NOTATION *(not examinable)*

We shall use the following standard notation

$\mathbb{N}$ is the set of positive integer numbers, $\mathbb{N} = \{1, 2, \ldots\}$.

$\mathbb{Z}$ is the set of integer numbers, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

$\mathbb{Q}$ is the set of rational numbers, $\mathbb{Q} = \{\frac{m}{n} : m, n \in \mathbb{Z}\}$.

$\mathbb{R}$ is the set of real numbers.

∞ is a shorthand for "infinity". It is not a proper number.

$\forall$ means "for all" or "for every",

$\exists$ means "there exists" or "there is",

The colon : in a mathematical formula means "such that".

REAL NUMBERS: AXIOMS *(not examinable)*

Real numbers obeys the following axioms.

(**A1**) $a + (b + c) = (a + b) + c$ for all $a, b, c \in \mathbb{R}$;

(**A2**) $a + b = b + a$ for all $a, b \in \mathbb{R}$;

(**A3**) there is a unique element in $\mathbb{R}$, denoted 0, such that $a + 0 = a$ for all $a \in \mathbb{R}$;

(**A4**) for every $a \in \mathbb{R}$, there is a unique element in $\mathbb{R}$, denoted $-a$, such that $a + (-a) = 0$;

(**A5**) $(a \times b) \times c = a \times (b \times c)$ for all $a, b, c \in \mathbb{R}$;

(**A6**) $a \times b = b \times a$ for all $a, b \in \mathbb{R}$;

(**A7**) there is a unique element in $\mathbb{R}$, denoted 1, such that $a \times 1 = a$ for all $a \in \mathbb{R}$;

(**A8**) for every nonzero $a \in \mathbb{R}$, there is a unique element in $\mathbb{R}$, denoted a^{-1} or $\frac{1}{a}$, such that $a \times a^{-1} = 1$;

(**A9**) $a \times (b + c) = a \times b + a \times c$ for all $a, b, c \in \mathbb{R}$.

Definition. Subtraction is defined by $a - b = a + (-b)$.

Definition. Division is defined by $\frac{a}{b} = a \times (b^{-1})$.

Remark. 0^{-1} does not exist. The expression $\frac{a}{0}$ has no meaning.

One "orders" two numbers by thinking of the larger as being the higher in order. Formally speaking, there is a relation $<$ between elements of $\mathbb{R}$ obeying the following axioms.

(**A10**) for any $x, y \in \mathbb{R}$, exactly one of the following is true: either $x = y$, or $x < y$, or $y < x$;

(**A11**) if $x < y$ and $y < c$ then $x < c$;

(**A12**) if $x < y$ then $x + c < y + c$ for all $c \in \mathbb{R}$;

(**A13**) if $x < y$ and $c > 0$ then $x\,c < y\,c$.

Definition. We write
$x > y$ if $y < x$,
$x \leqslant y$ if either $x < y$ or $x = y$ (or, in other words, if it is false that $x > y$) and
$x \geqslant y$ if $y \leqslant x$.

Remark. (**A1**)–(**A13**) are axioms and cannot be proved. All other known equalities and inequalities involving the addition and composition can be deduced from the above axioms.

MODULUS

For any $x \in \mathbb{R}$, the modulus (or absolute value) of x is defined by

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

We have $|a| = |-a|$ end $|ab| = |a|\,|b|$. If $r > 0$ then the inequality $|x| < r$ is equivalent to the pair of inequalities $-r < x$ and $x < r$. The estimate $|a - b| \leqslant |a| + |b|$ is usually called the triangle inequality. The triangle inequality implies that

$$|a_1 + a_2 + \cdots + a_n| \;\leqslant\; |a_1| + |a_2| + \cdots + |a_n|$$

for any collection of real numbers $a_1, a_2, \ldots, a_n$.

All the above results are proved by considering all possible cases of positive and negative a, b and $a - b$ and applying the axioms (**A1**)–(**A13**) (see the online lecture notes for details).

INTERVALS

It is convenient to identify real numbers with points on a straight line (which is usually called the real line). We fix an arbitrary point on the line, called the origin, and assume that this point represents the number 0. Negative numbers lie to the left of the origin and positive numbers lie to the right. The absolute value of a number coincides with the distance from the corresponding point on the line to the origin. The inequality $a < b$ means that b lies to the right of a, and the inequality $|a| < r$ is equivalent to saying that a is closer to the origin than r and $-r$.

Definition. Let $a, b \in \mathbb{R}$. Then

(1) the *open interval* (a, b) is the set of real numbers x such that $a < x$ and $b > x$ (the corresponding points on the line lie strictly between a and b);

(2) the *closed interval* $[a, b]$ is the set of real numbers x such that $a \leqslant x$ and $b \geqslant x$ (the corresponding points lie between a and b with the endpoints included);

(3) the half-open intervals $(a, b]$ and $[a, b)$ are defined in a similar manner, the square bracket indicates that the point is included in the set.

We shall also consider the following infinite intervals.

(4) The open intervals (a, ∞) and $(-\infty, b)$ are the sets of numbers x such that $x > a$ and, respectively, $x < b$.

(5) The closed intervals $[a, \infty)$ and $(-\infty, b]$ are the sets of numbers x such that $x \geqslant a$ and, respectively, $x \leqslant b$.

The open intervals $(0, \infty)$ and $(-\infty, 0)$ are traditionally denoted by $\mathbb{R}_+$ and $\mathbb{R}_-$ (these are the sets of positive and negative numbers).

BOUNDED SETS

A set of real numbers S is said to be

bounded from above if $S \subseteq (-\infty, b]$ for some real number b,

bounded from below if $S \subseteq [a, \infty)$ for some real number a,

bounded if it is bounded from below and from above (or, in other words, if $S \subseteq [a, b]$ for some $a, b \in \mathbb{R}$).

The numbers a and b are said to be the lower and upper bounds for the set S. Note that these numbers are certainly not unique (if they exist).

Definition. The smallest number b for which $S \subseteq (-\infty, b]$ is said to be the *least upper bound* or *supremum* of the set S (abbreviated *l.u.b.* or *sup*). The largest a satisfying $S \subseteq [a, \infty)$ is called the *greatest lower bound* or *infimum* of the set S (abbreviated *g.l.b.* or *inf*).

Remark. Let b be the l.u.b. of a set S. Then, for any $\delta > 0$, we can find a real number x lying in S such that $b - \delta < x \leqslant b$. Indeed, we have $x \leqslant b$ for all $x \in S$ because b is an upper bound. If the inequality $b - \delta < x$ is not true for any $x \in S$ then we have $b - \delta \geq x$ for all $x \in S$. But this means that $b - \delta$ is an upper bound for S, which is not possible because this number is smaller than b.

In a similar way one can prove the following: if a is the g.l.b. of a set S then, for any $\delta > 0$, we can find a real number x lying in S such that $a \leqslant x < a + \delta$.

REAL NUMBERS: COMPLETENESS AXIOM

It is not clear a priori whether every bounded set of real numbers has a l.u.b. or a g.l.b. It turns out that these statements cannot be proved or disproved. The following is an axiom (it is often called the *completeness axiom*).

(**A14**) every bounded from above set of real numbers has a least upper bound.

