > 1$; such an M exists by the Archimedian Principle. $\square$

The process of taking limits of functions behaves in much the same way as the process of taking limits of sequences. For example, the limit of a function at a point is unique - we prove this using the Heine definition.

Theorem 4.1.2. *If* $\lim\limits_{x \to a} f(x) = A$ *and* $\lim\limits_{x \to a} f(x) = B$ *then* $A = B$.

Proof. Let $x_n \to a$ as $n \to \infty$. Then $f(x_n) \to A$ as $n \to \infty$ and $f(x_n) \to B$ as $n \to \infty$. From the uniqueness of the limit of a sequence it follows that $A = B$. $\square$

Sometimes it is more convenient to use the Cauchy definition.

Theorem 4.1.3. *Let* $\lim\limits_{x \to a} f(x) = A$. *Let* $B > A$. *Then*

$$(\exists \delta > 0)(\forall x \in D(f)) \ \Big[(0 < |x-a| < \delta) \Rightarrow \big(\ f(x) < B \ \big)\Big].$$

Proof. Using $\varepsilon = B - A > 0$ in the Cauchy definition of the limit we have

$$(\exists \delta > 0)(\forall x \in D(f)) \; \Big[(0 < |x-a| < \delta) \Rightarrow \big(\; |f(x) - A| < B - A \; \big) \Big]$$

which implies that $f(x) - A < B - A$, or $f(x) < B$. □

The following two theorems can be proved using the Heine definition for the limit of a function at a point, and corresponding properties for the limit of a sequence. The proofs are left as exercises.

Theorem 4.1.4. *Let $\lim\limits_{x\to a} f(x) = A$ and $\lim\limits_{x\to a} g(x) = B$. Then*

(i) $\lim\limits_{x\to a} \big(f(x) + g(x) \big) = A + B$;

(ii) $\lim\limits_{x\to a} \big(f(x) \cdot g(x) \big) = A \cdot B$.

If in addition $B \neq 0$, then

(iii) $\lim\limits_{x\to a} \left(\dfrac{f(x)}{g(x)} \right) = \dfrac{A}{B}$.

Theorem 4.1.5. *Let $\lim\limits_{x\to a} f(x) = A$ and $\lim\limits_{x\to a} g(x) = B$. Suppose that*

$$(\exists \delta > 0) \; \Big[\{ x \in \mathbb{R} \,|\, 0 < |x-a| < \delta \} \subseteq D(f) \cap D(g) \Big]$$

and

$$\big(0 < |x-a| < \delta \big) \Rightarrow \big[f(x) \leq g(x) \big].$$

Then $A \leq B$; that is,

$$\lim_{x\to a} f(x) \leq \lim_{x\to a} g(x).$$

Using Theorem 4.1.4 one can easily compute limits of some functions.

Example 4.1.4.

$$\lim_{x\to 2} \frac{x^3 + 2x^2 - 7}{2x^3 - 4} = \frac{\lim\limits_{x\to 2} x^3 + \lim\limits_{x\to 2} 2x^2 - \lim\limits_{x\to 2} 7}{\lim\limits_{x\to 2} 2x^3 - \lim\limits_{x\to 2} 4} = \frac{8 + 8 - 7}{16 - 4} = \frac{3}{4}.$$

Definition 4.1.4. (**one-sided limits**)

(i) Let f be defined on an interval $(a, d) \subseteq \mathbb{R}$. We say that $f(x) \to b$ as $x \to a+$ and write $\lim\limits_{x\to a+} f(x) = b$ if

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in (a, d)) \; \Big[(0 < x - a < \delta) \Rightarrow \big(\; |f(x) - b| < \varepsilon \; \big) \Big].$$

(ii) Let f be defined on an interval $(c, a) \subseteq \mathbb{R}$. We say that $f(x) \to b$ as $x \to a-$ and write $\lim\limits_{x\to a-} f(x) = b$ if

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in (c, a)) \; \Big[(-\delta < x - a < 0) \Rightarrow \big(\; |f(x) - b| < \varepsilon \; \big) \Big].$$

4.2 Continuous functions

Definition 4.2.1. Let $a \in \mathbb{R}$. Let a function f be defined in a neighborhood of a. Then the function f is said to be **continuous at** a if

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

The above definition may be reformulated in the following way.

Definition 4.2.2. A function f is said to be **continuous at a point** $a \in \mathbb{R}$ if f is defined on an interval (c, d) containing a and

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in (c, d)) \; \Big[(|x - a| < \delta) \Rightarrow \big(\; |f(x) - f(a)| < \varepsilon \; \big) \Big].$$

Note the difference between this definition and the definition of the limit: the function f needs to be defined at a.

Using the above definition, it is easy to specify what it means for a function f to be discontinuous at a point a. A function f is **discontinuous** at $a \in \mathbb{R}$ if either

$$f \text{ is not defined in any neighbourhood } (c, d) \text{ containing } a,$$

or if

$$(\exists\, \varepsilon > 0)(\forall\, \delta > 0)(\exists\, x \in (c, d)) \; \Big[(|x - a| < \delta) \wedge \big(\; |f(x) - f(a)| \geq \varepsilon \; \big) \Big].$$

An equivalent way to define continuity at a point is to use the Heine definition of the limit.

Definition 4.2.3. A function f is said to be **continuous at a point** $a \in \mathbb{R}$ if f is defined on an interval (c, d) containing a and for any sequence $(x_n)_{n \in \mathbb{N}}$ such that

(i) $(\forall n \in \mathbb{N}) \; \Big[x_n \in (c, d) \Big]$,

(ii) $x_n \to a$ as $n \to \infty$,

we have

$$\lim_{n \to \infty} f(x_n) = f(a).$$

Example 4.2.1. Let $c \in \mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = c$ for any $x \in \mathbb{R}$. Then f is continuous at any point in $\mathbb{R}$.

Example 4.2.2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x$ for any $x \in \mathbb{R}$. Then f is continuous at any point in $\mathbb{R}$.

The following theorem follows easily from the definition of continuity and properties of limits.

Theorem 4.2.1. *Let f and g be continuous at $a \in \mathbb{R}$. Then* *(i)* $f+g$ *is continuous at* a.

(ii) $f \cdot g$ *is continuous at* a.

Moreover, if $g(a) \neq 0$, *then*

(iii) $\dfrac{f}{g}$ *is continuous at* a.

It is a consequence of this last theorem and the preceeding two examples that the **rational function**

$$f(x) = \frac{a_0 x^n + a_1 x^{n-1} + \cdots + a_n}{b_0 x^m + b_1 x^{m-1} + \cdots + b_m}$$

is continuous at every point of its domain of definition.

Theorem 4.2.2. *Let* g *be continuous at* $a \in \mathbb{R}$ *and* f *continuous at* $b = g(a) \in \mathbb{R}$. *Then* $f \circ g$ *is continuous at* a.

Proof. Fix $\varepsilon > 0$. Since f is continuous at b,

$$(\exists \delta > 0)(\forall y \in D(f)) \, \Big[(|y-b| < \delta) \Rightarrow \big(\, |f(y) - f(b)| < \varepsilon \, \big) \Big].$$

Fix this $\delta > 0$. From the continuity of g at a,

$$(\exists \gamma > 0)(\forall x \in D(g)) \, \Big[(|x-a| < \gamma) \Rightarrow \big(\, |g(x) - g(a)| < \delta \, \big) \Big].$$

From the above, it follows that

$$(\forall \varepsilon > 0)(\exists \gamma > 0)(\forall x \in D(g)) \, \Big[(|x-a| < \gamma) \Rightarrow \big(\, |f(g(x)) - f(g(a))| < \varepsilon \, \big) \Big].$$

This proves continuity of $f \circ g$ at a. $\square$

Another useful characterization of continuity of a function f at a point a is the following: f is **continuous at a point** $a \in \mathbb{R}$ if and only if

$$\lim_{x \to a-} f(x) = \lim_{x \to a+} f(x) = f(a).$$

In other words, the one-sided limits exist, are equal and equal the value of the function at a.

Theorem 4.2.3. *Let* f *be continuous at* $a \in \mathbb{R}$. *Let* $f(a) < B$. *Then there exists a neighbourhood of* a *such that* $f(x) < B$ *for all points* x *belonging to this neighbourhood.*

Proof. Take $\varepsilon = B - f(a)$ in Definition 4.2.2. Then

$$(\exists \delta > 0)(\forall x \in (c,d)) \, \Big[(|x-a| < \delta) \Rightarrow \big(\, |f(x) - f(a)| < \varepsilon \, \big) \Big]. \quad \square$$

Remark 4.2.1. A similar fact is true for the case $f(x) > B$.

Definition 4.2.4. Let $f : D(f) \to \mathbb{R}$, $D(f) \subseteq \mathbb{R}$. f is said to be **bounded** if $Ran(f)$ is a bounded subset of $\mathbb{R}$.

Example 4.2.3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \frac{1}{1+x^2}$. Then $Ran(f) = (0, 1]$, so f is bounded.

Definition 4.2.5. Let $A \subseteq D(f)$. f is said to be **bounded above** on A if

$$(\exists K \in \mathbb{R})(\forall x \in A)\bigl(f(x) \leq K\bigr).$$

Remark 4.2.2. Boundedness below and boundedness on a set are defined analogously.

Theorem 4.2.4. *If f is continuous at a, then there exists $\delta > 0$ such that f is bounded on the interval $(a - \delta, a + \delta)$.*

Proof. Since $\lim\limits_{x \to a} f(x) = f(a)$,

$$(\exists \delta > 0)(\forall x \in D(f))\ \bigl[(|x - a| < \delta) \Rightarrow \bigl(|f(x) - f(a)| < 1\bigr)\bigr].$$

So on the interval $(a - \delta, a + \delta)$,

$$f(a) - 1 < f(x) < f(a) + 1. \quad \square$$

4.3 Continuous functions on a closed interval

In the previous section we dealt with functions which were continuous at a point. Here we consider functions which are continuous at every point of an interval $[a, b]$. We say that such functions are continuous on $[a, b]$. In contrast to the previous section, we shall be interested here in the global behaviour of such functions. We first define this notion formally.

Definition 4.3.1. Let $[a, b] \subseteq \mathbb{R}$ be a closed interval and $f : D(f) \longrightarrow \mathbb{R}$ a function with $[a, b] \subseteq D(f)$. We say that f is a **continuous** function on $[a, b]$ if:

(i) f is continuous at every point of (a, b), and

(ii) $\lim_{x \to a^+} f(x) = f(a)$ and $\quad \lim_{x \to b^-} f(x) = f(b)$.

Our first theorem describes the intermediate-value property for continuous function.

Theorem 4.3.1. (**intermediate value theorem**) *Let f be a continuous function on a closed interval $[a, b] \subseteq \mathbb{R}$. Suppose that $f(a) < f(b)$. Then*

$$\bigl(\forall \eta \in [f(a), f(b)]\bigr)\bigl(\exists x \in [a, b]\bigr)\Bigl(\ f(x) = \eta\ \Bigr).$$

Proof. If $\eta = f(a)$ or $\eta = f(b)$ then there is nothing to prove. Fix $\eta \in (f(a), f(b))$. Let us introduce the set

$$A := \{x \in [a, b] \mid f(x) < \eta\}.$$

The set A is not empty, as $a \in A$. The set A is bounded above (by b). Therefore, the supremum $\xi := \sup A$ exists and $\xi \in [a, b]$. Our aim is to prove that $f(\xi) = \eta$. We do this by ruling out the possibilities, $f(\xi) < \eta$ and $f(\xi) > \eta$. Note that $\xi \in (a,b)$ by continuity of f at a and b.

First, let us assume that $f(\xi) < \eta$. Then, by Theorem 4.2.3, we have that

$$(\exists\, \delta > 0)(\forall x \in D(f))\ \Big[\big(x \in (\xi - \delta, \xi + \delta)\big) \Rightarrow \big(f(x) < \eta\big)\Big].$$

Therefore,

$$(\exists\, x_1 \in \mathbb{R})\ \Big[\big(x_1 \in (\xi, \xi + \delta)\big) \wedge \big(f(x_1) < \eta\big)\Big].$$

In other words,

$$(\exists\, x_1 \in \mathbb{R})\ \Big[\big(x_1 > \xi\big) \wedge \big(x_1 \in A\big)\Big].$$

This contradicts the fact that ξ is an upper bound of A.

Next, let us assume that $f(\xi) > \eta$. Then, by the remark after Theorem 4.2.3, we have that

$$(\exists\, \delta > 0)\big[\ D(f) \cap (\xi - \delta, \xi + \delta) \subseteq A^c\big] \quad (A^c \text{ is the complement of } A).$$

This contradicts the fact that $\xi = \sup A$. □

The next theorem establishes boundedness for continuous functions defined on a closed interval.

Theorem 4.3.2. *(***boundedness theorem***) Let f be continuous on $[a,b]$. Then f is bounded on $[a,b]$.*

Proof. Let us introduce the set

$$A := \{x \in [a,b] \,\big|\, f \text{ is bounded on } [a,x]\ \}.$$

Note that

1. $A \neq \varnothing$ as $a \in A$;
2. A is bounded (by b).

This means that $\xi := \sup A$ exists. We prove

STEP 1: $\xi = b$.

