A FRIENDLY INTRODUCTION TO ANALYSIS

Maung Min-Oo

April 3, 2004

Preface

Analysis is the rigorous and more advanced study of the techniques and results used in calculus. As is well-known, although calculus was "invented" by Newton and Leibniz in the 17th century a more precise and rigorous foundation was laid only about 200 years later by Cauchy, Weierstrass Bolzano Dedekind, Cantor and others. The main difficulty was to give a precise definition of the set of real numbers and understand the notions of infinities, infinitesimals and limits. Of course, such attempts go back to ancient times as can be seen in the paradoxes of Zeno of Elea, 5th century B.C.

The main difficulty of modern-day undergraduate students in understanding rigorous analysis is the lack of basic training in the nature and structure of logical arguments and mathematical proofs. This is partly due to the fact that disciplines like classical Euclidean Geometry, which generations of human beings have studied, are no longer taught in schools. It is claimed that after the Bible, Euclid's Elements is the most widely read book of western civilization. In any case, proofs not only "scare" many students, but it is also sometimes hard for students to appreciate why mathematicians go through all that trouble to "prove" what seems like obvious facts. I do admit that the notation and style of most analysis text books is quite unappealing, especially to students whose only encounter with mathematics is a "standardized calculus text-book" where proofs are not rigorous, definitions are not clearly stated and the emphasis is more on a rather superficial (cook-book-style) mastery of certain techniques. It is therefore not the fault of the students that they are never exposed to some of the finer points and perhaps the true nature of mathematical reasoning in the first year at a University. Education can then degenerate into blindly following a set of technical instructions (given by instructors (sic)) rather than an awe-inspiring and mind-opening experience that we all strive for in life. I admit I'm also not so sure that the prescribed material that I will be teaching you in this course is really that mind-boggling either! (On a personal note, I should add that in the early 70's in Germany, where I was a student, there were no credits for courses and also no tuition fees. I never had to take a written test or examination. All examinations I took were oral and one-on-one with my professors. It was challenging and a lot of fun!)

These notes are written in an informal and personal style of a "friendly" mathematician, whose motto is: *"It's better to be approximately right than exactly wrong"*. The notes are by no means a substitute for my live lectures (which by the way are in general much more entertaining!). They are aimed to give a somewhat more organized and abbreviated outline of the material to be discussed in the lectures. The notes (and also the lectures) are not supposed to be comprehensive. The art of teaching (and learning) consists in choosing carefully what is relevant and beautiful. My aim is to let you discover (McMaster motto!) rather than cover the maximum amount of material. I therefore strongly advise every student to read at least one other recommended book in Real Analysis. Some are available on reserve in the Thode library. I apologize for not including too many exercises in the notes, but I will be making up on that by giving you more exercises and problems during the term. I also plan to post extra supplementary material and interesting links related to the course web site. My last (in my opinion, unforgivable) mistake is the omission of any pictures in these notes. I am a geometer and I think visually, but I couldn't bring myself to spend the time to make the appropriate ps files etc. I will certainly draw some pictures on the blackboard during the lectures and hopefully, if I am allowed to teach this course again and if these notes ever become publishable (you will then have to pay for them!), I will include some pictures, although, Analysis is still not Geometry!

Yours truly,
Min-Oo

Contents

Preface **i**

0 Preliminaries **1**

0.1 Some elementary propositional logic . . . 1
0.2 Some naive set theory . . . 2
0.3 Natural Numbers and the Principle of Induction . . . 4
0.4 Rational Numbers . . . 4

1 Real Numbers **7**

1.1 Field Axioms . . . 7
1.2 Axioms of Order . . . 10
1.3 The Completeness Axiom . . . 13
1.4 Remarks on orders of Infinity . . . 16
1.5 Exercises . . . 18
1.5.1 Short solutions to some of the exercise problems . . . 19

2 Sequences and Series **21**

2.1 Sequences . . . 21
2.2 Convergence . . . 22
2.3 Monotonic sequences . . . 25
2.4 Cauchy sequences . . . 28
2.5 Series . . . 30
2.6 Convergence criteria for series . . . 32
2.7 Power Series . . . 34
2.8 The Exponential Function . . . 36
2.9 Exercises . . . 39
2.9.1 Short solutions to some of the exercises . . . 42

3 Topology of the real line $\mathbb{R}$ 45
3.1 Open sets and closed sets 45
3.2 Connected sets 46
3.3 Compact sets 46
3.4 Elementary topology of $\mathbb{R}^n$ 48
3.5 Exercises 49
3.5.1 Solutions to some exercise problems 50

4 Continuity 51
4.1 Continuous functions 51
4.2 Continuity and Connectedness 53
4.3 Continuity and Compactness 55
4.3.1 The Contraction Mapping Principle 57
4.4 Uniform Continuity 57
4.5 Exercises 59
4.5.1 Short solutions to some of the exercise problems . . . 60

5 Differentiation 63
5.1 Definition and basic properties 63
5.2 Local properties of the derivative 66
5.3 Some global properties of the derivative 67
5.3.1 The Logarithm 70
5.4 Differentiation in $\mathbb{R}^n$ 70
5.4.1 Inverse Function Theorem 71
5.5 Exercises 74
5.5.1 Hints and short solutions 75

6 Integration 77
6.1 Definition of the Riemann Integral 77
6.2 Basic properties of the Integral 79
6.3 Fundamental Theorem of Calculus 80
6.4 Improper Integrals 81
6.4.1 The Gamma Function 82
6.4.2 Stirling's Formula 83
6.4.3 The Euler-Maclaurin Summation Formula 85
6.5 Exercises 88
6.5.1 Hints and short solutions 90

7 Metric spaces and Uniform Convergence **93**
7.1 Definitions and examples 93
7.2 Uniform Convergence 95
7.2.1 Uniform convergence of power series 98
7.3 Weierstrass approximation theorem 98
7.4 Exercises 100
7.4.1 Hints and short solutions to the exercises 102

Chapter 0

Preliminaries

0.1 Some elementary propositional logic

First we address some simple matters of the basic logic used in mathematics especially for definitions and proofs.

A proposition is a statement which can (only) be either true or false. For example, the proposition "$1+2=2+1$" is true, while the proposition "There is a rational number x such that $x^2 = 2$ " is false.

If p and q are propositions then the proposition $p \wedge q$ (p AND q) is true if both p and q are true and false otherwise. The proposition $p \vee q$ (p OR q) is true if either p or q (or both) is (are) true and is false otherwise. The proposition $\neg p$ (read NOT p) is true if p is false and is false if p is true. In some books this is also written as $\sim p$.

The proposition $p \Rightarrow q$ (p implies q or if p then q) is only false if p is true and q is false and is true otherwise. In particular, if p is false then $p \Rightarrow q$ is true. In other words, $p \Rightarrow q$ has the same truth value as $\neg p \vee q$. The proposition $p \Leftrightarrow q$ (read p is equivalent to q or p if and only if q) is true if either both p and q are true or both p and q are false and is false otherwise. In other words, $(p \Leftrightarrow q) \Leftrightarrow (p \Rightarrow q) \wedge (q \Rightarrow p)$. We sometimes write "iff" as an abbreviation for if and only if.

Many of the propositions we will encounter in this course are constructed using the quantifiers $\forall$ and $\exists$. The symbol $\forall$ stands for for all or for every or for each and the symbol $\exists$ stands for there exists or there is. To make the expressions with these quantifiers easier to read we also use the phrase "such that " (often abbreviated to "s.t.")quite often. For example:

$$\forall m \in \mathbb{Z}\ \exists n \in \mathbb{Z}\, s.t.\ n > m$$

which we can read as: For every integer m there is some integer n such that $n > m$. This is obviously true. However, the statement:

$$\exists n \in \mathbb{Z}\ \forall m \in \mathbb{Z}\, s.t.\ n > m$$

which we can read as: There is some integer n such that for every integer m we have $n > m$ is definitely false. So the negation:

$$\forall n \in \mathbb{Z}\ \exists m \in \mathbb{Z}\, s.t.\ n \leq m$$

is true. Note how we switched the quantifiers to negate a given proposition. The main message here is that the *order* in which the quantifiers are written is extremely important. It is easy to fall into logical blunders (even for discussions in everyday life!) if one is not careful with quantifiers.

0.2 Some naive set theory

We will ignore all the subtle paradoxical difficulties in defining sets and think of a set naively as a collection of objects called its elements. We use the notation: $x \in S$ if x is an element of a set S (x belongs to S) and $x \notin S$ if x is not an element of S, (x does not belong to S).
The simplest way to define a set is by listing all its elements. We use the notation: $S = \{\pi, e, 2.71828, \sqrt{2}, 1.4142\}$ to denote the set of five (distinct) real numbers: $\pi, e, 2.71828, \sqrt{2}$ and 1.4142 . This notation is of course, not always applicable especially if we have a set with an infinite number of elements. In such cases, sets are defined by naming the property which distinguishes elements of the set from objects which are not in the set. We use the notation: $\{x \big| \, p(x)\}$ (or sometimes $\{x \, : \, p(x)\}$), where $p(x)$ is a statement about the object x which can only be true or false. The set consists of all objects (of a bigger set) for which $p(x)$ is true. For example, $\{x \in \mathbb{Q} \,\big|\, x^2 < 2\}$ denotes the set of all rational numbers whose square is strictly less than 2. We will assume some familiarity with the following examples:

(i) the natural numbers $\mathbb{N} = \{\, 1, 2, 3, \dots \,\}$,

(ii) the integers $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$, and

(iii) the rational numbers (or fractions) $\mathbb{Q} = \{\frac{n}{d} \,\big|\, n, d \in \mathbb{Z}, d > 0\}$.

All these sets lie inside the set of real numbers $\mathbb{R}$ which is what this course is mainly about. Here are some more standard notations that we will use for sets:

If A, B are two sets, we say that A is a subset of B and write $A \subset B$ if every element of A is also an element of B. For example, $A \subset A$ and $\emptyset \subset A$ for any set A , where $\emptyset$ denotes the empty set which does not contain any elements.

$A \cup B = \{x \,|\, (x \in A) \vee (x \in B)\}$ and $A \cap B = \{x \,|\, (x \in A) \wedge (x \in B)\}$ will denote the union and intersection of two sets.

$A \times B$ denotes the Cartesian product which is the set of all ordered pairs: $\{(a, b)| \; a \in A, b \in B\}$.

One of the important concepts in Mathematics is the idea of a function (or a map) from a set to another set. We will be mainly concerned with functions from intervals in $\mathbb{R}$ to $\mathbb{R}$. It took a long time for mathematicians to clarify exactly what should be meant by a function. A simple operational definition is as follows:

A map or a function f from a set A to another set B, written $f : A \to B$ is a rule that assigns to every element $a \in A$ a *unique* element $f(a) \in B$. We write $a \mapsto b = f(a)$ and say that a is mapped to b. The set A is called the domain of f and the set of all $b \in B$ with $f(a) = b$ for some $a \in A$ is called the range (or image) of f.

The graph of a function f is then defined to be the subset

$$\{(a, f(a)) \in A \times B \,\big|\, a \in A\}$$

of the Cartesian product. (In fact, a more rigorous way to define a function is through its graph).

f is said to be *one-one* (or injective) if two distinct elements of A never map to the same element of B. In other words: $f : A \to B$ is one-one iff $(\forall\, x, y \in A)(f(x) = f(y) \Rightarrow x = y)$. f is said to be *onto* (or surjective) if every element of B has some element of A mapped to it. In other words; $f : A \to B$ is onto iff $\forall\, b \in B \,\exists\, a \in A \, s.t. \; f(a) = b$. A map which is both one-one and onto is called a *bijection* (or a one-one correspondence).

If $f : A \to B$ is a map, the *image* of a set $X \subset A$ is defined by $f(X) = \{b \in B \,|\, \exists\, x \in X \, s.t. \, b = f(x)\}$ and the *inverse image* of a set $Y \subset B$ is defined by $f^{-1}(Y) = \{a \in A \,|\, f(a) \in Y\}$.

0.3 Natural Numbers and the Principle of Induction

Natural numbers are used for counting and ordering. The main property (or axiom, if you wish) about $\mathbb{N}$ that we will use is the Induction Principle

Axiom 0.3.1 Principle of Induction
In order to prove the proposition $P(n)$ for all $n \geq n_0$. it is sufficient to show the following:

N1 $P(n_0)$ *is true*

N2 *(induction step)* $\forall\, k \geq n_0$, *we have* $P(k) \Rightarrow P(k+1)$

We also use the inductive property of the natural numbers for recursive definitions such as in the following definition of the factorial.

$0! = 1$

$\forall\, k \geq 0\ (k+1)! = (k+1)k!$

0.4 Rational Numbers

The rational numbers $\mathbb{Q}$ are all the fractions a/b with $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ ($b \neq 0$), where we identify two expressions a/b and c/d as defining the same rational number if $ad = bc$ (this is the usual way we cancel fractions). We add, subtract, multiply and divide rational numbers the same way as we did in elementary school (Of course, we do not divide by zero!) and find out that the set of rational numbers $\mathbb{Q}$ form, what is known as a *field* (which we will define in the next chapter). It just means that if $r, s \in \mathbb{Q}$ then $r+s, r-s, r.s$ and also r/s (provided$s \neq 0$ are all in $\mathbb{Q}$ and the operations satisfy the usual commutative, associative and distributive rules that we are familiar with from elementary school. The numbers 0 and 1 of course, play a special role. 0 is the “identity element” for addition and 1 is the “identity element” for multiplication. Identity just meaning that it does not change any element that it operates on.

The Greeks believed a long time ago that fractions were sufficient to describe all “real” phenomena. However, the Pythagoreans, a philosophical school founded by Pythagoras of Samos (569 to 475 BC) discovered the following result.

Proposition 0.4.1 *The real number $\sqrt{2}$ is not a rational number.*

Proof (by contradiction) Suppose that $\sqrt{2} = a/b$ with $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. We may as well assume that a, b have no common factors else we could cancel them out. Then $2b^2 = a^2$ and so a is even. But then a^2 and hence $2b^2$ is divisible by 4 and so b^2 is even. But then b is also even and so a and b do have a common factor, viz. 2 . Thus we arrive at a contradiction. Hence $\sqrt{2}$ is not in $\mathbb{Q}$.

QED

The Pythagoreans realized that one could easily construct a line segmentof length $\sqrt{2}$ by elementary geometry (for example, the diagonal of a unit square) and so they were forced to the conclusion that rational numbers were not sufficient to describe their geometric system. Mathematics need real numbers to describe reality (or at least the shadow of it!).

Chapter 1

Real Numbers

1.1 Field Axioms

The basic algebraic fact about the real numbers is that they form a field. Of course, there are many other fields and the theory of fields (for example, Galois theory) is an elegant algebraic subject. However, this is a course in analysis, so we will simply assume that the student is familiar with the operations of addition, subtraction, multiplication and division for real numbers and the basic rules that they satisfy (commuativity, associativity, distributivity etc.) For the sake of completeness, we will give the formal algebraic definition of a field.

Definition 1.1.1 *A* field *is a set* $\mathbb{F}$ *equipped with two operations* $+ : \mathbb{F}\times\mathbb{F} \to \mathbb{F}$ *(addition) and* $\cdot\ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ *(multiplication) satisfying the following axioms:*

F1 $\forall\, x, y, z \in \mathbb{F}; x + (y + z) = (x + y) + z$ *and* $x \cdot (y \cdot z) = (x \cdot y) \cdot z$

F2 $\forall\, x, y \in \mathbb{F}; x + y = y + x$ *and* $x \cdot y = y \cdot x$

F3 $\forall\, x, y, z \in \mathbb{F}; x \cdot (y + z) = (x \cdot y) + (x \cdot z)$

F4 $\exists\, 0 \in \mathbb{F}$ *s.t.* $\forall x \in \mathbb{F}; 0 + x = x$

F5 $\exists\, 1 \in \mathbb{F}\,, 1 \neq 0$ *s.t.*$\forall x \in \mathbb{F}; 1 \cdot x = x$

F6 $\forall\, x \in \mathbb{F}\, \exists (-x) \in \mathbb{F}$ *s.t.* $(-x) + x = 0$

F7 $\forall\, x \in \mathbb{F}$ *with* $x \neq 0$, $\exists x^{-1} \in \mathbb{F}$ *such that* $(x^{-1}) \cdot x = 1$

One normally omits the symbol $\cdot$ for multiplication and simply juxtapose. We also write $x - y$ for $x + (-y)$. Our main algebraic axiom about $\mathbb{R}$ is therefore:

Axiom 1.1.1 *The real numbers form a field under the usual rules of addition and multiplication.*

Other well-known examples of fields are:

1. $\mathbb{Q}$ is a field, but $\mathbb{N}$ and $\mathbb{Z}$ do not satisfy all the field axioms. (Can you see which of the axioms fail for $\mathbb{N}$ and for $\mathbb{Z}$?)

2. However integers modulo a prime number p denoted by $\mathbb{Z}/p\mathbb{Z}$ form a field. (The main axiom to check is the last one about the existence of the multiplicative invers. can you see that this would fail if p is not a prime, say $p = 12$)

3. The complex numbers $\mathbb{C}$ form a field. (Do you know any other fields?)

The field axioms trivially imply some algebraic operational rules that we are all familiar with from elementary school. A few selected examples of such rules are the following:

(i) The commutative and the associative laws can be extended by induction to any finite number of elements. The same is true for the distributive law.

(ii) The "neutral" elements 0 and 1 are unique and the same is true for both "inverses" $-x$ and x^{-1}.

(iii) $0\,x = 0$ for all x.
Proof: $0x = (0+0)x = 0x + 0x$. Now add the inverse $-(0x)$ to both sides of the equation and use associativity of multiplication: $0 = -(0x) + 0x = -(0x) + (0x + 0x) = (-(0x) + 0x) + 0x = 0 + 0x = 0x$
This seems like splitting hairs but that is what we are learning here!

(iv) Simple rules such as $-(-a) = a$; $(a^{-1})^{-1} = a$ for $a \neq 0$; $(-a)\,b = a.(-b) = -ab$; $(-a)(-b) = ab$ etc.
Sample proof: $0 = 0a = a0 = a(b + (-b)) = ab + a(-b)$ and so $-(ab) = 0 + (-(ab)) = -(ab) + ((ab) + (a(-b)) = 0 + a(-b) = a(-b)$.

(v) One important fact is the cancellation rule (or in fancier words, the absence of zero divisors in a field!): $ab = 0$ if and only if $a = 0 \vee b = 0$. This implies: $ac = ab \wedge a \neq 0 \Rightarrow c = b$. (This is, for example, not true in $\mathbb{Z}/12\mathbb{Z}$) .
Proof: If $ab = 0$ and suppose $a \neq 0$. Then $\exists\, a^{-1}$ so that $a^{-1}a = 1$, but then $0 = a^{-1}0 = a^{-1}(ab) = (a^{-1}a)b = 1b = b$. *QED*

(vi) Powers are defined inductively by: $x^0 = 1$, $x^{n+1} = x.x^n$ for $n \in \mathbb{N}$. We also define $x^{-n} = (x^{-1})^n = (x^n)^{-1}$

A more important result is the binomial formula:

Proposition 1.1.1

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

for any $n \in \mathbb{N}$, where the binomial coefficients are defined by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Proof (by induction)
For $n = 1:\ (x+y)^1 = x + y = \binom{1}{0} x^1 y^0 + \binom{1}{1} x^0 y^1$

Assume: $(x+y)^k = \sum_{j=0}^{k} \binom{k}{j} x^{k-j} y^j$

Then

$$\begin{aligned}
(x+y)^{k+1} &= (x+y) \sum_{j=0}^{k} \binom{k}{j} x^{k-j} y^j \\
&= \sum_{j=0}^{k} \binom{k}{j} \left(x^{k-j+1} y^j + x^{k-j} y^{j+1} \right) \\
&= \sum_{j=0}^{k} \binom{k}{j} x^{k-j+1} y^j + \sum_{j=1}^{k} \binom{k}{j-1} x^{k-j+1} y^j \\
&= \sum_{j=0}^{k+1} \binom{k+1}{j} x^{k+1-j} y^j
\end{aligned}$$

where we use the Pascal Triangle relation: $\binom{k}{j} + \binom{k}{j-1} = \binom{k+1}{j}$.

This relation follows by a simple calculation with factorials from the definition of the binomial coefficients, but if we use the fact that $\binom{n}{k}$ represents the number of different ways of choosing k objects from n objects (that's why we read "n choose k" for $\binom{n}{k}$), it can be seen immediately from the following combinatorial argument: If we have to choose a committee of j members from k students (that's you) and 1 professor (that's me), then there are only two mutually exclusive possibilities: either I'm on the committee or I'm not.

1.2 Axioms of Order

There is a non-empty subset of real numbers called the positive numbers (notation: $a > 0$) satisfying the following axioms:

O1 Trichotomy: For any $a \in \mathbb{R}$ exactly one of $a > 0$, $a = 0$, $-a > 0$ is true.

O2 If $a, b > 0$ then $a + b > 0$ and $a \cdot b > 0$

A field with a relation called an ordering satisfying the three properties above is called an *ordered field*. The field $\mathbb{Q}$ of rationals is an ordered field (with the usual ordering) but the field $\mathbb{C}$ of complex numbers is not an ordered field (under any ordering). Our next basic axiom about $\mathbb{R}$ is therefore:

Axiom 1.2.1 *The real numbers form an ordered field.*

We will write $a > b$ or $b < a$ for $a-b > 0$. It follows from the axioms that the ordering $>$ is *transitive*, i.e. , if $a > b$ and $b > c$, then $a > c$ (because $a - c = (a - b) + (b - c)$) We also write $a \geq b$ to mean the same thing as $a > b \vee a = b$. ($a \leq b$ is equivalent to $-a \geq -b$ because $b-a = -a-(-b)$).

The axioms trivially imply some simple and familiar properties such as:

(i) $1 > 0$.

(ii) $x > 0, y < 0 \Rightarrow xy < 0$; $x < 0, y < 0 \Rightarrow xy > 0$.

(iii) If $a > b \Rightarrow a + c > b + c$. If $c > 0$ and $a > b$, then $ac > bc$ (because $ac - bc = (a - b)c$, but if $c < 0$, then $a > b \Rightarrow bc > ac$.

(iv) If $0 < a < b$, then $a^{-1} > b^{-1} > 0$, but if $a < b < 0$, then $0 > a^{-1} > b^{-1}$.

(v) The square x^2 of a real number $x \neq 0$ is always strictly positive: $x^2 > 0$ (because $(-x)(-x) = x^2$). Of course $0^2 = 0$.

The following notation will be used for intervals on the real line $\mathbb{R}$: An *open* interval: $(a, b) = \{x \in \mathbb{R} |\, a < x < b\}$ does not contain its end points. A *closed* interval: $[a, b] = \{x \in \mathbb{R} |\, a \leq x \leq b\}$ contains both end points. We will also use some other types of intervals: $[a, b), (a, b], (-\infty, b], (a, \infty)$, etc. Their meaning should be clear from the notation, for example: $[a, b) = \{x \in \mathbb{R} |\, a \leq x < b\}$ $(-\infty, b] = \{x \in \mathbb{R} |\, x \leq b\}$.

We define the *absolute value* of a real number to be:

Definition 1.2.1 $|x| = x$ *if* $x \geq 0$ *and* $|x| = -x$ *if* $x < 0$

Then we have $-|x| \leq x \leq |x|$. (in fact, it is true that either $|x| = x$ or $x = |x|$. It can never happen that $-|x| < x < |x|$. If we add the two inequalities for x and y, we obtain: $x + y \leq |x| + |y|$ and also $-x - y \leq |x| + |y|$. This proves the following basic:

Proposition 1.2.1 *(Triangle Inequality)*

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

for all $x, y \in \mathbb{R}$.

Of course, it can happen now that the strict inequality holds, e.g., $|2 + (-1)| = 1$ is strictly less than $|2| + |-1| = 2 + 1 = 3$.

Other very useful inequalities are the following:

Proposition 1.2.2 *(Bernoulli's Inequality)*

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

for every $x \geq -1$ and for all $n \in \mathbb{N}$

Proof: (by induction) For $n = 1$, $1 + x = 1 + 1.x$, so the inequality is true. Assuming that it is true for $n = k$, so that $(1 + x)^k \geq 1 + kx$, we have

$$(1+x)^{k+1} = (1+x)(1+x)^k \geq (1+x)(1+kx) = 1+(k+1)x+kx^2 \geq 1+(k+1)x$$

since $1 + x \geq 0$ and $kx^2 \geq 0$.

QED

Proposition 1.2.3 *(Cauchy-Schwarz Inequality)*

$$(x_1y_1 + \cdots x_ny_n)^2 \leq (x_1^2 + \cdots x_n^2)(y_1^2 + \cdots y_n^2)$$

for every $x_1, \cdots, x_n, y_1, \cdots, y_n \in \mathbb{R}$ and for all $n \in \mathbb{N}$. Moreover equality holds iff the x_i's are proportional to the y_i's.

Proof: The inequality is trivially true if all $x_1 = \cdots = x_n = 0$ or if $y_1 = \cdots = y_n = 0$, so we can assume w.l.o.g. (without loss of generality) by a simple scaling, that $x_1^2 + \cdots + x_n^2 = y_1^2 + \cdots + y_n^2 = 1$.

$$\sum_{i=1}^{n}(x_i - y_i)^2 \geq 0 \Rightarrow 2 - 2\sum_{i=1}^{n} x_iy_i \geq 0 \Rightarrow \sum_{i=1}^{n} x_iy_i \leq 1$$

$$\sum_{i=1}^{n}(x_i+y_i)^2 \geq 0 \Rightarrow 2+2\sum_{i=1}^{n} x_iy_i \geq 0 \Rightarrow -\sum_{i=1}^{n} x_iy_i \leq 1$$

Therefore

$$\Big|\sum_{i=1}^{n} x_iy_i\Big| \leq 1 = \big(\sum_{i=1}^{n} x_i^2\big)\big(\sum_{i=1}^{n} y_i^2\big)$$

Moreover, equality holds iff $x_i = y_i$ or $x_i = -y_i$ for all i.