Roughly speaking, (**A14**) means that there are no holes (gaps) in the real line. Note that the completeness axiom would not be true if we considered only rational numbers. Indeed, the set of rational numbers which are strictly smaller that $\sqrt{2}$ is bounded from above, but its l.u.b. $\sqrt{2}$ is not a rational number.

Proposition. Every bounded from below set $S \subseteq \mathbb{R}$ has a g.l.b.

Proof. Let $-S$ be the set of numbers $-x$ where $x \in S$. This set is bounded from above and, by (**A14**), has a l.u.b. a. Then $-a$ is the g.l.b. of S.

Proposition. The l.u.b. of a set S is unique.

Proof. If there are two distinct upper bounds for S then, by (**A10**), one of them is smaller than the other. But then the larger number is not the smallest upper bound for S and therefore is not a l.u.b.

Remark. The g.l.b and l.u.b. of the set S may not belong to S. For instance, if $S = [1, 2)$ then g.l.b of S is 1, and the l.u.b is 2. Since the number 2 is not included in the interval S, this set does not contain its l.u.b.

If g.l.b. belongs to S then it is the minimal element of the set S, which is denoted by $\min S$. If l.u.b belongs to S then it is the maximal element of the set S, which is denoted by $\max S$.

Example. The open interval $S = (-1, 1)$ does not have the minimal or maximal element. But its g.l.b and the l.u.b. are well defined and are equal to -1 and 1 respectively.

Theorem (Archimedean Property). For every given $x \in \mathbb{R}$, there exists $n \in \mathbb{N}$ such that $n > x$.

Proof. If the statement is not true then x is an upper bound for the set $\mathbb{N}$, that is, $\mathbb{N}$ is bounded from above. But then, by (**A14**), the set $\mathbb{N}$ has a l.u.b y. Since y is the l.u.b., $y - 1$ is not an upper bound, which means that $y - 1 < n$ for some positive integer n. But then $y < n + 1$ which contradicts to the fact that y is an upper bound for $\mathbb{N}$.

SEQUENCES

Definition. A *sequence* of real numbers is a "listing" $a_1, a_2, a_3, \ldots$ of numbers $a_n \in \mathbb{R}$, labelled by positive integers $n \in \mathbb{N}$.

The sequence $a_1, a_2, \ldots$ is usually denoted by $\{a_n\}_{n\in\mathbb{N}}$ or just $\{a_n\}$. A sequence can be thought of as a countable ordered set of real numbers; the word "countable" means that elements of the set can be enumerated by positive integers. Traditionally, one assumes that a sequence has infinitely many members. A finite collection of real numbers is sometimes called a *finite* sequence (but then the use of the word "finite" is compulsory).

There is no requirements that the elements of a sequence are distinct numbers. It may well happen that $a_k = a_j$ for two distinct "labels" k and j. For instance, one can consider the constant sequence $a, a, a, \ldots$, where all elements a_n are equal to the same number a. It is also a properly defined sequence.

Example. A sequence with an arbitrary a_1 and other a_n defined by $a_{n+1} = a_1 + n\,d$ with some $d \in \mathbb{R}$ is called an *arithmetic progression.*

Example. A sequence with an arbitrary a_1 and other a_n defined by $a_{n+1} = a_1\,d^n$ with some $d \in \mathbb{R}$ is called a *geometric progression.*

Note that in these two examples the sequences are defined by the simple recurrence relations $a_{n+1} = a_n + d$ and $a_{n+1} = a_n\,d$. One can consider more complicated recurrence relations (that is, equations defining each term of the sequence as a function of the preceding terms).

Example. There is a sequence $\{a_n\}$ that contains all rational numbers as its members. Indeed, the set $\mathbb{Q}$ of rational numbers $\frac{k}{m}$ can be represented as a union of finite sets $\mathbb{Q}_j = \{\frac{k}{m} : |k| + |m| = j\}$, where $j = 1, 2, \ldots$ The required sequence is obtained by enumerating elements of the set $\mathbb{Q}_1$, then elements of the set $\mathbb{Q}_2$(starting from $n_1 + 1$ where n_1 is the number of elements in $\mathbb{Q}_1$), then elements of the set $\mathbb{Q}_2$(starting from $n_2 + 1$ where n_2 is the number of elements in $\mathbb{Q}_1 \bigcup \mathbb{Q}_2$), and so on.

Remark. A sequence cannot contain all real numbers; in other words, the set of real numbers is uncountable. Moreover, every non-degenerate interval is an uncountable set. For the interval $(0, 1)$ this can be proved as follows. Consider an arbitrary sequence of numbers a_n lying in $(0, 1)$. Each of this numbers can be written as a decimal fraction, so that $a_1 = 0.a_{11}a_{12}a_{13}\ldots$, $a_2 = 0.a_{21}a_{22}a_{23}\ldots$ and so on. Choose a decimal faction $0.b_1b_2b_3\ldots$ in such a way that $b_k \neq a_{kk}$ for all $k = 1, 2, \ldots$ Then $b_k \in (0, 1)$ but $b \neq a_n$ for any $n = 1, 2, \ldots$

Definition. We shall say that the sequence $\{a_n\}$ is
nondecreasing (or increasing) if $a_{n+1} \geqslant a_n$ (or $a_{n+1} > a_n$) for all n;
nonincreasing (or decreasing) if $a_{n+1} \leqslant a_n$ (or $a_{n+1} < a_n$) for all n.

PRINCIPLE OF INDUCTION

Let $\{a_n\}$ be a sequence. Suppose that, for each $n \in \mathbb{N}$, we have a statement $P(a_n)$ about the number a_n such that

(1) $P(a_1)$ is true;

(2) for every $k \in \mathbb{N}$, the truth of $P(a_k)$ implies the truth of $P(a_{k+1})$.

Then $P(a_n)$ is true for all a_n.

Example. $1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$ for all $n \in \mathbb{N}$.

Proof. Consider the above identity as a statement $P(n)$. Clearly, $P(1)$ is true. Assume that $P(k)$ is true. Then

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

A simple calculation shows that the right hand side coincides with

$$\frac{(k+1)((k+1)+1)(2(k+1)+1)}{6} .$$

Thus we have $P(k+1)$. Now, using the induction principle, we see that $P(n)$ is true for all $n \in \mathbb{N}$.

You must remember and be able to use all the definitions and theorems stated in this week's notes. Their proofs can be found in the CM115 lecture notes. *The proofs are not examinable.*

CONVERGENT SEQUENCES

Let $\{a_n\}$ be a sequence.

The Crucial Definition. We say that $\{a_n\}$ converges to a number a and write $\lim_{n\to\infty} a_n = a$ or $a_n \to a$ if

$$\forall \varepsilon > 0 \; \exists n_\varepsilon \in \mathbb{N} \; : \; n \geqslant n_\varepsilon \Rightarrow |a_n - a| < \varepsilon.$$

In usual words, this means the following: for every positive ε there is a positive integer n_ε such that $|a_n - a| < \varepsilon$ whenever $n \geqslant n_\varepsilon$. The number a is called the *limit* of $\{a_n\}$. A sequence which has a limit is called a *convergent* sequence.

One can reformulate the above definition in "geometric" terms.