First note that that $\xi > a$. For, by (left-)continuity of f at a,

$$(\exists\, \delta > 0)(\forall\, x \in D(f))\,\big[(0 \le x - a < \delta) \Rightarrow \big(\ |f(x) - f(a)| < 1\ \big)\big].$$

So f is bounded on $[a, a + \delta/2]$, for example, and $\xi \geqslant a + \delta/2$. This shows that $\xi > a$.

Suppose that $\xi < b$. By Theorem 4.2.4,

$$(\exists\, \delta > 0)\big[\ f \text{ is bounded on } \ (\xi - \delta, \xi + \delta)\ \big]. \tag{4.3.3}$$

By definition of the supremum,

$$\big(\exists\, x_1 \in (\xi - \delta, \xi]\ \big)\big[\ f \text{ is bounded on } \ [a, x_1]\ \big].$$

Also, from (4.3.3), it follows that

$$\big(\exists\, x_2 \in (\xi, \xi + \delta)\ \big)\big[\ f \text{ is bounded on } \ [x_1, x_2]\ \big].$$

Therefore f is bounded on $[a, x_2]$ where $x_2 > \xi$, which contradicts the fact that ξ is the supremum of A. This proves that $\xi = b$.

(Note that this does not complete the proof, as the supremum may not belong to the set: it is possible that $\xi \notin A$).

STEP 2: $b \in A$.

From the continuity of f at b, it follows that

$$(\exists\, \delta_1 > 0)\big[\ f \text{ is bounded on } \ (b - \delta_1, b]\ \big].$$

By definition of the supremum,

$$(\exists x_3 \in (b - \delta_1, b]\ )\big[\ f \text{ is bounded on } \ [a, x_3]\ \big].$$

Therefore f is bounded on $[a, b]$. □

The last theorem asserted that the range $Ran(f)$ of a continuous function f restricted to a closed interval is a bounded subset of $\mathbb{R}$. Consequently, $Ran(f)$ possesses a supremum and infimum. The next theorem asserts that, in fact, these values are attained; in other words, there exist points in $[a, b]$ where the function attains its maximum and minimum values.

Theorem 4.3.3. *Let f be continuous on $[a, b] \subseteq \mathbb{R}$. Then*

$$(\exists\, y \in [a, b])(\forall\, x \in [a, b])\,\big(f(x) \le f(y)\big).$$

(In other words, $f(y) = \max\limits_{x \in [a,b]} f(x)$.) Similarly,

$$(\exists\, z \in [a, b])(\forall\, x \in [a, b])\,\big(f(x) \ge f(z)\big).$$

Proof. Let us introduce the set of values attained by f on $[a, b]$, namely,

$$F := \{f(x) \,\big|\, x \in [a, b]\}.$$

Then $F \neq \varnothing$ and by Theorem 4.3.2, F is bounded. Therefore the supremum $\alpha = \sup F$ exists. We prove that there exists $y \in [a, b]$ such that $f(y) = \alpha$. We do this by contradiction.

Suppose, on the contrary, that

$$(\ \forall\, x \in [a, b]\ )\big(f(x) < \alpha).$$

Define

$$g(x) := \frac{1}{\alpha - f(x)}, \; x \in [a,b].$$

Since the denominator is never zero on $[a,b]$, g is continuous and, by Theorem 4.3.2, g is bounded on $[a,b]$.
At the same time, by definition of the supremum, for any $\varepsilon > 0$,

$$(\exists\, x \in [a,b])\big(f(x) > \alpha - \varepsilon\big);$$

in other words, $\alpha - f(x) < \varepsilon$. So $g(x) > \dfrac{1}{\varepsilon}$. This proves that

$$(\forall\, \varepsilon > 0)(\exists\, x \in [a,b])\Big(g(x) > \frac{1}{\varepsilon}\Big).$$

Therefore g is unbounded on $[a,b]$, which contradicts the above. □

4.4 Inverse functions

In this section, we discuss the inverse function of a continuous function. Recall that a function possesses an inverse if and only if it is a bijection.

It turns out that a continuous bijection defined on an interval $[a,b]$ is monotone (that is, either strictly increasing or strictly decreasing) (see Exercises). The notions of (strictly) increasing and (strictly) decreasing parallel the corresponding notions for sequences. Here is a precise definition.

Definition 4.4.1. A function f defined on $[a,b]$ is said to be **increasing** on $[a,b]$ if

$$(\forall x_1, x_2 \in [a,b]) \Big[(x_1 \leqslant x_2) \Rightarrow \big(f(x_1) \leqslant f(x_2)\big)\Big].$$

It is said to be **strictly increasing** on $[a,b]$ if

$$(\forall x_1, x_2 \in [a,b]) \Big[(x_1 < x_2) \Rightarrow \big(f(x_1) < f(x_2)\big)\Big].$$

The notions of **decreasing** and **strictly decreasing** function are defined similarly.

Theorem 4.4.1. *Let f be continuous and strictly increasing on $[a,b]$. Let $f(a) = c,\ f(b) = d$. Then there exists a function $g : [c,d] \to [a,b]$ which is continuous and strictly increasing such that*

$$(\forall y \in [c,d]) \Big[f(g(y)) = y\Big].$$

Proof. Let $\eta \in [c,d]$. By Theorem 4.3.1,

$$(\exists\, \xi \in [a,b]) \Big[f(\xi) = \eta\Big].$$

There is only one such value ξ since f is increasing. (Prove this.) The inverse function g is defined by

$$\xi = g(\eta).$$

It is easy to see that g is strictly increasing. Indeed, let $y_1 < y_2$, $y_1 = f(x_1)$ and $y_2 = f(x_2)$. Suppose also that $x_2 \leq x_1$. Since f is strictly increasing it follows that $y_2 \leq y_1$. This is a contradiction.

Now let us prove that g is continuous. Let $y_0 \in (c,d)$. Then

$$\big(\exists x_0 \in (a,b)\big)\,\big[y_0 = f(x_0)\big],$$

or, in other words,

$$x_0 = g(y_0).$$

Let $\varepsilon > 0$. We assume also that ε is small enough such that $[x_0 - \varepsilon, x_0 + \varepsilon] \subseteq (a,b)$. Let $y_1 = f(x_0 - \varepsilon),\ y_2 = f(x_0 + \varepsilon)$. Since g is increasing, we have that

$$\big[\ y \in (y_1, y_2)\ \big] \Rightarrow \big[\ x = g(y) \in (x_0 - \varepsilon, x_0 + \varepsilon)\ \big].$$

Take $\delta = \min\{y_2 - y_0, y_0 - y_1\}$. Then

$$\big[\ |y - y_0| < \delta\ \big] \Rightarrow \big[\ |g(y) - g(y_0)| < \varepsilon\ \big].$$

Continuity at the endpoints of the interval can be established similarly. $\square$

Chapter 5

Differential Calculus

In this chapter the notion of a *differentiable function* is developed, and several important theorems about differentiable functions are presented.

5.1 Definition of derivative. Elementary properties

Definition 5.1.1. Let f be defined in a δ-neighbourhood $(a-\delta, a+\delta)$ of $a \in \mathbb{R}$ $(\delta > 0)$. We say that f is **differentiable** at a if the limit

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

exists in $\mathbb{R}$. This limit, denoted by $f'(a)$, is called the **derivative** of f at a.

Example 5.1.1. Let $c \in \mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = c$. Then f is differentiable at any $x \in \mathbb{R}$ and $f'(x) = 0$.

Proof.

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

□

Example 5.1.2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x$. Then f is differentiable at any $x \in \mathbb{R}$ and $f'(x) = 1$.

Proof.

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = \lim_{h \to 0} \frac{a + h - a}{h} = 1.$$

□

Example 5.1.3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Then f is differentiable at any $x \in \mathbb{R}$ and $f'(x) = 2x$.

Proof.

$$\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}=\lim_{h\to 0}\frac{(a+h)^2-a^2}{h}=\lim_{h\to 0}\frac{2ah+h^2}{h}=\lim_{h\to 0}(2a+h)=2a.$$

□

Example 5.1.4. Let $n\in\mathbb{N}$ and $f:\mathbb{R}\to\mathbb{R}$ be defined by $f(x)=x^n$. Then f is differentiable at any $x\in\mathbb{R}$ and $f'(x)=nx^{n-1}$.

This may be proved using mathematical induction together with the product rule (see below). It is left as an exercise.

Example 5.1.5. Let $f:\mathbb{R}\to\mathbb{R}$ be defined by $f(x)=|x|$. Then f is differentiable at any $x\in\mathbb{R}-\{0\}$. But f is not differentiable at 0.

Proof. If $x>0$, then

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=1.$$

If $x<0$, then

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=-1.$$

Therefore, the derivative does not exist at 0, as

$$\lim_{h\to 0+}\frac{f(x+h)-f(x)}{h}\neq\lim_{h\to 0-}\frac{f(x+h)-f(x)}{h}.$$

□

Note that the function f in the above example is continuous at 0: thus, continuity does **not** imply differentiability. However, the converse is true.

Theorem 5.1.1. *If f is differentiable at a, then f is continuous at a.*

Proof.

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

Hence,

$$\begin{aligned}\lim_{x\to a}\big(f(x)-f(a)\big)&=\lim_{h\to 0}\big(f(a+h)-f(a)\big)\\&=\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot h=f'(a)\cdot\lim_{h\to 0}h=0.\end{aligned}$$

Therefore,

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

□

Remark 5.1.1. If f is differentiable at $a \in \mathbb{R}$ then there exists a function $\alpha(x)$ such that $\lim_{x\to a} \alpha(x) = 0$ and

$$f(x) = f(a) + f'(a)(x-a) + \alpha(x)(x-a).$$

Indeed, define

$$\alpha(x) := \frac{f(x) - f(a)}{x-a} - f'(a).$$

Then $\alpha(x) \to 0$ as $x \to a$ and $f(x) = f(a) + f'(a)(x-a) + \alpha(x)(x-a)$.

This enables one to re-interpret the formula in the above Remark as follows. If f is differentiable at $a \in \mathbb{R}$, then one can write for the value of $f(x = a+h)$, that is "near" a:

$$f(a+h) = f(a) + f'(a)h + o(h),$$

where the notation $o(h)$ reads as "little o of h", and denotes *any function* which has the following property: $\lim_{h\to 0} \frac{o(h)}{h} = 0$. (A more explicit expression for the quantity $o(h)$ can be obtained, under the assumption of the existence of higher order derivatives of f using the Taylor theorem in Section 5.3.) The latter formula can be taken as well for the definition of differentiability, the derivative $f'(a)$ being defined as the *coefficient*, multiplying h in the above expression for $f(a+h)$, the quantity $o(h)$ being referred to as the "error term", since as long as $f'(a) \neq 0$, the absolute value of $o(h)$ is much smaller than that of $f'(a)h$ and becomes negligible as $h \to 0$. So, $f'(a)h$ can be understood as a *linear function* of h, approximating the quantity $\Delta f = f(a+h) - f(a)$. This viewpoint enables one, in principle, to extend the notion of differentiability to functions of more than one real variable.

Theorem 5.1.2. *If f and g are differentiable at a, then $f+g$ is also differentiable at a, and*

$$(f+g)'(a) = f'(a) + g'(a).$$

The proof is left as an exercise.

Theorem 5.1.3. (**product rule**). *If f and g are differentiable at a, then $f \cdot g$ is also differentiable at a, and*

$$(f \cdot g)'(a) = f'(a) \cdot g(a) + f(a) \cdot g'(a).$$

Proof.

$$\begin{aligned}
&\lim_{h\to 0} \frac{(f\cdot g)(a+h) - f\cdot g(a)}{h} = \lim_{h\to 0} \frac{f(a+h)g(a+h) - f(a)g(a)}{h} \\
=&\lim_{h\to 0}\left[\frac{f(a+h)[g(a+h) - g(a)]}{h} + \frac{[f(a+h) - f(a)]g(a)}{h}\right] \\
=&\lim_{h\to 0} f(a+h) \lim_{h\to 0} \frac{g(a+h) - g(a)}{h} + \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} \lim_{h\to 0} g(a) \\
=&f'(a)\cdot g(a) + f(a)\cdot g'(a).
\end{aligned}$$

□

Theorem 5.1.4. *If g is differentiable at a and $g(a) \neq 0$, then $\phi = 1/g$ is also differentiable at a, and*

$$\phi'(a) = (1/g)'(a) = -\frac{g'(a)}{[g(a)]^2}.$$

Proof. The result follows from

$$\frac{\phi(a+h) - \phi(a)}{h} = \frac{g(a) - g(a+h)}{hg(a)g(a+h)}.$$

□

Theorem 5.1.5. *(***quotient rule***). If f and g are differentiable at a and $g(a) \neq 0$, then $\phi = f/g$ is also differentiable at a, and*

$$\phi'(a) = \left(\frac{f}{g}\right)'(a) = \frac{f'(a) \cdot g(a) - f(a) \cdot g'(a)}{[g(a)]^2}.$$

Proof. This follows from Theorems 5.1.3 and 5.1.4. □

Theorem 5.1.6. *(***chain rule***) If g is differentiable at $a \in \mathbb{R}$ and f is differentiable at $g(a)$, then $f \circ g$ is differentiable at a and*

$$(f \circ g)'(a) = f'(g(a)) \cdot g'(a).$$

Proof. By definition of the derivative and Remark 5.1.1, we have

$$f(y) - f(y_0) = f'(y_0)(y - y_0) + \alpha(y)(y - y_0),$$

where $\alpha(y) \to 0$ as $y \to y_0$. Replace y and y_0 in the above equality by $y = g(x)$ and $y_0 = g(a)$, and divide both sides by $x - a$, to obtain

$$\frac{f(g(x)) - f(g(a))}{x - a} = f'(g(a))\frac{g(x) - g(a)}{x - a} + \alpha(g(x))\frac{g(x) - g(a)}{x - a}.$$

By Theorem 5.1.1, g is continuous at a. Hence $y = g(x) \to g(a) = y_0$ as $x \to a$, and $\alpha(g(x)) \to 0$ as $x \to a$. Passing to the limit $x \to a$ in the above equality yields the required result. □

Example 5.1.6. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = (x^2 + 1)^{100}$. Then f is differentiable at every point in $\mathbb{R}$ and $f'(x) = 100(x^2 + 1)^{99} \cdot 2x$.