QED

Proposition 1.2.4 *(Geometric Mean / Arithmetic Mean Inequality)*
If $x_1, \dots, x_n$ are positive real numbers then

$$\Big(\frac{x_1+\cdots+x_n}{n}\Big)^n \geq x_1\cdots x_n$$

Moreover, strict inequality holds unless all the numbers are equal.

Proof (by induction): For $n = 1$ the inequality is trivially true. We assume the inequality is true for any n positive numbers, i.e. $(AM_n)^n \geq (GM_n)^n$ where the arithmetic mean of n positive numbers $x_1 \cdots, x_n$ is defined by $AM_n = \frac{x_1+\cdots+x_n}{n}$ and the geometric mean is defined by $(GM_n)^n = (x_1 \cdots x_n)$. (I am just avoiding taking the nth root!).
We want to show the inequality holds for $n+1$ positive real numbers and we may assume w.l.o.g. that $x_1 \leq \ldots \leq x_n \leq x_{n+1}$. Now

$AM_{n+1} = \frac{x_1+\cdots+x_{n+1}}{n+1} = \frac{n}{n+1}AM_n + \frac{1}{n+1}x_{n+1} = AM_n\big(1+\frac{1}{n+1}(\frac{x_{n+1}}{AM_n}-1)\big)$

Therefore:

$\Big(\frac{AM_{n+1}}{AM_n}\Big)^{n+1} = \Big(1+\frac{1}{n+1}(\frac{x_{n+1}}{AM_n}-1)\Big)^{n+1} \geq 1+(\frac{x_{n+1}}{AM_n}-1) = \frac{x_{n+1}}{AM_n}$

by the Bernoulli inequality. Hence, by the induction hypothesis

$(AM_{n+1})^{n+1} \geq (AM_n)^n\, x_{n+1} \geq (GM_n)^n\, x_{n+1} = (GM_{n+1})^{n+1}$

QED

1.3 The Completeness Axiom

The main analytic property which distinguishes $\mathbb{R}$ from $\mathbb{Q}$ is the completeness axiom. There are different ways to introduce this axiom depending on whether one wants to assume the Archimedean property (see below) as an axiom or not, but for the sake of simplicity (remember this course is supposed to be a friendly introduction to analysis!), we will use the least upper bound property which then automatically implies the Archimedean property.

Definition 1.3.1 *An* upper bound *of a non-empty subset A of $\mathbb{R}$ is an element $b \in \mathbb{R}$ with $b \geq a$ for all $a \in A$. A is said to be bounded from above if A has an upper bound. A* lower bound *of a non-empty subset A of $\mathbb{R}$ is defined analogously as an element $b \in \mathbb{R}$ with $b \leq a$ for all $a \in A$. A is said to be bounded from below if A has a lower bound. A is said to be bounded if it has both upper and lower bounds.*

Definition 1.3.2 *An element $M \in \mathbb{R}$ is called a* least upper bound *or* supremum *of a non-empty set A, written $lub(A)$ or $sup(A)$, if M is an upper bound of A and if b is any upper bound of A then $b \geq M$. M (if it exists!) is uniquely defined by this property.*

In other words, M is the upper bound of A such that any other upper bound of A (different from M) is strictly larger than M. If A is not bounded from above we often write $sup(A) = +\infty$

Definition 1.3.3 *An element $m \in \mathbb{R}$ is called a* greatest lower bound *or* infimum *of a non-empty set A, written $glb(A)$ or $inf(A)$, if m is a upper bound of A and if b is any lower bound of A then $b \leq m$. m (if it exists!) is uniquely defined by this property.*

In other words, m is the lower bound of A such that any other lower bound of A (different from m) is strictly smaller than M. If A is not bounded from below we often write $inf(A) = -\infty$

We can now state **The Least Upper Bound Property**:

Axiom 1.3.1 *If a non-empty subset A of $\mathbb{R}$ has an upper bound, it has a least upper bound.*

An ordered field that satisfies the least upper bound property is called a *complete ordered field*. So to sum up, all we assume about the real numbers $\mathbb{R}$ is that it is a complete ordered (Archimedean) field. (In fact, one can show that up to " isomorphism of ordered fields", that $\mathbb{R}$ is the only complete (Archimedean) ordered field).

Note that the ordered field $\mathbb{Q}$ is not complete. For example, the set $A = \{q \in \ Q | \, q^2 < 2\}$ is bounded but does not have a least upper bound in $\mathbb{Q}$. We will see why in a little while. We first list some easy consequences of the completeness axiom. First of all by just changing signs (and flipping the inequalities) it is obvious that any subset $A \subset \mathbb{R}$ which has a lower bound has a greatest lower bound. If $B = \{x \in \mathbb{R} | -x \in A\}$. Then B is bounded from above iff A is bounded from below and $glb(A) = -lub(B)$.

Before we state the next property we need to describe more precisely how the $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are imbedded into the reals. First of all $\mathbb{N}$ is identified with multiples of the unit element 1 in $\mathbb{R}$, i.e. $n = 1 + \cdots + 1$ where the sum is over n terms. The key fact here is that this is an injective (one to one) map. (This is for example, not true for finite fields). After that $\mathbb{Z}$ and $\mathbb{Q}$ can be imbedded in the obvious fashion, so $p/q \in \mathbb{Q}$ is identified with $p \cdot q^{-1}$ since $\mathbb{R}$ is a field. An important property about the real numbers which follows from our axioms is the following:

Proposition 1.3.1 . *$\forall x \in \mathbb{R} \; \exists \, n \in \mathbb{N}$ such that $n > x$.*

Equivalently: $\forall a > 0 \; \exists \, n \in \mathbb{N}$ such that $1/n < a$.

Proof: This is equivalent to saying that $\mathbb{N}$ is not bounded above. This seems like a very obvious fact, but we will prove it from the axioms. Suppose $\mathbb{N}$ were bounded above. Then it would have a least upper bound,M say. But then $M - 1$ is not an upper bound and so there is an integer $n \in \mathbb{N}$ with $n > M - 1$. But then $n + 1 > M$ contradicting the fact that M is an upper bound for all of $\mathbb{N}$. *QED*

The above property of the real numbers is called the **Archimedean property of the Reals** and is been attributed to the famous Greek mathematician Archimedes (287 to 212 BC) and appears in Book V of The Elements of Euclid.

We also deduce the following important fact:

Proposition 1.3.2 . *Between any two distinct real numbers there is at least one (and hence infinitely many) rational numbers.*

Proof: Let $a, b \in \mathbb{R}$ with (say) $a < b$. Choose $n \in \mathbb{N}$ so that $1/n < b - a$. Then look at integer multiples of $1/n$. Since these are unbounded, we may choose the first such multiple with $m/n > a$. We claim that $m/n < b$. If not, then since $(m-1)/n < a$ and $m/n > b$ we would have $1/n > b - a$.

QED

A set A with the property that an element of A lies in every interval (a, b) of $\mathbb{R}$ is called *dense* in $\mathbb{R}$. We have just proved that the rationals $\mathbb{Q}$ are dense in $\mathbb{R}$. The irrationals ($= \mathbb{R} \cap \mathbb{Q}^c$) are also dense in $\mathbb{R}$.

We now prove the result we stated earlier.

Proposition 1.3.3 *The real number $\sqrt{2}$ exists.*

Proof : We will get the existence of $\sqrt{2}$ as the least upper bound of the set $A = \{q \in \ Q | \, q^2 < 2\}$. We know that A is bounded above. 2 is an upper bound for A . (Why?) Let $b = lub(A)$. We now prove that $b^2 < 2$ and $b^2 > 2$ both lead to contradictions and so we must have $b^2 = 2$ (by the trichotomy axiom for the ordering).

So first suppose that $b^2 > 2$. By the Archimedean property we can choose an $n \in \mathbb{N}$ such that $n > \frac{2b}{b^2-2}$ so that $b^2 - \frac{2b}{n} > 2$. Then $(b - \frac{1}{n})^2 = b^2 - \frac{2b}{n} + (\frac{1}{n})^2 > 2$, since a square is always non-negative. Thus $b - \frac{1}{n}$ is an upper bound of A, contradicting the assumption that b is the least upper bound.

Similarly, if $b^2 < 2$, then $a = b + \frac{2-b^2}{b+2} = \frac{2b+2}{b+2}$ is $> b$ and is in A since $a^2 = \frac{4b^2+8b+4}{b^2+4b+4} = 2 - \frac{2(2-b^2)}{(b+2)^2} < 2$. This contradicts the fact that b is an upper bound of A.

QED

We end this section by reminding you that real numbers can be defined by decimal expansions. Given the decimal expansion $n + \sum_{i>0} a_i \, 10^{-i}$ of a positive real number, the set of rational approximations $q_k = n + \sum_{0<i\leq k} a_i \, 10^{-i}$ with $k \in \mathbb{N}$ form a bounded set and so it has a least upper bound. This is by definition then the real number defined by the decimal expansion. Every real number has a unique decimal expansion – except that decimals that terminate in a sequence of 9's should be truncated. For example 1.29999... = 1.3 . Rational numbers have decimal expansions which repeat periodically (or terminate). This follows from the Euclidean algorithm. Of course, the reason that we use the number 10 as our base for decimals probably stems from the "accidental" fact that humans have ten fingers! One could also use any other positive integer as a base. Computers use the binary system with base

2 (and also base 16). The terniary system (base 3) is useful in describing the Cantor set, which we will encounter in Chapter 3. The Cantor set consists of all points in $[0, 1]$ whose terniary decimal expansion does not contain 2. For example $0.12 = 5/9$ in base 3, so is not in the Cantor set.

1.4 Remarks on orders of Infinity

It was not until the 19th century that mathematicians realized mainly through the pioneering work of Georg Cantor (1845-1918) that infinity comes in different orders.

Definition 1.4.1 *A set is said to be* countable *if it can be put into one-one correspondence with* $\mathbb{N}$*.*

Intuitively this means that we can count off all the elements of a countable set. Examples and some basic properties for countable sets follows:

1. The sets $\mathbb{Z}$ and $\mathbb{Q}$ are countable.

2. A subset of a countable set is either finite or countable.

3. A countable union of finite or countable sets is finite or countable and also any (finite) Cartesian product of countable sets is countable.

However:

Theorem 1.4.1 *(Cantor) The set of real numbers* $\mathbb{R}$ *is not countable.*

Proof:
We will show that the interval $(0, 1) \subset \mathbb{R}$ is not countable. The method of proof used below is the famous Cantor diagonalisation argument.
Suppose we could write down the decimal expansions of *all* the real numbers in (0, 1) in a countable list: $0.a_1a_2a_3...$, $0.b_1b_2b_3...$, $0.c_1c_2c_3...$, ...
Now define a decimal number: $x = x_1x_2x_3...$ by $x_1 = a_1 + 1$ if $a_1 \leq 7$ and $x_1 = 0$ if $a_1 = 8$ or 9, $x_2 = b_2 + 1$ if $b_2 \leq 7$ and $x_2 = 0$ if $b_2 = 8$ or 9, etc. Then the decimal expression of x differs from the nth element of the countable list in the nth decimal place (and does not end in recurring 9's). Hence it represents an element of the interval $(0, 1)$ which is distinct from all members of the list and hence $(0, 1)$ is uncountable.

QED

A real number is called *algebraic* if it is a root of a polynomial with rational (or integer) coefficients. All other real numbers are called *transcendental*.

Proposition 1.4.1 *The set of algebraic numbers is countable. Hence there are uncountably many transcendental numbers.*

Proof: Since a polynomial of degree n with rational coeficients has $(n+1)$ coefficients, such polynomials can be put into one-one correspondence with $\mathbb{Q}^{n+1}$. This is countable since $\mathbb{Q}$ is countable and so there are only countably many polynomials of all degrees. Such a polynomial can have at most n roots and so there are only countably many such roots. Now, the set of algebraic numbers is the countable union of all roots of polynomials of all degrees and hence is a countable set.

Remark: Although there are uncountably many transcendental numbers, Liouville was the first to discover give a simple way to write down a decimal expression of a real number that is not algebraic. For example a decimal like 0.110001000... (with a 1 in the $n!$ place and 0 elsewhere) is transcendental. This follows from the fact that algebraic numbers cannot be approximated too well by rational numbers whose denominators are not too large. There are of course two famous transcendental numbers: e and π. That e is transcendental was established in 1873 by Hermite and in 1882, Lindemann finally proved that π is transcendental ending the dreams of all circle-squarers.

Another famous uncountable set with rather unusual properties that Cantor introduced and now named after him, is defined as follows:

Start with the unit interval $A_0 = [0,1]$. Now remove the middle third (open) segment $(\frac{1}{3}, \frac{2}{3})$. So what remains is the union of 2 closed intervals $A_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Now remove the middle third segment from each of these two closed intevals so that we get $A_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. Continuing in this manner ad infinitum, we obtain the Cantor set $C = \cup A_n$.

Some important facts about the Cantor set C:

(i) $x \in C$ iff the ternary (base 3) decimal expansion of x does not contain 1.

(ii) C is uncountable. (This follows from (i), so please prove it!)

(iii) C is closed and bounded and hence compact (see Chapter 3 for definitions).

1.5 Exercises

1. Find a formula for $\sum_{k=1}^{n}(2k-1)$ and prove your formula.

2. Show that: $n^2 < 2^n < n!$ for all natural numbers $n \geq 5$.

3. Show that $\sqrt{2}+\sqrt{3}$ is irrational.

4. Show that $0 < a < b \Rightarrow a^{-1} > b^{-1} > 0$ and $a < b < 0 \Rightarrow 0 > a^{-1} > b^{-1}$ for all $a, b \in \mathbb{F}$, where $\mathbb{F}$ is any ordered field.

5. Show that $||x|-|y|| \leq |x-y|$ for all $x, y \in \mathbb{R}$

6. Show that between any two distinct real numbers there is at least one (and hence infinitely many) irrational numbers.

7. Show that the number of subsets with exactly k elements of a set consisting of n elements ($k \leq n$) is $\binom{n}{k}$. What are the odds of winning the lottery "6 out of 49" ?

8. Show that

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

for any $x \in \mathbb{R}$ and for any $n \in \mathbb{N}$.

9. Show that $k!\binom{n}{k} \leq n^k$ for all $n \geq 1$ and $0 \leq k \leq n$

10. Prove the following identities:

(i) $\sum_{k=0}^{n}\binom{n}{k} = 2^n$

(ii) $\sum_{k=0}^{l}\binom{n}{k}\binom{m}{l-k} = \binom{n+m}{l}$

11. Show that

$$\left(1+\frac{1}{n}\right)^n \leq \sum_{k=0}^{n}\frac{1}{k!} < 3$$

for all $n \geq 1$.

12. Show that

$$(1+x)^n \geq \frac{1}{4}n^2x^2$$

for all real numbers $x \geq 0$ and for all natural numbers $n \geq 2$.

13. Show from first principles that $\forall n \in \mathbb{N}$ and $\forall\, y > 0$, $\exists$ a unique $x > 0$ such that $x^n = y$.

1.5.1 Short solutions to some of the exercise problems

1. $\sum_{k=1}^{n}(2k-1) = n^2$.
Proof by induction. For $n = 1$: $\sum_{k=1}^{1}(2k-1) = 1 = 1^2$.
Assume: $\sum_{k=1}^{n}(2k-1) = n^2$. Then $\sum_{k=1}^{n+1}(2k-1) = n^2+(2n+1) = (n+1)^2$
QED

2. By induction: For $n = 5$: $5^2 = 25 < 2^5 = 32$.
Assume: $k^2 < 2^k$. Then $(k+1)^2 = (1+\frac{1}{k})^2\, k^2 < 2\, k^2 < 2\, 2^k = 2^{k+1}$, since $(\frac{k+1}{k})^2 < (1+\frac{1}{k})^2 \le (\frac{6}{5})^2 < 2$ for all $k \ge 5$
Therefore: $n^2 < 2^n$ for all natural numbers $n \ge 5$.

Similarly: $2^5 = 32 < 5! = 120$.
Assume: $2^k < k!$. Then $2^{k+1} = 2\, 2^k < 2\, k! < (k+1)k! = (k+1)!$ for all $k \ge 5$.
Therefore: $2^n < n!$ for all natural numbers $n \ge 5$.

3. Suppose that $r = \sqrt{2} + \sqrt{3} > 0$ is rational. Then $r^{-1} = \sqrt{3} - \sqrt{2} \in \mathbb{Q}$ and therefore $\frac{1}{2}(r + r^{-1}) = \sqrt{2} \in \mathbb{Q}$, which contradicts what was proved in the lecture (and by the Greeks!).

4. $\forall\, x\ (x > 0 \Rightarrow x^{-1} > 0)$, since $x^{-1}x = 1 = 1^2 > 0$. So $b > a > 0 \Rightarrow a^{-1}b^{-1} = b^{-1}a^{-1} > 0$. Therefore by the order axioms: $b > a > 0 \Rightarrow (a^{-1}b^{-1})b = a^{-1} > (b^{-1}a^{-1})a = b^{-1} > 0$.
The case $a < b < 0$ is handled in a similar manner by applying the above argument to $-a > -b > 0$.

5. $x - y + y = x$. Therefore, by the triangle inequality: $|x-y| + |y| \ge |x|$. This implies: $|x-y| \ge |x| - |y|$. By interchanging the roles of x and y we also have: $|y-x| \ge |y| - |x|$. Combining these two inequalites, we have $|x-y| \ge ||x| - |y||$.

6. Given $a < b$, we can choose $n > \frac{10}{b-a}$, so that we have at least two rational numbers of the form $\frac{m+1}{n}, \frac{m+2}{n}$ inside the open interval (a,b). Now $\frac{m+\sqrt{2}}{n}$ is irrational and lies between a and b.

9. $k!\binom{n}{k} = n(n-1)(n-2)\cdots(n-k+1) \le n^k$.

10.
(i) Just apply the Binomial formula to $(1+1)^n$.

(ii) $(x+y)^n(x+y)^m = (x+y)^{n+m}$. Now apply the binomial formula to both sides:

$$\Big(\sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k\Big)\Big(\sum_{j=0}^{m}\binom{m}{j}x^{m-k}y^k\Big) = \sum_{l=0}^{n+m}\binom{n+m}{l}x^{n+m-l}y^l$$

and collect terms on the left hand side and look at the coefficient of $x^{n+m-l}y^l$.

11. By the binomial formula:

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n &= 1 + n\,\frac{1}{n} + \frac{n(n-1)}{2!}\,\frac{1}{n^2} + \cdots + \frac{n(n-1)\ldots 1}{n!}\,\frac{1}{n^n}\\ &< 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{n!}\\ &< 1 + 1 + \frac{1}{2} + \cdots + \frac{1}{2^n} < 1 + 2 = 3\end{aligned}$$

where we used the formula for the sum of a geometric series:

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} = \frac{1 - \frac{1}{2^{n+1}}}{1 - \frac{1}{2}} < 2$$

12. $(1+x)^n = 1 + n\,x + \frac{1}{2}n(n-1)\,x^2 + \cdots \ge \frac{1}{2}n(n-1)\,x^2 \ge \frac{1}{4}n^2x^2$ for all real numbers $x \ge 0$ and for all $n \ge 2$.

Chapter 2

Sequences and Series

We now come to one of the most fundamental concept in analysis: that of a limit of a sequence of real numbers. (It took mathematicians some time to settle on an appropriate definition!)

2.1 Sequences

Definition 2.1.1 *A sequence of real numbers is a map* $a : \mathbb{N} \to \mathbb{R}$.

We normally write $a_i = a(i)$ and think of a sequence as an (ordered) set of real numbers $(a_i)_{i \in \mathbb{N}} = (a_1, a_2, ..., a_k, ...)$. Sometimes, for notational convenience, we begin a sequence with a_0 instead of a_1 or for that matter, with any a_{n_0}.

Examples:

0. The simplest sequence is the constant sequence: $(a, a, \dots, a, \dots)$, where $a(i) = a$ for all i.

1. $a_n = \frac{1}{n}$ defines the sequence $(1, \frac{1}{2}, \frac{1}{3}, ...)$.

2. $a_n = b^n$ defines the geometric sequence $(b, b^2, b^3, \dots)$, where b is any real number.

3. $a_n = \frac{n}{2^n}$ defines $(\frac{1}{2}, \frac{1}{2}, \frac{3}{8}, ...)$.

4. $a_n = \left(1 + \frac{x}{n}\right)^n$, where x is any real number, defines the well known formula for compound interest.

5. One may define a sequence inductively by a recursive formula such as: $a_{n+2} = a_{n+1} + a_n$ with initial terms given by $a_1 = a_2 = 1$. This gives the Fibonacci sequence $(1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, \dots)$, first introduced by Fibonacci of Pisa (1170 to 1250).

6. Define $a_{n+1} = \frac{1}{2}(a_n + \frac{2}{a_n})$ with initial term $a_1 = 1$. This gives $(1, \frac{3}{2}, \frac{17}{12}, \frac{577}{408}, ...)$. This sequence can be written as a "continued fraction" and gives a very good rational approximation for $\sqrt{2}$. It was known (at least the first few terms) to the ancient Sumerians.

2.2 Convergence

Informally, we say that a sequence of real numbers converges to a limiting number a if all the terms of the sequence become arbitrarily close to a for *all* sufficiently large n. The exact definition now follows:

Definition 2.2.1 *We say that a sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ converges to to a limit $a \in \mathbb{R}$ (and write: $\lim_{n\to\infty} a_n = a$ or simply $\lim a_n = a$) if and only if the following is true:*
$\forall\, \epsilon > 0,\ \exists\, N \in \mathbb{N}$ *such that* $|a_n - a| < \epsilon$ *for all* $n \geq N$.

N, of course, would in general depend on ϵ, because the quantifier for N comes after that of ϵ.

An equivalent definition is as follows:

Definition 2.2.2 *A sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ converges to to a limit $a \in \mathbb{R}$ if and only if every ϵ-neighbourhood of a contains all but a finite number of the terms of the sequence.*

where we used the following:

Definition 2.2.3 *An ϵ-neighbourhood of a real number a is the open interval $(a - \epsilon, a + \epsilon) = \{x \in \mathbb{R} \,|\, |a - x| < \epsilon\}$*

A sequence which converges to a limit is said to be *convergent*. The limit of a convergent sequence is unique. A sequence that is not convergent is called *divergent*.

Examples:

0. The constant sequence , $(a, a, ..., a, ...)$ converges to a.

1. The sequence $a_n = \frac{1}{n}$ converges to 0.
This is because $\forall\, \epsilon > 0$, $|\frac{1}{n} - 0| = \frac{1}{n} < \epsilon$ for all $n \geq N > \frac{1}{\epsilon}$ and the existence of such an $N \in \mathbb{N}$ is guaranteed by the Archimedean property.

1 *bis.* More generally (as you know from first year calculus), the same argument proves that the sequence $a_n = \frac{P(n)}{Q(n)}$ converges to 0, where $P(n)$ and $Q(n)$ are polynomials provided the degree of P is strictly less than the degree of Q. In case they have the same degree the sequence converges to the ratio of the coefficients of the terms of highest degree and if $deg(P) > deg(Q)$ then the sequence diverges.

2. The geometric sequence $(b, b^2, b^3, \dots)$, converges to 0 if $|b| < 1$ and to 1 if $b = 1$. It is divergent otherwise. (This follows from Proposition 2.2.1 below).

3. The sequence $a_n = \frac{n}{2^n}$ converges to 0. We know from the Exercise 2 at the end of Chapter 2, that $n^2 < 2^n$, so $\frac{n}{2^n} < \frac{1}{n}$. Now we can use the same argument that we use for the sequence $(\frac{1}{n})$ in Example 1 above.

4. $a_n = \left(1 + \frac{x}{n}\right)^n$ converges to the exponential function e^x .(*See the exercises at the end of this Chapter*).

5. The Fibonacci sequence is obviously divergent, but the sequence of ratios of consecutive terms ($r_n = \frac{a_n}{a_{n+1}}$) is convergent. (Can you guess the limit of these ratios ?)

6. The sequence defined by $a_{n+1} = \frac{1}{2}(a_n + \frac{2}{a_n})$ and $a_1 = 1$ converges to $\sqrt{2}$. (This is one of the exercises at the end of this Chapter)

Here is a very basic limit, which we will use repeatedly. We prove it from first principles (i.e., we don't use logarithms and other stuff that we haven't defined yet).

Proposition 2.2.1 *Let* $0 < b < 1$. *Then* $\lim_{n\to\infty} b^n = 0$.

Proof: Since $0 < b < 1$, $b^{-1} = 1 + h$ with $h > 0$. Therefore, by the Bernoulli inequality we have:

$$b^n = \frac{1}{(1+h)^n} < \frac{1}{1+nh} < \frac{1}{nh} \to 0$$

as $n \to \infty$, by example 1 above, since h is a positive constant.

The following basic arithmetic rules about computing the limits of sums, differences, products and quotients of convergent sequences are rather obvious from the definitions and the trivial proofs will be left to the reader as an exercise (please do it!), but I will give ample hints below.

Proposition 2.2.2 *If a_n converges to a and b_n converges to b, then:*

(i) $a_n \pm b_n$ converges to $a \pm b$ respectively.

(ii) $a_n\, b_n$ converges to $a\, b$.

(iii) $\frac{a_n}{b_n}$ converges to $\frac{a}{b}$ provided $b \neq 0$. (This implies that $b_n \neq 0$ for n sufficiently large).

Proof (Sketch):
(i): Let $\epsilon > 0$ be given. Then $\exists\, n_1, n_2 \in \mathbb{N}$ such that $|a_n - a| < \frac{1}{3}\epsilon$ for all $n \geq n_1$and $|b_n - b| < \frac{1}{3}\epsilon$ for all $n \geq n_2$, by the definition of convergence. Now choose any $n_0 \in \mathbb{N}$ that is greater than both n_1 and n_2 (for example $n_0 = 10(n_1 + n_2)$ would do), then $n \geq n_0 \Rightarrow |(a_n + b_n) - (a + b)| = |(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b| < \frac{1}{3}\epsilon + \frac{1}{3}\epsilon < \epsilon$ by the triangle inequality.