The Crucial Definition: another version. The sequence $\{a_n\}$ converges to a number a if for every $\varepsilon > 0$ there are only finitely many elements a_n lying outside the open interval $(a - \varepsilon, a + \varepsilon)$.

Remark. There are divergent sequences, that is, the sequences which do not converge to a limit. For instance, the sequence $\{+1, -1, +1, -1, +1, \ldots\}$ does not converge.

Definition. We say that $\{a_n\}$ converges to $+\infty$ and write $a_n \to +\infty$ if

$$\forall R > 0 \; \exists n_R \in \mathbb{N} \; : \; n \geqslant n_R \Rightarrow a_n > R.$$

Similarly, we say that the sequence $\{a_n\}$ converges to $-\infty$ and write $a_n \to -\infty$ if

$$\forall R > 0 \; \exists n_R \in \mathbb{N} \; : \; n \geqslant n_R \Rightarrow a_n < -R.$$

Example. The sequence $\{1, 2, 3, 4 \ldots\}$ does not converge to a limit in $\mathbb{R}$ but converges to $+\infty$. The sequence $\{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$ converges to 0. Both these results follow from the Archimedean Property.

THEOREMS ABOUT CONVERGENT SEQUENCES

Convergence $\Rightarrow$ Boundedness. Every convergent sequence is bounded.

Monotone Convergence Theorem. Every bounded from above nondecreasing sequence a_n converges to its least upper bound (supremum). Every bounded from below nonincreasing sequence converges to its greatest lower bound (infimum).

Comparison Theorem. If $b_n \to 0$ and $|a_n| \leqslant b_n$ for all n then $a_n \to 0$.

The Sandwich Theorem. If $a_n \to c$, $b_n \to c$ and $a_n \leqslant c_n \leqslant b_n$ for all n then $c_n \to c$.

Theorem: algebraic rules for limits. Assume that $a_n \to a$ and $b_n \to b$. Then $a_n + b_n \to a + b$ and $a_n b_n \to ab$. If, in addition, $b \neq 0$ then $a_n/b_n \to a/b$.

The following examples illustrate how one can find limits with the use of above theorems.

Example. Let $a_1 = 2$ and $a_{n+1} = \frac{a_n^2+1}{2a_n}$ for all $n = 1, 2, 3, \ldots$ Then $a_n \to 1$.

Proof. Since $a_1 > 1$ and $\frac{b^2+1}{2b} \geqslant 1$ for all $b > 0$, we have $a_n \geqslant 1$ for all n. This implies that $a_{n+1} \leqslant a_n$, that is, the sequence is nonincreasing. By the Monotone Convergence Theorem, it converges to its g.l.b. a. Obviously, $\lim_{n\to\infty} a_n = \lim_{n\to\infty} a_{n+1}$. Therefore, by the above, $a = \frac{a^2+1}{2a}$. Solving this equation, we obtain $a = 1$.

Example. If $k \in \mathbb{N}$ then $\lim_{n\to\infty} n^k\, 2^{-n} = 0$.

Proof. Note first of all that $\lim_{n\to\infty} \frac{n+1}{n} = 1$. Consequently, $\lim_{n\to\infty} \left(\frac{n+1}{n}\right)^k = 1$ for all $k \in \mathbb{N}$. Let $a_n = n^k\, 2^{-n}$. Then $a_{n+1}/a_n = \frac{1}{2}\left(\frac{n+1}{n}\right)^k$. By the above, thee exists n_0 such that $a_{n+1}/a_n < \frac{2}{3}$ for all $n \geqslant n_0$. But then $a_{n_0+m} \leqslant \left(\frac{2}{3}\right)^m a_{n_0}$ for all $m = 1, 2, \ldots$ This estimate and the Comparison Theorem imply that $\lim_{n\to\infty} a_n = 0$.

Exercise. Let $k \in \mathbb{R}$ and $b > 1$. Prove that $\lim_{n\to\infty} n^k\, b^{-n} = 0$.

SUBSEQUENCES AND ACCUMULATION POINTS

Definition. A *subsequence* of a sequence is any sequence obtained by leaving out particular terms from the original sequence. A subsequence of the sequence $\{a_n\}$ is usually denoted by $\{a_{n_k}\}$ where $k \in \mathbb{N}$ is a new index, and n_k are the "numbers" of elements of the original sequence included in the subsequence.

Definition. A number c is said to be an *accumulation point* (or a *limit point*) of the sequence $\{a_n\}$ if there exists a subsequence $\{a_{n_k}\}$ such that $\lim_{k\to\infty} a_{n_k} = c$.

Exercise. Let $\{a_n\}$ be a convergent sequence and $\lim_{n\to\infty} a_n = a$. Prove that every subsequence of $\{a_n\}$ converges to a.

Corollary. A convergent sequence has a unique accumulation point, which coincides with its limit.

Proof. If a sequence converges to a then, by the above, all its subsequences converges to a. It follows that a is an accumulation point and there are no other accumulation points.

Remark. It may well happen that a subsequence converges to a limit but the sequence itself does not converge. A typical is $\{+1, -1, +1, -1, \ldots\}$. One can show that a sequence converges to a limit c if and only if **every** its subsequence converges to this limit.

Remark. A sequence may have many accumulation points. For example, if a sequence contains all rational numbers then every real number is its accumulation point.

Exercise. Show that c is an accumulation point if and only if any open interval of the form $(c - \varepsilon, c + \varepsilon)$ contains infinitely many elements of the sequence $\{a_n\}$.

The Bolzano–Weierstrass Theorem. Every bounded sequence has a convergent subsequence and, consequently, an accumulation point. If $a_n \in [a, b]$ for all n then the accumulation points of $\{a_n\}$ belong to the closed interval $[a, b]$.

CAUCHY SEQUENCES

Definition. We say that $\{a_n\}$ is a *Cauchy sequence* if

$$\forall \varepsilon > 0\ \exists n_\varepsilon \in \mathbb{N}\ :\ m, n \geqslant n_\varepsilon \Rightarrow |a_m - a_n| < \varepsilon$$

or, in usual words, if for all $\varepsilon > 0$ there exists n_ε such that $|a_m - a_n| < \varepsilon$ whenever $m, n \geqslant n_\varepsilon$.

In other words, $\{a_n\}$ is a Cauchy sequence if the distance $|a_n - a_m|$ between its elements goes to zero as $m, n \to \infty$ (but this is not a rigourous definition).

Cauchy's Theorem. Every Cauchy sequence converges and, conversely, every convergent sequence is Cauchy.

UPPER AND LOWER LIMITS

One says that $+\infty$ is an accumulation point of a sequence $\{a_n\}$ if there is a subsequence of $\{a_n\}$ which converges to $+\infty$. Similarly, $-\infty$ is an accumulation point of $\{a_n\}$ if there is a subsequence which converges to $-\infty$. This convention makes sense, even though $\pm\infty$ are not proper numbers. Note that an unbounded sequence must have a subsequence which converges either to $+\infty$, or to $-\infty$. Therefore the Bolzano–Weierstrass theorem implies the following

Corollary. Every sequence has a subsequence which converges either to a finite limit, or to $\pm\infty$.