In the next theorem, we establish a relation between the derivative of an invertible function and the derivative of the inverse function.

Theorem 5.1.7. *Let f be continuous and strictly increasing on (a, b). Suppose that, for some $x_0 \in (a, b)$, f is differentiable at x_0 and $f'(x_0) \neq 0$. Then the inverse function $g = f^{-1}$ is differentiable at $y_0 = f(x_0)$ and*

$$g'(y_0) = \frac{1}{f'(x_0)}.$$

Proof. According to Remark 5.1.1, we may write

$$\begin{aligned} y - y_0 &= f(g(y)) - f(g(y_0)) \\ &= f'(g(y_0))(g(y) - g(y_0)) + \alpha(g(y))(g(y) - g(y_0)), \end{aligned}$$

where $\alpha(g(y)) \to 0$ in case $g(y) \to g(y_0)$. However, as g is continuous at y_0, it follows that $g(y) \to g(y_0)$ as $y \to y_0$; hence, $\alpha(g(y)) \to 0$ as $y \to y_0$. Therefore, we have

$$\begin{aligned} \frac{g(y) - g(y_0)}{y - y_0} &= \frac{g(y) - g(y_0)}{f'(g(y_0))(g(y) - g(y_0)) + \alpha(g(y))(g(y) - g(y_0))} \\ &= \frac{1}{f'(g(y_0)) + \alpha(g(y))} \to \frac{1}{f'(g(y_0))} \end{aligned}$$

as $y \to y_0$. □

Example 5.1.7. Let $n \in \mathbb{N}$ and $f : \mathbb{R}_+ \to \mathbb{R}_+$ be defined by $f(x) = x^n$. Then the inverse function is $g(y) = y^{1/n}$. We know that $f'(x) = nx^{n-1}$. Hence, by the previous theorem,

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{nx^{n-1}} = \frac{1}{n} \cdot \frac{1}{y^{\frac{n-1}{n}}} = \frac{1}{n} \cdot y^{\frac{1}{n}-1}.$$

One-sided derivatives

In a manner similar to the definition of the one-sided limit, we may also define the **left** and **right derivatives** of f at a via

$$f'_-(a) = \lim_{h \to 0-} \frac{f(a+h) - f(a)}{h}, \quad f'_+(a) = \lim_{h \to 0+} \frac{f(a+h) - f(a)}{h}.$$

5.2 Theorems on differentiable functions

In this section, we investigate properties of differentiable functions on intervals.

Theorem 5.2.1. *Let f be a function defined on (a, b). If f attains its maximum (or minimum) at $(x_0, f(x_0))$ for some $x_0 \in (a, b)$, and f is differentiable at x_0, then $f'(x_0) = 0$.*

Note that we do not assume differentiability (or even continuity) of f at any other point apart from x_0.

Proof. We prove the theorem in the case that f attains its maximum at $(x_0, f(x_0))$; the proof is similar in the other case.
Let $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subseteq (a, b)$. For $0 < h < \delta$, $f(x_0 + h) \leqslant f(x_0)$ and

$$\frac{f(x_0 + h) - f(x_0)}{h} \leq 0,$$

so that

$$f'_+(x_0) = \lim_{h \to 0+} \frac{f(x_0 + h) - f(x_0)}{h} \leq 0.$$

For $-\delta < h < 0$, $f(x_0 + h) \leqslant f(x_0)$ and

$$\frac{f(x_0+h) - f(x_0)}{h} \geq 0,$$

so that

$$f'_-(x_0) = \lim_{h \to 0-} \frac{f(x_0+h) - f(x_0)}{h} \geq 0.$$

In sum,

$$0 \leqslant f'_-(x_0) = f'(x_0) = f'_+(x_0) \leqslant 0.$$

So $f'(x_0) = 0$. $\square$

Note that the converse statement is false. Here is a simple counter-example. Consider

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = x^3.$$

Then $f'(0) = 0$. But f does not attain its maximum or minimum at 0 on any interval containing 0.

Theorem 5.2.2. *(***Rolle's theorem***) If f is continuous on $[a, b]$ and differentiable on (a, b), and $f(a) = f(b)$, then*

$$\big(\exists\, x_0 \in (a, b)\big) \left[f'(x_0) = 0\right].$$

Proof. It follows from the continuity of f that f attains its maximum and minimum value on $[a, b]$. Suppose that f attains either its maximum or its minimum at an **interior** point, $x_0 \in (a, b)$, say. Then, by Theorem 5.2.1, $f'(x_0) = 0$, and the result follows. Now suppose that f attains neither its maxima nor its minima on (a, b) (the **interior** of $[a, b]$). This means that f attains both its maxima and its minima at the end-points of the interval, namely, a and b. However, $f(a) = f(b)$, so that the maximum and minimum values of f must be equal. Therefore, f is constant. Hence $f'(x) = 0$ for all $x \in (a, b)$. $\square$

Theorem 5.2.3. *(***mean value theorem***) If f is continuous on $[a, b]$ and differentiable on (a, b), then*

$$\Big(\exists\, x_0 \in (a, b)\Big) \left[\ f'(x_0) = \frac{f(b) - f(a)}{b - a}\ \right].$$

Proof. Set

$$g(x) := f(x) - \left[\frac{f(b) - f(a)}{b - a}\right](x - a).$$

Then g is continuous on $[a, b]$ and differentiable on (a, b), and

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

Moreover, $g(a) = f(a)$, and

$$g(b) = f(b) - \left[\frac{f(b) - f(a)}{b - a}\right](b - a) = f(a).$$

Therefore, by Rolle's theorem,

$$(\exists\, x_0 \in (a,b))\, [g'(x_0) = 0].$$

$\square$

Corollary 5.2.1. *If f is defined on an interval and $f'(x) = 0$ for all x in the interval, then f is constant there.*

Proof. Let a and b be any two points in the interval with $a \neq b$. Then, by the mean value theorem, there is a point x in (a,b) such that

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

But $f'(x) = 0$ for all x in the interval, so

$$0 = \frac{f(b) - f(a)}{b - a},$$

and consequently, $f(b) = f(a)$. Thus the value of f at any two points is the same and f is constant on the interval. $\square$

Corollary 5.2.2. *If f and g are defined on the same interval and $f'(x) = g'(x)$ there, then $f = g + c$ for some number $c \in \mathbb{R}$.*

The proof is left as an exercise.

Corollary 5.2.3. *If $f'(x) > 0$ (resp. $f'(x) < 0$) for all x in some interval, then f is increasing (resp. decreasing) on this interval.*

Proof. Consider the case $f'(x) > 0$. Let a and b be any two points in the interval, with $a < b$. By the mean value theorem, there is a point x in (a,b) such that

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

But $f'(x) > 0$ for all x in the interval, so that

$$\frac{f(b) - f(a)}{b - a} > 0.$$

Since $b - a > 0$, it follows that $f(b) > f(a)$, which proves that f is increasing on the interval. The case $f'(x) < 0$ is left as an exercise. $\square$

The next theorem is a generalisation of the mean value theorem. It is of interest because of its use in applications.

Theorem 5.2.4. *(***Cauchy mean value theorem***). If f and g are continuous on $[a,b]$ and differentiable on (a,b), then*

$$\big(\exists\, x_0 \in (a,b)\big)\Big[\ [f(b) - f(a)]g'(x_0) = [g(b) - g(a)]f'(x_0)\ \Big].$$

(If $g(b) \neq g(a)$, and $g'(x_0) \neq 0$, the above equality can be rewritten as

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(x_0)}{g'(x_0)}.$$

Note that if $g(x) = x$, we obtain the mean value theorem.)

Proof. Let $h : [a, b] \to \mathbb{R}$ be defined by

$$h(x) = [f(b) - f(a)]g(x) - [g(b) - g(a)]f(x).$$

Then $h(a) = f(b)g(a) - f(a)g(b) = h(b)$, so that h satisfies Rolle's theorem. Therefore,

$$\big(\exists x_0 \in (a, b)\big)\Big[\ 0 = h'(x_0) = [f(b) - f(a)]g'(x_0) - [g(b) - g(a)]f'(x_0)\ \Big].$$

$\square$

Theorem 5.2.5. *(***L'Hôpital's rule***) Let f and g be differentiable functions in the neighbourhood of $a \in \mathbb{R}$. Assume that*

(i) $g'(x) \neq 0$ *in some neighbourhood of a;*

(ii) $f(a) = g(a) = 0$*;*

(ii) $\lim_{x \to a} \frac{f'(x)}{g'(x)}$ *exists.*

Then

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}$$

Proof. By the Cauchy mean value theorem,

$$\frac{f(a + h) - f(a)}{g(a + h) - g(a)} = \frac{f'(a + th)}{g'(a + th)}$$

for some $0 < t < 1$. Now pass to the limit $h \to 0$ to get the result. $\square$

L'Hôpital's rule is a useful tool for computing limits.

Example 5.2.1. Let $m, n \in \mathbb{N}$, $0 \neq a \in \mathbb{R}$. Then

$$\lim_{x \to a} \frac{x^m - a^m}{x^n - a^n} = \lim_{x \to a} \frac{mx^{m-1}}{nx^{n-1}} = \frac{m}{n}a^{m-n}.$$

5.3 Approximation by polynomials. Taylor's Theorem

The class of polynomials consists of all those functions f which may be written in the form

$$f(x) = a_0\, x^0 + a_1\, x^1 + a_2\, x^2 + \cdots + a_n\, x^n$$

where $a_0, a_1, \ldots, a_n \in \mathbb{R}$ and $n \in \mathbb{N}$. In this section, we show that functions that are differentiable sufficiently many times can be approximated by polynomials. We use the notation $f^{(n)}(a)$ to stand for the n-th derivative of f at a.

Lemma 5.3.1. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is the polynomial function*

$$f(x) := a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where $n \in \mathbb{N}$ and $a_0, \ldots, a_n \in \mathbb{R}$. Then

$$a_k = \frac{1}{k!} f^{(k)}(0) \text{ for } k = 0, 1, \ldots, n.$$

Proof. a_k is computed by differentiating f k times and evaluating at 0. □

Thus, for a polynomial f of degree n, we have

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!} f''(0) + \cdots + \frac{x^n}{n!} f^{(n)}(0).$$

More generally, if $x = a + h$, where a is fixed, we have

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!} f''(a) + \cdots + \frac{h^n}{n!} f^{(n)}(a).$$

The next theorem describes how a sufficiently well-behaved function may be approximated by a polynomial.

Theorem 5.3.1. *(**Taylor's theorem**) Let $h > 0$ and $p \geq 1$. Suppose that f and its derivatives up to order $n-1$ are continuous on $[a, a+h]$ and that $f^{(n)}$ exists on $(a, a+h)$. Then there exists a number $t \in (0,1)$ such that*

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!} f''(a) + \cdots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(a) + R_n$$

where

$$R_n = \frac{h^n (1-t)^{n-p}}{p(n-1)!} f^{(n)}(a+th). \tag{5.3.1}$$

The polynomial

$$p(x) := f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!} f''(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a)$$

is called the **Taylor polynomial** for f of degree $n-1$ based at the point a.