(ii) First we use the fact that any convergent sequence is bounded, so that for example: $|a_n| \leq 1 + |a|$ for all sufficiently large n. The proof then follows the same pattern as above using the inequality:

$$|a_n\, b_n - ab| = |a_n(b_n - b) + (a_n - a)|b| \leq (1 + |a|)|b_n - b| + |a_n - a|(1 + |b|)$$

(iii) We first prove that $\frac{1}{b_n}$ converges to $\frac{1}{b}$ and then apply (ii).
Since $|b| > 0$, $\exists\, n_1 \in \mathbb{N}$ such that $|b_n - b| < \frac{|b|}{2}$ for all $n \geq n_1$. This implies (by the triangle inequality) that $|b_n| > \frac{|b|}{2}$ for all $n \geq n_1$.
Now let $\epsilon > 0$ be given. Then $\exists\, n_2 \in \mathbb{N}$ such that $|b_n - b| < \frac{|b|^2}{2}\epsilon$ for all $n \geq n_2$. Now choose any n_0 that is greater than both n_1 and n_2. Then

$$n \geq n_0 \Rightarrow |\frac{1}{b_n} - \frac{1}{b}| = |\frac{b - b_n}{b_n\, b}| \leq \frac{2}{|b|}\frac{1}{|b|}\frac{|b|^2}{2}\epsilon = \epsilon$$

QED

We will sometimes use the following notation:

Definition 2.2.4

$\lim_{n\to\infty} a_n = +\infty$ *iff* $\forall R\, \exists N \in \mathbb{N}$ *s.t.* $n \geq N \Rightarrow a_n \geq R$.
We write $\lim_{n\to\infty} a_n = -\infty$ *if* $\lim_{n\to\infty}(-a_n) = +\infty$

For example, $\lim_{n\to\infty} b^n = +\infty$ for all $b > 1$, but $\lim_{n\to\infty} b^n \neq -\infty$ for $b < -1$.

2.3 Monotonic sequences

Definition 2.3.1 *A sequence (a_n) is said to be bounded from above if the set of real numbers $A = \{a_1, a_2, \dots, a_n, \dots\}$ has an upper bound. It is said to be bounded from below if A has a lower bound and is said to be bounded if it has both upper and lower bounds.*

It follows from the definition that any convergent sequence is bounded (from both above and below).

Quick proof: Let a_n be a covergent sequence with $\lim a_n = a$. Since every ϵ-neighbourhood of a contains all but a finite number of points of the set $A = \{a_1, a_2, \dots, a_n, \dots\}$, A is contained in the union of a finite set and a bounded interval $(a - \epsilon, a + \epsilon)$. A finite set of real numbers is obviously bounded from above by its largest member and bounded from below by its smallest member.

Note: That a sequence (a_n) is not bounded from above does not necessarily imply that $\lim_{n\to\infty} a_n = +\infty$.

We now want to investigate what the completeness axiom tells us about the convergence of sequences. The example of the sequence $(1, 0, 1, 0, \dots)$ shows that bounded sequences do not necessarily have limits. We need the following.

Definition 2.3.2 *A sequence (a_n) is said to be monotonically increasing if $a_{n+1} \geq a_n$ for all $n \in \mathbb{N}$. It is said to be monotonically decreasing if $a_{n+1} \leq a_n$ for all $n \in \mathbb{N}$. It is said to be monotonic if it is either monotonically increasing or monotonically decreasing.*

Remark: We say the sequence is strictly monotonically increasing (respectively decreasing) if we have the strict inequality in the definition above.

The main result about monotonic sequences, which is an immediate consequence of the completeness axiom for real numbers is the following:

Theorem 2.3.1 *Every monotonically increasing sequence which is bounded from above is convergent. Similarly every monotonically decreasing sequence which is bounded from below is convergent*

Proof: Let (a_n) be a monotonically increasing sequence which is bounded from above. Let $a = lub(A)$, where $A = \{a_1, a_2, ..., a_n, ...\}$. We will prove that the sequence converges to a, its least upper bound. So let a be the least upper bound of the sequence. Given $\epsilon > 0$, we'll show that all except possibly a finite number of the terms of the sequence are in the interval $(a - \epsilon, a] \subset (a - \epsilon, a + \epsilon)$. Since a is an upper bound, we have $a_n \leq a$, but since a is the *least* upper bound, $a - \epsilon$ is *not* an upper bound of the sequence and therefore there must be a term a_{n_0} which is strictly larger than $a - \epsilon$. Now since the sequence is montonically increasing $a_n \geq a_{n_0}$ for all $n \geq n_0$. Thus we have shown that $\forall \epsilon > 0$, $\exists n_0 \in \mathbb{N}$ such that $a - \epsilon < a_n \leq a <$ for all $n \geq n_0$.

The statement about a monotonically decreasing sequences follow from the above proof since (a_n) is monotonically decreasing sequence and bounded from below if and only if $(-a_n)$ is monotonically increasing and bounded from above.

QED

Of course, there are convergent sequences which are *not* monotonic. For example, $a_n = 1 + \frac{(-1)^n}{n}$ is not monotonic but converges to 1. Also there are bounded sequences which are *not* convergent. For example, the sequence $(-1, 1, -1, 1, ...)$ defined by $a_n = (-1)^n$ is certainly not convergent even though it is bounded. We have seen some bounded sequences which do not converge. We can, however, say something about such sequences.

Definition 2.3.3 *A subsequence is an infinite ordered subset of a sequence.*

Proposition 2.3.1 *Any subsequence of a convergent sequence is convergent (to the same limit).*

The proof of the above proposition is trivial. The next result however is not totally obvious!

Proposition 2.3.2 *Every sequence (convergent or not) contains a monotonic subsequence.*

Proof: Let $M = \{m \in \mathbb{N} \,|\, a_n < a_m \;\forall\, n > m\}$. There are two possibilities:

Case 1. M is infinite. $M = \{m_1 < m_2 < \dots\}$. Then $b_k = a_{m_k}$ is a monotonically (strictly) decreasing sequence.

Case 2. M is finite. Let n_1 be larger than the maximum number in M. Now since $n_1 \notin M$, $\exists n_2 > n_1$ such that $a_{n_1} \geq a_{n_2}$. Furthermore $n_2 \notin M$ since it is larger than n_1. Thus we can continue this procedure to obtain a monotonically increasing sequence $a_{n_1} \geq a_{n_2} \geq a_{n_3} \geq \dots$.

QED

Combining this last proposition with the basic theorem about monotonic sequences (Theorem 2.3.1), we obtian the famous result attributed to the Czech mathematician and philosopher Bernard Bolzano (1781 to 1848) and the German mathematician Karl Weierstrass (1815 to 1897):

Theorem 2.3.2 Bolzano-Weierstrass Theorem
Every bounded sequence has a convergent subsequence.

Note: A bounded sequence may have many convergent subsequences (for example, a sequence consisting of a counting of the rationals has subsequences converging to every real number) or rather few (for example a convergent sequence has all its subsequences having the same limit). In fact we define:

Definition 2.3.4 *A point a is called an* accumulation point *of a sequence (a_n), if there is a subsequence that converges a.*

A point a is an accumulation point of (a_n) if every ϵ-neighbourhood of a contains infinitely many members of the sequence (not necessarily distinct points!). In other words, x is *not* an accumulation point of (a_n) only if $\exists \epsilon > 0$ such that $|a_n - x| \geq \epsilon$ for all n. Therefore the points that are not accumulation points of a sequence form an open set in the sense that every such point has a neighbourhood consisting only of points that are not accumulation points. From this it follows that the complementary set of accumulation points is closed (for the definition of closed and open sets, see next chapter). If A denotes the set of all accumulation points of a sequence (a_n), then $glb(A)$ and $lub(A)$ both belong to A (provided A is bounded).

Definition 2.3.5 *$limsup_{n\to\infty} a_n = lub(A)$ if the sequence is bounded from above. (Otherwise we set $limsup = \infty$).*
Similarly we define $liminf_{n\to\infty} a_n = glb(A)$ if the sequence is bounded from below. (Otherwise we set $liminf = -\infty$).

In simple language, limsup is the largest and liminf is the smallest accumulation point of a sequence. Moreover a sequence is convergent iff limsup is equal to liminf, so that there is exactly one (finite) accumulation point.

2.4 Cauchy sequences

One apparent problem with deciding whether a sequence is convergent or not, using the definition, is that one is supposed to know the limit first. An elegant way around this problem is to use a fundamental idea first introduced by the French mathematician Augustin Louis Cauchy (1789 to 1857).

Definition 2.4.1 *We say that a sequence* $(a_n)_{n\in\mathbb{N}}$ *is a* Cauchy sequence *if and only if the following is true:*
$\forall\,\epsilon > 0$, $\exists\, N \in \mathbb{N}$ *such that* $|a_{n_1} - a_{n_2}| < \epsilon$ *for all* $n_1, n_2 \geq N$.

Informally this means that any two terms (not just for two consecutive terms!) of a Cauchy sequence can be made arbitrarily close to each other if we go far enough in the sequence. Note that this definition does not use the limit of the sequence. Here are some immediate properties of Cauchy sequences:

1. Any Cauchy sequence is bounded.
Proof: Let (a_n) be a Cauchy sequence Setting "$\epsilon = 1$" in the definition we see that $\exists N$ such that $|a_{n_1} - a_{n_2}| < 1$ for all $n_1, n_2 \geq N$. In particular all $a_n \in (a_N - 1, a_N + 1)$ for all $n \geq N$ and hence the set $\{a_n : n \in \mathbb{N}\}$ is bounded.

2. Any convergent sequence is a Cauchy sequence.
Proof: Let (a_n) be a convergent sequence with $\lim a_n = a$. Given $\epsilon > 0$, $\exists\, N$ so that every a_n with $n \geq N$ lies in the $\frac{1}{3}\epsilon$ -neighbourhood of a. Any two points in that neighbourhood are obviously at a distance strictly less than ϵ apart.

The fundamental property about Cauchy sequences is the following

Theorem 2.4.1 *A sequence of real numbers is convergent if and only if it is a Cauchy sequence.*

Proof: By the proposition above, we only have to show that every Cauchy sequence is convergent, so let (a_n) be a Cauchy sequence. Since a Cauchy sequence is bounded it has a convergent subsequence $b_k = a_{n_k}$ by the Bolzano-Weierstrass theorem. Let b be the limit of the subsequence. Then given $\epsilon > 0$, $\exists n_K$ such that *all* terms of the subsequence b_k with $k \geq K$ lie in the $\frac{1}{3}\epsilon$ -neighbourhood of b. Now since (a_n) is Cauchy $\exists\, n_0$ such that all terms a_n with $n \geq n_0$ are at a distance less than $\frac{1}{3}\epsilon$ apart. Now choose

$N = max(n_0, n_K)$. Then all terms of the sequence a_n with $n \geq N$ are at a distance less than ϵ from b and hence $lim\, a_n = b$

Remarks: The fact that Cauchy sequences in $\mathbb{R}$ are the same as convergent sequences is called the Cauchy criterion for convergence. The completeness axiom to prove the last result is crucial. For example, any sequence of rational numbers converging to an irrational number is a Cauchy sequence that is trying to converge but cannot converge in $\mathbb{Q}$. In fact, Cantor (1845 - 1918) used the idea of a Cauchy sequence of rationals to give a constructive definition of the Real numbers. Spaces (not just $\mathbb{R}$) where all Cauchy sequences converge are called *complete*. In fact, one can formulate the completeness axiom for $\mathbb{R}$ in terms of Cauchy sequences (provided we assume the Archimedean property). Here are some equivalent formulations of the axiom:

C1 Every subset which is bounded from above has a least upper bound.

C2 Every bounded sequence has a convergent subsequence.

C3 Every Cauchy sequence is convergent.

The last property is a useful way of generalizing the idea of completeness to more general spaces. For the purpose of the next section we need the field of complex numbers, so a quick reminder:

Definition 2.4.2 *A complex number is simply a pair of real numbers written as $z = x + i\, y$ where the magic "imaginary" number i has the "unreal" property: $i^2 = -1$.*

We add and multiply two complex numbers as follows:

$(x_1 + i\, y_1) + (x_2 + i\, y_2) = (x_1 + x_2) + i\,(y_1 + y_2)$

$(x_1 + i\, y_1).(x_2 + i\, y_2) = (x_1x_2 - y_1y_2) + i\,(x_1y_2 + y_1x_2)$

The complex numbers are denoted by $\mathbb{C}$ and they form a field. The multiplicative inverse of a non-zero complex number z is given by I assume that you are familiar with simple arithmetic and algebraic properties of these numbers. For a complex number $z = x + iy$, $x = Re(z)$ is called the real part, $y = Im(z)$ is called the imaginary part. $\bar{z} = x - i\, y$ is the conjugate and $|z| = \sqrt{z\bar{z}}$ the absolute value (or modulus) of z. The complex numbers are denoted by $\mathbb{C}$ and they form a field. The multiplicative inverse of a non-zero complex number z is given by $z^{-1} = \frac{\bar{z}}{|z|^2}$. I assume that you are familiar with simple arithmetic and algebraic properties of these numbers. Using the distance $|z - w|$ between two complex numbers z and w, we can define a Cauchy sequence of complex numbers in exactly the same way as for real sequences. Using the Pythagorian identity $|x + i\,y|^2 = x^2 + y^2$ and the triangle inequality $|w+z| \leq |w| + |z|$ valid in $\mathbb{C}$, one easily observes that a sequence of complex numbers $(z_k) = (x_k + i\,y_k)$ is convergent iff the two real sequences (x_k) and (y_k) are both convergent. So we arrive at one main property that we need about complex numbers:

Proposition 2.4.1 *$\mathbb{C}$ is complete, i.e. every Cauchy sequence in $\mathbb{C}$ is convergent.*

Remark: The one property for $\mathbb{R}$ that is different from $\mathbb{C}$ is that we do not have an ordering for $\mathbb{C}$. There is no such thing as a positive complex number! On the other hand, there is an algebraic property that is crucial for $\mathbb{C}$ (but not true in $\mathbb{R}$), namely that $\mathbb{C}$ is algebraically closed, i.e., every polynomial (with complex coefficients) has a zero in $\mathbb{C}$. For example $z^2 + 1 = 0$ is solvable in $\mathbb{C}$ (but not in $\mathbb{R}$) and the two solutions are $\pm\, i$.

2.5 Series

Although computing with series of numbers (real or complex) is the bread and butter of everyday mathematics and is of great practical and theoretical importance, we will be brief in our treatment of series, since this is just a short one semester course.

Definition 2.5.1 *A* series *is simply a sequence of the form $s_n = \sum_{i=1}^{n} a_k = a_1 + \cdots + a_n$ where the a_i's numbers (real or complex).*
The series is said to be convergent iff the sequence of partial sums s_n is convergent and we write $\sum_{i=1}^{\infty} a_i = s$ if $\lim_{n\to\infty} s_n = s$. s is then called the sum of the infinite series. By definition, a divergent series *is a series that is not convergent.*

Examples:

1. The *geometric series* $a + a\,r + a\,r^2 + \cdots$ converges to the sum $\frac{a}{1-r}$ if $|r| < 1$. It is divergent otherwise (assuming $a \neq 0$). This follows from the identity:

$$\sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r}$$

and the fact that $r^{n+1} \to 0$ as $n \to \infty$ for $|r| < 1$.

2. $\sum_{k=1}^{\infty} \frac{1}{k^p}$, where $p > 0$ is an important series in number theory and defines the zeta function $\zeta(p)$. It converges for $p > 1$ and diverges for $p \leq 1$. For example $\zeta(2) = \frac{\pi^2}{6}$.

Proposition 2.5.1

$$\sum_{k=1}^{\infty} \frac{1}{k^p}$$

converges for $p > 1$

Proof:

$$\begin{aligned}
s_{2^n-1} &= 1 + \left(\frac{1}{2^p} + \frac{1}{3^p}\right) + \left(\frac{1}{4^p} + \cdots + \frac{1}{7^p}\right) + \cdots + \left(\frac{1}{\left(2^{n-1}\right)^p} + \cdots + \frac{1}{\left(2^n - 1\right)^p}\right) \\
&\leq 1 + 2\left(\frac{1}{2^p}\right) + 4\left(\frac{1}{4^p}\right) + \cdots + 2^{n-1}\left(\frac{1}{\left(2^{n-1}\right)^p}\right) \\
&= 1 + 2\,\frac{1}{2^{p-1}} + \frac{1}{4^{p-1}} + \cdots + \frac{1}{\left(2^{n-1}\right)^{p-1}} \\
&< \frac{1}{1 - 2^{1-p}}
\end{aligned}$$

Therefore s_n is a bounded from above (for $p > 1$), and since it is a monotonically increasing sequence (being the partial sum of a positive series it has to converge.

QED

Proposition 2.5.2

$$\sum_{k=1}^{\infty} \frac{1}{k^p}$$

diverges for $p \leq 1$

Proof: We first look at the *harmonic series* $\sum \frac{1}{k}$ where $p = 1$.

$$\begin{aligned} s_{2^n} &= 1 + \frac{1}{2} + \Big(\frac{1}{3} + \frac{1}{4}\Big) + \cdots + \Big(\frac{1}{2^{n-1}+1} + \cdots + \frac{1}{2^n}\Big) \\ &\geq \frac{3}{2} + 2\,\frac{1}{4} + \cdots + 2^{n-1}\Big(\frac{1}{2^n}\Big) \\ &= 1 + \frac{n}{2} \end{aligned}$$

Therefore s_n is unbounded and the harmonic series is divergent. For $0 < p < 1$, we have $\frac{1}{k^p} > \frac{1}{k}$ for all k, so $\sum_{k=1}^{n} \frac{1}{k^p} \geq \sum_{k=1}^{n} \frac{1}{k}$ for all n and hence $\sum \frac{1}{k^p}$ diverges for $0 < p < 1$.

QED

2.6 Convergence criteria for series

Proposition 2.6.1 (Cauchy criterium). $\sum a_i$ *converges iff* $\forall \epsilon > 0\ \exists n_0 \in \mathbb{N}$ *such that* $n > m > n_0 \Rightarrow |a_m + \cdots + a_n| < \epsilon$

This is just a restatement of the definition of a Cauchy sequence for the partial sums of a series.

Corollary 2.6.1 $\sum a_n$ *converges* $\Rightarrow \lim_{n\to\infty} a_n = 0$

Thi converse is not true. A counterexample is the harmonic series.

Proposition 2.6.2 (Basic Comparison Test)*:*
If $\exists K$ *such that* $0 \leq a_k \leq b_k$ *for all* $k \geq K$*, then* $\sum b_k$ *is convergent* $\Rightarrow$ $\sum a_k$ *is convergent.*

This follows from the fact that the partial sums of both series form monotonically increasing sequences and the partial sums for $\sum b_k$ dominate those of $\sum a_k$.

For ease of language, we will say that a statement is true for all sufficiently large k if $\exists\ K$ such that the statement is true for all $k \geq K$.

Definition 2.6.1 *A series* $\sum a_i$ *is said to be absolutely convergent if* $\sum |a_i|$ *is convergent.*

Since $|a_m + \cdots + a_n| < ||a_m| + \cdots + |a_n||$, it is easily seen, by the Cauchy criterium, that absolute convergence implies convergence, but the converse is not true.

Proposition 2.6.3 *If, for all sufficiently large* k *,* $|a_k| \leq C r^k$ *for some positive constants* C *and* $0 \leq r < 1$ *(*C *and* r *are independent of* k*!), then the series* $\sum a_k$ *is absolutely convergent. (*a_k *can be complex numbers).*

Proof: Use the basic comparison test to compare the series $\sum |a_k|$ with the convergent geometric series $C \sum r^k$.

QED

Corollary 2.6.2 (RatioTest)*:*
If $\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| = q < 1$*, then* $\sum a_k$ *is absolutely convergent.*

Proof: Choose $r = \frac{1}{2}(1+q)$. Then for all sufficiently large $k \geq K$, $\left|\frac{a_{k+1}}{a_k}\right| < r$ and hence $|a_k| \leq C r^k$ for all $k \geq K$ with $C = |a_K| r^{-K}$.

QED

Corollary 2.6.3 (Root Test)*:*
If $\lim_{k\to\infty} \Big(|a_k|\Big)^{\frac{1}{k}} = q < 1$*, then* $\sum a_k$ *is absolutely convergent.*

Proof: Choose $r = \frac{1}{2}(1+q)$. Thew for sufficiently large k , $\Big(|a_k|\Big)^{\frac{1}{k}} < r$ and so $|a_k| \leq r^k$ for all $k \geq K$.

QED

Definition 2.6.2 *An* alternating series *is a series of real numbers of the form* $\pm \sum_{k=1}^{\infty} (-1)^k a_k$ *, where all the numbers* $a_k \geq 0$.

Proposition 2.6.4 (Leibniz test for alternating series)*:*
An alternating series $\sum_{k=1}^{\infty} (-1)^{k-1} a_k$ *satisfying*

(i) $a_{k+1} \leq a_k$ *for all sufficiently large* k

(ii) $\lim_{k\to\infty} a_k = 0$

is convergent.

For your pleasure, the simple proof, which can be found in any elementary calculus text book, is left as an exercise at the end of this chapter. For example $\sum \frac{(-1)^{k-1}}{k}$ is convergent (to $log(2)$ although the harmonic series is divergent. Series which are convergent but not absolutely convergent are called *conditionally convergent*. These series are very sensitive to the order in which they are summed up. In fact the following rather surprising fact is true.

Proposition 2.6.5 *Let $\sum a_k$ be a conditionally convergent series of real numbers and let S be any real number. Then there is a rearrangement $\sum \tilde{a}_n$ such that $\sum \tilde{a}_n = S$.*

Proof: Let $\sum p_l$ be the series formed by the strictly positive terms of $\sum a_k$ and let $\sum q_m$ be the rest. Since $\sum a_k$ is convergent but *not* absolutely convergent, both series $\sum p_l$ and $\sum q_m$ are divergent (to $\pm\infty$ respectively). Let us assume (w.l.o.g.) that $S \geq 0$ Now we can always add enough terms of the positive series to get a smallest (first) partial sum that is strictly larger than S. After that we add up (or if you prefer subtract!) just enough terms from the other negative series to make the total strictly less than S. Now we know that all the terms $\to 0$ (since the original seres is convergent after all!), so by continuing this process, we arrive at a sequence of partial sums (with terms rearranged as above) converging to S.

QED

We end this section with a product formula that we need in the next section for power series.

Theorem 2.6.1 *(Cauchy product formula for series).*
Let $\sum_{k=0}^{\infty} a_n$ and $\sum_{k=0}^{\infty} b_n$ be two absolutely convergent series. Then their Cauchy-product, defined to be the series $\sum c_n$ with $c_n = \sum_{k=0}^{n} a_k \, b_{n-k}$ is absolutely convergent and $\big(\sum a_n\big)\big(\sum b_n\big) = \sum c_n$.

Proof: Since the two series are absolutely convergent, we can assume w.l.o.g. that all the terms are non-negative, so that all the partial sums are monotonically increasing. Since $s_n = \sum_{k=0}^{n} a_n \to S$ and $t_n = \sum_{k=0}^{n} b_n \to T$, the product sequence: $s_n \, t_n \to ST$. Therefore, by the Cauchy criterium $s_{2n}t_{2n} - s_n t_n \to 0$ as $n \to \infty$. Now $s_m t_m$ is the sum of all terms $a_i b_j$ with $i, j \leq m$. Let $u_m = \sum_{k=0}^{m} c_m$. Then u_m is the sum of all terms $a_i b_j$ with $i + j \leq m$. So $s_n t_n \leq u_{2n} \leq s_{2n} t_{2n}$ and hence, u_n converges to ST.

2.7 Power Series

Definition 2.7.1 *A* power series *is simply a series of the form*

$$\sum a_n \, (x - c)^n$$

where c and the a_n's are constants and x is a variable (real or complex).

Given a power series, the principal task is to determine the values of the variable x, for which the series converges and to study the sum of the power series as a function of x. Note that a power series always converges at $x = c$ to the value 0. The first main result states that a power series converges (absolutely) for all points inside a circle with center c in the complex plane and that it diverges outside this circle. (The circle could degenerate to a point or to the whole plane). We will mainly deal with the case $c = 0$, since the general case is just a simple shift.

Examples

1. The *geometric series* $a + a\,x + a\,x^2 + \cdots$ converges to the sum $\frac{a}{1-x}$ if $|x| < 1$. It is divergent otherwise (assuming $a \neq 0$).

2. The power series: $\sum_{k=0}^{\infty} \frac{1}{k!} x^k$ is one of the most important power series, since it defines the exponential function e^x for both real and and complex values of x. Since $\frac{1}{(k+1)!} x^{k+1} / \frac{1}{k!} x^k = \frac{x}{k+1}$, this series converges for all values of x (by the ratio test).

3. The power series: $\sum k!\, x^k$ converges only at $x = 0$ (again by the ratio test).

4. The series $\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k$ is convergent inside the unit circle $|x| < 1$. It arises by integrating a geometric series and defines the logarithm as we see later in the course).

Proposition 2.7.1

(i) If the power series $\sum a_n\, x^n$ converges for $x = x_0$ then it converges absolutely for all x such that $|x| < |x_0|$.

(ii) If the power series $\sum a_n\, x^n$ diverges for $x = x_0$ then it diverges for all x such that $|x| > |x_0|$.

Proof: We compare with a geometric series. If $\sum a_n x_0^n$ is convergent, $a_n x_0^n \to 0$ and so in particular, all the terms are bounded: $|a_n x_0^n| \leq C$ for some constant $C > 0$. Let $b = \frac{|x|}{|x_0|} < 1$. Then the geometric series $C \sum b^n$ is convergent and hence by the basic comparison theorem, so is the series $\sum a_n x^n$, since $|a_n x^n| = |a_n x_0^n| b^n \leq C b^n$. (ii) follows from a comparison with a divergent geometric series since $\frac{|x|}{|x_0|} > 1$.