Definition. The largest accumulation point of a sequence $\{a_n\}$ is called the upper limit of $\{a_n\}$ and is denoted $\limsup a_n$. The smallest accumulation point of $\{a_n\}$ is called the lower limit of $\{a_n\}$ and is denoted $\liminf a_n$.

In view of the above corollary, the upper and lower limits always exist, even if the sequence does not converge. If the sequence is bounded, both upper and lower limits are finite numbers. If the sequence is not bounded from above then $\limsup a_n = +\infty$. If the sequence is not bounded from below then $\liminf a_n = -\infty$.

Example. If $a_n = n$ then $\limsup a_n = \liminf a_n = +\infty$. If $a_n = -n$ then $\limsup a_n = \liminf a_n = -\infty$.

Theorem. A sequence $\{a_n\}$ converges to a number c if and only if $\limsup a_n = \liminf a_n = c$.

Proof. If $a_n \to c$ then c is the unique accumulation point of $\{a_n\}$ (see Week 2). Thus it is the smallest and the largest accumulation point at the same time, so that $\limsup a_n = \liminf a_n = c$.
On the other hand, if $\limsup a_n = \liminf a_n = c$ then c is the only accumulation point of $\{a_n\}$. This implies that there are only finitely many a_n outside any interval of the form $(c-\varepsilon, c+\varepsilon)$ (otherwise these elements would form a sequence lying outside $(c-\varepsilon, c+\varepsilon)$, which would have another accumulation point). This means that $a_n \to c$ (see Week 2).

Exercise. Show that the above theorem remains valid for $c = +\infty$ and $c = -\infty$

Warning. It may well happen that $\limsup(a_n + b_n) \neq \limsup a_n + \limsup b_n$ and/or $\liminf(a_n + b_n) \neq \liminf a_n + \liminf b_n$. For instance, if $a_n = (-1)^n$ and $b_n = (-1)^{n+1}$ then $\limsup a_n = \limsup b_n = 1$, $\liminf a_n = \liminf b_n = -1$ but $a_n + b_n = 0$ for all n.

Lemma. If $r > 0$ then $\limsup(r\, a_n) = r\, \limsup a_n$ and $\liminf(r\, a_n) = r\, \liminf a_n$. If $r < 0$ then $\limsup(r\, a_n) = r\, \liminf a_n$ and $\liminf(r\, a_n) = r\, \limsup a_n$.

Proof. Let $r \neq 0$. From the algebraic rules for limits it follows that a subsequence $\{a_{n_k}\}$ converges to a limit a if and only if $\{r\, a_{n_k}\}$ converges to the limit $r\, a$. Therefore c is an accumulation point of $\{a_n\}$ if and only if $r\, c$ is an accumulation point of the sequence $\{r\, a_n\}$. In other words, the accumulation points of the sequence $\{r\, a_n\}$ are obtained from the accumulation points of $\{a_n\}$ by multiplying them by the number r. If $r > 0$ then the multiplication maps the largest and smallest accumulation points of $\{a_n\}$ into the largest and smallest accumulation points of

$\{r\, a_n\}$. If $r < 0$ then the multiplication transfers the largest accumulation point of $\{a_n\}$ into the smallest accumulation point of $\{r\, a_n\}$, and the other way round.

Theorem. Let $s_k = \sup_{n \geq k} a_n$. Then the sequence $\{s_k\}$ converges (to a finite number or $\pm\infty$) and $\lim s_k = \limsup a_n$.

Proof. If the sequence $\{a_n\}$ is not bounded from above then $\limsup a_n = +\infty$ and $s_k = +\infty$ for all k.
Assume that $\{a_n\}$ is bounded from above. By definition, s_k is the least upper bound for the set $\{a_k, a_{k+1}, a_{k+2}, \ldots\}$. Obviously, it is also an upper bound for the smaller set $\{a_{k+1}, a_{k+2}, \ldots\}$. Since s_{k+1} is the least upper bound for $\{a_{k+1}, a_{k+2}, \ldots\}$, it follows that $s_{k+1} \leqslant s_k$. Thus $\{s_k\}$ is a nonincreasing sequence. By the monotone convergence theorem, it converge either to a finite number or to $-\infty$.
If $s_k \to -\infty$ then $a_n \to -\infty$ because $a_k \leqslant s_k$. In this case $\limsup a_n = \lim a_n = -\infty = \lim s_k$.
If s_k converge to a finite number c, it must be an accumulation point. Indeed, if c is not an accumulation point then there is an interval $(c-\varepsilon, c+\varepsilon)$ which contains only finitely many elements a_n. This implies that the elements $a_k, a_{k+1}, a_{k+2}, \ldots$ lie outside this interval for all sufficiently large k. It follows that $s_k \notin (c-\varepsilon, c+\varepsilon)$ and, consequently, s_k cannot converge to c.
Finally, if c is not the largest accumulation point then the sequence $\{a_n\}$ has another accumulation point b such that $b > c$. Let $\varepsilon >$ be small enough, so that $c < b - \varepsilon$. Since b is an accumulation point, there are infinitely many elements a_n in the interval $(b-\varepsilon, b+\varepsilon)$. But then, for every k, the set $\{a_k, a_{k+1}, a_{k+2}, \ldots\}$ has infinitely many elements in the interval $(b-\varepsilon, b+\varepsilon)$. It follows that its least upper bound s_k cannot be smaller than $b - \varepsilon$. Consequently, s_k does not converge to c. The obtained contradiction shows that c is the largest accumulation point.

In a similar way one can prove

Theorem. Let $r_k = \inf_{n \geq k} a_n$. Then the sequence r_k converges (to a finite number or $\pm\infty$) and $\lim r_k = \liminf a_n$.

SERIES

Series is just an infinite sum of the form $\sum_{n=1}^{\infty} a_n$ or $a_1 + a_2 + a_3 + \ldots$, where $a_n \in \mathbb{R}$.

Definition. We say that the series converges and write $\sum_{n=1}^{\infty} a_n = c$ if the sequence of finite partial sums $s_k = \sum_{n=1}^{k} a_n$ converges to c. Equivalently

$$\forall \varepsilon > 0 \ \exists k_\varepsilon \ : \ k \geqslant k_\varepsilon \Rightarrow \left| \sum_{n=1}^{k} a_n - c \right| < \varepsilon.$$

We say that a series is divergent if it is not convergent.

Example. The series $1 - 1 + 1 - 1 + 1 \ldots$ does not converge because its partial sums oscillate between -1 and 1.

Example. An infinite decimal fraction is an example of a convergent series. Indeed, when we write $a = 0.a_1 a_2 a_3 \ldots$ we actually mean that $a = \sum_{n=1}^{\infty} a_n\, 10^{-n}$.

Remark. The series $\sum_{n=1}^{\infty} a_n$ converges if and only if the series $\sum_{n=m}^{\infty} a_n$ converges. In other words the first few terms do not affect convergence, even though they change the value of the sum. Indeed, if s_k are the partial sums of the original series then the partial sums σ_k of the series $\sum_{n=m}^{\infty} a_n$ are equal to $s_k - a$ where $a = \sum_{n=1}^{m-1} a_n$. From the definition of convergence for sequences, it follows that the sequence $\{s_k\}$ converges if and only if the sequence $\{\sigma_k\}$ converges.