Proof. Set

$$R_n = f(a+h) - f(a) - hf'(a) - \frac{h^2}{2!}f''(a) - \cdots - \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a).$$

Define $g : [a, a+h] \to \mathbb{R}$ by

$$\begin{aligned} g(x) = f(a+h) - f(x) - (a+h-x)f'(x) - \frac{(a+h-x)^2}{2!}f''(x) - \frac{(a+h-x)^3}{3!}f^{(3)}(x) \\ - \cdots - \frac{(a+h-x)^{n-1}}{(n-1)!}f^{(n-1)}(x) - \frac{(a+h-x)^p R_n}{h^p}. \end{aligned}$$

Clearly $g(a+h) = 0$. From our definition of R_n, it follows that $g(a) = 0$. Therefore, we can apply Rolle's theorem to g on $[a, a+h]$. Hence, there exists $t \in (0,1)$ such that

$$g'(a+th) = 0.$$

Now

$$\begin{aligned} g'(x) &= -f'(x) - \big[(-1)f'(x) + (a+h-x)f''(x)\big] \\ &\quad - \left[-(a+h-x)f''(x) + \frac{(a+h-x)^2}{2!}f^{(3)}(x)\right] \\ &\quad - \left[-\frac{(a+h-x)^2}{2!}f^{(3)}(x) + \frac{(a+h-x)^3}{3!}f^{(4)}(x)\right] \\ &\quad - \cdots - \\ &\quad - \left[-\frac{(a+h-x)^{n-2}}{(n-2)!}f^{(n-1)}(x) + \frac{(a+h-x)^{n-1}}{(n-1)!}f^{(n)}(x)\right] \\ &\quad + \frac{p(a+h-x)^{p-1}}{h^p}R_n. \end{aligned}$$

Observe that all terms cancel apart from the last two. Therefore,

$$0 = g'(a+th) = -\frac{[h(1-t)]^{n-1}}{(n-1)!}f^{(n)}(a+th) + \frac{p(1-t)^{p-1}}{h}R_n.$$

From this we find that

$$R_n = \frac{h^n(1-t)^{n-p}}{p(n-1)!}f^{(n)}(a+th),$$

and this proves the theorem. □

Corollary 5.3.1. (**forms of the remainder**) *Let the conditions of Theorem 5.3.1 be satisfied. Then*

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a) + R_n, \tag{5.3.2}$$

where

(i) (**Lagrange form of the remainder**) $R_n = \dfrac{h^n}{n!}f^{(n)}(a+th)$ *for some* $t \in (0,1)$.

(ii) (**Cauchy form of the remainder**) $R_n = \dfrac{(1-s)^{n-1}h^n}{(n-1)!} f^{(n)}(a+sh)$ *for some* $s \in (0,1)$.

Proof. Put $p = n$ in (5.3.1) to get (i); put $p = 1$ to get (ii). □

Remark 5.3.1. Again, the number in (5.3.2) is called the **remainder term**. Note that $\lim_{h\to 0} \dfrac{R_n}{h^{n-1}} = 0$, which means that $R_n \to 0$ as $h \to 0$ faster than any other term in (5.3.2).

Example 5.3.1. Let $\alpha \in \mathbb{R}$. Let $f : (-1,1] \to \mathbb{R}$ be given by $f(x) = (1+x)^\alpha$. Since

$$\begin{aligned} f^{(n)}(x) &= \alpha(\alpha-1)\dots(\alpha-n+1)(1+x)^{\alpha-n}, \\ f^{(n)}(0) &= \alpha(\alpha-1)\dots(\alpha-n+1), \end{aligned}$$

Taylor's expansion of f takes the form

$$(1+x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}x^n + R_{n+1}(x),$$

where the remainder term in Lagrange form is

$$R_{n+1}(x) = \frac{\alpha(\alpha-1)\dots(\alpha-n)}{(n+1)!}(1+tx)^{\alpha-n-1}x^{n+1} \text{ for some } t \in (0,1).$$

Chapter 6

Series

In this chapter we continue to study series. We have already studied series of positive terms. Here, we shall discuss convergence and divergence of series which have infinitely many positive and negative terms.

6.1 Series of positive and negative terms

6.1.1 Alternating series

In Chapter 3 we have considered the general concept of series and studied specifically series, all whose terms (but possibly a finite number of terms) have the same – positive or negative – sign. Contrary to this, the general term *alternating series* applies to a series, containing infinitely many positive, as well as negative terms. Most often, in such series, the sign of each consecutive term is the opposite of the sign of the previous term, so in the notation for the series one encounters the factor $(-1)^n$.

In general, as will be somewhat shown below, alternating series are tricky. However, there is a specific case, described in the following theorem, where we can have a *sufficient* convergence criterion, which sounds very much like the general *necessary* one in Theorem 3.3.1 (but generally NOT sufficient, cf. the Harmonic series, Example 3.3.2).

Theorem 6.1.1. (Alternating series test) *Let $(a_n)_n$ be a sequence of positive numbers such that*

(i) $(\forall n \in \mathbb{N})\ \big(a_n \geq a_{n+1}\big)$*;*

(ii) $\lim\limits_{n\to\infty} a_n = 0$.

Then the series

$$a_1 - a_2 + a_3 - a_4 + \cdots + (-1)^{n-1}a_n + \cdots = \sum_{n=1}^{\infty}(-1)^{n-1}a_n$$

converges.

Proof. Denote by

$$s_n := a_1 - a_2 + a_3 - a_4 + \cdots + (-1)^{n-1}a_n = \sum_{k=1}^{n}(-1)^{k-1}a_k$$

the n-th partial sum of the alternating series. Now set

$$\begin{array}{rcl} t_n &:=& s_{2n} &=& a_1 - a_2 + a_3 - a_4 + \cdots + a_{2n-1} - a_{2n}, \\ u_n &:=& s_{2n-1} &=& a_1 - a_2 + a_3 - a_4 + \cdots + a_{2n-1}. \end{array}$$

A diagram indicates that t_n is increasing and u_n is decreasing. We now show this.
By (i),

$$t_{n+1} - t_n = s_{2n+2} - s_{2n} = a_{2n+1} - a_{2n+2} \geq 0.$$

Thus $(t_n)_n$ is increasing. Similarly

$$u_{n+1} - u_n = -a_{2n} + a_{2n+1} \leq 0,$$

so $(u_n)_n$ is decreasing.
Moreover, for each $n \in \mathbb{N}$,

$$u_n - t_n = a_{2n} \geq 0, \text{ i.e. } u_n \geq t_n.$$

Thus, the sequence $(u_n)_n$ is decreasing, and bounded below by t_1, so it converges to a limit $u = \lim_{n\to\infty} u_n$. Likewise, as $(t_n)_n$ is increasing and bounded above by u_1, it converges to a limit $t = \lim_{n\to\infty} t_n$. In addition, $u - t = \lim_{n\to\infty}(u_n - t_n) = \lim_{n\to\infty} a_{2n} = 0$, so t_n and u_n share a common limit $s := t = u$. So

$$s_{2n} = t_n \to s \text{ and } s_{2n+1} = u_n \to s \text{ as } n \to \infty.$$

Therefore, $s_n \to s$ as $n \to \infty$. □

Example 6.1.1. The series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$$

converges.

Example 6.1.2. The series

$$\sum_{n=1}^{\infty} \frac{x^n}{n}$$

converges if and only if $x \in [-1, 1)$.

Proof. First, let $x > 0$. Then

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{xn}{n+1} = x.$$

By the Ratio Test, if $x < 1$ then the series converges, while if $x > 1$, then it diverges. If $x = 1$, we obtain the harmonic series which is divergent.

Now let $x < 0$. Put $y := -x > 0$. We then have an alternating series

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

in case $y \leq 1$; for then, $\frac{y^n}{n} > \frac{y^{n+1}}{n+1}$ and $\lim_{n\to\infty} \frac{y^n}{n} = 0$. By Theorem 6.1.1, the series converges. If $y > 1$, on the other hand, then $\lim_{n\to\infty} \frac{y^n}{n} \neq 0$, and so the series diverges. □

6.1.2 Absolute convergence

Definition 6.1.1. (i) The series $\sum_{n=1}^{\infty} a_n$ is said to be **absolutely convergent** if $\sum_{n=1}^{\infty} |a_n|$ converges.

(ii) A convergent series which is not absolutely convergent is said to be **conditionally convergent**.

Theorem 6.1.2. *An absolutely convergent series is convergent.*

Proof. Let $\sum_{n=1}^{\infty} a_n$ be an absolutely convergent series. Define the **positive** and **negative parts** of a_n via

$$b_n := \begin{cases} a_n & \text{if} \quad a_n \geq 0, \\ 0 & \text{if} \quad a_n < 0, \end{cases}$$

$$c_n := \begin{cases} 0 & \text{if} \quad a_n \geq 0, \\ -a_n & \text{if} \quad a_n < 0, \end{cases}$$

respectively. Then b_n and c_n are non-negative. Moreover,

$$\begin{aligned} a_n &= b_n - c_n, \\ |a_n| &= b_n + c_n. \end{aligned}$$

Now, $b_n \leq |a_n|$ and $c_n \leq |a_n|$ for each n, and we know that $\sum_{n=1}^{\infty} |a_n| < +\infty$. Therefore, the series

$$\sum_{n=1}^{\infty} b_n \text{ and } \sum_{n=1}^{\infty} c_n$$

both converge by the comparison test. This means that

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

also converges. □

Example 6.1.3. (i) The series

$$1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{6} + \dots$$

is absolutely convergent as the series

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{6} + \dots$$

converges.

(ii) The series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

is conditionally convergent as the series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n} + \dots$$

diverges.

Example 6.1.4. The **exponential series**

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \dots$$

converges absolutely for each $x \in \mathbb{R}$.

Proof. This is an exercise. Consider the series

$$\sum_{n=1}^{\infty} \frac{|x|^n}{n!}$$

and use the Ratio Test to prove it converges for each $x \in \mathbb{R}$.

6.1.3 Rearranging series

Example 6.1.5. Let

$$s := 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

be the sum of the **alternating harmonic series**. Let

$$s_n := 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots + \frac{(-1)^{n-1}}{n}.$$

be the nth partial sum. Let us rearrange the series as follows:

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \dots$$

Denote by t_n the nth partial sum of this new series. Then

$$\begin{aligned} t_{3n} =& \left(1 - \frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{6} - \frac{1}{8}\right) + \cdots + \left(\frac{1}{2n-1} - \frac{1}{4n-2} - \frac{1}{4n}\right) \\ =& \frac{1}{2}\left(1 - \frac{1}{2}\right) + \frac{1}{2}\left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \frac{1}{2}\left(\frac{1}{2n-1} - \frac{1}{2n}\right) = \frac{1}{2}s_{2n}. \end{aligned}$$

Also,

$$\begin{aligned} t_{3n+1} &= t_{3n} + \frac{1}{2n+1}; \\ t_{3n+2} &= t_{3n+1} - \frac{1}{4n+2}. \end{aligned}$$

But $s_{2n} \to s$ as $n \to \infty$. Hence $t_{3n} \to \frac{1}{2}s$ as $n \to \infty$, and the same for t_{3n+1} and t_{3n+2}. In conclusion, therefore,

$$t_n \to \frac{1}{2}s \text{ as } n \to \infty.$$

Remark 6.1.1. The alternating harmonic series is conditionally convergent. The above Example shows that by rearranging the series, its sum may be changed. It is a remarkable fact that any conditionally convergent series can be rearranged in such a way that the rearranged series sums to any pre-assigned real number.

However, in the case of an absolutely convergent series, rearrangement does not alter the sum. We say that $\sum_{n=1}^{\infty} b_n$ is a **rearrangement** of $\sum_{n=1}^{\infty} a_n$ if there exists a bijection $\pi : \mathbb{N} \to \mathbb{N}$ with the property that

$$b_n = a_{\pi(n)}$$

for each $n \in \mathbb{N}$.

Theorem 6.1.3. (Rearrangement theorem) *Let $\sum_{n=1}^{\infty} a_n$ be an absolutely convergent series with sum s. Then any rearrangement of $\sum_{n=1}^{\infty} a_n$ is also convergent with sum s.*

Proof. First, let us consider the case where $a_n \geq 0$ for each n. Suppose that $b_1 + b_2 + b_3 + \ldots$ is a rearrangement of $a_1 + a_2 + a_3 + \ldots$. Let

$$\begin{aligned} s_n &:= a_1 + a_2 + a_3 + \cdots + a_n, \\ t_n &:= b_1 + b_2 + b_3 + \cdots + b_n, \end{aligned}$$

stand for the corresponding partial sums. The sequences $(s_n)_n$ and $(t_n)_n$ are increasing and $s_n \to s$ as $n \to \infty$. So $s_n \leq s$ for each n. Also, given any $n \in \mathbb{N}$,

$$t_n = b_1 + b_2 + \cdots + b_n = a_{\pi(1)} + a_{\pi(2)} + \cdots + a_{\pi(n)} \leqslant s_k$$

for some $k \in \mathbb{N}$. It follows that $t_n \leq s$ for each n. Hence the sequence $(t_n)_n$ is bounded above by s and so converges with limit t, say, where $t \leq s$.

We have established that $b_1+b_2+b_3+\dots$ converges with sum t. Now regard $a_1+a_2+a_3+\dots$ as a rearrangement of $b_1+b_2+b_3+\dots$ and apply the above argument to conclude that $s \le t$. These two estimates force $t = s$.