QED

The "largest" value R for such that the power series converges for all $|x| < R$ and diverges for all $|x| > R$, is called the *radius of convergence.* This can be computed by the formula $R^{-1} = limsup_{n\to\infty}|a_n|^{\frac{1}{n}}$ (by the root test) or by the formula $R^{-1} = lim_{n\to\infty}\frac{|a_{n+1}|}{|a_n|}$ (ratio test).

Remark: The ratio and the root tests are related by the following relation:

Proposition 2.7.2 *For any sequence of positive real numbers* (a_n) *, we have*

$$\liminf_{n\to\infty}\frac{a_{n+1}}{a_n} \leq \liminf_{n\to\infty}\sqrt[n]{a_n}$$

$$\limsup_{n\to\infty}\frac{a_{n+1}}{a_n} \geq \limsup_{n\to\infty}\sqrt[n]{a_n}$$

I will leave the proof of this proposition as an exercise for you (but please do it!). We will continue with the more important properties of using power series to define functions (for example differentiating and integrating power series) when we deal with uniform convergence in Chapter 7, but I cannot resist introducing the most important "transcendental" function: the exponential function, using a power series, in the next section.

2.8 The Exponential Function

Definition 2.8.1 *The exponential function is defined by:*

$$exp(z) = e^z = \sum_{n=0}^{\infty}\frac{1}{n!}\,z^n$$

for any $z \in \mathbb{C}$ *.*

Definition 2.8.2 *The number e is defined by:*

$$e = exp(1) = \sum_{n=0}^{\infty}\frac{1}{n!}$$

Besides the obvious normalization $exp(0) = 1$, the exponential function satisfies the following extremely important functional identity (In fact, it can be characterized by this property):

Theorem 2.8.1
$$exp(w+z) = exp(w)\, exp(z)$$
for all $z, w \in \mathbb{C}$

Proof: Multiplying the two series $\sum_{n=0}^{\infty} \frac{1}{n!}\, w^n$ and $\sum_{n=0}^{\infty} \frac{1}{n!}\, z^n$ by the Cauchy product formula we obtain the series $\sum_{n=0}^{\infty} c_n$, with

$$c_n = \sum_{k=0}^{n} \frac{w^{n-k}}{(n-k)!} \sum_{k=0}^{n} \frac{z^k}{k!} = \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} w^{n-k} z^k = \frac{1}{n!}(w+z)^n$$

by the binomial formula of Chapter 1.

QED

Corollary 2.8.1

(i) $exp(-z) = (exp(z))^{-1}$ *for all* $z \in \mathbb{C}$

(ii) $exp(x) > 0$ *for all* $x \in \mathbb{R}$

(iii) $exp(n) = e^n$ *for all* $n \in \mathbb{Z}$

(iv) $exp(\bar{z}) = \bar{e}xp(z)$ *for all* $z \in \mathbb{C}$

Definition 2.8.3 *The (real-valued) trigonometric functions are defined by:* $cos(y) = Re(e^{iy})$ *and* $sin(y) = Im(e^{iy})$ *for any* $y \in \mathbb{R}$ *.*

We therefore have Euler's famous formula
$$e^{i\,\theta} = \cos(\theta) + i\, \sin(\theta)$$
and also the power series expressions:

(i) $cos(x) = \frac{1}{2}(exp(ix) + exp(-ix)) = \sum_{k=0}^{\infty}(-1)^k \frac{1}{(2k)!}\, x^{2k}$

(ii) $sin(x) = \frac{1}{2i}(exp(ix) - exp(-ix)) = \sum_{k=0}^{\infty}(-1)^k \frac{1}{(2k+1)!}\, x^{2k+1}$

valid for any $x \in \mathbb{R}$.

All the other trigonometric functions (tangent, cotangent, secant and cosecant) can be expressed in terms of cosine and sine.

Trigonometric identities such as $\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$ $\sin(\alpha+\beta) = \cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta)$ follow from the fundamental functional equation for the exponential function. The fundamental identity: $(\cos\,\theta)^2 + (\sin\,\theta)^2$, which relates the trigonometric functions to the unit circle $x^2 + y^2 = 1$, follows from the fact that the conjugate of e^{ix} is e^{-ix} and so $|e^{ix}|^2 = 1$.

Closely related and very useful functions are the hyperbolic functions defined by:

$\cosh(x) = \frac{1}{2}(e^x + e^{-x})$
$\sinh(x) = \frac{1}{2}(e^x - e^{-x})$

These functions play an important role in geometry and physics.

The fundamental identity here is: $(\cosh\, t)^2 - (sinh\, t)^2 = 1$. This relates the hyperbolic functions to the hyperbola $x^2 - y^2 = 1$ in the same way that the trigonometric functions came from the circle. There are also transcendental functions that are related to the ellipse, called elliptic functions!

2.9 Exercises

0. Complete the proof of Proposition 2.2.2 about the arithmetic properties of limits.

1. Assume that $\lim_{k\to\infty} a_k = l$. Show that:

(i) $\lim_{k\to\infty} |a_k| = |l|$

(ii) If $l > 0$ then $\exists N \in \mathbb{N}$ such that $n > N \Rightarrow a_n > \frac{9}{10}\, l$

2. Let $a_k = \sqrt{k+10^3} - \sqrt{k}$; $b_k = \sqrt{k+\sqrt{k}} - \sqrt{k}$; and $c_k = \sqrt{k(1+10^{-3})} - \sqrt{k}$ for $k \in \mathbb{N}$. Show that $a_k > b_k > c_k$ for all $k < 10^6$, but $a_k \to 0$, $b_k \to 0.5$; and $c_k \to +\infty$ as $k \to \infty$.

3. Define a sequence recursively by the formulas: $a_1 = a$, $a_2 = b$ and $a_k = \frac{1}{2}(a_{k-1} + a_{k-2})$ for all $k > 2$. Show that a_n converges and find the limit.

4. Compute:

$$\sqrt{1+\sqrt{1+\sqrt{1+\ldots}}}$$

i.e., the limit of the sequence defined recursively by $a_{k+1} = \sqrt{a_k + 1}$ with $a_1 = 1$.

5. Compute:

$$1 + \cfrac{1}{1 + \frac{1}{1+\ldots}}$$

i.e., the limit of the sequence defined recursively by $a_{k+1} = 1 + \frac{1}{a_k}$ with $a_1 = 1$.

6. Compute the limits of the following sequences if they converge. Do not use logarithms or l'Hospital's rule and stuff like that! Verify your answers, i.e. prove that they are correct.

(i) $a_n = b^{\frac{1}{n}}$ with $b > 0$

(ii) $a_n = n^{\frac{1}{n}}$

(iii) $a_n = (b^n + c^n)^{\frac{1}{n}}$ with $b, c > 0$

(iv) $a_n = \frac{n!}{n^n}$

7. *Cauchy condensation test* Suppose that (a_n) is a monotonically decreasing sequence of positive real numbers such that the series $\sum a_n$ is convergent. Let $b_n = 2^n\, a_{2^n}$. Prove that the series $\sum b_n$ is also convergent.

8. Prove the following form of the Bolzano-Weierstrass theorem:
Every bounded infinite set has at least one accumulation point.
(An accumulation point of a set $X \subset \mathbb{R}$ is a point a such that every ϵ-neighbourhood of a contains at least one point of X distinct from a.)

9. Prove Leibniz' test for alternating series.

10. For each of the the following power series determine all values of x for which the series (i) converge absolutely; (ii) converges conditionally and (iii) diverges.

(i) $\sum \frac{x^n}{n^p}$ with $p > 0$

(ii) $\sum \frac{n^2}{3^n}(x+2)^n$

(iii) $\sum \frac{(n!)^2}{(2n)!}x^n$

(iv) $\sum 2^n \frac{n!}{n^n} x^n$

11. Prove proposition 2.7.2.

12. Suppose that $0 < r < 1$ and that a sequence (a_n) satisfies the contraction property: $|a_{n+1} - a_n| \leq r\,|a_n - a_{n-1}|$ for all $n \geq 10$. Prove that (a_n) is a Cauchy sequence and hence converges.

13. Sum the series

$$1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{9} + \frac{1}{15} + \frac{1}{25} + \frac{1}{27} + \cdots ,$$

where we sum over all reciprocals of integers whose only prime factors are 3 and 5. (*This is a product of two geometric series*).

14. Show that $\cos(2\theta) = \big(\cos(\theta)\big)^2 - \big(\sin(\theta)\big)^2$ and $\sin(2\theta) = 2\cos(\theta)\sin(\theta)$ from the definitions. Derive similar formulas for the hyperbolic functions.

15. Find a formula for the finite sum:

$$\frac{1}{2} + \cos\theta + \cos(2\theta) + \cdots + \cos(n\theta)$$

.

18. Show that e is irrational.

19. Show that $a_n = \left(1+\frac{1}{n}\right)^n$ is a bounded monotonically increasing sequence. What is the limit?

20*. Show that

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

for any $x \in \mathbb{R}$.

2.9.1 Short solutions to some of the exercises

2. If $k < 10^6$, then $\sqrt{k} < 10^3$ and $10^{-3}\,k < \sqrt{k}$.
So: $\sqrt{k + 10^{-3}\,k} < \sqrt{k + \sqrt{k}} < \sqrt{(k + 10^3)}$ and $a_k > b_k > c_k$.

$$a_k = \sqrt{(k + 10^3)} - \sqrt{k} = \frac{10^3}{\sqrt{k + 10^3} + \sqrt{k}} \to 0$$

$$b_k = \sqrt{k + \sqrt{k}} - \sqrt{k} = \frac{\sqrt{k}}{\sqrt{k + \sqrt{k}} + \sqrt{k}} = \frac{1}{\sqrt{1 + \frac{1}{\sqrt{k}}} + 1} \to \frac{1}{2}$$

$$c_k = \sqrt{k(1 + 10^{-3})} - \sqrt{k} = \frac{k\,10^{-3}}{\sqrt{k + 10^{-3}\,k} + \sqrt{k}} = \frac{10^{-3}\sqrt{k}}{\sqrt{1 + 10^{-3}}} \to \infty;$$

3. Let us assume first that $a = 0$ and $b = 1$. The general case follows by a translation and dilation of this interval. The recursive definition has the geometric meaning that the sequence is obtained by taking the midpoint of the preceding two points (beginning with 0 and 1). So you go back and forth, halving your step size everytime!
The sequence is therefore:

$$(0,\, 1,\, 1 - \frac{1}{2},\, 1 - \frac{1}{2} + \frac{1}{4},\, 1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{8}, \cdots)$$

.
By the formula for the sum of a geometric series:

$$a_k = 1 - \frac{1}{2} + \frac{1}{4} - \frac{1}{8} \cdots + (-\frac{1}{2})^{k-2} = \frac{1 - (-\frac{1}{2})^{k-1}}{1 - (-\frac{1}{2})} = \frac{2}{3}(1 - (-\frac{1}{2})^{k-1})$$

Therefore

$$\lim a_k = \frac{2}{3}$$

since $\lim(-\frac{1}{2})^{k-1} = 0$.
In the general case the limit is

$$a + \frac{2}{3}(b - a) = \frac{a + 2b}{3}$$

4. We first prove that the given sequence converges by showing that it is increasing and bounded from above.

To prove $a_{n+1} > a_n$ for all n, we proceed by induction. For $n = 1$: $a_2 = \sqrt{2} > 1 = a_1$. Assume $a_{k+1} > a_k$. Then $a_{k+2} = sqrt1 + a_{k+1} > a_{k+1} = sqrt1 + a_k$. We now show that $a_n \leq 2$ for all n by induction. For $n = 1$: $a_1 = 1 < 2$. Assume $a_k < 2$. Then $a_{k+1} = sqrt1 + a_k < sqrt1 + 2 < 2$.
Let $a = \lim a_n$. Then $0 \leq a \leq 2$ and satisfies $a = sqrt1 + a$. Therefore a is equal to the Golden Ratio:

$$\phi = \frac{\sqrt{5}+1}{2}$$

5. The limit here is also the Golden Ratio ϕ.

6.

(i) $\lim_{n\to\infty} b^{\frac{1}{n}} = 1$.
Proof: Assume first that $b > 1$. Then $a_n > 1$ and we put $a_n = 1 + h_n$ with $h_n > 0$. Then, by the Bernoulli inequality: $b = a_n^n = (1 + h_n)^n \geq 1 + nh_n$, so $0 < h_n \leq \frac{b-1}{n}$ and hence $h_n \to 0$. If $b < 1$, then since $\frac{1}{b} < 1$ and so $\lim_{n\to\infty}(\frac{1}{b})^{\frac{1}{n}} = 1$. If $b = 1$, there is nothing to prove!

(ii) We set $a_n = 1 + \epsilon_n$.
Now $n = (1 + \epsilon_n)^n > \frac{1}{2}n(n-1)\epsilon^2$ for all $n > 2$, by the binomial formula. So $\epsilon < sqrt\frac{2}{n-1}$ and hence $\epsilon_n \to 0$. Therefore $\lim_{n\to\infty} n^{\frac{1}{n}} = 1$.

(iii) $\lim_{n\to\infty}(b + c)^{\frac{1}{n}} = \max(b, c)$.
Proof: Assume w.l.o.g. that $b > c > 0$. $(b^n + c^n)^{\frac{1}{n}} = b(1 + q^n)^{\frac{1}{n}}$, where $0 < q = \frac{c}{b} < 1$. Now $(1+q^n)^{\frac{1}{n}} < (1+q)^{\frac{1}{n}}$ for $n > 1$ and so Set $(1+q^n)^{\frac{1}{n}} \to 0$ by (i).

(iv) $\lim_{n\to\infty} \frac{n!}{n^n} = 0$ since

$$\frac{n!}{n^n} = \frac{n}{n}\frac{n-1}{n}\cdots\frac{2}{n}\frac{1}{n} < \frac{1}{n}$$

12. By induction: $|a_{n+l+1} - a_{n+l}| \leq r\,|a_{n+l} - a_{n+l-1}| \leq \cdots r^(n + l - 1)\,|a_2 - a_1|$
By the triangle inequality: $|a_{n+k} - a_n| \leq \sum_{l=0}^{k-1} |a_{n+l+1} - a_{n+l}| \leq |a_2 - a_1| \sum_{l=0}^{k-1} r^{n+l-1} \leq \frac{r^n}{1-r}|a_2 - a_1|$ for all $n > 10$ and *for all* k! Since $r^n \to 0$, this implies that a_n is a Cauchy sequence.

13. By the Cauchy product formula of two absolutely converging series we have:

$$\Big(\sum_{k=0}^{\infty}\frac{1}{3^k}\Big)\Big(\sum_{l=0}^{\infty}\frac{1}{5^l}\Big) = \Big(\sum_{k,l=0}^{\infty}\frac{1}{3^k\,5^l}\Big)$$

which is the given series and therefore its sum is $\frac{1}{1-\frac{1}{3}}\frac{1}{1-\frac{1}{5}} = \frac{15}{8}$.

Remark: It follows from this that the number of primes is infinite, since the harmonic series diverges and each integer is a (unique) product of primes!

15.

$$\begin{aligned}
\frac{1}{2}+\sum_{k=1}^{n}\cos(k\theta) &= \frac{1}{2}\sum_{k=0}^{n}(e^{i\,k\theta}+e^{-i\,k\theta})\\
&= \frac{1}{2}\sum_{k=0}^{n}e^{i\,k\theta}+\frac{1}{2}e^{-i\,n\theta}\sum_{k=0}^{n}e^{i\,(n-k)\theta}\\
&= \frac{1}{2}(1+e^{-i\,n\theta})\frac{1-e^{i\,(n+1)\theta}}{1-e^{i\,\theta}}\\
&= \frac{1}{2}e^{-i\,n\frac{\theta}{2}}\big(e^{i\,n\frac{\theta}{2}}+e^{-i\,n\frac{\theta}{2}}\big)\,\frac{e^{i\,(n+1)\frac{\theta}{2}}\big(e^{-i\,(n+1)\frac{\theta}{2}}-e^{i\,(n+1)\frac{\theta}{2}}\big)}{e^{i\,\frac{\theta}{2}}(e^{-i\,\frac{\theta}{2}}-e^{i\,\frac{\theta}{2}})}\\
&= \frac{\cos(n\frac{\theta}{2})\sin((n+1)\frac{\theta}{2})}{\sin(\frac{\theta}{2})}
\end{aligned}$$

Chapter 3

Topology of the real line $\mathbb{R}$

3.1 Open sets and closed sets

Definition 3.1.1 *A subset* $U \subset \mathbb{R}$ *is said to be* open *if* $\forall a \in U, \exists\, \delta > 0$ *such that* $(a-\delta, a+\delta) \subset U$ *. A subset* $A \subset \mathbb{R}$ *is said to be* closed *if its complement* A^c *is open.*

We decree the empty set to be open (this actually follows logically from the definition). Open intervals on the real line are open and closed intervals are closed. It is trivial to see that any union of open sets is open and hence any intersection of closed sets is also closed. However, in general, only finite intersections of open sets are open and only finite unions of closed sets are closed. For example, the union of all the closed intervals $[\frac{1}{n}, 1-\frac{1}{n}]$ is the open interval $(0,1)$ and the intersection of all open intervals $(-\frac{1}{n}, 1+\frac{1}{n})$ is the closed interval $[0,1]$.

Definition 3.1.2 *The interior of* $A \subset \mathbb{R}$ *is the union of all open sets contained in* A*. It is always open and is denoted by* $int(A)$ *(sometimes also by* $\mathring{A}$*). The closure of* $A \subset \mathbb{R}$ *is the intersection of all closed sets that contain* A*. It is obviously closed and is denoted by* $\bar{A}$ *(sometimes also by* $clo(A)$ *.*

If U is open then $\mathring{U} = U$ and if A is closed then $\bar{A} = A$. The closure of a set A consists of A and all its limit (or accumulation) points where a limit point of A is a point p such that every neighbourhood of p contains a point distinct from p which is in A. The limit of any sequence (a_n) contained in A, if it exists, belongs to the closure of A.

3.2 Connected sets

Definition 3.2.1 *A subset* $A \subset \mathbb{R}$ *is called* disconnected *if there are two disjoint open sets* U *and* V *with* $U \cap A \neq \emptyset, V \cap A \neq \emptyset$ *such that* $A \subset U \cup V$. *A* connected *set is a set which is not disconnected.*

The fundamental fact about connected sets on $\mathbb{R}$ is

Theorem 3.2.1 *If* A *is a connected set on the real line and if* $a, b \in A$ *with* $a < b$*, then the whole interval* $[a, b] \subset A$. *Any interval* $[a, b] \subset \mathbb{R}$ *is connected. In fact, any connected set of* $\mathbb{R}$ *is an interval (not necessarily bounded or closed).*

Proof: If $\exists\, x \in (a, b)$ such that $x \notin A$, then $(-\infty, x)$ and $(x, +\infty)$ are two disjoint non-empty open sets such that would disconnect A, i.e , $A \subset (-\infty, x) \cup (x, +\infty)$. Now let $[a, b] \subset U \cup V$ with U, V two non-empty disjoint open subsets of $\mathbb{R}$. Then $\exists\, x \in U \cap [a, b]$, $y \in V \cap [a, b]$, where we can assume that $x < y$. Let $z = lub(U \cap [x, y])$. If $z \in U$, then $z < y$ (since $y \in V$) and so there is an open neighbourhood of z contained in $U \cap [x, y]$, This contradicts the fact that z is the least upper bound. On the other hand, if $z \in V$ then $x < z$ and so there is an open neighbourhood of z contained in $V \cap [x, y]$ contradicting the fact that z is the least upper bound.

QED

3.3 Compact sets

Definition 3.3.1 *A subset of* $\mathbb{R}$ *is called* compact *if it is closed and bounded.*

Let K be a compact set in $\mathbb{R}$. Then since K is bounded, by the completeness axiom, K has a unique least upper bound $M = sup(K)$ and a unique greatest lower $m = inf(K)$. Suppose $M \notin K$. Now K^c is open, so there exists an open neighbourhood of M: $(M - \epsilon, M + \epsilon) \subset K^c$ implying that $M - \frac{1}{2}\epsilon$ is an upper bound for K which contradicts the fact that M is the *least* upper bound. Therefore $M \in K$. Similarly the greatest lower bound m also belongs to K. We have thus proved:

Proposition 3.3.1 *For any compact set* K, $sup(K), inf(K) \in K$

Let (a_n) be sequence contained in a compact set K. Since $K$is bounded, by the Bolzano-Weierstrass theorem, (a_n) has a convergent subsequence whose limit point belongs to $\overline{K} = K$, since K is closed. We have thus proved:

Proposition 3.3.2 *Every sequence of points in a compact set K has a subsequence that converges to a limit point in K.*

This property is sometimes used as the definition of compact sets in $\mathbb{R}$. There is however a more general and useful (albeit rather abstract) characterization of compact sets which we would like to introduce:

Definition 3.3.2 *An* open cover *of a subset $A \subset \mathbb{R}$ is a union of open sets $\bigcup\{U_\omega |\, \omega \in \Omega\}$ that contains A. A* finite subcover *of the cover $\bigcup\{U_\omega | \omega \in \Omega\}$ is a finite subset $\{\omega_1, \omega_2, ..., \omega_n\} \subset \Omega$ such that $U_{\omega_1} \cup U_{\omega_2} \cup ... \cup U_{\omega_n}$ still contains A.*

Let us, for the moment call sets with the property that *every* open cover has a finite subcover "*kompakt*" (German for compact!). First of all such a set is necessarily bounded, because $U_n = (-n, n), n \in \mathbb{N}$ is an open cover for any subset of $\mathbb{R}$ and unless the set is bounded we will never be able to find a finite subcover. Secondly, a "kompakt" set has to be closed because if p is a limit point of the set which does not belong to the set, then we can form the open cover $A_n^c, n \in \mathbb{N}$, where A_n is the closed interval $[p - \frac{1}{n}, p + \frac{1}{n}]$. This certainly covers A since it covers everything in $\mathbb{R}$ except the point p. In fact we now prove the key result:

Theorem 3.3.1 *A subset $K \subset \mathbb{R}$ is compact if and only if every open cover of K has a finite subcover.*

Proof: Let K be a compact set and let $\bigcup\{U_\omega |\, \omega \in \Omega\}$ be an open cover of K. As a preliminary step, we find a countable subcover $U_{\omega_1} \cup U_{\omega_2} \cup ... \cup U_{\omega_n} ...$ that still covers K. This can be achieved by using the fact that the countable set $\mathbb{Q}$ is dense in $\mathbb{R}$ and so the countable set of intervals of the form $(r - \frac{1}{n}, r + \frac{1}{n})$, where $r \in \mathbb{Q}$ and $n \in \mathbb{N}$, form an open cover for all of $\mathbb{R}$. Each point $x \in K$ is contained in such an interval $(r - \frac{1}{n}, r + \frac{1}{n})$ and we choose a definite U_ω which contains it. So by a change of notation we can assume that $\bigcup\{U_k : k \in \mathbb{N}\}$ covers K. Assume, on the contrary, that there is no finite n such that $K \subset U_1 \cup ... \cup U_n$. Then $\forall n, \exists x_n \in K$ such that $x_n \notin U_1 \cup ... \cup U_n$. The sequence (x_n) has a subsequence that converges to a limit point $x \in K$.

Since x is in the open set U_N for some N, and since $x_n \notin U_N$ for all $n > N$, we obtain a contradiction.

QED

3.4 Elementary topology of $\mathbb{R}^n$

As everyone knows $\mathbb{R}^n$ denotes the Euclidean vector space of all n -tuples of real numbers: $\{x = (x_1, x_2, ..., x_n) | \, x_i \in \mathbb{R}\}$, where $n \in \mathbb{N}$.

We denote the distance in $\mathbb{R}^n$ by $d(x, y) = ||x - y||$, where the norm (= the length) $||v||$ of a vector v is defined as usual by the scalar product $||v||^2 =< v, v >$. (For simplicity of notation we will denote vectors by simple letters. $n = 1$ is a special case!).

The open ball (or disk) of radius $r > 0$ with centre a in $\mathbb{R}^n$ is the set $B_r(a) = \{x \in \mathbb{R}^n | \, d(x, a) < r\}$. Sometimes we also use the closed ball defined by $\overline{B}_r(a) = \{x \in \mathbb{R}^n | \, d(x, a) \leq r\}$

Definition 3.4.1 *A subset* $A \subset \mathbb{R}^n$ *is said to be* open *if* $\forall a \in A, \, \exists \, \delta > 0$ *such that* $B_\delta(a) \subset A$.
A subset $A \subset \mathbb{R}^n$ *is said to be* closed *if its complement* A^c *is open.*

Any union of open sets is open and hence any intersection of closed sets is also closed, but in general, only finite unions of open sets are open and only finite intersections of closed sets are closed. This basic property of open sets leads to the following definition of an abstract topological space where every other "superficial" geometric property is stripped away until we are left with the bare bones of the topology of open subsets.

Definition 3.4.2 *A* topological space *is a set* X *with a distinguished collection* $\{U_\tau\}$ *of subsets called open sets, containing the empty set and the set* X *itself with the property that any union and any finite intersection of open sets is open.*

Definition 3.4.3 *A subset* A *in a topological space is said to be* connected *if it is not contained in a disjoint union of two non-empty open subsets*

Definition 3.4.4 *An* open cover *of a subset* A *in a topological space is a union of open sets* $\bigcup\{U_\omega | \, \omega \in \Omega\}$ *that contains* A.
A finite subcover *of the cover* $\bigcup\{U_\omega | \omega \in \Omega\}$ *is a finite subset* $\{\omega_1, \omega_2, ..., \omega_n\} \subset \Omega$ *such that* $U_{\omega_1} \cup U_{\omega_2} \cup ... \cup U_{\omega_n}$ *still contains* A.