Example (geometric progression). If $|b| < 1$ then the series $\sum_{n=0}^{\infty} b^n$ converges and $\sum_{n=0}^{\infty} b^n = (1-b)^{-1}$. Indeed, by induction in m we obtain

$$\sum_{n=0}^{m} b^n = (1-b)^{-1} - b^{m+1}(1-b)^{-1}.$$

If $|b| < 1$ then $b^{m+1} \to 0$ and, consequently, $\displaystyle\sum_{n=0}^{m} b^n \;\to\; (1-b)^{-1}$ as $m \to \infty$.

Theorem. If the series $\sum_{n=1}^{\infty} a_n$ converges then $\lim_{n\to\infty} a_n = 0$.

Proof. Since the series converges, its partial sums s_k form a convergent sequence $\{s_k\}$. By Cauchy's theorem, $\{s_k\}$ is a Cauchy sequence, that is, $|s_m - s_k| \to 0$ as $k, m \to \infty$. Taking $k = m-1$, we see that $a_m = s_m - s_{m-1} \to 0$ as $m \to \infty$.

Remark. The converse is FALSE: it may well happen that $\lim_{n\to\infty} a_n = 0$ but the series $\sum_{n=1}^{\infty} a_n$ diverges. For instance, the series $\sum_{n=1}^{\infty} 1/n$ diverges.

Theorem. If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are convergent series then
$\sum_{n=1}^{\infty}(a_n + b_n) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n$ and $\sum_{n=1}^{\infty} c\, a_n = c \sum_{n=1}^{\infty} a_n$ for all $c \in \mathbb{R}$.

Proof. Let $s_m = \sum_{n=1}^{m} a_n$ and $\sigma_m = \sum_{n=1}^{m} b_n$. Then from the algebraic rules for limits it follows that

$$\lim_{m\to\infty} \sum_{n=1}^{m} (a_n + b_n) = \lim_{m\to\infty} (s_m + \sigma_m) = \lim_{m\to\infty} s_m + \lim_{m\to\infty} \sigma_m = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n\,,$$

which means that $\sum_{n=1}^{\infty}(a_n + b_n) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n$. Similarly,

$$\sum_{n=1}^{\infty} c\, a_n = \lim_{m\to\infty} \sum_{n=1}^{m} c\, a_n = \lim_{m\to\infty} c\, s_n = c \lim_{m\to\infty} s_n = c \sum_{n=1}^{\infty} a_n\,.$$

ABSOLUTELY CONVERGENT SERIES

Definition. A series $\sum_{n=1}^{\infty} a_n$ is said to be absolutely convergent if the positive series $\sum_{n=1}^{\infty} |a_n|$ converges.

Instead of saying that the positive series is absolutely convergent, one often writes $\sum_{n=1}^{\infty} |a_n| < \infty$. Obviously, a positive series $\sum_{n=1}^{\infty} a_n$ is convergent if and only if it is absolutely convergent (in this case $a_n = |a_n|$).

Absolute Convergence Theorem. Every absolutely convergent series is convergent. Or: if $\sum_{n=1}^{\infty} |a_n| < \infty$ then $\sum_{n=1}^{\infty} a_n$ converges.

Proof. Let $s_k = \sum_{n=1}^{k} a_n$ and $\sigma_k = \sum_{n=1}^{k} |a_n|$. If $m \geqslant k$ then, using the triangle inequality for the modules, we obtain

$$|s_m - s_k| \;=\; \left|\sum_{n=k+1}^{m} a_n\right| \;\leqslant\; \sum_{n=k+1}^{m} |a_n| \;=\; \sigma_m - \sigma_k\,.$$

Since the series $\sum_{n=1}^{\infty} |a_n|$ converges, the sequence $\{\sigma_k\}$ converges. Therefore it is a Cauchy sequence, that is, $\sigma_m - \sigma_k \to 0$ as $m, k \to \infty$. The above inequality implies that $|s_m - s_k| \to 0$ as $m, k \to \infty$, which means that $\{s_k\}$ is also a Cauchy sequence. Finally, by Cauchy's theorem, the sequence $\{s_k\}$ converges.

Comparison Theorem for Series. If $|a_n| \leqslant b_n$ for all sufficiently large n and the series $\sum_{n=1}^{\infty} b_n$ is convergent then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

Proof. The words "for all sufficiently large n" mean that $|a_n| \leqslant b_n$ for all $n \geqslant n_0$, where n_0 is some fixed positive integer.
Denote $\sigma_k = \sum_{n=1}^{k} |a_n|$ and $\sigma'_k = \sum_{n=1}^{k} b_n$. Then $0 \leqslant (\sigma_m - \sigma_k) \leqslant (\sigma'_m - \sigma'_k)$ whenever $m \geqslant k \geqslant n_0$. The right hand side of this inequality converges to 0 as $m, k \to \infty$ because the sequence $\{\sigma'_k\}$ is convergent (Cauchy's theorem). Therefore $|\sigma_m - \sigma_k| \to 0$ as $m, k \to \infty$ and, consequently, $\{\sigma_k\}$ is a Cauchy sequence. By Cauchy's theorem, the sequence $\{\sigma_k\}$ converges, which means that the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

Definition. We say that the integral $\int_1^{\infty} f(x)\,\mathrm{d}x$ converges and $\int_1^{\infty} f(x)\,\mathrm{d}x = c$ if $\lim_{k\to\infty}\left(\int_1^{k} f(x)\,\mathrm{d}x\right) = c$, where c is a finite number.

Integral Comparison Theorem. If $f(x)$ is a positive nonincreasing function on $[1, \infty)$ then the sum $\sum_{n=1}^{\infty} f(n)$ converges if and only if the integral $\int_1^{\infty} f(x)\,\mathrm{d}x$ converges. (This is not stating that $\sum_{n=1}^{\infty} f(n) = \int_1^{\infty} f(x)\,\mathrm{d}x$.)

Proof. Denote $c_n = \int_n^{n+1} f(x)\,\mathrm{d}x$. Obviously,

$$\int_0^{k} f(x)\,\mathrm{d}x \;=\; \sum_{n=1}^{k-1} \int_n^{n+1} f(x)\,\mathrm{d}x \;=\; \sum_{n=1}^{k-1} c_n\,.$$

It follows that the integral $\int_0^{\infty} f(x)\,\mathrm{d}x$ converges if and only if the series $\sum_{n=1}^{k} c_n$ converges.
Since f is nonincreasing, $f(n+1) \leqslant f(x) \leqslant f(n)$ for all $x \in [n, n+1]$. Therefore we obviously have

$$f(n+1) \;=\; \int_n^{n+1} f(n+1)\,\mathrm{d}x \;\leqslant\; \int_n^{n+1} f(x)\,\mathrm{d}x$$

and

$$f(n) \;=\; \int_n^{n+1} f(n)\,\mathrm{d}x \;\geqslant\; \int_n^{n+1} f(x)\,\mathrm{d}x\,,$$

that is, $f(n+1) \leqslant c_n \leqslant f(n)$.
If the series $\sum_{n=1}^{\infty} f(n)$ converges then, by the comparison theorem, the series

$\sum_{n=1}^{k} c_n$ is also convergent. On the other hand, the series $\sum_{n=1}^{k} c_n$ converges then, by the comparison theorem, the series $\sum_{n=1}^{\infty} f(n+1)$ converges. The latter is equivalent to the convergence of the series $\sum_{n=1}^{\infty} f(n)$ because

$$\sum_{n=1}^{\infty} f(n+1) \;=\; f(2)+f(3)+f(4)+\dots \;=\; \sum_{n=2}^{\infty} f(n).$$

Remark. The "obvious" estimates and equalities used in the proof will be discussed in the end of the course.