We now turn to the general case. Define the positive and negative parts of a_n via

$$a_n^+ := \begin{cases} a_n & \text{if} \quad a_n \ge 0, \\ 0 & \text{if} \quad a_n < 0, \end{cases}$$

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

Then

$$0 \le a_n^+ \le |a_n| \text{ and } 0 \le a_n^- \le |a_n|.$$

By the comparison test, both

$$\sum_{n=1}^{\infty} a_n^+ \quad \text{and} \quad \sum_{n=1}^{\infty} a_n^-$$

converge. Set $a^+ := \sum_{n=1}^{\infty} a_n^+$ and $a^- := \sum_{n=1}^{\infty} a_n^-$. Then

$$\begin{aligned} s &= \sum_{n=1}^{\infty} a_n \\ &= \sum_{n=1}^{\infty} (a_n^+ - a_n^-) \\ &= a^+ - a^-. \end{aligned}$$

Now let $b_1+b_2+b_3+\dots$ be a rearrangement of $a_1+a_2+a_3+\dots$. Denote by b_n^+ resp. b_n^- the positive resp. negative parts of b_n. Then $b_1^+ + b_2^+ + \dots$ is a rearrangement of $a_1^+ + a_2^+ + \dots$. From the first part of the proof, we have that

$$b_1^+ + b_2^+ + \dots = a_1^+ + a_2^+ + \dots;$$

and similarly,

$$b_1^- + b_2^- + \dots = a_1^- + a_2^- + \dots.$$

But $b_n = b_n^+ - b_n^-$ for each n. Therefore,

$$\begin{aligned} b_1 + b_2 + \dots &= \sum_{n=1}^{\infty} b_n \\ &= \sum_{n=1}^{\infty} (b_n^+ - b_n^-) \\ &= \sum_{n=1}^{\infty} b_n^+ - \sum_{n=1}^{\infty} b_n^- \\ &= \sum_{n=1}^{\infty} a_n^+ - \sum_{n=1}^{\infty} a_n^- \\ &= a^+ - a^- \\ &= s. \end{aligned}$$

$\square$

6.1.4 Multiplication of series

Theorem 6.1.4. **(Product series)** *Let* $\sum_{n=1}^{\infty} a_n$ *and* $\sum_{n=1}^{\infty} b_n$ *be absolutely convergent series with sums* a *and* b *respectively. Then the* **product series**

$$a_1\, b_1 + a_2\, b_1 + a_2\, b_2 + a_1\, b_2 + a_1\, b_3 + a_2\, b_3 + \dots \tag{6.1.1}$$

is absolutely convergent with sum $a\, b$.

Notice that the above arrangement of terms may be thought of as describing a path through the lattice $\mathbb{N} \times \mathbb{N}$ as below.

$$\begin{array}{ccccccccc}
a_1\, b_1 & & a_1\, b_2 & \rightarrow & a_1\, b_3 & & a_1\, b_4 & \rightarrow \\
\downarrow & & \uparrow & & \downarrow & & \uparrow & \\
a_2\, b_1 & \rightarrow & a_2\, b_2 & & a_3\, b_3 & & a_4\, b_4 & \\
& & & & \downarrow & & \uparrow & \\
a_3\, b_1 & \leftarrow & a_3\, b_2 & \leftarrow & a_3\, b_3 & & a_4\, b_4 & \\
\downarrow & & & & & & \uparrow & \\
a_4\, b_1 & \rightarrow & a_4\, b_2 & \rightarrow & a_4\, b_3 & \rightarrow & a_4\, b_4 &
\end{array}$$

But any other arrangement (or order) would yield the identical sum $a\, b$, thanks to absolute convergence and the last result.

Proof. Denote by w_n the n-th term in the series above; so $w_n = a_k\, b_l$ for some $k, l \in \mathbb{N}$. We first prove that the series $\sum_{n=1}^{\infty} |w_n|$ converges. Let S_n stand for the n-th partial sum of this series. Let m be the largest k or l amongst those products $|a_k\, b_l|$ that appear in S_n. Then

$$S_n \leq (|a_1| + |a_2| + \cdots + |a_m|)\,(|b_1| + |b_2| + \cdots + |b_m|) \leq \sum_{n=1}^{\infty} |a_n| \sum_{n=1}^{\infty} |b_n| < +\infty$$

So $(S_n)_n$ is bounded above. This proves absolute convergence of the product series.

It remains to prove that the sum of the series is ab. Let s denote the sum of the series (6.1.1) and s_n its n-th partial sum; so, $s = \lim_{n\to\infty} s_n$. Notice that

$$s_{n^2} = (a_1 + a_2 + \cdots + a_n)(b_1 + b_2 + \cdots + b_n).$$

This implies that

$$\lim_{n\to\infty} s_{n^2} = ab.$$

But, of course, $\lim_{n\to\infty} s_{n^2} = \lim_{n\to\infty} s_n = s$. This completes the proof. $\square$

Example 6.1.6. Set

$$e(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \dots$$

We know from Example 6.1.4 that the series on the right-hand side of the above equality is absolutely convergent for all $x \in \mathbb{R}$. Hence any two such series can be multiplied together in the obvious way and the order of the terms can be changed without affecting the sum. Therefore for all $x, y \in \mathbb{R}$ we have

$$\begin{aligned} e(x)e(y) =& (1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots)(1 + \frac{y}{1!} + \frac{y^2}{2!} + \dots + \frac{y^n}{n!} + \dots) \\ =& 1 + x + y + \frac{y^2}{2} + xy + \frac{x^2}{2} + \dots \\ =& 1 + \frac{(x+y)}{1!} + \frac{(x+y)^2}{2!} + \dots + \frac{(x+y)^n}{n!} + \dots = e(x+y), \end{aligned}$$

where we used the observation that the terms of degree n in x and y are

$$\frac{x^n}{n!} + \dots + \frac{x^k}{k!}\frac{y^{n-k}}{(n-k)!} + \dots + \frac{y^n}{n!} = \frac{(x+y)^n}{n!}.$$

6.2 Power series

Definition 6.2.1. A **power series** is a series of the form

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots = \sum_{n=0}^{\infty} a_n x^n,$$

where the **coefficients** a_n and variable x are real numbers.

Example 6.2.1. The exponential series

$$1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \ldots$$

is a power series which converges for all $x \in \mathbb{R}$ (see Example 6.1.6).

Example 6.2.2. The power series

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

converges only for $x = 0$.

Indeed, if $x \neq 0$ then $\lim_{n\to\infty} n^n x^n \neq 0$.

Example 6.2.3. The geometric series

$$1 + x + x^2 + \cdots + x^n + \cdots = \sum_{n=0}^{\infty} x^n$$

is a power series and converges if and only if $x \in (-1, 1)$.

Example 6.2.4. The power series

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n-1}\frac{x^n}{n} + \cdots = \sum_{n=1}^{\infty} (-1)^{n-1}\frac{x^n}{n}$$

converges if and only if $x \in (-1, 1]$ (see Example 6.1.1).

Theorem 6.2.1. *Let $r \in \mathbb{R}$ be such that the series $\sum_{n=0}^{\infty} a_n r^n$ converges. Then the power series $\sum_{n=0}^{\infty} a_n x^n$ is absolutely convergent for all $x \in (-|r|, |r|)$ (that is, $|x| < |r|$).*

Proof. If $r = 0$ the sum of the series is 0. So we may take $r \neq 0$. As $\sum_{n=0}^{\infty} a_n r^n$ converges, it follows that $\lim_{n\to\infty} a_n\, r^n = 0$. So the sequence $(a_n\, r^n)_{n\in\mathbb{N}}$ is bounded. Specifically,

$$|a_n\, r^n| \leq K$$

for some $0 < K < \infty$ and all $n \in \mathbb{N}$. Let $x \in \mathbb{R}$ with $|x| < |r|$. Put $y := |x|/|r|$, so $0 \leq y < 1$. We have

$$|a_n\, x^n| = |a_n| \cdot |x|^n = |a_n| \cdot |r^n| \cdot y^n = |a_n\, r^n| \cdot y^n \leq K\, y^n.$$

Therefore, the series $\sum_{n=0}^{\infty} |a_n\, x^n|$ converges by comparison with the convergent geometric series $K(1 + y + y^2 + \cdots + y^n + \dots)$. $\square$

Theorem 6.2.2. *Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series. Then exactly one of the following possibilities occurs.*

(i) The series converges only when $x = 0$.

(ii) The series converges absolutely for all $x \in \mathbb{R}$.

(iii) There exists $r > 0$ such that the series is absolutely convergent for all $x \in \mathbb{R}$ such that $|x| < r$ and is divergent for all $x \in \mathbb{R}$ such that $|x| > r$.

Proof. Put

$$E := \{x \in \mathbb{R} :\ x \geqslant 0 \text{ and } \sum_{n=0}^{\infty} a_n\, x^n \text{ converges}\}.$$

Observe that $0 \in E$. If $E = \{0\}$ then (i) is true. Note that if $E \ni x > 0$ then the series converges for all $y \in \mathbb{R}$ such that $0 \leq |y| \leq x$ by the last result.

Suppose that E is not bounded above. Let y be any non-negative real number. As y is not an upper bound for E, we can find $x \in E$ with $y < x$. By Theorem 6.2.1, $y \in E$. Thus E consists of the set of all non-negative real numbers, $E = [0, \infty)$. So, with the help of Theorem 6.2.1, we see that the series converges absolutely for each $x \in \mathbb{R}$. In short, case (ii) holds.

Finally, let us suppose that E is bounded above and contains at least one positive number. Then E possesses a supremum, $r := \sup E$, and $r > 0$. Let $x \in (-r, r)$. We can find $y \in E$ satisfying $|x| < y$. This entails that the series converges absolutely at x, again by Theorem 6.2.1. Now suppose that $|x| > r$. Choose $y \in \mathbb{R}$ with $r < y < |x|$; so $y \notin E$. The series is not convergent at y, hence by Theorem 6.2.1, it is not convergent at x. In short, case (iii) holds. $\square$

Definition 6.2.2. The **radius of convergence** R of the power series

$$a_0 + a_1 x + a_2 x^2 + \dots$$

is defined as follows.

(i) If the series converges for all $x \in \mathbb{R}$, then $R = \infty$.

(ii) If the series converges for $x = 0$ only, then $R = 0$.

(iii) If the series converges absolutely for $|x| < r$ and diverges for $|x| > r$ for some $0 < r < \infty$, then $R = r$.

Example 6.2.5. The power series

$$1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

has infinite radius of convergence, $R = \infty$.

Proof. Set $y = x^2$. Then the series may be rewritten

$$a_0 + a_1 y + a_2 y^2 + \dots \text{ where } a_n = \frac{(-1)^n}{(2n)!}.$$

We then compute

$$\lim_{n\to\infty} \frac{|a_{n+1}y^{n+1}|}{|a_n y^n|} = \lim_{n\to\infty} \frac{y}{(2n+1)(2n+2)} = 0$$

for any $y \geqslant 0$. By the Root Test, the series is absolutely convergent. □

Example 6.2.6. The series

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

has infinite radius of convergence, $R = \infty$.

The proof is similar to the previous one.

Example 6.2.7. Find all $x \in \mathbb{R}$ such that the series

$$1 - \frac{x}{3} + \frac{x^2}{5} - \frac{x^3}{7} + \dots = \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{2n+1}$$

is convergent.

We use the Ratio Test.

$$\lim_{n\to\infty} \frac{|a_{n+1}x^{n+1}|}{|a_n x^n|} = \lim_{n\to\infty} \frac{|x|(2n+1)}{2n+3} = |x|.$$

Therefore the series is absolutely convergent if $|x| < 1$. The series diverges if $|x| > 1$ since the general term does not converge to zero. (So the radius of convergence R is given by $R = 1$.) For $x = 1$ we have the alternating series

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots,$$

which converges. For $x = -1$ we have the series

$$1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \dots = \sum_{n=0}^{\infty} \frac{1}{2n+1},$$

which diverges by comparison with the harmonic series.

Chapter 7

Elementary functions

We now use power series to strictly define the Exponential, Logarithmic, and Trigonometric functions and describe their properties.

7.1 Exponential function

Definition 7.1.1. We define the **exponential function** $\exp : \mathbb{R} \to \mathbb{R}$ via the power series

$$\exp(x) := 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}. \tag{7.1.1}$$

(The power series is convergent for all $x \in \mathbb{R}$ - see Example 6.2.1).

Theorem 7.1.1. (multiplicative property) *For all $x, y \in \mathbb{R}$,*

$$\exp(x) \cdot \exp(y) = \exp(x + y).$$

The proof is given in Example 6.1.6.

The following properties follow from the definition and this last theorem.

Corollary 7.1.1. *(i)* $\exp(0) = 1$.

(ii) $\exp(-x) = 1/\exp(x)$ *for all* $x \in \mathbb{R}$.

(iii) $\exp(x) > 0$ *for all* $x \in \mathbb{R}$.

Theorem 7.1.2. *The exponential function $f(x) = \exp(x)$ is everywhere differentiable and*

$$f'(a) = \exp(a)$$

for each $a \in \mathbb{R}$.

Proof. By the multiplicative property $\exp(x+y) = \exp(x)\cdot\exp(y)$ we have

$$\frac{\exp(a+h)-\exp(a)}{h} = \exp(a)\frac{\exp(h)-1}{h}.$$

From the power series representation of the exponential function, we have

$$\frac{\exp(h)-1}{h} = 1 + \frac{h}{2!} + \frac{h^2}{3!} + \cdots = 1 + g(h).$$

For $|h| < 2$, we may estimate

$$|g(h)| \leq \frac{|h|}{2} + \frac{|h|^2}{4} + \cdots + \frac{|h|^n}{2^n} + \cdots = \frac{|h|/2}{1-(|h|/2)} \to 0 \text{ as } h \to 0.$$

So finally,

$$f'(a) = \lim_{h\to 0}\frac{\exp(a+h)-\exp(a)}{h} = \exp(a). \quad \square$$

It follows from Theorem 5.1.1 that

Corollary 7.1.2. *The exponential function* $\exp(\cdot)$ *is everywhere continuous.*

Theorem 7.1.3. *(i)* $\exp(\cdot)$ *is a strictly increasing function.*

(ii) $\exp(x) \to +\infty$ *as* $x \to +\infty$.