Definition 3.4.5 *A subset A in a topological space is said to be* compact *if every open cover of A has a finite subcover.*

The key theorem about compact sets in $\mathbb{R}^n$ is the following generalization of the Bolzano-Weierstrass property:

Theorem 3.4.1 *A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.*

3.5 Exercises

1. Determine for each of the following sets wheter they are (i) open, (ii) closed, (iii) connected, (iv) bounded and (v) compact:

(i) $(-\infty, 1] \cup (0, 5] \subset \mathbb{R}$

(ii) $(-\infty, 1] \cap [0, 5) \subset \mathbb{R}$

(iii) $\{x \ : \ |x| \geq 2\} \subset \mathbb{R}$

(iv) $\{z \ : \ 0 < |z| \leq 1\} \subset \mathbb{C}$

(v) $\{z \ : \ |z| > 2\} \subset \mathbb{C}$

2. Give an example of an open cover of the set $(-1, +1)$ which does not admit any finite subcover.

3. Show that a finite union and an arbitrary intersection of compact sets is again compact

4. Let $(K_n)_{n\in\mathbb{N}}$ be a sequence of compact subsets of the real line satisfying the property: $K_{k+1} \subset K_k$, Such a sequence is called a nested sequence. Prove that the intersection $\bigcap\{K_n | n \in \mathbb{N}\}$ is non-empty. (For the sake of simplicity you can assume that each K_n is a closed interval $[a_n, b_n]$).

6*. Show that any (non-empty) open subset of $\mathbb{R}$ is a countable union of disjoint open intervals.

3.5.1 Solutions to some exercise problems

1.

(i) $(-\infty, 1] \cup (0, 5] = (-\infty, 5]$ is not open, closed, connected, not bounded and not compact.

(ii) $(-\infty, 1] \cap [0, 5) = [0, 1]$ is not open, closed, connected, bounded and compact.

(iii) $\{x \ : \ |x| \geq 2\} = (-\infty, -2] \cup [2, +\infty)$ is not open, closed, not connected, not bounded and not compact.

(iv) $\{z \ : \ 0 < |z| \leq 1\} \subset \mathbb{C}$ is not open, not closed, connected, bounded and not compact.

(v) $\{z \ : \ |z| > 2\} \subset \mathbb{C}$ is open, not closed, connected, not bounded and not compact.

2. $\{(-1+\frac{1}{n}, 1-\frac{1}{n}) \mid n \in \mathbb{N}\}$ form an open cover of $(-1, +1)$ without a finite subcover. The union of any finite subcollection of this nested sequence of open intervals is of the form $(-1+\frac{1}{N}, 1-\frac{1}{N})$ which does not cover $(-1, +1)$.

4. Let $K_n = [a_n, b_n]$ be a nested sequence of compact intervals of the real line such that $K_{k+1} \subset K_k$, for all k. The sequence (a_k) is an increasing sequence which is bounded from above by b_1 and hence it converges to a limit a. Similarly (b_k) is a decreasing sequence which is bounded from below by a_1 and hence it converges to a limit b. Now $a \leq b$ since $a_k < b_k$ for all k and hence the intersection id the non empty closed interval $[a, b]$.

Chapter 4

Continuity

It took mathematicians some time to settle on an appropriate definition of this key concept in analysis. Intuitively continuity means that there are no sudden unexpected jumps. A continuous function is supposed to have a graph with no breaks. Continuity can be defined in several different ways, depending on the degree of abstraction and generality. The "easiest" definition is topological and a continuous function between two topological spaces is just a map with the property that the inverse image of any open set is also open. Remember in a topological space every other "superfluous" structure has been stripped away so that the only fundamental concept left is that of an open set, so everything has to be defined in terms of open sets.

4.1 Continuous functions

We will begin with the usual ϵ, δ-definition for functions on the real line and show how it can be reformulated purely in terms of open sets.

Definition 4.1.1 *A function $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, is said to be continuous at a point $a \in A$ if and only if the following holds:*
$\forall\, \epsilon > 0, \exists\, \delta > 0$ *such that* $x \in A \cap (a - \delta, a + \delta) \;\Rightarrow\; |f(x) - f(a)| < \epsilon$

If f is continuous at each point of a subset B of A, then f is said to be continuous on B.

Definition 4.1.2 *We say that a subset B of A is relatively open (or relatively closed) in A if it is the intersection of an open (respectively closed) set and A.*

Proposition 4.1.1 *A function $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, is continuous on A if and only if for every open subset $U \subset \mathbb{R}$ the inverse image $f^{-1}(U) = \{a \in A \,|\, f(a) \in U\}$ is relatively open in A.*

Proof: Assume first that f is continuous and let $U \in \mathbb{R}$ be open. If $f^{-1}(U)$ is empty it is open. Otherwise, let $a \in f^{-1}(U)$ so that $f(a) \in U$. Since U is open, $\exists\, \epsilon > 0$ such that $B_\epsilon(f(a)) = (f(a) - \epsilon, f(a) + \epsilon) \subset U$. Therefore, $\exists\, \delta > 0$ such that $f(A \cap B_\delta(a)) \subset B_\epsilon(f(a)) \subset U$, so that $A \cap B_\delta(a) \subset f^{-1}(U)$ and hence $f^{-1}(U)$ is relatively open in A .
Now suppose that the inverse image (under f) of every open set is relatively open in A . Now let $a \in A$ and $\epsilon > 0$ be given. Then since $B_\epsilon(f(a))$ is open, its inverse image: $V = f^{-1}(B_\epsilon(f(a)))$, which contains the point a, is relatively open in A. Therefore $\exists\, \delta > 0$ such that $B_\delta(a) \subset V$ which implies that $f(A \cap B_\delta(a)) \subset f(A \cap V) \subset B_\epsilon(f(a))$.

QED

Now you can see why the characterization of continuity by open sets is not only more general, but also by far the more elegant description. (The ϵ, δ definition has been the main reason why a lot of students dislike analysis and quit studying mathematics!). By using the definition that closed sets are nothing but complements of open sets one can now easily prove the following:

Proposition 4.1.2 *A function $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, is continuous on A if and only if for every closed subset $B \subset \mathbb{R}$ the inverse image $f^{-1}(B)$ is relatively closed.*

A simple corollary of the above proposition is the following useful fact:

Proposition 4.1.3 *$f : A \to \mathbb{R}$, where $A \subset \mathbb{R}$, is continuous at a point $a \in A$ if and only if whenever (a_n) is a sequence in A converging to a, the sequence $(f(a_n))$ converges to $f(a)$.*

It is also convenient to have a definition of a limit of a function

Definition 4.1.3 *Let $f : A \to \mathbb{R}$ be a function and $a \in \bar{A}$ (closure of A). Then we write* $\lim_{x \to a} f(x) = l$ *if and only if, whenever (a_n) is a sequence in A converging to a, the sequence $f(a_n)$ converges to l. In other words:*

$\forall\, \epsilon > 0, \exists\, \delta > 0$ *such that* $x \in A \cap (a - \delta, a + \delta) \;\Rightarrow\; |f(x) - l| < \epsilon$

If f and g are real valued functions, we define the function $f + g$ by $(f + g)(x) = f(x) + g(x)$ for all x in the common domain of definition.

Similarly, we may define the difference, product and quotient of functions. (The quotient is not defined at points where the denominator is zero). The following arithmetic properties for continuous functions then follow easily from the corresponding properties for limits

Proposition 4.1.4 *If f and g are continuous, then so are $f \pm g, f.g$ and f/g whenever they are defined .*

Another basic property is the following:

Proposition 4.1.5 *The composite $f \circ g$ of two continuous functions is continuous, wherever the composition is defined.*

Proof: This follows from the fact that $(f \circ g)^{-1}(U) = g^{-1}(f^{-1}(U))$ and the elegant proposition 4.1.1.

Examples

0. Any constant function $f(x) = c$ is continuous everywhere.

1. The identity function defined by $f(x) = x$ is continuous everywhere.

2. Hence by the above propositions, any polynomial function is continuous everywhere and any rational function (a ratio of two polynomial functions) is continuous at all points where the denominator is not zero.

3. We will see later that the elementary "transcendental " functions like sin, cos, exp, log, ... are all continuous on their domains of definition.

4. The functions defined by $f(r) = 1$ for $r \in \mathbb{Q}$ and $f(x) = 0$ for x irrational is defined everywhere but is discontinuous at every point, because $\mathbb{Q}$ is dense in $\mathbb{R}$.

5. Let f be the function defined by $f(x) = 0$ if x is irrational and $f(\frac{m}{n}) = \frac{1}{n}$ for a rational number $\frac{m}{n}$ (in lowest terms). Then f is discontinuous at every rational point, but continuous at every irrational point.

4.2 Continuity and Connectedness

The main result of this section is the following important

Theorem 4.2.1 *The image of a connected set under a continuous map is connected.*

Proof: Let A be a connected set and let f be continuous on A. If $f(A)$ were disconnected, then by definition, there are two disjoint non-empty open sets U and V such that $f(A) \subset U \cup V$. Since f is continuous, $f^{-1}(U)$ and $f^{-1}(V)$ are two non-empty disjoint open sets such that $A \subset f^{-1}(U) \cup f^{-1}(V)$ which contradicts the assumption that A is connected.

QED

Using the fact that intervals on the real line are connected we obtain the "classical" Intermediate Value Theorem of first year calculus.

Theorem 4.2.2 *Let f be real-valued function continuous on a (connected) interval I containing the two points a and b. Then for any value y which lies between the two real numbers $f(a)$ and $f(b)$ there exists a point x between a and b such that $f(x) = y$.*

Since this is one of the key results, let me give you another "more elementary" proof of the above theorem.

Alternate Proof: By a some simple translation (i.e., a vertical shift of the graph of the function), we can assume w.l.o.g. that $I = [a,b]$, $f(a) < 0$, $f(b) > 0$ and that $y = 0$. We then define two monotonic sequences (one increasing and the other decreasing) (a_k) and (b_k) such that $f(a_k) \leq 0$ and $f(b_k) \geq 0$ for all $k \in \mathbb{N}$ recursively as follows:
First put: $a_1 = a$, $b_1 = b$. Now suppose a_k, b_k are defined. Let $c_k = \frac{1}{2}(a_k + b_k)$ be the mid point of $[a_k, b_k]$. If $f(c_k) \leq 0$ set $a_{k+1} = c_k$ and $b_{k+1} = b_k$. If $f(c_k) > 0$ set $a_{k+1} = a_k$ and $b_{k+1} = c_k$.
We then have: $a = a_1 \leq \cdots \leq a_k \leq a_{k+1} < b_{k+1} \leq b_k \leq \cdots \leq b_1 = b$. Moreover $b_{k+1} - a_{k+1} = \frac{1}{2}(b_k - a_k)$, so that by induction $b_{k+1} - a_{k+1} = \frac{1}{2^k}(b-a) \to 0$ as $k \to \infty$ and hence $\lim a_k = \lim b_k = x$, say. Since f is continuous, $\lim f(a_k) = f(x) \leq 0$ and $\lim f(b_k) = f(x) \geq 0$ and therefore $f(x) = 0$.

QED

Remark: The method used in the above proof can be implemented numerically (say on a computer) to find solve an equation of the form $f(x) = 0$. It is called the bisection method. The method of false position uses a similar idea, but this time instead of taking the midpoint of the interval at each interation step, one uses the intersection point of the $x-$axis with the straight line joining $(a_k, f(a_k))$ and $(b_k, f(b_k))$, i.e.

$$c_k = (b_k f(a_k) - a_k f(b_k))/(f(a_k) - f(b_k))$$

.

Another simple consequence is the following:

Proposition 4.2.1 *If $f : [a,b] \to \mathbb{R}$ is continuous and injective then f is either strictly increasing (i.e. $x < y \Rightarrow f(x) < f(y)$) or strictly decreasing.*

Proof: We assume w.l.o.g. that $f(a) < f(b)$. Let $x < y\,;\, x, y \in [a,b]$. Define $g(t) = f(y(t)) - f(x(t))$, where $x(t) = a + t(x-a)$ and $y(t) = b - t(b-y)$ for $t \in [0,1]$, so that $\forall\, t \in (0,1)\, y(t) - x(t) = (1-t)(b-a) + t(y-x) > 0$. $g[0,1] \Rightarrow \mathbb{R}$ is a difference of composition of continuous functions and hence is continuous. At $t = 0$: $g(0) = f(b) - f(a) > 0$. If $g(1) = f(y) - f(x) < 0$, then $\exists\, t^* \in (0,1)$ such that $g(t^*) = 0$, by the Intermediate Value Theorem applied to g. This contradicts the fact that f is injective since $y(t^*) > x(t^*)$, so $f(x) < f(y)$.

QED

An important consequence is the following:

Corollary 4.2.1 *If $f : [a,b] \to [c,d]$ is continous and bijective, then the inverse function $g = f^{-1} : [c,d] \to [a,b]$ is continuous (and monotonic).*

Proof: We may assume w.l.o.g. that f is strictly increasing. Let $x \in (a,b)$, $y = f(x) \in (c,d)$ and let $\epsilon > 0$ be given. f maps the ϵ-neighbourhood of x (in $[a,b]$) bijectively onto the interval $U = (f(x+\epsilon), f(x-\epsilon))$ containing y in $[c.d]$. Pick $\delta = \frac{1}{10} min\big(f(x+\epsilon) - y\,,\, y - f(x-\epsilon)\big)$ so that the δ-neighbourhood of y lies inside U. Then $g = f^{-1}$ will map this neighbourhood back into the given ϵ-neighbourhood U of $x = g(y)$.

QED

4.3 Continuity and Compactness

The main result of this section is the following extremely important

Theorem 4.3.1 *The image of a compact set under a continuous map is compact*

Proof: Let K be a compact set and let f be continuous on K. Let $\bigcup\{U_\omega | \,\omega \in \Omega\}$ be an open cover of $f(K)$. Then since f is continuous, $\bigcup\{f^{-1}(U_\omega) | \,\omega \in \Omega\}$ is an open cover of the compact set K and so has a finite subcover: $K \subset f^{-1}(U_{\omega_1}) \cup f^{-1}(U_{\omega_2}) \cup ... \cup f^{-1}(U_{\omega_N})$. It is now obvious that $f(K) \subset U_{\omega_1} \cup U_{\omega_2} \cup ... \cup U_{\omega_N}$.

QED

As you can see from the proof given above, the theorem is valid in any topological space, but for real-valued functions it boils down to the following "Extreme Value Theorem" of first year calculus:

Proposition 4.3.1 *Let f be real-valued function continuous on a closed and bounded set K. Then f attains its absolute minimum and maximum values on K*

Proof: Since $f(K)$ is a bounded set of $\mathbb{R}$, both its least upper bound: $sup(f(K))$ and its greatest lower bound $inf(f(K))$ exist and since $f(K)$ is closed they are both elements of $f(K)$. *QED*

Combining the two basic theorems, we can now say that the continuous image of a closed, connected and bounded interval $[a, b]$ is also a closed, connected and bounded interval $[m, M]$ where m is the absolute minimum and M is the absolute maximum value of f that are attained on $[a, b]$.

As an application of the above ideas and theorems let us state and prove the following beautiful result.

Theorem 4.3.2 Fixed Point Theorem
Let $f : [a, b] \to [a, b]$ be a continuous map. Then $\exists\, p \in [a, b]$ such that $f(p) = p$, i.e. p is a fixed point of the map f.

Proof: We assume that $f(a) \neq a$ and $f(b) \neq b$ because otherwise, either a or b is a fixed point. Since f maps $[a, b]$ to $[a, b]$ we then must have $f(a) > a$ and $f(b) < b$. Let g be defined by $g(x) = f(x) - x$. g is then a continuous function such that $g(a) > 0$ and $g(b) < 0$, so by the Intermediate Value Theorem $\exists\, p \in (a, b)$ such that $g(p) = 0$ which means that $f(p) = p$.

QED

The above is an example of an existence result that mathematicians find extremely useful. There is a generalization of the above theorem for certain domains in $\mathbb{R}^n$ (or even to more general spaces) called the Brouwer Fixed Point Theorem.

Continuing in the same vein, I would also like to mention another very important existence result result that is amazingly simple to state and prove but its appearance is ubiquitous like dandelions.

4.3.1 The Contraction Mapping Principle

Theorem 4.3.3 *: Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous map with the following contraction property:*
$\exists\, C \in (0,1)$ such that $|f(x) - f(y)| \le C|x - y|$ for all $x, y \in \mathbb{R}$.
Then, $\exists$ a unique fixed point $p \in \mathbb{R}$ such that $f(p) = p$.

Proof:

Step 0. We first show that f is continuous:
$\forall \epsilon > 0\ \exists\, \delta = \frac{\epsilon}{C}$ such that $|x - y| < \delta \Rightarrow |f(x) - f(y)| \le C|x - y| < \epsilon$ for all x, y.

Step 1. Construction of a sequence converging to the fixed point:
Define: $x_1 = 1$ and recursively: $x_{k+1} = f(x_k)$.

Step 2. Proof that the sequence is Cauchy:
$|x_{k+2} - x_{k+1}| = |f(x_{k+1} - f(x_k)| \le C|x_{k+1} - x_k|$ and hence by induction: $|x_{k+2} - x_{k+1}| \le C^k|x_2 - x_1|$ for all $k \in \mathbb{N}$. Therefore:

$$|x_{k+m+1} - x_{k+1}| \le \sum_{i=1}^{m} |x_{k+i+1} - x_{k+i}| < |x_2 - x_1| \sum_{i=1}^{m} C^{k+i-1} < |x_2 - x_1| \frac{C^k}{1-C}$$

Since $\lim_{k\to\infty} C^k = 0$, (x_k) is a Cauchy sequence and hence converges to a limit $\lim x_k = p$.

Step 3. Proof that x_k converges to a fixed point:
Since f is continuous, $p = \lim x_{k+1} = \lim f(x_k) = f(\lim x_k) = f(p)$.

Step 4. Proof that the fixed point is unique:
If p and q are two fixed points, then $|p - q| = |f(p) - f(q)| \le C|p - q|$ with $0 < C < 1$ so $p = q$.

QED

Remark: From the proof we can see that the theorem holds in any complete metric space (see Chapter 7 for the definition of a metric space).

4.4 Uniform Continuity

Definition 4.4.1 *A function $f : A \to \mathbb{R}$, where $A \subset \mathbb{R}^n$, is said to be* uniformly continuous *on A if and only if the following holds:*
$\forall\, \epsilon > 0, \exists\, \delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for all $x, y \in A$ satisfying $||x - y|| < \delta$.

The key thing to note here is that the choice of δ should depend only on ϵ, the function f and the set A; it has to be independent of the points x and y.

Examples:

1 The function $f(x) = x^2$ is not uniformly continuous on $A = \mathbb{R}$, but it is on any bounded interval. It is not uniformly continuous on $\mathbb{R}$, since $\exists\, \epsilon = 2$ such that for every $\delta > 0$ we can find two points $x = n$ and $y = n + \frac{1}{n}$, such that $|x - y| = \frac{1}{n} < \delta$, but with $|(n + \frac{1}{n})^2 - n^2| = 2 + \frac{1}{n^2} > 2$.

2. The function $f(x) = sin(x)$ is uniformly continuous on all of $\mathbb{R}$. (The easiest way to prove this is to use the mean value theorem from the next chapter)

3. The function $f(x) = \frac{1}{x}$ is not uniformly continuous on the interval $(0, 1]$ but is uniformly continuous on $[1, \infty)$. It is not uniformly continuous on $(0, 1]$, since $\exists\, \epsilon = 1$ such that for every $\delta > 0$ we can find two points $\frac{1}{n}$ and $\frac{1}{n+1}$ in $(0, 1)$, satisfying $|\frac{1}{n} - \frac{1}{n+1}| = \frac{1}{n(n+1)} < \delta$, but with $|f(\frac{1}{n}) - f(\frac{1}{n+1})| = 1$. It is uniformly continuous in $(0, \infty)$, because $|\frac{1}{x} - \frac{1}{y}| = \frac{|x-y|}{|xy|} < |x - y|$ for all $x, y \in (0, \infty)$, so for any $\epsilon > 0$, we can choose $\delta = \epsilon$ uniformly independent of x and y.

The main result of this section is the following:

Theorem 4.4.1 *A continuous function on a compact set is uniformly continuous.*

Proof : Let f be a continuous function defined on a compact set K. If f is not uniformly continuous then there exists an $\epsilon > 0$ such that for each k with $\delta = \frac{1}{k}$, we can find two points $x_k, y_k \in K$ so that $||x_k - y_k|| < \delta = \frac{1}{k}$ but with $|f(x_k) - f(y_k)| \geq \epsilon$. By the Bolzano-Weierstraas property there exists convergent subsequences x_{k_l}, y_{k_l} converging to $x^* \in K$ and $y^* \in K$ respectively. $x^* = y^*$ since $||x_k - y_k|| < \frac{1}{k}$ for all k. Since f is a continuous function $f(x_{k_l}) \to f(x^*)$ and $f(y_{k_l}) \to f(y^*) = f(x^*)$ contradicting the fact that there exists a fixed $\epsilon > 0$ with $|f(x_k) - f(y_k)| \geq \epsilon$ for all k.

QED

4.5 Exercises

1. Prove that every polynomial of odd degree with real coefficients has at least one real root.

2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function. Show that the zero-set of f, i.e., the set $\{x \in \mathbb{R}^n \mid f(x) = 0\}$ is closed in $\mathbb{R}^n$.

3. Let $f, g : [a, b] \to \mathbb{R}$ be two continuous functions such that $f(r) = g(r)$ for all rational $r \in [a, b] \cap \mathbb{Q}$. Show that $f(x) = g(x)$ for all $x \in [a, b]$.

4. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 1$ if $x \in \mathbb{Q}$ and $f(x) = 0$ otherwise. Show that $\lim_{x\to 0} f(x)$ does not exist, but that $\lim_{x\to 0} x\, f(x) = 0$.

5. Let $f : K \to \mathbb{R}$ be continuous on a compact interval K. Suppose that $\forall\, x \in K\ \exists y \in K$ such that $|f(y)| \leq \frac{1}{2}|f(x)|$. Prove that $\exists a \in K$ such that $f(a) = 0$.

6. If $f, g : [a, b] \to \mathbb{R}$ are continuous real-valued functions, show that $\max(f, g)$ is also continuous. ($\max(f, g)$ is obviously the function defined by: $\max(f, g)(x) = \max(f(x), g(x))$).

7. Show that the function $f : (0, \infty) \to \mathbb{R}\,; f(x) = \sqrt{x}$ is uniformly continuous but that the function $g : (0, \infty) \to \mathbb{R}\,; g(x) = x^2$ is not uniformly continuous.

8*. Show that there does not exist continuous function $f : \mathbb{R} \to \mathbb{R}$ which takes on each value (that it takes on) exactly twice. (in other words the pre-image of a point under f is either empty or consists of exactly two points).

9. Is the composition of two uniformly continuous functions uniformly continuous? Justify your answer.

10. Show that if $f : \mathbb{R} \to \mathbb{R}$ is uniformly continuous and if (x_n) is a Cauchy sequence, then $f(x_n)$ is also a Cauchy sequence.

11. Show that a uniformly continuous real valued function defined on a bounded (but not necessarily compact!) set $B \subset \mathbb{R}^n$ is bounded on B.

4.5.1 Short solutions to some of the exercise problems

1. This follows from the Intermediate Value Theorem.

2. $\{0\}$ is a closed set in $\mathbb{R}$, because its complement is the union of two open intervals: $(\infty, 0) \cup (0, \infty)$. Since the inverse image of a closed set under a continuous map is closed, $f^{-1}(\{0\}) = \{x \in \mathbb{R}^n \mid f(x) = 0\}$ is closed in $\mathbb{R}^n$.

3. The function $h(x) = f(x) - g(x)$ is continuous and $h(r) = 0$ for all $r \in \mathbb{Q} \cap [a, b]$. Suppose that there exists an irrational number $x \in [a, b]$, such that $h(x) \neq 0$. Let $\epsilon = \frac{1}{2}|h(x)| > 0$. Then $\exists \delta > 0$ such that $|y - x| < \delta \Rightarrow |h(y) - h(x)| < \frac{1}{2}|h(x)| \Rightarrow |h(y)| > \frac{1}{2}|h(x)| > 0$ for all $y \in [a, b]$. This is a contradiction, since every neighbourhood of x contains a rational point r with $h(r) = 0$.

Another way to see it is because the the set of points where $f - g = 0$ is a closed set (since it is the inverse image of the closed set consisting of the single point $\{0\}$ under the continuous map $f - g$). Now the smallest closed set containingg $\mathbb{Q}$ is its closure which is all of $\mathbb{R}$.

4. Suppose $\lim_{x\to 0} f(x)$ exists and is equal to some real number a. Then for $\epsilon = \frac{1}{2}$ (say), $\exists \delta > 0$ such that $|f(x) - a| < \frac{1}{2}$ for all $x \in (-\delta, +\delta)$. Now each such neighbourhood contains both an irrational point s and a rational point r; but then $|f(s) - f(r)| = 1 > \frac{1}{2}$. Contradiction!

Since $f(x)$ is either zero or one, $|x\,f(x)| \leq |x|$ for all x. Therefore: $\forall \epsilon > 0$, we choose $\delta = \epsilon$ to have $|x - 0| < \delta \Rightarrow |xf(x) - 0| \leq |x| < \epsilon$. Therefore $\lim_{x\to 0} x\,f(x) = 0$.

6. Since $|f|$ is a continuous function, it attains its absolute minimum value at a certain point $a \in [0, 1]$. If $|f(a)| > 0$, we obtain a contradiction from the assumption that there is another point y, where $|f(y)| \leq \frac{1}{2}|f(a)|$. Therefore $f(a) = 0$.