Example. The series $\sum_{n=1}^{\infty} 1/n^{\alpha}$ converges if $\alpha > 1$ and diverges if $\alpha \leqslant 1$.

TWO CONVERGENCE TESTS

nth-root Test Theorem. Let $\limsup |a_n|^{1/n} = c$. If $c < 1$ then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent. If $c > 1$ then the series diverges. (If $c = 1$ nothing can be said.)

Proof. Assume that $c < 1$. Then there exists a positive number b such that $c < b < 1$. From the definition of the upper limit it follows that $|a_n|^{1/n} \leqslant b$ for all sufficiently large n or, in other words, $|a_n| \leqslant b^n$ for all $n \geqslant m$, where m is some positive integer. Since the series $\sum_{n=1}^{\infty} b^n$ converges, by the Comparison Theorem the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

Assume now that $c > 1$. Then there exists a positive number b such that $1 < b < c$. From the definition of the limit it follows that $|a_n|^{1/n} \geqslant b$ for infinitely many values of n. In other words, there exist positive integers $n_1 < n_2 < n_3 \ldots$ such that $|a_{n_k}| \geqslant b^{n_k}$. Since $n_k \to \infty$ and, therefore, $b^{n_k} \to \infty$ as $k \to \infty$, we see that a_n do not converge to 0 as $n \to \infty$. This implies that the series $\sum_{n=1}^{\infty} a_n$ diverges.

Ratio Test Theorem. Assume that $a_n \neq 0$ for all n, and that $\lim_{n\to\infty} |a_{n+1}/a_n|$ exists and is equal to c. If $c < 1$ then the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent. If $c > 1$ then the series diverges. (If $c = 1$ nothing can be said.)

Proof. Assume that $c < 1$. Then there exists a positive number b such that $c < b < 1$. From the definition of the limit it follows that $|a_{n+1}/a_n| \leqslant b$ for all sufficiently large n or, in other words, $|a_{n+1}| \leqslant b\,|a_n|$ for all $n \geqslant n_0$, where n_0 is some positive integer depending on b. Then we have

$$|a_{n_0+1}| \leqslant b\,|a_{n_0}|,\ |a_{n_0+2}| \leqslant b^2\,|a_{n_0}|,\ |a_{n_0+3}| \leqslant b^3\,|a_{n_0}|,\ \ldots,\ |a_{n_0+j}| \leqslant b^j\,|a_{n_0}|,\ \ldots,$$

which implies that $|a_n| \leqslant |a_{n_0}|\, b^{-n_0}\, b^n$ for all $n \geqslant n_0$. Since the series $\sum_{n=1}^{\infty} b^n$ converges, by the Comparison Theorem the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent.

Assume now that $c > 1$. Then there exists a positive number b such that $1 < b < c$. From the definition of the limit it follows that $|a_{n+1}/a_n| \geqslant b$ for all sufficiently large n or, in other words, $|a_{n+1}| \geqslant b\,|a_n|$ for all $n \geqslant n_0$, where n_0 is some positive integer depending on b. Then we have

$$|a_{n_0+1}| \geqslant b\,|a_{n_0}|,\ |a_{n_0+2}| \geqslant b^2\,|a_{n_0}|,\ |a_{n_0+3}| \geqslant b^3\,|a_{n_0}|,\ \ldots,\ |a_{n_0+j}| \geqslant b^j\,|a_{n_0}|,\ \ldots,$$

which implies that $a_{n_0+k} \to +\infty$ as $k \to \infty$. In this case the series $\sum_{n=1}^{\infty} a_n$ diverges because a_n do not converge to 0 as $n \to \infty$.

Remark. Note that we cannot replace sup with lim sup in the ratio test theorem. For instance, the series $b_1 + 2b_1 + b_2 + 2b_2 + b_3 + 2b_3 + \ldots$ converges whenever the series $\sum_{n=1}^{\infty} b_n$ converges. But for the first series $\limsup |a_{n+1}/a_n| \geqslant 2$ because $a_{n+1} = 2a_n$ for all odd n.

CONDITIONAL CONVERGENCE

Definition. If a series converges but is not absolutely convergent, one says that the series converges conditionally.

The following theorem shows that the sum of a conditionally convergent series depends on the order of summation.

Riemann Series Theorem. Let $\sum_{n=1}^{\infty} a_n$ be a conditionally convergent series, and let b be an arbitrary real number or $\pm\infty$. One can always rearrange the terms a_n is such a way that the new series converges to b.

Definition. A series of the form $\sum_{n=1}^{\infty}(-1)^n a_n$, where $a_n \geqslant 0$, is called an alternating series. The alternating series may well converge even if it is not absolutely convergent, that is, even if $\sum_{n=1}^{\infty} |a_n| = \infty$

Theorem. If the sequence $\{a_n\}$ is nonincreasing and $a_n \to 0$ then the alternating series $\sum_{n=1}^{\infty}(-1)^n a_n$ converges.

Proof. Let $s_m = \sum_{n=1}^{m}(-1)^n a_n$ be the partial sums. Our goal is to show that the sequence $\{s_m\}$ converges.

Consider first the even values of m, that is, put $m = 2k$ where $k = 1, 2, \ldots$ Since $s_{2(k+1)} - s_{2k} = a_{2k+2} - a_{2k+1} \leqslant 0$, the numbers s_{2k} form a nonincreasing sequence. This sequence is bounded from below by $-a_1$ because

$$s_{2k} = -a_1 + a_2 - a_3 + a_4 - a_5 + \ldots + a_{2k} = -a_1 + (a_2 - a_3) + (a_4 - a_5) + \cdots + a_{2k}$$

and all the terms in the right hand side with the exception of $-a_1$ are nonnegative. Now the Monotone Convergence Theorem implies that s_{2k} converge to a limit $s \in \mathbb{R}$ as $k \to \infty$.

It remains to show that the whole sequence $\{s_m\}$ converges to the same limit. In order to do this, we need to prove that for any $\varepsilon > 0$ there exists m_ε such that $|s - s_m| < \varepsilon$ for all $m \geqslant m_\varepsilon$.

Since $s_{2k} \to s$, we know that there exists a positive integer m'_ε such that $|s - s_m| < \varepsilon/2$ for all even integers $m \geqslant m'_\varepsilon$. (Here we use the definition of convergence with $\varepsilon/2$ instead of ε, which is justified because ε is an arbitrary positive number.)