(iii) $\exp(x) \to 0$ *as* $x \to -\infty$.

(iv) $Ran(\exp(\cdot)) = (0, +\infty)$.

(v) $\frac{\exp(x)}{x^k} \to +\infty$ *as* $x \to +\infty$ *for each* $k \in \mathbb{N}$.

Proof. Exercise. $\square$

Definition 7.1.2. The **base of the natural logarithms** e is defined by

$$e := \exp(1) = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!} + \cdots = \sum_{n=0}^{\infty}\frac{1}{n!}.$$

The approximate value of e is

2.718281828459045

It may be shown that e is irrational (see Exercises 14, Q7).

It follows from the multiplicative property that $\exp(n) = e^n$ for each $n \in \mathbb{N}$. In fact, more is true.

Theorem 7.1.4. *Let* $r \in \mathbb{Q}$. *Then*

$$\exp(r) = e^r.$$

Proof. First, let $n \in \mathbb{N}$ and set $r = -n$. Then

$$\exp(r) = \exp(-n) = 1/\exp(n) = 1/e^n = e^{-n} = e^r.$$

Now let $r = p/q$ where $p, q \in \mathbb{N}$ and $q > 0$. Then

$$\begin{aligned} \exp(r)^q &= \exp(p/q)^q \\ &= \exp(\frac{p}{q} \cdot q) \text{ (by multiplicative property)} \\ &= \exp(p) \\ &= e^p. \end{aligned}$$

Hence

$$\exp(r) = \exp(p/q) = e^{p/q} = e^r.$$

It remains to prove the result for $r \in \mathbb{Q}$, $r < 0$. Use Corollary 7.1.1, (ii). □

Definition 7.1.3. If x is irrational, define

$$e^x := \exp(x).$$

The monotonicity and continuity properties of the exponential function entail that

$$e^x = \sup\{e^p \,|\, p \text{ is rational and } p < x\}.$$

Remark 7.1.1. As $f = \exp(\cdot)$ is strictly increasing and $\mathrm{Ran}(f) = (0, +\infty)$, the function

$$f = \exp(\cdot) : \mathbb{R} \to (0, +\infty);\ x \to \exp(x)$$

is a bijection. (Note change of co-domain.)

7.2 The logarithm

Definition 7.2.1. The function $f : \mathbb{R} \to (0, +\infty);\ x \to e^x$ is a bijection. Its inverse $g : (0, +\infty) \to \mathbb{R}$ is called the **logarithm function**. We write $g(x) = \log(x)$. Thus, for $x > 0$,

$$y = \log(x) \text{ if and only if } e^y = x.$$

Theorem 7.2.1. *The function* $\log : (0, +\infty) \to \mathbb{R}$ *is a differentiable (and so continuous) increasing function with the following properties.*

(i) $\log(x\,y) = \log(x) + \log(y)$ *for all* $x, y \in (0, +\infty)$.

(ii) $\log(x\,/\,y) = \log(x) - \log(y)$ *for all* $x, y \in (0, +\infty)$.

(iii) $\log(1) = 0$.

(iv) $(\log x)' = 1/x$.

(v) $\lim_{x \to +\infty} \log x = +\infty$.

(vi) $\lim_{x\downarrow 0} \log x = -\infty$.

(vii) $\frac{\log x}{x^k} \to 0$ *as* $x \to +\infty$ *for each* $k \in \mathbb{N}$.

Proof. These results follow from properties of the exponential function. Let us show (i). Given $x, y > 0$, we have

$$\begin{aligned}
\exp(\log(x\,y)) &= x\,y \ \text{ (inverse property)} \\
&= \exp(\log(x)) \cdot \exp(\log(y)) \ \text{ (inverse property)} \\
&= \exp(\log(x) + \log(y)) \ \text{ (multiplicative property of exponential)}\ .
\end{aligned}$$

Property (i) now follows from injectivity of the exponential function. The proof of (ii) is similar.

As for property (iv), we can argue as follows. Let $y = g(x) = \log(x)$ for $x > 0$; so $x = f(y) = \exp(y)$. By Theorem 5.1.7,

$$g'(x) = \frac{1}{f'(y)} = \frac{1}{\exp(y)} = \frac{1}{x}.$$

Parts (iii) and (v) - (vii) are left as exercises. $\square$

There is a simple representation of $\log(1 + x)$ as a power series.

Theorem 7.2.2. *For* $x \in (-1, 1)$,

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n-1}\frac{x^n}{n} + \cdots = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^n}{n}.$$

One proof of this result depends upon a version of Taylor's Theorem in which the remainder term R_n is written as an integral (the Lagrange or Cauchy forms of the remainder are not adequate). It is beyond the scope of this course.

We are now able to define arbitrary powers of nonnegative numbers (till now it was only clear how to do this for rational powers).

Definition 7.2.2. Let $a > 0$. Then for any $x \in \mathbb{R}$ we define

$$a^x := e^{x \log(a)}.$$

Given $a > 0$ and $x, y \in \mathbb{R}$, the **laws for indices**

$$\begin{aligned}
a^x \cdot a^y &= a^{x+y} \\
(a^x)^y &= a^{xy}
\end{aligned}$$

can be verified using properties of the exponential and logarithm functions.

7.3 Trigonometric functions

In this section, we sketch an approach to defining the trigonometric functions.

Definition 7.3.1. The trigonometric functions $\sin : \mathbb{R} \to \mathbb{R}$ and $\cos : \mathbb{R} \to \mathbb{R}$ are defined via the following power series.

$$\sin x := x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \tag{7.3.2}$$

$$\cos x := 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \tag{7.3.3}$$

These series are absolutely convergent for all $x \in \mathbb{R}$ (see Example 6.2.6 and Example 6.2.5).

The next theorem may be proved using the above definitions by multiplying and adding the corresponding series. We skip the details. The complex exponential (**Euler's formula**) can be used to give a slicker proof.

Theorem 7.3.1. *For all $x, y \in \mathbb{R}$,*

$$\sin(x + y) = \sin x \cos y + \cos x \sin y; \tag{7.3.4}$$

$$\cos(x + y) = \cos x \cos y - \sin x \sin y; \tag{7.3.5}$$

$$\sin^2 x + \cos^2 x = 1. \tag{7.3.6}$$

From the above it follows that

$$(\forall x \in \mathbb{R})\big[(-1 \le \sin x \le 1) \wedge (-1 \le \cos x \le 1)\big].$$

Theorem 7.3.2. *The functions* $\sin : \mathbb{R} \to \mathbb{R}$ *and* $\cos : \mathbb{R} \to \mathbb{R}$ *are everywhere differentiable and*

$$(\sin x)' = \cos x, \quad (\cos x)' = -\sin x.$$

The proof uses Theorem 7.3.1 and is somewhat similar to the proof of Theorem 7.1.2. We omit the details.

Periodicity of trigonometric functions

Theorem 7.3.3. *There exists a smallest positive constant $\frac{1}{2}\alpha$ such that*

$$\cos \frac{1}{2}\alpha = 0 \quad \textit{and} \quad \sin \frac{1}{2}\alpha = 1.$$

Proof. If $0 < x < 2$ then

$$\sin x = \left(x - \frac{x^3}{3!}\right) + \left(\frac{x^5}{5!} - \frac{x^7}{7!}\right) + \dots > 0$$

Therefore, for $0 < x < 2$ we have that

$$(\cos x)' = -\sin x < 0,$$

so $\cos x$ is decreasing. Notice that

$$\cos x = \left(1 - \frac{x^2}{2!}\right) + \left(\frac{x^4}{4!} - \frac{x^6}{6!}\right) + \dots$$

It is then easy to see that $\cos\sqrt{2} > 0$. Another rearrangement

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \left(\frac{x^6}{6!} - \frac{x^8}{8!}\right) - \dots$$

shows that $\cos\sqrt{3} < 0$. This proves the existence of a smallest number $\sqrt{2} < \frac{1}{2}\alpha < \sqrt{3}$ such that $\cos\frac{1}{2}\alpha = 0$. Since $\sin\frac{1}{2}\alpha > 0$ it follows from (7.3.6) that $\sin\frac{1}{2}\alpha = 1$. $\square$

Theorem 7.3.4. *Let α be the number defined in Theorem 7.3.3. Then for all $x \in \mathbb{R}$,*

$$\sin(x + \frac{1}{2}\alpha) = \cos x, \quad \cos(x + \frac{1}{2}\alpha) = -\sin x; \tag{7.3.7}$$

$$\sin(x + \alpha) = -\sin x, \quad \cos(x + \alpha) = -\cos x; \tag{7.3.8}$$

$$\sin(x + 2\alpha) = \sin x, \quad \cos(x + 2\alpha) = \cos x. \tag{7.3.9}$$

The proof is an easy consequence of (7.3.5) and (7.3.6). The above theorem entails that sin and cos are periodic with period 2α. (Of course, with further analysis, it may be shown that $\alpha = \pi$.)

Chapter 8

The Riemann Integral

In this chapter we study one of the approaches to theory of integration. A more detailed exposition is developed in the F.T.A. course.

8.1 Definition of integral

Let us consider a function $f : [a, b] \to \mathbb{R}$ defined on a closed interval $[a, b] \subset \mathbb{R}$. We are going to measure the area under the graph of the curve $y = f(x)$ on $\mathbb{R} \times \mathbb{R}$ between the vertical lines $x = a$ and $x = b$. In order to visualize the picture you may think of the case $f(x) \geq 0$. The approach we undertake is to approximate the area by means of the sum of rectangles dividing the interval $[a, b]$.

In the following we always assume that f is bounded on $[a, b]$.

8.1.1 Definition of integral and integrable functions

Definition 8.1.1. Let $a < b$. A partition of the interval $[a, b]$ is a finite collection of points in $[a, b]$ one of which is a and one of which is b.

The point of a partition can be numbered $x_0, x_1, \ldots, x_n$ so that

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

Definition 8.1.2. Let $P = \{x_0, \ldots, x_n\}$ be a partition on $[a, b]$. Let

$$m_i = \inf\{f(x)\,|, x_{i-1} \leq x \leq x_i\},$$
$$M_i = \sup\{f(x)\,|, x_{i-1} \leq x \leq x_i\}.$$

The lower sum of f for P is defined as

$$L(f, P) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}).$$

The upper sum of f for P is defined as

$$U(f,P) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}).$$

Definition 8.1.3. The integral sum of f for the partition P and $\{\xi_i \in [x_{i-1}, x_i] \,|\, (x_i)_{i=0}^n \in P\}$ is

$$\sigma(f,P,\xi) = \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}).$$

Denote

$$m := \inf_{x\in[a,b]} f(x), \;\; M := \sup_{x\in[a,b]} f(x).$$

The following inequality is obviously true

$$(\forall i \in \{1,2,\dots,n\}) \left[m \le m_i \le f(\xi_i) \le M_i \le M\right].$$

Therefore we have

$$m(b-a) = \sum_{i=1}^{n} m(x_i - x_{i-1}) \le \sum_{i=1}^{n} m_i(x_i - x_{i-1}) \le \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$
$$\le \sum_{i=1}^{n} M_i(x_i - x_{i-1}) \le \sum_{i=1}^{n} M(x_i - x_{i-1}) = M(b-a),$$

which implies that for every partition and for every choice $\{\xi_i \in [x_{i-1}, x_i] \,|\, (x_i)_{i=0}^n \in P\}$ the following inequality holds

$$m(b-a) \le L(f,P) \le \sigma(f,P,\xi) \le U(f,P) \le M(b-a). \tag{8.1.1}$$

In other words, the sets of real numbers

$$\{L(f,P)\,|\,P\}, \;\; \{U(f,P)\,|\,P\}$$

are bounded.

Definition 8.1.4. The upper integral of f over $[a,b]$ is

$$J := \inf\{U(f,P)\,|\,P\},$$

the lower integral of f over $[a,b]$ is

$$j := \sup\{L(f,P)\,|\,P\}.$$

(The infimum and the supremum are taken over all partitions on $[a,b]$.)

Definition 8.1.5. A function $f : [a,b] \to \mathbb{R}$ is called Riemann integrable if

$$J = j.$$

The common value is called integral of f over $[a,b]$ and is denoted by

$$\int_a^b f(x)dx.$$

8.1.2 Properties of upper and lower sums

Lemma 8.1.1. *Let P and Q be two partitions of $[a, b]$ such that $P \subset Q$. Then*

$$L(f, P) \leq L(f, Q),$$
$$U(f, P) \geq U(f, Q).$$

(The partition Q is called a refinement of P.)