7. $\sqrt{x}$ is continuous on $[0, \infty)$ and hence it is uniformly continuous on any compact subset, for example on $[0, 1]$. Now, if $x, y \geq 1$, then

$$|\sqrt{x} - \sqrt{y}| = \frac{|x - y|}{|\sqrt{x} + \sqrt{y}|} \leq \frac{1}{2}|x - y|$$

This shows that $\sqrt{x}$ is uniformly continuous on $[1, \infty)$. (Choose $\delta = 2\epsilon$)!

However the function $g(x) = x^2$ is not uniformly continuous on $(0, \infty)$. Consider the two sequences $x_n = n$ and $y_n = n + \frac{1}{n}$. $|x_n - y_n| = \frac{1}{n}$, but $|(x_n)^2 - (y_n)^2| = 2 + \frac{1}{n^2} \geq 2$. So there exists "$\epsilon$" $= 2 > 0$ such that for each $\delta > 0$, we can find points x_n and y_n in $(0, \infty)$ with $|x_n - y_n| = \frac{1}{n} < \delta$ and $|(x_n)^2 - (y_n)^2| \geq 2$

11. Let f be uniformly continuous on a bounded set B. Suppose, on the contrary that f is not bounded. Then there exists a sequence (x_n) in B, such that $\lim_{n\to\infty} f(x_n) = \infty$ (or $-\infty$). Since (x_n) is a bounded sequence, by Bolzano-Weierstrass, it has a subsequence: $y_k = x_{n_k}$ which is convergent and hence Cauchy. We want to show now that $f(y_k)$ is also a Cauchy sequence and hence convergent, contradicting the fact that $\lim_{n\to\infty} f(x_n) = \infty$ (or $-\infty$).
So let $\epsilon > 0$ be given. Then since f is uniformly continuous, $\exists\, \delta > 0$, such that $|y_{k+l} - y_k| < \delta \Rightarrow |f(y_{k+l}) - f(y_k)| < \epsilon$ for any $k, l \in \mathbb{N}$. Since (y_k) is a Cauchy sequence, $\exists\, K \in \mathbb{N}$ such that $k \geq K \Rightarrow |y_{k+l} - y_k| < \delta \Rightarrow |f(y_{k+l}) - f(y_k)| < \epsilon$ for all l.

Chapter 5

Differentiation

5.1 Definition and basic properties

Definition 5.1.1 *Let f be a real-valued defined in an open neighbourhood U of $a \in \mathbb{R}$. f is said to be* differentiable *at a if there exists a real number $f'(a)$ called the derivative of f at a, such that the folowing holds: $\forall\, \epsilon > 0, \exists\, \delta > 0$ s. t. $\forall x \in U$, $|x-a| < \delta \;\Rightarrow\; |f(x) - f(a) - f'(a)(x-a)| < \epsilon\, |x-a|$.*
If f is differentiable at each point of an open set A, then we say that f is differentiable in A.

In other words, f is differentiable at a point a if there is a number c called the derivative of f at a such that the error function $r(x) = f(x) - (f(a) + c(x-a))$, that remains after approximating f near a by the linear (affine) function $l(x) = f(a) + c(x-a)$ satisfies:

$$\lim_{x \to a,\, x \neq a} \frac{r(x)}{x-a} = 0$$

.

It follows that if f is differentiable at a, then f is continuous at a, since near a, f differs from a simple (linear) function $l(x)$, which is definitely continuous, by a term $r(x)$ that goes to 0 (even faster than $|x-a|\,!$) as $x \to a$. This suggests the notation:

Definition 5.1.2 *We say that a function $\phi(x)$ is $o(|x-a|)$ (read: little-oh) as $x \to a$ if $\lim_{x\to a} \frac{\phi(x)}{x-a} = 0$*

We say that $\phi(x)$ is $O(|x-a|)$ (read: big-oh) as $x \to a$ if the ratio $\frac{|\phi(x)|}{|x-a|} = l$ stays bounded as $x \to a$

The following arithmetic properties for derivatives then follow easily from the corresponding properties for limits

Proposition 5.1.1 *If f and g are differentiable at a, then so are $f+g, f-g, f.g$ and f/g (provided $g(a) \neq 0$) and we have the following formulas:*

(i) $(f \pm g)'(a) = f'(a) \pm g'(a)$

(ii) (Product Rule) $(f.g)'(a) = f'(a).g(a) + f(a).g'(a)$

(ii) (Quotient Rule) $(\frac{f}{g})'(a) = \frac{f'(a).g(a) - f(a).g'(a)}{(g(a))^2}$

Proof:

(i) This is quite trivial.

(ii)
$f(x)g(x) - f(a)g(a) = (f(x) - f(a))g(a) + f(x)(g(x) - g(a)) = (c_1(x-a) + r_1(x))g(a) + f(a)(c_2(x-a) + r_2(x))$, where $c_1 = f'(a)$, $c_2 = g'(a)$ and r_1, r_2 are the corresponding remainder terms. It follows that
$f(x)g(x) - (f(a)g(a) + (c_1 g(a) + f(a)c_2)(x-a) = r_1(x)g(a) + f(a)r_2(x)$ is $o(|x-a|)$ since r_1 and r_2 are $o(|x-a|)$. Therefore, $f.g$ is differentiable at a with derivative $f'(a).g(a) + f(a).g'(a)$.

(iii)
The special case where f is constant $= 1$ follows easily from the calculation:

$$\lim_{h \to 0} \frac{1}{h}\left(\frac{1}{g(x+h)} - \frac{1}{g(x)}\right) = \lim_{h \to 0} \frac{1}{g(x+h)g(x)} \frac{g(x) - g(x+h)}{h}$$

The general case then follows from the product rule.

QED

Another extremely important formula is the following:

Proposition 5.1.2 The Chain Rule
If g is differentiable at a and f is differentiable at $g(a)$, then the composite $f \circ g$ is differentiable at a with derivative given by:

$$(f \circ g)'(a) = f'(g(a)).\, g'(a)$$

Proof:
Let $b = g(a)$, $y = g(x)$, $c_1 = f'(a)$, and $c_2 = f'(b)$.
$f(y) - f(b) = c_1(y-b) + r_1(y)$, where $r_1(y)$ is $o(|y-b|)$ and $y - b = c_2(x-a) + r_2(x)$ with $r_2(x)$ in $o(|x-a|)$.
Since $f(y) - f(b) = c_1(c_2(x-a) + r_2(x)) + r_1(y)$, the chain rule now follows because $|y-b|$ is $O(|x-a|)$.

QED

Examples

0. The constant functions $f(x) = c$ are all differentiable everywhere. (Their derivatives are dead zero everywhere !)

1. The "identity map"' $f(x) = x$ is differentiable. ($f(x) - f(a) = x - a$, so the derivative is dead constant $= 1$ everywhere !)

1(bis). By the elementary arithmetic for derivatives (the last two propositions) we see that all polynomials and rational functions (wherever they are defined) are differentiable and their derivatives can be computed easily and mechanically (by Maple for example!) by the above rules.

2. As you all know, the basic transcendental function $exp(x) = e^x$ is also differentiable everywhere. However, this fact does not follow from the rational arithemetic rules above and has to be proved. The most natural place to do this in Chapter 7, when we discuss uniform convergence of power series, but I will sketch an elementary proof that the derivative of e^x at $x = a$ is equal to e^a :
Proof: By the functional equation $e^{x+y} = e^x \, e^y$, we have:

$$\frac{e^{a+h} - e^a}{h} = e^a \, \frac{e^h - 1}{h}$$

but by the definition of the exponential function: $e^h - 1 = \sum_{k=1}^{\infty} \frac{h^k}{k!} = h + r(h)$ where $r(h) = h^2 \sum_{k=2}^{\infty} \frac{h^{k-2}}{k!}$ is obviously $o(|h|)$, by the estimates we learned in Chapter 2.

5.2 Local properties of the derivative

Suppose f is differentiable at a and that $c = f'(a) > 0$. Let $l(x) = f(a) + c(x-a)$. By the definition of the derivative $\exists$ a δ-neighbourhood $U$of a, where we have $|f(x) - l(x)| < \frac{1}{10}c|x-a|$. This implies (for every $x \in U$) $f(x) > l(x) - \frac{1}{10}c(x-a) = f(a) + \frac{9}{10}c(x-a)$ which is $> f(a)$ if $x > a$ and $<$ $f(a)$ if $x < a$. So there exists $\delta > 0$ such that $a - \delta < x < a \Rightarrow f(x) < f(a)$ and $a < x < a + \delta \Rightarrow f(x) > f(a)$. It follows that f is strictly increasing in a neighbourhood where the derivative is strictly positive. We can apply the same argument to $-f$ to show

Proposition 5.2.1 *A differentiable function is strictly increasing in an open interval where the derivative is strictly positive (everywhere in that open interval) and is strictly decreasing if the derivative is strictly negative.*

As you well remember from first year, a *relative* or *local maximum* of a function f is a point a such that $f(x) \leq f(a)$ for all x in a neighbourhood of a. Similarly, a *relative* or *local minimum* is a point b where $f(x) \geq f(b)$ for all x in a neighbourhood of b. We say a is a *strict relative maximum* (resp. minimum) if we have strict inequalities in the above definition (except of course at $x = a$!). A *relative extremum* is then the collective term used for either a relative maximum or a relative minimum. Since f can neither be strictly increasing nor strictly decreasing near a relative extremum, we obtain as a corollary of the above proposition the following familiar fact:

Corollary 5.2.1 *The derivative of a differentiable function vanishes at relative extrema.*

In fact we can say more if we assume that not only f but also it's derivative f' is differentiable near a relative extremum a

Corollary 5.2.2 *Suppose both f and f' are differentiable in a neighbourhood of a and assume $f'(a) = 0$. Then*

(i) $f''(a) > 0 \ \Rightarrow\ a$ is a strict relative minimum.

(ii) $f''(a) < 0 \ \Rightarrow\ a$ is a strict relative maximum.

Proof: Just apply the proposition to the derivative of f.

QED

Another related important result is the behaviour of a differentiable function at a point where the derivative does not vanish. We will prove the general form of this result, called the Inverse Function Theorem in $\mathbb{R}^n$ in the next section, but for now, let me derive a simple formula for the derivative of the inverse function (not the reciprocal!) of a continuously differentiable function at a point where the derivative is non-zero (so it is either positive or negative).

Definition 5.2.1 *f is asaid to be continuously differentiable or C^1 in an open interval if f is differentaible at all points of the interval and the derivative f' is continuous in that interval.*

Proposition 5.2.2 *Suppose that f is continuously differentiable in a neighbourhood of a and that $f'(a) \neq 0$. Then f is bijective in a neighbourhood of a and the inverse function $g = f^{-1}$ is differentiable at $b = g(a)$ with derivative given by*

$$g'(b) = \frac{1}{f'(a)}$$

Proof: If we can assume that the inverse function is differentiable at b then the formula in fact follows from the chain rule. We know that f is strictly increasing (let's say) in a neighbourhood of a so it maps a small interval U around a bijectively onto a small interval V around $b = f(a)$. To prove that the inverse function is differentiable, let $y = f(x)$ so that $x = g(y)$ for $x \in U$, $y \in V$. Now

$$\frac{g(y) - g(b)}{y - b} = \frac{x - a}{f(x) - f(a)} = \Big(\frac{f(x) - f(a)}{x - a}\Big)^{-1}$$

provided $x \neq a$ so that $f(x) \neq f(a)$. (This is where we used the fact that $f'(a) \neq 0$). Now take the limit on both sides and use the fact that $x \to a \Leftrightarrow y \to b$ since f is continuous.

QED

5.3 Some global properties of the derivative

Proposition 5.3.1 Rolle's Theorem

Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function which is differentiable at all points of the open interval (a, b) and assume that $f(a) = f(b)$. Then $\exists\, x \in (a, b)$ where $f'(x) = 0$.

Proof Since $[a, b]$ is compact and f is continuous on a compact interval $[a, b]$, f attains both its absolute maximum and its absolute minimum value at some point in $[a, b]$. If this point is either a or b, then the function is constant on the whole interval because of our assumption $f(a) = f(b)$. A constant function has derivative zero everywhere. If f is not constant either the absolute minimum or the absolute minimum occurs at an interior point in (a, b), but there the function is differentiable and by the corollary from the last section the derivative vanishes (i.e. it is zero) at that point.

QED

By applying Rolle's Theorem to the function

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a)$$

which obviously satisfies $g(a) = g(b)(= f(a))$ and the other continuity and differentiability requirements of Rolle's theorem, provided f does, we easily derive the following major result of this section:

Theorem 5.3.1 Mean Value Theorem

Suppose that $f : [a, b] \to \mathbb{R}$ *is a continuous function which is differentiable at all points of the open interval* (a, b) *. Then* $\exists\, x \in (a, b)$ *where*

$$f'(x) = \frac{f(b) - f(a)}{b - a}$$

.

The following (very important) corollaries are immediate consequences of the Mean Value Theorem: (I remember telling you how they follow from the mean value theorem in Math 1A03!)

Corollary 5.3.1 *Suppose* f *is continuous on* $[a, b]$ *and differentiable in* (a, b)*. If* $f'(x) = 0$ *forall* $x \in (a, b)$ *, then* f *is constant on* $[a, b]$ *.*

Corollary 5.3.2 *Suppose* f *is continuous on* $[a, b]$ *and differentiable in* (a, b)*. If* $f'(x) > 0$ *forall* $x \in (a, b)$ *, then* f *is strictly increasing on* $[a, b]$ *, i.e.* $\forall x_1, x_2 \in [a, b]\ x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$ *.*

Corollary 5.3.3 *Suppose* f *is continuous on* $[a, b]$ *and differentiable in* (a, b)*. If* $f'(x) < 0$ *forall* $x \in (a, b)$ *, then* f *is strictly decreasing on* $[a, b]$ *.*

Obviously if we weaken the assumption to $f(x) \geq 0$ (resp. $f(x) \leq 0$), we have to drop the word "strictly". We can only claim that the function is non-decreasing (resp. non-increasing). We also note that in contrast to the last section, the statements here are global in the sense that it applies to the whole interval $[a, b]$. We of course also assume more. The condition on the derivative is not just at a point! Applying this kind of argument to the derivative of f we have the following global result about convexity, but first the definition:

Definition 5.3.1 *A real valued function f (not necessarily differentiable) is said to be* convex *on $[a,b]$ if the following inequality holds:*

$$f\big((1-t)\,x\,+\,t\,y\big) \leq (1-t)\,f(x)\,+\,t\,f(y)$$

for all $t \in [0,1]$ and $x, y \in [a,b]$.

If z is the point $(1-t)\,x + t\,y$, then $t = \frac{z-x}{y-x}$ and $1 - t = \frac{y-z}{y-x}$. So convexity implies

(i) $f(z) - f(x) \leq t\,\big(f(y) - f(x)\big)$

(ii) $f(z) - f(y) \leq (1-t)\,\big(f(x) - f(y)\big)$

This proves

Proposition 5.3.2 *If f is convex on $[a,b]$, then for any three points $x < z < y$ in the interval $[a,b]$,*

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(z)}{y - z}$$

Proof: Just put $t = \frac{z-x}{y-x}$ and $1 - t = \frac{y-z}{y-x}$ in (i) and (ii) above.

Now by the mean value theorem: $\frac{f(z)-f(x)}{z-x} = f\,'(p)$ and $\frac{f(y)-f(z)}{y-z} = f\,'(q)$ for some $p \in (x,z)$ and $q \in (z,y)$. So if $f\,'(p) \leq f\,'(q)$ for all $p < q$ in $[a,b]$, f would be convex. We have thus proved the following propositions and its corollary:

Proposition 5.3.3 *Suppose f is continuous on $[a,b]$ and differentiable in (a,b). If $f\,'$ is monotonically increasing in (a,b) , then f is convex on $[a,b]$.*

Corollary 5.3.4 *Suppose f is continuous on $[a,b]$ and twice differentiable in (a,b). If $f\,''(\xi) \geq 0$ for all $\xi \in (a,b)$, then f is convex on $[a.b]$.*

5.3.1 The Logarithm

The natural logarithm is the inverse function of the exponential function. As we have seen above, the exponential functions $exp(x) = e^x$ has derivative $exp\,'(x) = exp(x)$, which is always positive since it is a square ($e^x = e^{\frac{x}{2}}\, e^{\frac{x}{2}} = \left(e^{\frac{x}{2}}\right)^2 > 0$). Moreover, since $\lim_{x\to+\infty} e^x = +\infty$ (because $e^x > 1 + x$ for all $x > 0$ by definition) and $\lim_{x\to-\infty} e^x = 0$ (because $e^{-x} = (e^x)^{-1}$), $exp : \mathbb{R} \to (0,\infty)$ is a strictly increasing function bijective map. Its inverse is called the natural logarithm.

Definition 5.3.2 *The* (natural) logarithm $\log : (0,\infty) \to \mathbb{R}$ *is defined to be the inverse funsction of the exponential function. It therefore satisfies* $\log(e^x) = x$ *for all* $x \in \mathbb{R}$ *and* $e^{\log(y)} = y$ *for all* $y > 0$ *.*

By what we know about inverse functions in general, it now follows that $\log : (0,\infty) \to \mathbb{R}$ is a strictly increasing bijective and differentiable map with derivative given by:

$$\log\,'(y) = \frac{1}{y}$$

5.4 Differentiation in $\mathbb{R}^n$

The derivative of a function of several variables is a linear map, so we will use the notation $L(\mathbb{R}^n;\mathbb{R}^m)$ for the vector space of all linear maps from R^n to $\mathbb{R}^m$.

Definition 5.4.1 *Let f be a R^m-valued defined in an open neighbourhood U of $a \in \mathbb{R}^n$. f is said to be* differentiable *at a with derivative $f\,'(a) \in L(\mathbb{R}^n;\mathbb{R}^m)$ if and only if the following holds:* $\forall\,\epsilon > 0, \exists\,\delta > 0$ *such that* $\forall x \in U\ ,\ \|x - a\| < \delta \ \Rightarrow\ \|f(x) - f(a) - f\,'(a)(x - a)\| < \epsilon\|x - a\|$

In other words, f is differentiable at a point a if there is a linear map $A = f\,'(a)$ called the derivative of f at a such that the error function $r(x) = f(x) - (f(a) + A(x - a))$, that remains after approximating f near a by the linear (affine) function $l(x) = f(a) + A(x - a)$ is in $o(\|x - a\|)$, i.e., it satisfies:

$$\lim_{x\to a,\, x\neq a} \frac{r(x)}{\|x - a\|} = 0$$

Differentiability obviously implies continuity as in the scalar case $n = 1$. If f is differentiable at each point of an open set U, then we say that f is differentiable in U. f is said to be C^1 on U, if the map $a \mapsto f'(a)$ is continuous in U.

We have the usual arithmetical rules for the derivative as in the case of $n = 1$:

Proposition 5.4.1 *If f and g are differentiable at a, then so are $f+g, f-g, < f, g >$, wherever they are defined. (Here $<, >$ is the scalar product). The following formulas hold:*

(i) $(f \pm g)'(a) = f'(a) \pm g'(a)$

(ii) (Product Rule) $< f, g >'(a)(v) =< f'(a)(v), g(a) > + < f(a), g'(a)(v) >$

Proposition 5.4.2 The Chain Rule

If g is differentiable at a and f is differentiable at $g(a)$, then the composite $f \circ g$ (if defined) is differentiable at a with derivative given by:

$$(f \circ g)'(a) = f'(g(a)) \circ g'(a)$$

All the proofs are appropriate vectorial modifications (like using $\|, \|$ instead of $|, |$) of the proofs I gave you for the scalar case and does not involve any fundamentally new ideas, so I will skip them. Instead I will now state and prove the Inverse Function Theorem and its Corollary the Implicit Function Theorem.

5.4.1 Inverse Function Theorem

Theorem 5.4.1 *Let $f : U \to \mathbb{R}^n$ be a C^1 map, where U is an open set in $\mathbb{R}^n$, containing a. If $A = f'(a) = df_a$ is invertible, then there exists a neighbourhood V of a such that f restricted to V maps V bijectively onto $W = f(V)$. Moreover the inverse function $g = f^{-1}$ is continuously differentiable on W with derivative given by the Chain Rule (so that $dg_{f(x)} = (df_x)^{-1}$).*

Proof:

Step 0.
By two translations, we may assume that $a = 0$ and $f(a) = 0$. Furthermore by composing f with a linear map $A^{-1} = (df_0)^{-1}$ we may also assume that $df_0 = I$, where I is the identity matrix. Now let $h(x) = f(x) - x$. Then $dh_0 = 0$.

Step 1.
Since f and hence h are C^1, $\exists\ r > 0$, s.t. $\|dh_x\| < \frac{1}{2}$, for all $x \in B_{3r}(0)$, where we define $\|A\|$ to be the maximum length of all column vectors of A. Applying the Mean Value Theorem to the components of h we find that for every $x_1, x_2 \in B_{3r}(0)$, $|f(x_1) - x_1 - f(x_2) + x_2| = |h(x_1) - h(x_2)| < \frac{1}{2}|x_1 - x_2|$. So if $f(x_1) = f(x_2)$, then $x_1 = x_2$. Therefore f is injective on $B_{3r}(0)$

Step 2.
To prove surjectivity, we pick a point y in $B_r(0)$ and define (inductively and constructively!) a sequence as follows:

$$x_0 = 0\,, \qquad x_1 = y\,, \qquad x_{k+1} = y - h(x_k) = y - f(x_k) + x_k$$

Step 3.
As in Step 1, we have: $|x_{k+1} - x_k| = |h(x_k) - h(x_{k-1})| < \frac{1}{2}|x_k - x_{k-1}$, provided $x_k\,, x_{k+1} \in B_{2r}$ By induction we show that $x_n \in B_{2r}(0)$, and hence that $|x_{n+1} - x_n| < \frac{1}{2}|x_n - x_{n-1}|$ for all n.
For $n = 0$: $x_0 = 0, x_1 = y \in B_r(0) \subset B_{2r}(0)$ and since $\|dh\| < \frac{1}{2}$ in $B_{3r}(0)$, $|x_2 - x_1| = |h(x_1) - h(x_0)| < \frac{1}{2}|x_1 - x_0|$.
Assume $x_k \in B_{2r}(0)$, for all $k \leq n$. Then $|x_{k+1} - x_k| < \frac{1}{2}|x_k - x_{k-1}| < \cdots < 2^{-k}|x_1 - x_0|$ for all $k \leq n$ and hence by summing up the geometric series we obtain: $|x_{n+1}| = |x_{n+1} - x_0| \leq \sum_{k=0}^{n} |x_{k+1} - x_k| < 2|x_1 - x_0| < 2r$.

Step 4.
We claim now that x_k is a Cauchy sequence and hence converges (This was actually proved in Problem # 1 of your Assignment #2 !). By the same argument as in Step 3, $|x_{n+k} - x_n| \leq 2^{-n}|x_1 - x_0| = |y|$, so x_n is a Cauchy sequence and it converges to a limit x, which is in B_{2r}, since $|x_n - x_0| < 2|x_1 - x_0| = 2|y|$ for all n.

Step 5.
We have thus proved that f restricted to the inverse image of $B_r(0)$ is a bijective map. The differentiability of the inverse function follows in a similar fashion as in the case of a single variable and the chain rule then gives the formula for the derivative of the inverse function.

Corollary 5.4.1 *(*Implicit Function Theorem*) Suppose* $f = (f_1, \dots, f_m) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ *is continuously differentiable in an open neighbourhood of* (a, b) *and let* $f(a, b) = 0$ *. Let* A *be the* $m \times m$ *matrix* $\left(\frac{\partial}{\partial x_{n+k}} f_k(a, b)\right)_{k=1,\dots,m}$ *. If* $\det(A) \neq 0$ *, then there exists an open neighbourhood* U *of* a *and an open neighbourhood* V *of* b *such that* $\forall\, x \in A$ *, there exists a unique* $y = g(x) \in B$ *satisfying the implicit equation:* $f(x, g(x)) = 0$ *. Moreover* g *is differentiable in* A *.*

(*Sketch of proof*): Apply the inverse function theorem to the function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$, defined by $F(x, y) = (x, f(x, y))$. The inverse of F is of the form $F^{-1}(x, y) = (x, G(x, y))$. Then $g(x) = G(x, 0)$ will satisfy the equation $f(x, g(x)) = 0$.

5.5 Exercises

1. Show that $f : (0, \infty) \to \mathbb{R}$ defined by $f(x) = x^n e^x$; $n > 0$ attains its absolute maximum value at the point $x = n$.

2. Let $f : \mathbb{R} \to \mathbb{R}$ satisfy the inequality:

$$|f(x) - f(y)| \leq C(x-y)^2$$

for all $x, y \in \mathbb{R}$, where C is a fixed positive constant. Prove that f is a constant function.

3. Suppose f is a differentiable real-valued function defined on $\mathbb{R}$ satisfying $|f'(x)| \leq 10$ for all $x \in \mathbb{R}$. Show that the function $F(x) = x + \frac{1}{100} f(x)$ is an injective (one-to-one) map.

4. Let $f_{(n)}$ denote the nth iterate of f, i.e. $f_{(n)} = f \circ f \circ \ldots \circ f$ (n-times). Express the derivative of $f_{(n)}$ in terms of f' and prove that if $m \leq |f'| \leq M$, then $m^n \leq |f_{(n)}{}'| \leq M^n$.

5. Let $p_1, p_2, \ldots, p_n$ be positive numbers satisfying $\sum_{k=1}^n p_k = 1$ and let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Prove *Jensen's inequality*:

$$f\Big(\sum_{k=1}^n p_k\, x_k\Big) \leq \sum_{k=1}^n p_k\, f(x_k)$$

for all real numbers $x_1, x_2, \ldots, x_n$.

6.

Let $f : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function with $f(0) = 0, f(1) = 1$, and with $f'(0) = f'(1) = 0$. Prove that $\exists\, x \in [0,1]$ with $|f''(x)| \geq 4|$.
In more physical terms; a particle which travels a unit distance in unit time and starts and ends with zero velocity (like landing on Mars!) has at least at some time an acceleration ≥ 4 (in absolute value).