On the other hand, since $a_m \to 0$ and $s_{m+1} - s_m = a_m$, we have $|s_{m+1} - s_m| \to 0$. From the definition of convergence it follows that there exists a positive integer m''_ε such that $|s_{m+1} - s_m| < \varepsilon/2$ for all $m \geqslant m''_\varepsilon$. Now, if $m \geqslant m'_\varepsilon$ and $m \geqslant m''_\varepsilon$ and m is odd, we have

$$|s - s_m| \leqslant |s - s_{m+1}| + |s_{m+1} - s_m| < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

because $m + 1$ is even and $m + 1 > m'_\varepsilon$. Thus, if we take $m_\varepsilon = m'_\varepsilon = m''_\varepsilon$ then the estimate $|s - s_m| < \varepsilon$ holds for all $m \geqslant m_\varepsilon$. This proves the theorem.

POWER SERIES

Definition. The series of the form $\sum_{n=1}^{\infty} a_n x^n$ is called a power series. Here a_n are fixed coefficients and x is considered as a parameter, so that the sum of the series is a function of x.

Theorem. Let $\hat{R} = \left(\limsup |a_n|^{1/n}\right)^{-1}$. Then the series $\sum_{n=1}^{\infty} a_n x^n$ is absolutely convergent for $|x| < \hat{R}$ and is divergent for $|x| > \hat{R}$.

Note that $\hat{R}$ may be $+\infty$. The nonnegative "number" $\hat{R}$ is called the *radius of convergence* of the power series. The theorem states that the series absolutely converges inside the open interval $(-\hat{R}, \hat{R})$ and diverges outside the closed interval $[-\hat{R}, \hat{R}]$. It does not give any information about the end points, where $|x| = \hat{R}$.

Proof of the Theorem. Denote $b_n := a_n x^n$. We have

$$\limsup |b_n|^{1/n} = \limsup(|a_n|^{1/n}\,|x|) = c\,|x|$$

where $c := \limsup |a_n|^{1/n}$. Clearly, $\limsup |b_n|^{1/n} < 1$ if and only if $|x| < c^{-1}$, and $\limsup |b_n|^{1/n} > 1$ if and only if $|x| > c^{-1}$. Therefore the Theorem follows from the n-th Root Test.

Remark. Some authors define the radius of convergence $\hat{R}$ by saying that the series is absolutely convergent for all $x \in (-\hat{R}, \hat{R})$ and divergent for $x \notin [\hat{R}, \hat{R}]$. The above theorem shows that this definition makes sense, and also give an explicit formula for $\hat{R}$.

Theorem. Assume that $a_n \neq 0$ and that the sequence of positive numbers $|a_n|/|a_{n+1}|$ converges to a limit c. Then the radius of convergence of the series $\sum_{n=1}^{\infty} a_n x^n$ coincides with c.

Proof. Denote $b_n := a_n x^n$. We have

$$\lim |b_{n+1}/b_n| = \lim |a_{n+1}/a_n|\,|x| = \lim \frac{|x|}{|a_n/a_{n+1}|} = c^{-1}\,x.$$

Ny the ratio test, the series converges if and only if $c^{-1}\,|x| < 1$, that is, if and only if $|x| < c$. Therefore $\hat{R} = c$.

You should memorize the following expansions and the range of x for which the series converge.

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + ... \quad \text{for all } x \in \mathbb{R}$$

$$\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - ... \quad \text{for all } x \in \mathbb{R}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - ... \quad \text{for all } x \in \mathbb{R}$$

$$(1+x)^\alpha = 1 + \alpha\,x + \frac{\alpha(\alpha-1)}{2!}\,x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}\,x^3 ... \quad \text{if } |x| < 1$$

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - ... \quad \text{if } |x| < 1$$

Exercise. Using a suitable test, show that $\hat{R} = \infty$ for the first three series, and $\hat{R} = 1$ for the last two series.

FUNCTIONS OF A REAL VARIABLE

Let $\Omega \subseteq \mathbb{R}$ be a nonempty subset of the real line. A (real-valued) function f on Ω is a mapping $\Omega \to \mathbb{R}$, that is, an association $x \mapsto f(x)$ of each element x of Ω to some real number $f(x)$ which is called the value of the function f at the point x. Further on, we shall usually be assuming that Ω is a nondegenerate interval (which may coincide with $\mathbb{R}$).

The set Ω is called the *domain of definition* of the function. The set of all its values $f(x)$ (when x runs over Ω) is said to be the *range* of f. Finally, the set of points $(x, y) \in \mathbb{R}^2$ such that $y = f(x)$ is called the *graph* of the function f. The graph can be thought of as a curve line in the two dimensional space $\mathbb{R}^2$ whose intersection with every vertical straight line passing through $(x, 0)$ consists of one point with coordinates $(x, f(x))$. Every function is uniquely defined by its graph.

Some functions are given by "nice" explicit formulae. For example, $f(x) = x^2+x+1$ is a function on the whole real line, $f(x) = (\sqrt{x})^{-1}$ is a real-valued function on the positive half-line $(0, +\infty)$, and so on. But it is not always the case. For instance,

$$f(x) \;=\; \begin{cases} 1\,, & \text{if } x \text{ is rational,} \\ 0\,, & \text{if } x \text{ is irrational,} \end{cases}$$

is a well defined function on the real line (do not try to sketch its graph, it is impossible!). Another important example is the function

$$f(x) \;=\; \begin{cases} 1\,, & \text{if } x \in A \\ 0\,, & \text{if } x \notin A, \end{cases}$$

where A is an arbitrary subset of $\mathbb{R}$. This function is called the *characteristic function* of the set A and is often denoted by ḩi_A.

BOUNDED FUNCTIONS

Definition. We say that a function is bounded if its range (the set of its values) is a bounded subset of $\mathbb{R}$.

If a function f is bounded then its range has finite g.l.b. m and l.u.b. M and $m \leqslant f(x) \leqslant M$ for all x from the domain of definition. The numbers m and M are usually denoted $\inf f$ and $\sup f$ and are called the greatest lower bound and the least upper bound of the function f, respectively.

If $f(x) = m$ for some x from the domain of definition, one says that f attains its minimum value m. If $f(x) = M$ for some x from the domain of definition, one says that f attains its maximum value M. It is not always the case.

Example. The function $f(x) = x^{-1}$ on the interval $[1, \infty)$ attains the maximum value 1 but does not take the value $\inf f = 0$.

Example. The function $f(x) = (1-x)\sin(1/x)$ is bounded on $(0, 1]$ with $\inf f = -1$ and $\sup f = 1$. However, it does not take the values ± 1 anywhere, and there is no obvious way of defining it at $x = 0$.

CONTINUOUS FUNCTIONS

Definition. Let f be a function on an interval I (which may coincide with $\mathbb{R}$), and let $c \in I$.

$\lim_{x\to c} f(x) = y$ means that $\forall \varepsilon > 0\, \exists \delta_\varepsilon > 0 \,:\, 0 < |x - c| < \delta_\varepsilon \Rightarrow |f(x) - y| < \varepsilon$.

$\lim_{x\to c+0} f(x) = u$ means that $\forall \varepsilon > 0\, \exists \delta_\varepsilon > 0 \,:\, c < x < c + \delta_\varepsilon \Rightarrow |f(x) - u| < \varepsilon$.