Proof. First let us consider a particular case. Let P' be a partition formed from P by adding one extra point, say $c \in [x_{k-1}, x_k]$. Let

$$m'_k = \inf_{x \in [x_{k-1}, c]} f(x), \quad m''_k = \inf_{x \in [c, x_k]} f(x).$$

Then $m'_k \geq m_k$, $m''_k \geq m_k$, and we have

$$\begin{aligned} L(f, P') &= \sum_{i=1}^{k-1} m_i(x_i - x_{i-1}) + m'_k(c - x_{k-1}) + m''_k(x_k - c) + \sum_{i=k+1}^{n} m_i(x_i - x_{i-1}) \\ &\geq \sum_{i=1}^{k-1} m_i(x_i - x_{i-1}) + m_k(x_k - x_{k-1}) + \sum_{i=k+1}^{n} m_i(x_i - x_{i-1}) = L(f, P). \end{aligned}$$

Similarly one obtains that

$$U(f, P') \leq U(f, P).$$

Now to prove the assertion one has to add to P consequently a finite number of points in order to form Q. □

Proposition 8.1.1. *Let P and Q be arbitrary partitions of $[a, b]$. Then*

$$L(f, P) \leq U(f, Q).$$

Proof. Consider the partition $P \cup Q$. The by Lemma 8.1.1 we have

$$L(f, P) \leq L(f, P \cup Q) \leq U(f, P \cup Q) \leq U(f, Q). \quad \square$$

Theorem 8.1.1. *$J \geq j$.*

Proof. Fix a partition Q. Then by Proposition 8.1.1

$$(\forall P)\ L(f, P) \leq U(f, Q).$$

Therefore

$$j = \sup\{L(f, P) \mid P\} \leq U(f, Q).$$

And from the above

$$(\forall Q) j \leq U(f, Q).$$

Hence

$$j \leq \inf\{U(f, Q) \mid Q\} = J. \quad \square$$

Remark. Integrability of f means by definition $j = J$. If $j < J$ we say that f is not Riemann integrable.

Example 8.1.1. Let $f : [a, b] \to \mathbb{R}$ is defined by $f(x) = C$. Then

$$(\forall P) \left[L(f, P) = U(f, P) = C(b - a)\right].$$

Hence

$$J = j = C(b - a).$$

Example 8.1.2. The Dirichlet function $D : [0, 1] \to \mathbb{R}$ is defined as $D(x) = 1$ if x is rational and $D(x) = 0$ is x is irrational.
Then

$$(\forall P) \left[L(D, P) = 0\right] \wedge \left[U(D, P) = 1\right].$$

Hence the Dirichlet function is not Riemann integrable.

8.2 Criterion of integrability

Theorem 8.2.1. *A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if for any $\varepsilon > 0$ there exists a partition P of $[a, b]$ such that*

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

Proof. 1) Necessity. Let $J = j$, i.e. let us assume that f is integrable. Fix $\varepsilon > 0$. Then

$$(\exists P_1) \left[L(f, P_1) > j - \varepsilon/2\right].$$

Also

$$(\exists P_2) \left[U(f, 2_1) < J + \varepsilon/2\right].$$

Let $Q = P_1 \cup P_2$. Then

$$j - \varepsilon/2 < L(f, P_1) \leq L(f, Q) \leq U(f, Q) \leq U(f, P_2) < J + \varepsilon/2.$$

Therefore (since $J = j$)

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

2) Sufficiency. Fix $\varepsilon > 0$. Let P be a partition such that

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

Note that

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

Therefore it follows that

$$(\forall \varepsilon > 0) \left(J - j < \varepsilon\right).$$

This implies that $J = j$. ∎

8.3 Integrable functions

The following definition is used in the proof of the next theorems and will be also used in the subsequent sections.

Definition 8.3.1. Let P be a partition of $[a, b]$. The length of the greatest subinterval of $[a, b]$ under the partition P is called the norm of the partition P and is denoted by $\|P\|$, i.e.

$$\|P\| := \max_{1 \leq i \leq n} (x_i - x_{i-1}).$$

Theorem 8.3.1. *Let $f : [a, b] \to \mathbb{R}$ be monotone. Then f is Riemann integrable.*

Proof. Without loss of generality assume that f is increasing, so that $f(a) < f(b)$. Fix $\varepsilon > 0$. Let us consider a partition P of $[a, b]$ such that

$$\|P\| < \delta = \frac{\varepsilon}{f(b) - f(a)}.$$

For this partition we obtain

$$U(f, P) - L(f, P) = \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1})$$

$$= \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))(x_i - x_{i-1}) < \delta \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))$$

$$= \delta(f(b) - f(a)) = \varepsilon. \quad \square$$

Theorem 8.3.2. *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is Riemann integrable.*

Proof. The proof of this theorem requires the concept of *Uniform continuity*, developed in F.T.A. Fix $\varepsilon > 0$. Since f is continuous on a closed interval, it is uniformly continuous (See section 4.4). Therefore for $\frac{\varepsilon}{b-a}$ there exists $\delta > 0$ such that

$$(\forall x_1, x_2 \in [a, b])\big[(|x_1 - x_2| < \delta) \Rightarrow (|f(x_1) - f(x_2)| < \frac{\varepsilon}{b - a}.$$

Hence for every partition P with norm $\|P\| < \delta$ we have

$$U(f, P) - L(f, P) = \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) < \frac{\varepsilon}{b - a} \sum_{i=1}^{n} (x_i - x_{i-1}) = \varepsilon. \quad \square$$

8.4 Elementary properties of the integral

Note: Proofs in small print are not part of the course, and are relevant only for F.T.A.

Theorem 8.4.1. *Let $a < c < b$. Let f be integrable on $[a, b]$. Then f is integrable on $[a, c]$ and on $[c, b]$ and*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Conversely, if f is integrable on $[a, c]$ and on $[c, b]$ then it is integrable on $[a, b]$.

Proof. Suppose that f is integrable on $[a, b]$. Fix $\varepsilon > 0$. Then there exists a partition $P = \{x_0, \ldots, x_n\}$ of $[a, b]$ such that

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

We can assume that $c \in P$ so that $c = x_j$ for some $j \in \{0, 1, \ldots, n\}$ (otherwise consider the refinement of P adding the point c). Then $P_1 = \{x_0, \ldots, x_j\}$ is a partition of $[a, c]$ and $P_2 = \{x_j, \ldots, x_n\}$ is a partition of $[c, b]$. Moreover,

$$L(f, P) = L(f, P_1) + L(f, P_2), \quad U(f, P) = U(f, P_1) + U(f, P_2).$$

Therefore we have

$$[U(f, P_1) - L(f, P_1)] + [U(f, P_2) - L(f, P_2)] = U(f, P) - L(f, P) < \varepsilon.$$

Since each of the terms on the left hand side is non-negative, each one is less than ε, which proves that f is integrable on $[a, c]$ and on $[c, b]$. Note also that

$$L(f, P_1) \le \int_a^c f(x)dx \le U(f, P_1),$$
$$L(f, P_2) \le \int_c^b f(x)dx \le U(f, P_2),$$

so that

$$L(f, P) \le \int_a^c f(x)dx + \int_c^b f(x)dx \le U(f, P).$$

This is true for any partition of $[a, b]$. Therefore

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Now suppose that f is integrable on $[a, c]$ and on $[c, b]$. Fix $\varepsilon > 0$. Then there exists a partition P_1 of $[a, c]$ such that

$$U(f, P_1) - L(f, P_1) < \varepsilon/2.$$

Also there exists a partition P_2 of $[c, b]$ such that

$$U(f, P_2) - L(f, P_2) < \varepsilon/2.$$

Let $P = P_1 \cup P_2$. Then

$$U(f, P) - L(f, P) = [U(f, P_1) - L(f, P_1)] + [U(f, P_2) - L(f, P_2)] < \varepsilon. \quad \square$$

The integral $\int_a^b f(x)dx$ was defined only for $a < b$. We add by definition that

$$\int_a^a f(x)dx = 0 \text{ and } \int_a^b f(x)dx = -\int_b^a f(x)dx \text{ if } a > b.$$

With this convention we always have that

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Theorem 8.4.2. *Let f and g be integrable on $[a, b]$. Then $f + g$ is also integrable on $[a, b]$ and*

$$\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx.$$

Proof. Let $P = \{x_0, \ldots, x_n\}$ be a partition of $[a, b]$. Let

$$\begin{aligned} m_i =& \inf\{f(x) + g(x) \,|\, x_{i-1} \le x \le x_i\}, \\ m_i' =& \inf\{f(x) \,|\, x_{i-1} \le x \le x_i\}, \\ m_i'' =& \inf\{g(x) \,|\, x_{i-1} \le x \le x_i\}. \end{aligned}$$

Define M_i, M_i', M_i'' similarly. The following inequalities hold

$$\begin{aligned} m_i \ge& m_i' + m_i'', \\ M_i \le& M_i' + M_i''. \end{aligned}$$

Therefore we have

$$\begin{aligned} L(f, P) + L(g, P) \le& L(f + g, P), \\ U(f + g, P) \le& U(f, P) + U(g, P). \end{aligned}$$

Hence for any partition P

$$L(f, P) + L(g, P) \le L(f + g, P) \le U(f + g, P) \le U(f, P) + U(g, P),$$

or otherwise

$$U(f + g, P) - L(f + g, P) \le [U(f, P) - L(f, P)] + [U(g, P) - L(g, P)].$$

Fix $\varepsilon > 0$. Since f and g are integrable there are partitions P_1 and P_2 such that

$$\begin{aligned} U(f, P_1) - L(f, P_1) < \varepsilon/2, \\ U(g, P_2) - L(g, P_2) < \varepsilon/2. \end{aligned}$$

Thus for the partition $P = P_1 \cup P_2$ we obtain that

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

This proves that $f + g$ is integrable on $[a, b]$.

Moreover,

$$\begin{aligned} L(f, P) + L(g, P) \le L(f + g, P) \le& \int_a^b [f(x) + g(x)]dx \\ \le& U(f + g, P) \le U(f, P) + U(g, P), \end{aligned}$$

and

$$L(f, P) + L(g, P) \le \int_a^b f(x)dx + \int_a^b g(x)dx \le U(f, P) + U(g, P).$$

Therefore it follows that

$$\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx. \quad \square$$

Theorem 8.4.3. *Let f be integrable on $[a, b]$. Then, for any $c \in \mathbb{R}$, cf is also integrable on $[a, b]$ and*

$$\int_a^b cf(x)dx = c\int_a^b f(x)dx.$$

Proof. The proof is left as an exercise. Consider separately two cases: $c \ge 0$ and $c \le 0$. $\square$

Theorem 8.4.4. *Let f, g be integrable on $[a, b]$ and*

$$(\forall x \in [a, b])\,(f(x) \le g(x)).$$

Then

$$\int_a^b f(x)dx \le \int_a^b g(x)dx.$$

Proof. For any partition P of $[a, b]$ we have

$$L(f, P) \le L(g, P) \le \int_a^b g(x)dx.$$

The assertion follows by taking supremum over all partitions. □

Corollary 8.4.1. *Let f be integrable on $[a, b]$ and there are $M, m \in \mathbb{R}$ such that*

$$(\forall x \in [a, b])\,(m \le f(x) \le M).$$

Then

$$m(b - a) \le \int_a^b f(x)dx \le M(b - a).$$

Corollary 8.4.2. *Let f be continuous on $[a, b]$. Then there exists $\theta \in [a, b]$ such that*

$$\int_a^b f(x)dx = f(\theta)(b - a).$$

Proof. From Corollary 8.4.1 it follows that

$$m \le \frac{1}{b - a}\int_a^b f(x)dx \le M,$$

where $m = \min_{[a,b]} f(x)$, $M = \max_{[a,b]} f(x)$. Then by the Intermediate Value Theorem we conclude that there exists $\theta \in [a, b]$ such that

$$f(\theta) = \frac{1}{b - a}\int_a^b f(x)dx. \quad \square$$

Theorem 8.4.5. *Let f be integrable on $[a, b]$. Then $|f|$ is integrable on $[a, b]$ and*

$$\Big|\int_a^b f(x)dx\Big| \le \int_a^b |f(x)|dx.$$

Proof. Note that for any interval $[\alpha, \beta]$

$$\sup_{[\alpha,\beta]} |f(x)| - \inf_{[\alpha,\beta]} |f(x)| \le \sup_{[\alpha,\beta]} f(x) - \inf_{[\alpha,\beta]} f(x). \tag{8.4.2}$$

Indeed,

$$(\forall x, y \in [\alpha, \beta])\left(f(x) - f(y) \le \sup_{[\alpha,\beta]} f(x) - \inf_{[\alpha,\beta]} f(x)\right), \text{ so that}$$
$$(\forall x, y \in [\alpha, \beta])\left(|f(x)| - |f(y)| \le \sup_{[\alpha,\beta]} f(x) - \inf_{[\alpha,\beta]} f(x)\right),$$

which proves (8.4.2) by passing to the supremum in x and y.