7. Let $f : [0,1] \to [0,1]$ be continuous on $[0,1]$ and differentiable in the interior $(0,1)$ with $f(0) = 0$. Assume that $|f'(x)| \leq 10|f(x)|$ for all $x \in (0,1)$. Prove that $f(x) = 0$ for all $x \in [0,1]$.

8. Let $f : \mathbb{R} \to \mathbb{R}$ be a strictly increasing and convex function which is three times differentiable. Assume that: $f(0) = 0$. Starting with an initial value $x_1 > 0$, we now define inductively a sequence by the formula (*Newton's method*):

$$x_{n+1} = x_n - \frac{f(x_n}{f'(x_n)}$$

Prove that $\lim_{n\to\infty} x_n = 0$.

9. The Legendre polynomial of order n is defined by:

$$P_n(x) = \frac{1}{2^n\, n!}\frac{d^n}{dx^n}\Big((x^2-1)^n\Big)$$

Prove that:
(i) P_n has exactly n distinct zeros in $(-1, +1)$
(ii) P_n satisfies the differential equation:

$$(1-x^2)P_n{}''(x) - 2x\,P_n{}'(x) + n(n+1) = 0$$

5.5.1 Hints and short solutions

6. First we prove a little lemma:
If a twice differentiable function h satisfies:$h(0) = h'(0) = 0$ and $h''(t) < 4$, for all $t \in [0,1]$, then $h(\frac{1}{2}) < \frac{1}{2}$.
Proof of Lemma: Since $h''(t) < 4$, the function $h'(t) - 4t$ is strictly decreasing and hence $h'(t) - 4t < h'(0) - 0 = 0$ for $0 < t < 1$. This implies that $h(t) - 2t^2$ is also strictly decreasing and hence $h(\frac{1}{2}) - \frac{1}{2} < h(0) - 0 = 0$. q.e.d.
If $|f''(t)| < 4$ for all $t \in [0,1]$, then both $f(t)$ and $g(t) = f(1) - f(1-t)$ satisfy the assumptions of the lemma and hence $f(\frac{1}{2}) < \frac{1}{2}$ and $f(1) - f(\frac{1}{2}) < \frac{1}{2}$ contradicting the fact that $f(1) = 1$!

Chapter 6

Integration

6.1 Definition of the Riemann Integral

Let f be a bounded (not necessarily continuous) real valued function defined on a closed (and bounded) interval $[a, b] \subset \mathbb{R}$ or more generally, on a multidimensional rectangle $R = \prod_{i=1}^{n}[a_k, b_k] \subset \mathbb{R}^n$. We denote the length (or more generally the n-dimensional volume) of R by $|R|$. ($|R|$ is just the product of all the side lengths of R and is always positive). A *partition* $\mathcal{R}$ of R is a finite collection of subintervals (subrectangles) $\{R_i \subset R\}_{i=1,\cdots,N}$, whose interiors are all mutually disjoint and whose union is R. A *refinement* of a partition $\mathcal{R}$ is a partition $\mathcal{S}$ such that each subrectangle S_j of $\mathcal{S}$ is contained in a subrectangle R_i of $\mathcal{R}$. For each interval (or rectangle) R_i, we set $m(f, R_i) = inf(f(R_i))$ and $M(f, R_i) = sup(f(R_i))$. These numbers are finite since f is bounded. Now define:

$US(f, \mathcal{R}) = \sum_{i=1}^{N} M(f, R_i)|R_i|$

$LS(f, \mathcal{R}) = \sum_{i=1}^{N} m(f, R_i)|R_i|$

$UI(f) = inf\{US(f, \mathcal{R}) \,|\, \mathcal{R} \text{ is a partition of } R\}$ and

$LI(f) = sup\{LS(f, \mathcal{R}) \,|\, \mathcal{R} \text{ is a partition of } R\}$

These are all (finite) real numbers for any bounded function defined on a closed and bounded rectangle. US stands for upper sum, LS for lower sum, UI for upper integral and LI for lower integral. It is obvious from the definitions that $US \geq LS$ for any partition and that $LS(f, \mathcal{R}) \leq LS(f, \mathcal{S})$ and $US(f, \mathcal{R}) \geq US(f, \mathcal{S})$ if $\mathcal{S}$ is a refinement of $\mathcal{R}$, since $inf(A) \leq inf(B)$ and $sup(A) \geq sup(B)$ if $B \subset A$ and volume is an additive function on rectangles. Given two partitions $\mathcal{R}$ and $\mathcal{S}$, we can define their join (or common refinement) $\mathcal{R} \vee \mathcal{S}$ to be the partition consisting of all intersections $R_i \cap S_j$. $\mathcal{R} \vee \mathcal{S}$ is a refinement of both $\mathcal{R}$ and $\mathcal{S}$. By looking at the upper and lower sums of this common refinement we arrive at the following:

Proposition 6.1.1 *$UI(f) \geq LI(f)$ for any bounded function f defined on a bounded rectangle.*

Definition 6.1.1
We say that f is Riemann-integrable *on R if $UI(f) = LI(f)$ and the common value is defined to be the integral of f on R:*

$$\int_R f = UI(f) = LI(f)$$

The function defined by $f(r) = 1$ for $r \in \mathbb{Q}$ and $f(x) = 0$ is discontinuous at every point and is not Riemann-integrable on any closed interval, because the upper sums are always the length of the interval and all the lower sums are zero. Next year in Math 3A03 you will learn that this function is integrable in the more general Lebesgue sense and that the integral is zero, because although $\mathbb{Q}$ is dense in $\mathbb{R}$, it is still a negligible set from the point of view of Lebesgue measure.

We will show that all continuous functions are integrable on a closed bounded interval.

Proposition 6.1.2 *If f is continuous on a closed and bounded rectangle $R \subset \mathbb{R}^n$, then f is integrable.*

Proof: We know that f is bounded and uniformly continuous on the compact set R. Let $\epsilon > 0$ be given. Then $\exists \delta > 0$ such that $||x - y|| < \delta \Rightarrow |f(x) - f(y)| < \frac{\epsilon}{|R|}$. Now let $\mathcal{P}$ be a partition where every subrectangle has all side lengths less than $\frac{\delta}{10\sqrt{n}}$ so that any two points in a subrectangle are at a distance less than δ apart. This means that the minimum and maximum values taken by f on each of these subrectangles are less than $\frac{\epsilon}{|R|}$ apart, so that the difference between the upper sum and lower sum for this partition is less than ϵ which was arbitrary, proving that the upper and the lower integrals coincide.

QED

Now let f be a bounded function defined on an arbitrary bounded set $A \subset \mathbb{R}^n$. We extend the definition of f to a rectangle R which contains

A by simply setting $f(x) = 0$ for all points $x \notin A$. We then say that f is integrable on A if this exetended function is integrable on R and define the integral to be the integral over R of the extended function. The integral, if it exists, does not depend on the extension. However, there are complicated bounded sets A, for which even constant functions are not integrable.

6.2 Basic properties of the Integral

The three main arithmetic properties of the integral are the following:

Proposition 6.2.1 *If f and g are (Riemann-)integrable functions on a bounded set A, then*

(Linearity) $\int_A (c_1 f + c_2 g) = c_1 \int_A f + c_2 \int_A g$

(Monotonicity) If $f(x) \geq 0$ for all $x \in A$, then $\int_A f \geq 0$. In fact if $f(x) > 0$ for all $x \in A$, then $\int_A f > 0$

Proposition 6.2.2 *(Additivity) If A_1 and A_2 are disjoint sets, on which f is integrable, then $\int_{A_1 \cup A_2} f = \int_{A_1} f + \int_{A_2} f$*

We will leave the proof of these simple properties as an exercise for you.

Monotonicity is an important property and we state some immediate consequences:

M1 If $f(x) \geq g(x)$ for all $x \in A$, then $\int_A f \geq \int_A g$.

M2 If $m \leq f(x) \leq M$ for all $x \in A$, then $m|A| \leq \int_A f \leq M|A|$.

M3 $\left| \int f(x) \right| \leq \int |f|$ provided $|f|$ is integrable.

Another property that follows immediately from the above properties is the following Mean Value Theorem for integrals.

Theorem 6.2.1 *If f is continuous on $[a, b]$ then there exists a value $c \in (a, b)$ such that*

$$\int_a^b f(t)\, dt = f(c)\,(b-a)$$

Proof: Let $m = f(p)$ be the absolute minimum value of f on $[a, b]$ and $M = f(q)$ be its absolute maximum value. Then by monotonicity: $m = f(p) \leq \frac{1}{b-a} \int_a^b f(t)\, dt \leq M = f(q)$. Now since f is continuous, by the Intermediate Value Theorem, there exists a point c (between p and q) such that $f(c) = \frac{1}{b-a} \int_a^b f(t)\, dt$

QED

6.3 Fundamental Theorem of Calculus

We state and prove now the fundamental theorem of calculus, which relates integration and differentation for functions of a single variable. The formulation of the corresponding theorems in higher dimensions (for example, Stokes' theorem that you learned last term in in Math 2A03) are a lot more elaborate although fundamentally the proofs are based on the following one dimensional fundamental theorem of calculus.

Theorem 6.3.1 *If f is continuous on $[a, b]$ then the function defined by*

$$F(x) = \int_a^x f(t)\, dt$$

for $x \in [a, b]$ is continuous on $[a, b]$ and differentiable in (a, b) with derivative given by $F'(x) = f(x)$.

Proof: Let $x \in (a, b)$. For $h \neq 0$ sufficiently small we have, by the additivity property and the mean value theorem for integrals:

$$F(x + h) - F(x) = \int_x^{x+h} f = h\, f(c_h)$$

for some c_h , where $c_h \in (x, x + h)$, if $h > 0$ and $c_h \in (x + h, x)$ if $h < 0$. Now since f is continuous $f(c_h) \to f(x)$ as $h \to 0$. Therefore

$$lim_{h \to 0} \frac{1}{h} (F(x + h) - F(x)) = f(x)$$

It is easy to see from the estimate above that not only is F continuous on $[a, b]$, but in fact, it is Lipschitz continuous.

QED

The function $F(x) = \int_a^x f(t)\,dt$ is then what is known as an "antiderivative" in first year calculus, since it satisfies $F'(x) = f(x)$. It is also sometimes called the indefinite integral. Two antiderivatives differ by a constant on any connected interval. This follows from the mean value theorem for derivatives, since two antiderivatives have the same derivative! Perhaps a more familiar form of the fundamental theorem that you remember from first year calculus is the following:

Corollary 6.3.1 *Let $F : [a,b] \to \mathbb{R}$ be an antiderivative for $f : [a,b] \to \mathbb{R}$, in the sense that F is continuous on $[a,b]$, differentiable in (a,b) with $F'(x) = f(x)$. Then*

$$\int_a^b f(t)\,dt = F(b) - F(a)$$

6.4 Improper Integrals

Improper integrals are integrals for unbounded sets and/or unbounded functions. We will restrict ourselves in these notes to the real line.

Let (a,b) be an open interval in $\mathbb{R}$. To deal with unbounded intervals, we will allow a to be $-\infty$ and b to be ∞. Let $f : (a,b) \to \mathbb{R}$ be a function not necessarily bounded or continuous. Then:

Definition 6.4.1 *f is said to be integrable on (a,b) if f is* locally-integrable*, i.e. f is integrable on every closed and bounded subinterval of (a,b) and if*

$$\int_a^b f = \lim_{c \to a^+} \left(\lim_{d \to b^-} \int_c^d f \right)$$

exists (as a finite limit).

If the limit exists we say that the improper integral is convergent. The order of the limits in the definition actually does not matter.

Here are some familiar examples from first year calculus:

1. $\int_0^1 x^{-p} dx$ is convergent iff $p < 1$.
2. $\int_1^\infty x^{-p} dx$ is convergent iff $p > 1$.

3. $\int_0^\infty e^{-x} dx$ converges to 1.

4. $\int_{-\infty}^\infty e^{-x^2} dx$ converges to $\sqrt{\pi}$.

In order to test other improper integrals we often use the following:

Proposition 6.4.1 (Comparison Test)

Suppose f and g are locally integrable on (a,b). If $0 \le f(x) \le g(x)$ for all $x \in (a,b)$, and if g is integrable on (a,b)then f is also integrable on (a,b) and we have $\int_a^b f \le \int_a^b g$.

As a corollary it follows that if $|f|$ is integrable it follows that f is integrable but the converse is not . Functions f with $|f|$ integrable are called *absolutely integrable.* Otherwise (if f is integrable without $|f|$ being integrable) we say that f is *conditionally integrable.*

6.4.1 The Gamma Function

The gamma function is the continuous extension of the familiar factorial function $n!$, which is defined for positive integers only.

Definition 6.4.2

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt \quad (x > 0)$$

The improper integral is easily seen to be convergent at both end points. At $t = \infty$, the integral converges, by comparison with the convergent integral $\int_1^\infty t^{-2} \, dt$, since $\lim_{t\to\infty} t^2 \, t^{x-1} e^{-t} = 0$ for all $x > 0$.
At $t = 0$, the integral converges because $t^{x-1} e^{-t} \le t^{x-1}$ for $t > 0$ and the integral $\int_0^1 t^{x-1}$ is convergent since $x > 0$.
Moreover, the Gamma function is always positive and $\Gamma(1) = 1$, since $\int_0^\infty e^{-t} \, dt = 1$.

Since $\frac{d}{dt}\left(t^x \, e^{-t}\right) = x \, t^{x-1} \, e^{-t} - t^x \, e^{-t}$, it follows from the fundamental theorem of calculus that:

$$x \int_a^b t^{x-1} \, e^{-t} \, dt - \int_a^b t^x \, e^{-t} \, dt = b^x \, e^{-b} - a^x \, e^{-a}$$

Letting $a \to 0_+$ and $b \to +\infty$ to compute the indefinite integral we get $\lim_{a\to 0_+} a^x \, e^{-a} = 0$ $\lim_{b\to+\infty} b^x \, e^{-b} = 0$ (for $x > 0$). This proves the following fundamental functional equation satisfied by the Gamma function, which makes it the continuous interpolation of the factorial!

Proposition 6.4.2 *The gamma function* $\Gamma(x)$ *is positive and satisfies the functional equation:*

$$\Gamma(x+1) = x\,\Gamma(x)$$

We therefore have $n! = \Gamma(n+1)$ for $n \in N$. Another important fact is that

$$\Gamma(\frac{1}{2}) = \int_0^{\infty} \sqrt{t}\, e^{-t}\, dt = 2\int_0^{+\infty} e^{-x^2}\, dx \;= \sqrt{\pi}$$

which follows from the substitution $t = x^2$.

6.4.2 Stirling's Formula

If a_n and b_n are two sequences of real numbers we will use the notation $a_n \sim b_n$ to mean $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Theorem 6.4.1

$$n! \;\sim \sqrt{2\pi}\, n^n\, n^{\frac{1}{2}}\, e^{-n}$$

Proof: Let

$$d_n = \log(n!) + n - (n + \frac{1}{2})\log(n)$$

Then

$$\begin{aligned} d_n - d_{n+1} &= -\log(n+1) - 1 - (n+\frac{1}{2})\log(n) + (n+\frac{3}{2})\log(n+1) \\ &= (n+\frac{1}{2})\log(\frac{n+1}{n}) - 1 \\ &= \frac{1}{2x}\log(\frac{1+x}{1-x}) - 1 \end{aligned}$$

where $x = (2n+1)^{-1} > 0$.

By a power series expansion for the logarithm we obtain (for sufficiently small and positive x):

$$
\begin{aligned}
\frac{1}{2x}\log\left(\frac{1+x}{1-x}\right)-1 &= \frac{1}{x}\left(x+\frac{x^3}{3}+\frac{x^5}{5}+\cdots\right)-1 \\
&= \frac{x^2}{3}+\frac{x^4}{5}+\cdots \\
&< \frac{x^2}{3}(1+x^2+x^4+\cdots) \\
&= \frac{x^2}{3(1-x^2)} \\
&= \frac{1}{12}\left(\frac{1}{n}-\frac{1}{n+1}\right) \qquad \text{since } x=\frac{1}{2n+1}
\end{aligned}
$$

Therefore

$$0 < d_n - d_{n+1} < \frac{1}{12}\left(\frac{1}{n}-\frac{1}{n+1}\right)$$

This simultaneously shows two important properties of the sequence d_n :

(i) d_n is strictly decreasing

(ii) $(d_n - \frac{1}{12n})$ is strictly increasing

It follows from (ii) that d_n is bounded from below and so by (i) is a convergent sequence with limit $\lim d_n = d$.

We claim that $e^d = \sqrt{2\pi}$.
The proof relies on the famous product formula of John Wallis (1616-1703):

$$\frac{2.2.4.4.6.6.\ldots(2n)(2n)}{1.3.3.5.5.7.\ldots(2n-1)(2n+1)} \to \frac{\pi}{2}$$

as $n \to \infty$

(*The derivation of this formula is one of the exercise problems at the end of this chapter*)
Taking the square root we get

$$\frac{2.4.6.\ldots 2n}{3.5.7.\ldots(2n-1)}\frac{1}{\sqrt{2n+1}} \to \sqrt{\frac{\pi}{2}}$$

but the left hand side is

$$\frac{2^2.4^2.6^2.\ldots(2n)^2}{1.2.3.4.5.6.\ldots(2n-1)(2n)}\frac{1}{\sqrt{2n+1}} = \frac{2^{2n}\,(n!)^2}{(2n)!}\frac{1}{\sqrt{2n}}\frac{1}{\sqrt{1+(2n)^{-1}}}$$

Taking logarithms we find

$$2\log(n!) - \log((2n)!) + 2n\log(2) - \frac{1}{2}\log(n) \;\to\; \log(\sqrt{\pi})$$

On the other hand, by the definition of d_n:

$$e^{2d_n - d_{2n}} = 2\log(n!) - \log((2n)!) + 2n\log(2) - \frac{1}{2}(\log(n) + \log(2))$$

Therefore $d = \lim(2d_n - d_{2n}) = \log(\sqrt{2\pi})$

6.4.3 The Euler-Maclaurin Summation Formula

Let $P(t)$ be a polynomial of degree n and $f(x)$ be a function which is continuously differentiable $(n+1)$-times for all x in an open interval containing a point$a$. Then by differentation we find for $0 \le t \le 1$:

$\frac{d}{dt}\sum_{k=1}^{n}(-1)^k P^{(n-k)}(t) f^{(k)}(a + t(x-a))(x-a)^k$
$= -P^{(n)} f\,'(a + t(x-a))(x-a) + (-1)^n P(t) f^{(n+1)}(a + t(x-a))(x-a)^{n+1}$.

where for a function F, $F^{(k)}$ denotes the k-th derivative of F.
Since $P^{(n)}(t)$ is a constant $c = P^{(n)}$, we obtain by integrating from $t = 0$ to $t = 1$, the following identity due to Darboux:

Proposition 6.4.3 *(***Darboux Formula***)*

$P^{(n)}(0)\big(f(x) - f(a)\big)$
$+\sum_{k=1}^{n}(-1)^k\big(P^{(n-k)}(1) f^{(k)}(x) - P^{(n-k)}(0) f^{(k)}(a)\big)(x-a)^k$
$= (-1)^n (x-a)^{n+1} \int_0^1 P(t) f^{(n+1)}(a + t(x-a))\, dt$

This innocuous looking little formula is quite powerful! For example, if we use the polynomials: $P(t) = (t-1)^n$, we obtain Taylor's formula that you all learn in first year:

Proposition 6.4.4 *(***Taylor-Maclaurin's Formula***)*

$$f(x) - f(a) = \sum_{k=1}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + R_{n+1}$$

where the remainder term is given by:

$R_{n+1} = (-1)^{n+1}(x-a)^{n+1} \int_0^1 (1-t)^n f^{(n+1)}(a + t(x-a))\, dt$

Functions for wchich the Taylor-Maclaurin formula remains true (in a neighbourhood of a) when $n \to \infty$ are called *analytic.*

To obtain the Euler-Maclaurin summation formula we first introduce a very important sequence of rational numbers named after Bernoulli. The Bernoulli numbers B_{2j} , are defined as coefficients in the power series expansion:

$$\frac{x}{e^x - 1} = 1 - \frac{x}{2} + \sum_{j=1}^{\infty} (-1)^{j-1} B_{2j} \frac{x^{2j}}{(2j)!}$$

They can also be defined through:

$$\frac{1}{2}\, x\, cot(\frac{x}{2}) = 1 - \sum_{j=1}^{\infty} \frac{1}{(2j)!} B_{2j}\, x^{2j}$$

(Maple or some other software will compute them for you: $B_2 = \frac{1}{6}, B_4 = \frac{1}{30}, B_6 = \frac{1}{42}, B_8 = \frac{1}{30}, B_{10} = \frac{5}{66}$ etc.)

These numbers are intimately related to a family of polynomials also named after Bernoulli. They are defined as the coefficients which you obtain when the function $x\frac{e^{tx}-1}{e^x-1}$ is expanded in a Taylor Maclaurin power series

$$x\frac{e^{tx} - 1}{e^x - 1} = \sum_{1}^{\infty} \beta_n(t) \frac{x^n}{n!}$$

The Taylor-Maclaurin summation formula is derived by using the Bernoulli polynomials in the Darboux formula. These polynomials satisfy the important recursive relation:

$$\beta_n(t+1) - \beta_n(t) = nt^{n-1}$$

This is easily checked from the definition of the polynomials by putting $t+1$ for t and then subtracting.

Multiplying the two powers series for $e^{xt} - 1$ and $\frac{t}{e^t-1}$, we find that these polynomials are related to the Bernoulli numbers as follows:

$$\beta_n(t) = t^n - \frac{1}{2} n\, t^{n-1} + \binom{n}{2} B_2\, t^{n-2} - \binom{n}{4} B_4\, t^{n-4} + \cdots$$

This implies:
$\beta^{(n-2j-1)}(0) = 0\,, \;\; \beta_n^{(n-2j)}(0) = (-1)^{j-1} \frac{n!}{(2j)!}\, B_{2j}\,,$

$\beta_n^{(n-1)}(0) = -\frac{1}{2}\, n!\,,\ \ \beta_n^{(n)}(0) = n!\,.$

Moreover by differentiating the recurrence relation $\beta_n(t+1) = \beta_n(t)$, we have: $\beta_n^{(n-k)}(1) = \beta_n^{(n-k)}(0)$ for all k. Putting these all together for β_{2n} in the Darboux formula we obtain:

$$f(x) - f(a) - \tfrac{1}{2}\big(f'(x) + f'(a)\big)(x-a)$$
$$+\textstyle\sum_{j=1}^{n-1} \frac{(-1)^(j-1)}{(2j)!} B_{2j} \left(f^{(2j)}(x) - f^{(2j)}(a)\right)(x-a)^{2j}$$
$$= R_{n+1} = \tfrac{1}{(2n)!}(x-a)^{2n+1} \textstyle\int_0^1 \beta_{2n}(t) f^{(2n+1)}(a + t(x-a))\, dt$$

Applying this formula to a derivative , i.e if we write f' instead of f and assuming that we can neglect the error term R_{n+1}, we obtain (by using the fundamental theorem of calculus!):

$$\textstyle\int_a^{a+h} f(t)\, dt = \frac{1}{2} h\big(f(a) + f(a+h)\big)$$
$$+\textstyle\sum_{j=1}^{n-1} \frac{(-1)^{j-1}}{(2j)!} B_{2j} h^{2j} \left(f^{(2j)}(a+h) - f^{(2j)}(a)\right)$$

where we have replaced x by $a+h$. Now for the special case of $a \in \mathbb{N}$ and $h = 1$, we can add up all these formulas to obtain the famous:

Theorem 6.4.2 *(***Euler-Maclaurin Summation Formula***)*

$$\sum_{k=m}^{n} f(k) - \int_m^n f(x)\, dx \;=\; \frac{1}{2}(f(m)+f(n)) - \sum_{j=1}^{\infty} \frac{(-1)^j}{(2j)!} B_{2j} \left(f^{(2j-1)}(m) - f^{(2j-1)}(n)\right)$$

6.5 Exercises

1. Suppose $f : [0,1] \to \mathbb{R}$ is integrable. Show that

(i) $exp\Big(\int_0^1 f(x)\,dx \Big) \le \int_0^1 exp(f(x))\,dx$

(ii) $\Big(\int_0^1 |f(x)|^p\,dx \Big)^{\frac{1}{p}} \le \int_0^1 |f(x)|\,dx$ for $0 < p < 1$

(iii) $\Big(\int_0^1 |f(x)|^p\,dx \Big)^{\frac{1}{p}} \le \int_0^1 |f(x)|\,dx$ for $p > 1$

2. Suppose $f : [0,1] \to \mathbb{R}$ is continuous. Show that as $N \to \infty$

$$\frac{1}{N} \sum_{k=1}^{N} f(\frac{k}{N}) \quad \to \quad \int_0^1 f$$

3. By using the sequence of partitions $0 < a < \cdots < a\,q^n = b$, where $q = \sqrt[n]{\frac{b}{a}}$, show that

$$\lim_{n\to\infty} n\Big(\sqrt[n]{\frac{b}{a}} - 1 \Big) = \int_a^b \frac{1}{x}\,dx = \log(\frac{b}{a})$$

4.

(i) Show that

$$\int_0^1 \log(1+x)\,dx = 2\,\log 2 - 1$$

(ii) Deduce that

$$\frac{1}{n} \log\Big(\frac{(2n)!}{n^n\,n!} \Big) \to \log\big(\frac{4}{e} \big)$$

as $n \to \infty$

5. Prove that the following limit exists:

$$\gamma = \lim_{n\to\infty} \Big(\sum_{k=1}^{n} \frac{1}{k} - \log(n) \Big)$$

The limit, denoted by $\gamma \approx 0.57721\ldots$, is called Euler's constant is closely related to the Gamma function, for example $\Gamma'(1) = -\gamma$. It is still unknown whether γ is irrational)

6. (i) Compute $S_n = \int_0^{\frac{\pi}{2}} \sin^{\frac{n}{2}}(x)\, dx$. (Hint: $n\, S_n = (n-1)\, S_{n-2}$).