It follows from (8.4.2) that for any partition of $[a, b]$

$$U(|f|, P) - L(|f|, P) \leq U(f, P) - L(f, P),$$

which proves the integrability of $|f|$ by the criterion of integrability, Theorem 8.2.1. The last assertion follows from Theorem 8.4.4. $\square$

Theorem 8.4.6. *Let $f : [a, b] \to \mathbb{R}$ be integrable and $(\forall x \in [a, b])(m \leq f(x) \leq M)$. Let $g : [m, M] \to \mathbb{R}$ be continuous. Then $h : [a, b] \to \mathbb{R}$ defined by $h(x) = g(f(x))$ is integrable.*

Proof. Fix $\varepsilon > 0$. Since g is uniformly continuous on $[m, M]$, there exists $\delta > 0$ such that $\delta < \varepsilon$ and

$$(\forall t, s \in [m, M])\big[(|t - s| < \delta) \Rightarrow (|g(t) - g(s)| < \varepsilon)\big].$$

By integrability of f there exists a partition $P = \{x_0, \dots, x_n\}$ of $[a, b]$ such that

$$U(f, P) - L(f, P) < \delta^2. \tag{8.4.3}$$

Let $m_i = \inf_{[x_{i-1}, x_i]} f(x)$, $M_i = \sup_{[x_{i-1}, x_i]} f(x)$ and $m_i^* = \inf_{[x_{i-1}, x_i]} h(x)$, $M_i^* = \sup_{[x_{i-1}, x_i]} h(x)$. Decompose the set $\{1, \dots, n\}$ into two subset : $(i \in A) \Leftrightarrow (M_i - m_i < \delta)$ and $(i \in B) \Leftrightarrow (M_i - m_i \geq \delta)$.

For $i \in A$ by the choice of δ we have that $M_i^* - m_i^* < \varepsilon$.

For $i \in B$ we have that $M_i^* - m_i^* \leq 2K$ where $K = \sup_{t \in [m, M]} |g(t)|$. By (8.4.3) we have

$$\delta \sum_{i \in B} (x_i - x_{i-1}) \leq \sum_{i \in B} (M_i - m_i)(x_i - x_{i-1}) < \delta^2,$$

so that $\sum_{i \in B} (x_i - x_{i-1}) < \delta$. Therefore

$$\begin{aligned} U(h, P) - L(h, P) &= \sum_{i \in A} (M_i^* - m_i^*)(x_i - x_{i-1}) + \sum_{i \in B} (M_i^* - m_i^*)(x_i - x_{i-1}) \\ &< \varepsilon(b - a) + 2K\delta < \varepsilon[(b - a) + 2K], \end{aligned}$$

which proves the assertion since ε is arbitrary. $\square$

Corollary 8.4.3. *Let f, g be integrable on $[a, b]$. Then the product fg is integrable on $[a, b]$.*

Proof. Since $f + g$ and $f - g$ are integrable on $[a, b]$, $(f + g)^2$ and $(f - g)^2$ are integrable on $[a, b]$ by the previous theorem. Therefore

$$fg = \frac{1}{4}[(f + g)^2 - (f - g)^2] \text{ is integrable on } [a, b]. \quad \square$$

8.5 Integration as the inverse to differentiation

Theorem 8.5.1. *Let f be integrable on $[a,b]$ and let F be defined on $[a,b]$ by*

$$F(x) = \int_a^x f(t)dt.$$

Then F is continuous on $[a,b]$.

Proof. By the definition of integrability f is bounded on $[a,b]$. Let $M = \sup_{[a,b]} |f(x)|$. Then for $x, y \in [a,b]$ we have

$$\big|F(x) - F(y)\big| = \big|\int_x^y f(t)dt\big| \leq M|x-y|,$$

which proves that F is uniformly continuous on $[a,b]$. □

If in the previous theorem we in addition assume that f is continuous, we can prove more.

Theorem 8.5.2. *Let f be integrable on $[a,b]$ and let F be defined on $[a,b]$ by*

$$F(x) = \int_a^x f(t)dt.$$

Let f be continuous at $c \in [a,b]$. Then F is differentiable at c and

$$F'(c) = f(c).$$

Proof. Let $c \in (a,b)$. Let $h > 0$. Then

$$\frac{F(c+h) - F(c)}{h} = \frac{1}{h} \int_c^{c+h} f(t)dt.$$

By Corollary 8.4.2 there exists $\theta \in [c, c+h]$ such that

$$\int_c^{c+h} f(t)dt = f(\theta)h.$$

Hence we have

$$\frac{F(c+h) - F(c)}{h} = f(\theta).$$

As $h \to 0$, $\theta \to c$, and due to continuity of f we conclude that $\lim_{h\to 0} f(\theta) = f(c)$. The assertion follows. The case $h < 0$ is similar. The cases $c = a$ and $c = b$ are similar (In these cases one talks on one-sided derivatives only.) □

Theorem 8.5.3. *Let f be continuous on $[a,b]$ and $f = g'$ fore some function g defined on $[a,b]$. Then for $x \in [a,b]$*

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

Proof. Let
$$F(x) = \int_a^x f(t)dt.$$
By Theorem 8.5.2 the function $F - g$ is differentiable on $[a, b]$ and $F' - g' = (F - g)' = 0$. Therefore by Corollary 5.2.2 there is a number c such that
$$F = g + c.$$
Since $F(a) = 0$ we have that $g(a) = -c$. Thus for $x \in [a, b]$
$$\int_a^x f(t)dt = F(x) = g(x) - g(a). \quad \square$$

The next theorem is often called *The Fundamental Theorem of Calculus.*

Theorem 8.5.4. *Let f be integrable on $[a, b]$ and $f = g'$ for some function g defined on $[a, b]$. Then*
$$\int_a^b f(x)dx = g(b) - g(a).$$

Proof. Let $P = \{x_0, \ldots, x_n\}$ be a partition of $[a, b]$. By the Mean Value Theorem there exists a point $t_i \in [x_{i-1}, x_i]$ such that
$$g(x_i) - g(x_{i-1}) = g'(t_i)(x_i - x_{i-1}) = f(t_i)(x_i - x_{i-1}).$$
Let
$$m_i = \inf_{[x_{i-1},x_i]} f(x), \ M_i = \sup_{[x_{i-1},x_i]} f(x).$$
Then
$$m_i((x_i - x_{i-1}) \le f(t_i)(x_i - x_{i-1}) \le M_i(x_i - x_{i-1}), \quad \text{that is}$$
$$m_i((x_i - x_{i-1}) \le g(x_i) - g(x_{i-1}) \le M_i(x_i - x_{i-1}).$$
Adding these inequalities for $i = 1, \ldots, n$ we obtain
$$\sum_{i=1}^n m_i((x_i - x_{i-1}) \le g(b) - g(a) \le \sum_{i=1}^n M_i(x_i - x_{i-1}),$$
so that for any partition we have
$$L(f, P) \le g(b) - g(a) \le U(f, P),$$
which means that
$$g(b) - g(a) = \int_a^b f(x)dx. \quad \square$$

8.6 [Integral as the limit of integral sums - in F.T.A.]

Let $f : [a, b] \to \mathbb{R}$. We defined the integral sums of f in Definition 8.1.3.

Definition 8.6.1. A number A is called the limit of integral sums $\sigma(f, P, \xi)$ if

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall P)(\forall (\xi_i) \in P)\left[(\|P\| < \delta) \Rightarrow \big(|\sigma(f, P, \xi) - A| < \varepsilon\big)\right].$$

In this case we write

$$\lim_{\|P\| \to 0} \sigma(f, P, \xi) = A.$$

The next theorem shows that the Riemann integral can be equavalently defined via the limit of the integral sums.

Theorem 8.6.1. *Let $f : [a, b] \to \mathbb{R}$. The f is Riemann integrable if and only if $\lim_{\|P\| \to 0} \sigma(f, P, \xi)$ exists. In this case*

$$\lim_{\|P\| \to 0} \sigma(f, P, \xi) = \int_a^b f(x)dx.$$

Proof. First, assume that f is Riemann integrable. Then we know that f is bounded, so that there is a constant C such that $|f(x)| \le C$ for all $x \in [a, b]$. Fix $\varepsilon > 0$. Then there exists a partition P_0 of $[a, b]$ such that

$$U(f, P_0) - L(f, P_) < \varepsilon/2.$$

Let m be the number of points in the partition P_0. Choose $\delta = \dfrac{\varepsilon}{8mC}$. Then for any partition P_1 such that $\|P_1\| < \delta$ and $P = P_0 \cup P_1$ we have

$$\begin{aligned} U(f, P_1) &= U(f, P) + \big(U(f, P_1) - U(f, P)\big) \\ &\le U(f, P_0) + \big(U(f, P_1) - U(f, P)\big) \\ &\le U(f, P_0) + 2C\|P_1\|m < U(f, P_0) + \varepsilon/4. \end{aligned}$$

Similarly,

$$L(f, P_1) > L(f, P_0) - \varepsilon/4.$$

Therefore we get

$$L(f, P_0) - \varepsilon/4 < L(f, P_1) \le U(f, P_1) < U(f, P_0) + \varepsilon/4.$$

Hence

$$U(f, P_1) - L(f, P_1) < \varepsilon,$$

which together with the inequalities

$$L(f, P_1) \le \int_a^b f(x)dx \le U(f, P_1),$$

$$L(f, P_1) \le \sigma(f, P_1, \xi) \le U(f, P_1)$$

leads to

$$\Big|\int_a^b f(x)dx - \sigma(f, P_1, \xi)\Big| < \varepsilon.$$

Now suppose that $\lim_{\|P\| \to 0} \sigma(f, P, \xi) = A$. Fix $\varepsilon > 0$. Then there exists $\delta > 0$ such that if $\|P\| < \delta$ then

$$A - \varepsilon/2 < \sigma(f, P, \xi) < A + \varepsilon/2.$$

Choose P as above. Varying (ξ_i) take sup and inf of $\sigma(f, P, \xi)$ in the above inequality. We obtain

$$A_\varepsilon/2 \le L(f, P) \le U(f, P) \le A + \varepsilon/2.$$

By the criterion of integrability this shows that f is Riemann integrable. ☐

8.7 Improper integrals. Series

Integrals over an infinite interval

Definition 8.7.1. The improper integral $\int_a^\infty f(x)dx$ is defined as

$$\int_a^\infty f(x)dx = \lim_{A\to\infty}\int_a^A f(x)dx.$$

We use the same terminology for improper integrals as for series, that is, an improper integral may converge or diverge.

Example 8.7.1.

$$\int_1^\infty \frac{1}{x^2}dx = 1.$$

Indeed,

$$\int_1^A \frac{1}{x^2}dx = 1 - \frac{1}{A} \to 1 \text{ as } A \to \infty.$$

Theorem 8.7.1. $\displaystyle\int_1^\infty \frac{1}{x^k}dx$ *converges if and only is* $k > 1$*.*

Integrals of unbounded functions

Example 8.7.2.

$$\int_0^1 \frac{dx}{\sqrt{x}} = \lim_{\delta\to 0}\int_\delta^1 \frac{dx}{\sqrt{x}} = \lim_{\delta\to 0}(2 - 2\sqrt{\delta}) = 2.$$

The notion of the improper integral is useful for investigation of convergence of certain series. The following theorem is often called the integral test for convergence of series.

Theorem 8.7.2. *Let* $f : [1,\infty) \to \mathbb{R}$ *be positive and increasing. Then the integral* $\int_1^\infty f(x)dx$ *and the series* $\sum_{n=1}^\infty f(n)$ *both converge or both diverge.*

Proof. Since f is monotone it is integrable on any finite interval (Theorem 8.3.1). For $n-1 \leq x \leq n$ we have

$$f(n) \leq f(x) \leq f(n-1).$$

Integrating the above inequality from $n-1$ to n we obtain

$$f(n) \leq \int_{n-1}^n f(x)dx \leq f(n-1).$$

Adding these inequalities on interrvals $[1,2], [2,3], \ldots, [n-1,n]$ we have

$$\sum_{k=2}^n f(k) \leq \int_1^n f(x)dx \leq \sum_{k=1}^{n-1} f(k).$$

Now the assertion easily follows. □

As an application of the above theorem we consider the following

Example 8.7.3. The series
$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$$
converges if $\alpha > 1$.

Proof. By Thoerem 8.7.2 it is enough to prove that
$$\int_1^\infty \frac{dx}{x^\alpha} < \infty.$$
Indeed,
$$\int_1^\infty \frac{dx}{x^\alpha} = \lim_{A\to\infty} \int_1^A \frac{dx}{x^\alpha} = \frac{1}{\alpha-1}\left(1 - \frac{1}{A^{\alpha-1}}\right) = \frac{1}{\alpha-1} < \infty. \qquad \square$$

Example 8.7.4. The series
$$\sum_{n=1}^{\infty} \frac{1}{n(\log n)^\alpha}$$
converges if $\alpha > 1$ and diverges if $\alpha \le 1$.

Proof. Left as an exercise.

8.8 The constant π

In Chapter 7 we defined the trigonometric functions sin and cos by the series and proved that they are periodic with period 2ϖ. Here we show that ϖ is the same constant as π know from elementary geometry.

Consider a circle $x^2 + y^2 \le 1$, which is centered at the origin with radius 1. It is known that its area is π.

The area of the semi-circle can be obtained by
$$\int_{-1}^{1} \sqrt{1-x^2}dx = \int_0^\varpi \sin^2\theta d\theta = \frac{1}{2}\int_0^\varpi (1-\cos 2\theta)d\theta = \frac{1}{2}\varpi$$
where we used the substitution $x = \cos\theta$.

Hence $\varpi = \pi$.