(ii) Deduce Wallis' formula:

$$\frac{2.2.4.4.6.6.\ldots(2n)(2n)}{1.3.3.5.5.7.\ldots(2n-1)(2n+1)} \;\to\; \frac{\pi}{2}$$

7. Using an appropriate change of variables, derive the following alternate expressions for the Gamma function:

$$\Gamma(x) = 2 \int_0^{\infty} t^{2x-1}\, e^{-t^2}\, dt = \int_0^1 \left(\log(\frac{1}{t}\right)^{x-1} dt$$

8. Prove that the logarithm of the Gamma function is convex.

9. Find the first four non-vanishing terms of the Taylor-Maclaurin series for the following functions about the point $x = 0$:

(i) $f(x) = \int_0^x \log(1+t)\, dt$

(ii) $f(x) = \exp(\sin(x))$

(iii) $f(x) = \log(\frac{\sin(x)}{x})\;\; f(0) = 0$

(iv) $f(x) = x\, \cot(x)\;\; f(0) = 1$

6.5.1 Hints and short solutions

3(i) $\frac{d}{dx}\big((1+x)\log(1+x)-x\big)=\log(1+x)$. Therefore by the fundamental theorem of calculus: $\int_0^1 \log(1+x)\,dx = 2\,\log 2 - 1\ = \log\left(\frac{4}{e}\right)$, since $1\log 1 - 0 = 0$.

(ii) Choose the partition $0<\frac{1}{n}<\ldots<\frac{n-1}{n}<1$ and form the Riemann sum for the integral $\int_0^1 \log(1+x)\,dx$:
$\frac{1}{n}\sum_{k=1}^{n}\log(1+\frac{k}{n}) = \frac{1}{n}\,\log\left(\prod_{k=1}^{n}\frac{k+n}{n}\right)$, but $\prod_{k=1}^{n}\frac{k+n}{n}=\frac{(2n)!}{n!\,n^n}$ and the Riemann sum converges to the integral as $n\to\infty$. Therefore $\log\left(\frac{(2n)!}{n^n\,n!}\right)^{\frac{1}{n}}\to\log\left(\frac{4}{e}\right)$ as $n\to\infty$.

4. Let $\gamma_n = 1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\log(n)$. Then $\gamma_{n+1}-\gamma_n=\frac{1}{n+1}-\int_n^{n+1}\frac{dt}{t}>0$, since $\frac{1}{t}$ is a strictly decreasing function. This shows that γ_n is a monotonically decreasing sequence. Moreover the sequence is bounded from below by 0, because $\sum_{k=1}^{n-1}\frac{1}{k}$ is a upper Riemann sum for $\int_1^n\frac{dt}{t}=\log(n)$ and hence $d_n>\frac{1}{n}>0$. Therefore γ_n converges to a limiting value γ, because it is bounded from below by 0.

\# 5.

For $n\geq 2$, we have:

$$\begin{aligned}\frac{d}{dx}\big(\sin^{n-1}(x)\,\cos(x)\big) &= (n-1)\sin^{n-2}(x)\,\cos^2(x)-\sin^n(x)\\ &= (n-1)\sin^{n-2}(x)\big(1-\sin^2(x)\big)-\sin^n(x)\\ &= (n-1)\sin^{n-2}(x)-n\,\sin^n(x)\end{aligned}$$

Therefore by the fundamental theorem of calculus:

$$(n-1)\int_0^{\frac{\pi}{2}}\sin^{n-2}(x)\,dx = n\int_0^{\frac{\pi}{2}}\sin^n(x)$$

since $\sin(0)=\cos(\frac{\pi}{2})=0$. By induction, we obtain:

$$S_{2n}=\frac{(2n-1).(2n-3)\ldots 3.1}{(2n).(2n-2)\ldots 4.2}S_0=\frac{(2n-1).(2n-3)\ldots 3.1}{(2n).(2n-2)\ldots 4.2}\frac{\pi}{2}$$

$$S_{2n+1}=\frac{(2n).(2n-2)\ldots 4.2}{(2n+1).(2n-1)\ldots 5.3}I_1=\frac{(2n).(2n-2)\ldots 4.2}{(2n+1).(2n-1)\ldots 5.3}\,1$$

Since $0 \leq \sin x \leq 1$, for $x \in [0, \frac{\pi}{2}]$ we have $0 < S_n \leq S_{n+1} \leq S_{n+2}$ and because $n\,S_n = (n-1)\,S_{n-2}$. $\lim_{n\to\infty} \frac{S_{2n+1}}{S_{2n}} = 1$. This proves Wallis' formula:

$$\lim_{n\to\infty} \frac{2.2.4.4.6.6\ldots 2n.2n}{1.3.3.5.5.7\ldots(2n-1).(2n+1)} = \frac{\pi}{2}$$

Chapter 7

Metric spaces and Uniform Convergence

The central idea of convergence is fundamental in a much more general context. In particular, in this chapter we want to discuss convergence of functions. In order to do this in a systematic way, it is convenient to work in the framework of general metric spaces. Although topological spaces provide the most general framework for discussions of continuity and convergence, it is necessary for many problems in analysis to work with the more concrete spaces called metric spaces, where the notion of a distance is the key idea.

7.1 Definitions and examples

Definition 7.1.1 *A* metric *on a set X is an assignment of a non-negative number, called the distance $d(x, y) \in [0, \infty)$ to every pair of points x, y in X satisfying the following axioms:*

M1.(Positivity) $\forall\ x, y \in X,\ d(x, y) \geq 0$ *and* $d(x, y) = 0$ *iff* $x = y$.

M2.(Symmetry) $\forall\ x, y \in X,\ d(x, y) = d(y, x)$

M3.(Triangle inequality) $\forall\ x, y, z \in X,\ d(x, y) + d(y, z) \geq d(x, z)$.

A metric space *is a set X, together with a metric.*

Exactly as for $\mathbb{R}^n$, we define an open ball with centre a and radius $r > 0$ in a metric space to be the set $B_r(a) = \{\, x \in X \mid d(x, a) < r \,\}$ and we call $U \subset X$ open if $\forall\, a \in U\, \exists\, r > 0$ such that $B_r(a) \subset U$. This defines a topology for X in the sense of Chapter 3 , so metric spaces are special cases of topological spaces and all the general definitions and theorems, using open sets as the fundamental concept in Chapters 3 and 4 about connectedness, compactness and continuity hold for metric spaces (in particular!).

Definition 7.1.2 *A sequence* (x_n) *of points in a metric space* (X, d) *is said to converge to a limit* $x \in X$ *if and only if the sequence of real numbers* $d(x_n, x)$ *converges to* 0 *.*

Definition 7.1.3 *A sequence* (x_n) *of points in a metric space* (X, d) *is said to be a Cauchy sequence if* $\forall \epsilon > 0, \exists N$ *, such that* $n \geq N \Rightarrow d(x_{n+k}, x_n) < \epsilon$ for all $k \in \mathbb{N}$.

As was the case for real numbers, it is easy to see that every convergent sequence is a Cauchy sequence, but the converse need not be true in a general metric space.

Definition 7.1.4 *A metric space is said to be complete if every Cauchy sequence is convergent.*

Examples

0. The Euclidean vector space $\mathbb{R}^n$ with the usual metric $d(x, y) = ||x - y||$ is the mother of all metric spaces. (Warning: the four dimensional Minkowski space-time $\mathbb{R}^{1,3}$ of special relativity where the distance is defined by an inner product that is *not* positive-definite is not a metric space.) Similarly, $\mathbb{C}^n$ is usually given the same metric as $\mathbb{R}^{2n}$. All these spaces are complete metric spaces.

1. The vector space $\mathbb{R}^n$ can also be equipped with other non-Euclidean metrics. Important examples are defined by the L^p norms : $||x||_p$ defined by $\big(||x||_p\big)^p = \sum_{i=1}^n |x_i|^p$ (for $p > 0$). The distance is then given by: $d(x, y) = ||x - y||_p$. (You will check all the axioms as an exercise at the end of this chapter. The first two properties of the metric are easy to check. The triangle inequality is a bit harder.) One can extend this definition to infinite sequences of real numbers provided we restrict to those sequences whose L_p-norm $\sum_{i=1}^\infty |x_i|^p$ is finite. All these spaces are also complete.

2. (The most important metric for the purposes of this section!).
Let $X = C^0[a,b]$ be the set of all real-valued continuous functions on the compact interval $[a,b]$. Since all continuous functions on a compact set are bounded we can define a metric on this space by:

$$d(f,g) = sup(\{|f(x) - g(x)| \mid x \in [a,b]\})$$

. The corresponding norm $||f|| = d(f,0)$ is called the *supremum norm.* We will see in the next section that this is a complete metric space.

3. Using integrals, we can also define L_p norms on $C^0[a,b]$ as follows:

$$\left(||f||_p\right)^p = \int_a^b |f(x)|^p$$

but I will leave the study of these spaces to your next course in analysis.

Remarks

0. A sequence in $\mathbb{R}^n$ (or $\mathbb{C}^n$)converges iff each of its components converges. The situation for infinite-dimensional spaces is more delicate. For example each component of the shift sequence in "$\mathbb{R}^\infty$": $a_1 = (1,0,0,...), a_2 = (0,1,0,...), a_3 = (0,0,1,0,...), ...$ converges to 0, but it would not be wise to say that the whole sequence converges to 0.

1. For the finite dimensional spaces $\mathbb{R}^n$, convergence with respect to the L^p-norma is the same for all p, but this is quite different when we consider infinite dimensional function spaces. You will learn all of this next year, hopefully!

2. Convergence in the metric space $X = C^0[a,b]$ equipped with the metric $d(f,g) = sup(\{|f(x) - g(x)| \mid x \in [a,b]\})$ is our main concern for this section and convergence in this space is given a special name:

Definition 7.1.5 *A sequence of functions* (f_n) *in* $C^0[a,b]$ *is said to converge* **uniformly** *to a function* f *if and only if* $\forall\, \epsilon > 0, \exists N$ *, such that* $n \geq N \Rightarrow |f(x) - g(x)| < \epsilon$ for all $x \in [a,b]$.

7.2 Uniform Convergence

Definition 7.2.1 *A sequence of functions* (f_n) *defined on a set* A *is said to converge* **uniformly** *to a function* f *if and only if* $\forall\, \epsilon > 0, \exists N$ *, such that* $n \geq N \Rightarrow |f_n(x) - f(x)| < \epsilon$ for all $x \in A$.

This is in contrast to the following much weaker notion of convergence we could have used:

Definition 7.2.2 *A sequence of functions* (f_n) *defined on a set* A *is said to converge* **pointwise** *to a function* f *if for every point* $x \in A$ *and* $\forall \epsilon > 0, \exists N$ *, (depending possibily on* x*), such that* $n \geq N \Rightarrow |f_n(x) - f(x)| < \epsilon$ *.*

Remarks

0. The sequence of functions $f_n(x) = x^n$ converges pointwise but not uniformly in $[0, +1]$. The limit function f is obviously discontinuous at the end point 0 .

1. The sequence of functions $f_n(x) = \frac{nx}{1+|nx|}$ converges pointwise but not uniformly in $\mathbb{R}$.

Proposition 7.2.1 *If a sequence of continuous functions* $f_k : [a, b] \to \mathbb{R}$ *converges uniformly to* $f : [a, b] \to \mathbb{R}$ *then* f *is continuous on* $[a, b]$ *.*

Proof:

The classic proof of this (which thousands of mathematicians had to learn in the last 100 years) is actually very simple and is based on the fact that: $|f(x)-f(y)| \leq |f(x)-f_k(x)|+|f_k(x)-f_k(y)|+|f_k(y)-f(y)|$ (triangle inequality). All three terms on the right-hand-side are made to go to 0 by using the assumptions of the theorem, so let's just do what others have done before us!

Let $x \in [a, b]$ and $\epsilon > 0$ be given. Then because we have uniform convergence we can find a K such that the first term $|f(x) - f_K(x)|$ and the last term: $|f_K(y)-f(y)|$ are *simultaneously* less than (say) $\frac{1}{10}\epsilon$. Now the second term : $|f_K(x)-f_K(y)|$ can also be made less than $\frac{1}{10}\epsilon$ by choosing y sufficiently close to x , (i.e. $\exists\, \delta > 0\, \mathrm{s.}t.\, |x-y| < \delta \Rightarrow |f_K(x)-f_K(y)| < \frac{1}{10}\epsilon$), since f_K is a continuous at x .
Adding the three terms together finishes the proof, since $\frac{3}{10} < 1$!

QED

One nice way to paraphrase what we just proved is to say that

$$\lim_{y\to x} \lim_{n\to\infty} f_k(y) = \lim_{n\to\infty} \lim_{y\to x} f_k(y)$$

,
so uniform continuity implies commutativity of two different limits.

Corollary 7.2.1 $C^0[a,b]$ *is a complete metric space with respect to the metric defined by the supremum norm.*

The proof given above can be strengthened to show that:

Proposition 7.2.2 *If a sequence of uniformly continuous functions* $f_k : [a,b] \to \mathbb{R}$ *converges uniformly to* $f : [a,b] \to \mathbb{R}$ *then* f *is uniformly continuous on* $[a,b]$.

More importantly, we can relate uniform convergence to differentiation and integration:

Proposition 7.2.3 *Let* $f_k : A \to \mathbb{R}$ *be a sequence of continuous functions converging uniformly on a compact set* A*. Then*

(i) $\int_a^b f_k \to \int_a^b f$ *as* $k \to \infty$ *for any interval* $[a,b] \subset A$.

(ii) If each f_k *is continuously differentiable and if* $f_k^{'} \to g$ *uniformly on* A*, then* $f^{'} = g$.

Proof:

(i) Let $\epsilon > 0$ be given. Then $\exists\, K$ such that $k \geq K \Rightarrow$
$\sup_{x\in A} |f_k(x) - f(x)| < \frac{\epsilon}{b-a} \Rightarrow \left| \int_a^b (f_k - f) \right| \leq \int_a^b |f_k - f| < \epsilon$

(ii) By (i): $f_k^{'} \to g$ uniformly on $A \Rightarrow \int_a^x f_k^{'} \to \int_a^x g$ for any $x \in A$. By the fundamental theorem of calculus: $f_k(x) - f(a) = \int_a^x f_k^{'}$ and since $f_k(x) \to f(x)$ we have $\int_a^x g = f(x) - f(a)$. Therefore again by the fundamental theorem of calculus $g(x) = f^{'}(x)$

QED

Remarks

1. The sequence of functions f_n defined on $[0,1]$, whose graph consists of isosceles triangles with base $[0, \frac{2}{n}$ and of height n all have area 1, so that $\int_0^1 f_n = 1$, for all n, but the limit function is 0 everywhere except at 0.

2. The sequence of functions

$$f_n(x) = \frac{x}{1 + n\, x^2}$$

converges uniformly to a limit function f, but

$$\lim_{n\to\infty} f\,'(0) \neq f\,'(0)$$

7.2.1 Uniform convergence of power series

The main message of this section is that on any compact set which is strictly inside its radius of convergence a power series converges uniformly and therefore we can integrate and differentiate power series term by term on any compact set inside its radius of convergence. The standard way to establish this is to prove a general, but very useful comparison theorem, called the Weierstrass M-Test for a series of functions (not necessarily a power series)

Proposition 7.2.4 *(*Weierstrass M-test*) Suppose $\sum M_n$ is convergent series of strictly positive numbers. If a sequence of continuous functions (f_n) defined on a compact set A satisfies $|f_n(x)| \leq M_n$ for all $x \in A$, then the series of functions $\sum f_n(x)$ is uniformly and absolutely convergent on A.*

Proof: The partial sums $S_n(x) = \sum_{k=1}^{n} f_k(x)$ form a Cauchy sequence in the complete metric space $C^0(A)$, since for any $n > m$:

$$|S_n(x) - S_m(x)| \leq \sum_{k=m}^{n} M_k$$

for all $x \in A$ (and M_n is a Cauchy sequence).

QED

Corollary 7.2.2 *Suppose a power series $\sum a_n x^n$ converges for $|x| < R$ (with $R > 0$). Then it converges uniformly on $|x| \leq \rho$ for each $\rho < R$.*

Proof: Apply the Weierstrass M-Test: $|a_n x^n| \leq M_n = |a_n \rho^n|$.

7.3 Weierstrass approximation theorem

Theorem 7.3.1 *If f is a continuous function on $[a, b]$, then there exists a sequence of polynomials (P_n) which converges uniformly to f.*

Proof: After applying a translation and a dilation, we may assume w.l.o.g. that $[a, b] = [0, 1]$ and also that $max\{\, |f(x)| \,|\, x \in [0, 1] \,\} = 1$.

Define the **Bernstein polynomials** that approximate f by the formula:

$$BP_n(f,x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k\,(1-x)^{n-k}$$

For each f, BP_n is obviously a polynomial of degree n (in x. They satisfy the "binomial" identity:

$$BP_n(1,x) = \sum_{k=0}^{n} \binom{n}{k} x^k\,(1-x)^{n-k} = (1-x+x)^n = 1$$

This implies:

$$BP_n(f,x) - f(x) = \sum_{k=0}^{n} \Big(f\left(\frac{k}{n}\right) - f(x)\Big)\binom{n}{k} x^k\,(1-x)^{n-k}$$

Let $\epsilon > 0$ be given. Since f is uniformly continuous on $[a,b]$, $\exists\, \delta > 0$ such that $|x-y| < \delta \Rightarrow |f(x) - f(y) < \frac{1}{2}\epsilon$.
We now split the sum into two pieces: a sum $\sum{}'$ where we sum over all k which satisfy $|x - \frac{k}{n}| < \delta$ and $\sum{}''$ being the sum of the rest.
By uniform continuity the first sum is less than $\frac{1}{2}\epsilon \sum \binom{n}{k} x^k\,(1-x)^{n-k} < \frac{1}{2}\epsilon$.

To estimate the second sum we borrow some ideas from probabilty theory! We think of it as the probability of a random variable to lie in a certain range. The random variable X is given by the binomial distribution of n Bernoulli trials with "success" probability x. The mean (expected value) of X is $n\,x$ and its variance $(E((X-E(X))^2)$ is $n\,x(1-x)$ (see an elementary text book on probability). Since max $|f| \leq 1$, the second sum is bounded from above by $\sum{}'' \binom{n}{k} x^k\,(1-x)^{n-k}$, where we sum only over those values of k which satisfy $|x - \frac{k}{n}| \geq \delta$. We now interpret this sum as the probability that the random variable differs from its mean by more than $n\,\delta$ (i.e. a "tail estimate"). Using now the well known Tchebychev's (or Markov's) inequality:

$$Prob(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

where μ is the mean and σ is the variance of the random variable X, we can estimate the second sum to be less than $\frac{x(1-x)}{n^2\delta^2} \leq \frac{1}{4n^2\delta^2}$. which can be made less than $\frac{1}{2}\epsilon$ for n sufficiently large.

QED

7.4 Exercises

\# 1 Show that each of the following defines a complete metric on $\mathbb{R}^n$:

(i)

$$d(x,y) = \sum_{k=1}^{n} |x_k - y_k|$$

(ii)

$$d(x,y) = \max_{1\le k\le n} |x_k - y_k|$$

\# 2 Let (X,d) be a metric space. Show that

$$\rho(x,y) = \frac{d(x,y)}{1+d(x,y)}$$

is a metric on X . Is every subset of X bounded in the new metric ρ ?

\# 3. For each of the following sequence of functions defined on the closed interval $[0,1]$, determine whether the convergence is uniform or just point wise and calculate the limit functions. Are the limit functions continuous?

(i)

$$f_n(x) = \frac{n\,x}{1+|n\,x|}$$

(ii)

$$f_n(x) = n\,x\,(x-1)^n$$

(iii)

$$f_n(x) = \frac{x}{1+n\,x^2}$$

\# 4. Let $[[x]]$ denote the distance from x to the nearest integer (sketch the graph!). Show that the function:

$$f(x) = \sum_{k=1}^{\infty} \frac{[[3^k\,x]]}{3^n}$$

is continuous for all $x \in \mathbb{R}$, but is **not** differentiable at any point!

5. Give an example of a sequence of continuous functions f_n defined on $[0,1]$ that converges pointwise to zero but such that

$$\lim_{n\to\infty}\int_0^1 f_n(x)\,dx = 1$$

6. Recall that the Fibonacci numbers are defined by: $a_{n+1} = a_n + a_{n-1}$ with $a_0 = a_1 = 1$.

(i) Show that

$$\sum_{n=0}^{\infty} a_n\, x^n \;=\; \frac{-1}{x^2+x-1}$$

for all $|x| < \frac{1}{2}$.

(ii) Deduce the following formula for the Fibonacci numbers:

$$a_{n-1} = \frac{1}{\sqrt{5}}\Big(\big(\frac{1+\sqrt{5}}{2}\big)^n - \big(\frac{1-\sqrt{5}}{2}\big)^n\Big)$$

7*. Prove that:

$$\int_0^1 x^{-x}\,dx = \sum_{n=1}^{\infty} n^{-n}$$

7.4.1 Hints and short solutions to the exercises

2. Except for the triangle inequality, the other axioms are trivially satisfied, since d satisfies them! The triangle inequality follows from the following fact: If a, b, c are positive real numbers then

$$a \leq b + c \Rightarrow \frac{a}{1+a} \leq \frac{b}{1+b} + \frac{c}{1+c}$$

which I leave as an easy exercise for you. All sets are bounded in the new metric, since $\rho(x, y) \leq 1$ for all x, y.

3.

(i) $f_n(x) = \frac{n\,x}{1+|n\,x|}$

This sequence converges pointwise to the function $f(x) = +1$ for $0 < x \leq 1$ and $f(0) = 0$. The convergence is not uniform since the limit function is not continuous.

(ii)$f_n(x) = n\,x\,(x-1)^n$ This sequence converges pointwise to the constant function $f(x) = 0$ for every $x \in [0,1]$. Although the limit function is continuous, the convergence is **not** uniform since the maximum value of $f_n(x)$ on the interval $[0,1]$ (occurring at $x = \frac{1}{n+1}$) is $\frac{n}{n+1}\big(1 - \frac{1}{n+1}\big)^n$ which approaches $e > 0$ as $n \to \infty$.

(iii)$f_n(x) = \frac{x}{1+n\,x^2}$ This sequence converges pointwise to the constant function $f(x) = 0$ for every $x \in [0,1]$. The convergence is uniform since and the maximum value of $f_n(x)$ on the interval $[0,1]$ which occurs at $x = \frac{1}{\sqrt{n}}$ is $\frac{1}{2\sqrt{n}}$ which approaches zero as $n \to \infty$. The limit function has to be continuous (as it is obviously!). Note that the derivative at zero of all the f_n's is $f'(0) = 1$ and so they do not converge to $f'(0) = 0$!

6. Let $f(x) = \sum_{n=0}^{\infty} a_n\,x^n$. Then

$$\begin{aligned}
(1 - x - x^2) f(x) &= \sum_{n=0}^{\infty} a_n\,x^n - \sum_{n=0}^{\infty} a_n\,x^{n+1} - \sum_{n=0}^{\infty} a_n\,x^{n+2} \\
&= \sum_{n=0}^{\infty} a_n\,x^n - \sum_{n=1}^{\infty} a_{n-1}\,x^n - \sum_{n=2}^{\infty} a_{n-2}\,x^n \\
&= 1 + x - x + \sum_{n=2}^{\infty} \big(a_n - a_{n-1} - a_{n-2}\big) x^n \\
&= 1
\end{aligned}$$

Convergence is guaranteed for all $|x| < \frac{1}{\alpha}$, where $\alpha = \lim |\frac{a_{n+1}}{a_n}| = \frac{1+\sqrt{5}}{2}$ is the golden ratio and $\alpha < 2$.

(ii) Since the roots of the polynomial $x^2 + x - 1$ are $-\alpha = \frac{-1-\sqrt{5}}{2}$ and $\frac{1}{\alpha} = \frac{-1+\sqrt{5}}{2}$, we have a partial fraction decomposition:

$$\sum_{n=0}^{\infty} a_n\, x^n = \frac{-1}{x^2 + x - 1} = \frac{1}{\sqrt{5}}\Big(\frac{1}{x+\alpha} - \frac{1}{x - \frac{1}{\alpha}}\Big)$$

Expanding the right hand side into two geometric series:

$$\frac{1}{\alpha(1+\frac{x}{\alpha})} = \sum_{n=0}^{\infty}(-1)^n \alpha^{-n-1}\, x^n \qquad\qquad \frac{\alpha}{1-\alpha\, x} = \sum_{n=0}^{\infty} \alpha^{n+1}\, x^n$$

and comparing coefficients we get the formula for the Fibonacci numbers: $\sqrt{5}\, a_n = \alpha^{n+1} - (-1)^{n+1}\alpha^{-n-1} = \left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}$.

7. *Prove that:*

$$\int_0^1 x^{-x}\, dx = \sum_{n=1}^{\infty} n^{-n}$$

$x^{-x} = e^{-x\, log(x)} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} x^k \big(log(x)\big)^k$, so all we have to prove is that

$$\frac{(-1)^k}{k!} \int_0^1 x^k \big(log(x)\big)^k\, dx = (k+1)^{-(k+1)}$$

The improper integral is convergent, because $x \log x \to 0$ as $x \to 0$. Moreover, since $\frac{d}{dx}\Big(x^{k+1}\big(\log(x)\big)^l\Big) = (k+1)\, x^k(\log(x))^l + l\, x^k(log(x))^{l-1}$,

$$\int_0^1 x^k \big(log(x)\big)^l = -\frac{l}{k+1}\int_0^1 x^k \big(log(x)\big)^{l-1}$$

Inductively, we see that

$$\int_0^1 x^k \big(log(x)\big)^k\, dx = (-1)^k \frac{k!}{(k+1)^{(k+1)}}$$