$\lim_{x\to c-0} f(x) = v$ means that $\forall \varepsilon > 0\, \exists \delta_\varepsilon > 0 \,:\, c - \delta_\varepsilon < x < c \Rightarrow |f(x) - v| < \varepsilon$.

Here y, u, v are some real numbers. The number y is called the limit (or limit value) of f at the point c. The numbers u and v are said to be the right and the left limit, respectively.

Note that the limit of f at a given point c may not exist. Even if the limit exists, it may not coincide with the value of f at the point c. The same is true about the right and left limits.

It is clear from the definition that $\lim_{x\to c} f(x)$ exists and is equal to y if and only if the right and the left limits exist and $u = v = y$. The right limit is not defined at the right end point of I, and the left limit is not defined at the left end point.

Definition of continuity. We say that f is continuous at an interior point c of the interval I if $\lim_{x\to c} f(x) = f(c)$ or equivalently, if
$\forall \varepsilon > 0\, \exists \delta_\varepsilon > 0 \,:\, |x - c| < \delta_\varepsilon \Rightarrow |f(x) - f(c)| < \varepsilon\,.$

If the interval contains the right (left) end point c, we say that f is continuous at c if the left (right) limit of f at the point c exists and coincides with $f(c)$.

A function f defined on an interval I is continuous if $\lim_{x\to c} f(x) = f(c)$ for all $c \in I$. If c is an end point of I and c is included in I then we consider the right or the left limit at c. If c does not belong to I, we do not impose any condition on the right or left limit of $f(x)$ at c.

Remark. Clearly, if f is continuous on I then it is also continuous on any smaller interval $I_1 \subset I$.

Remark. Roughly speaking, continuity means that the graph is a continuous line. But this is not a proper definition, since the notion of a "continuous line" is introduced with the use of the above definition of continuity.

Example. $f(x) = \sqrt{x}$ is a continuous function on the nonnegative half-line $[0, +\infty)$. Indeed, if $c = 0$ then $|f(x) - f(c)| = \sqrt{x} < \varepsilon$ whenever $|x - c| = x < \varepsilon^2$, that is, we can take $\delta = \varepsilon^2$. If $c \neq 0$ then

$$\sqrt{x} - \sqrt{c} = (\sqrt{x} + \sqrt{c})^{-1}(x - c) \leqslant c^{-1/2}(x - c).$$

In this case $|\sqrt{x} - \sqrt{c}| < \varepsilon$ whenever $|x - c| < c^{1/2}\varepsilon$, that is, we can take $\delta = c^{1/2}\varepsilon$.

Example. The function $f(x) = 1/x$ is continuous on the half-open interval $(0, 1]$ but is not defined on $[0, 1]$. It is impossible to extend f to a continuous function on $[0, 1]$

Example. Let $f(x) = \sin(1/x)$ for $x \neq 0$ and $f(0) = a$ where a is some real number. Then, for each $a \in \mathbb{R}$, the function $f(x)$ is not continuous at $x = 0$. Indeed, for any fixed number $b \in [-1, 1]$, any open interval about the origin contains infinitely many points at which f takes the value b.

Theorem. A function f is continuous at c if and only if for every sequence of points $x_n \in I$ such that $\lim_{n\to\infty} x_n = c$ we have $\lim_{n\to\infty} f(x_n) = f(c)$.

Proof. Assume first that f is continuous at c, and let $\{x_n\}$ be a sequence which converges to c. In view of the former, for every $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that $|f(x) - f(c)| < \varepsilon$ whenever $|x - c| < \delta_\varepsilon$. On the other hand, since $x_n \to c$, there exists n_ε such that $|x_n - c| < \delta_\varepsilon$ for all $n \geqslant n_\varepsilon$. Then $|f(x_n) - f(c)| < \varepsilon$ for all $n \geqslant n_\varepsilon$. By definition of convergence, this means that $f(x_n) \to f(c)$.

Assume now that f is not continuous at c. Then there exists $\varepsilon > 0$ such that for all $\delta > 0$ we can find x'_δ such that $|x'_\delta - c| < \delta$ but $|f(x'_\delta) - f(c)| \geqslant \varepsilon$. Let us take $\delta = 1/n$ and denote by x_n the corresponding x'_δ. Then $x_n \to c$ because $|x_n - c| < 1/n \to 0$. On the other hand, $|f(x_n) - f(c)| \geqslant \varepsilon$ for all n, which shows that the sequence $f(x_n)$ does not converge to $f(c)$. Thus if $f(x_n) \to f(c)$ for all sequences $\{x_n\}$ such that $x_n \to c$ then the function f must be continuous at c.

The above theorem and the algebraic rules for the limits of sums, products and quotients of sequences immediately imply

Theorem. Let f and g be continuous at a point c functions. Then the functions $f+g$ and $f\,g$ are continuous at c. If $g(c) \neq 0$ then the quotient $\frac{f}{g}$ is also continuous at c.

Example. A polynomial is a continuous function on $\mathbb{R}$. Indeed, in view of the above theorem, it is sufficient to prove that the function $f(x) = x^n$ is continuous for every $n \in \mathbb{N}$. We have $x^n - c^n = (x - c)\,(x^{n-1} + x^{n-2}c + \cdots + x\,c^{n-2} + c^{n-1})$. It follows that

$$|x^n - c^n| \leqslant |x - c|\,(|x|^{n-1} + |x|^{n-2}|c| + \cdots + |x|\,|c|^{n-2} + |c|^{n-1}).$$

If $|x-c| < 1$ then $|x|^{n-1-k}|c|^k < (|c|+1|)^n$ for all $k = 0, \ldots, n-1$ and, consequently, $|x^n - c^n| < n\,(|c| + 1|)^n |x - c|$. Thus we have $|x^n - c^n| < \varepsilon$ whenever $|x - c| < n\,(|c| + 1|)^n\,\varepsilon$, that is, we can take $\delta = n\,(|c| + 1|)^n\,\varepsilon$ in the definition of continuity.

Example. Let $f(x)$ be a function on the interval $[0, 1]$ defined as follows. If x is an irrational number then $f(x) = 0$. If x is rational then it can be uniquely represented as $x = p/q$ with nonnegative coprime integers p and q (that is, integers p and q without a common divisor). In this case we define $f(x) = 1/q$.

The function f is discontinuous at all rational points and is continuous at all irrational points. Indeed, if c is rational then $f(c) \neq 0$ but $f(x_n) = 0$ for any sequence of irrational numbers $x_n \to c$, that is, $f(x_n) \not\to f(c)$. On the other hand, if c is irrational and $\{x_n\}$ is a sequence of rational numbers $x_n = p_n/q_n \to c$ with coprime p_n, q_n then $q_n \to \infty$ (otherwise the sequence would contain infinitely many elements with the same denominator q_n which would not converge to the irrational c). Therefore $f(x_n) \to 0 = f(c)$.

There will be no lectures on the week starting 7 November. Use the opportunity to revise the material covered in lectures. Some of you may need to go through solutions to the last year class test and exercise sheets to see what you got wrong and why.

DO NOT LEAVE THIS UNTIL THE SPRING VACATION!

CONTINUOUS FUNCTION II

Recall that intervals of the form (a,b), $(-\infty,b)$, $(a,+\infty)$ and $(-\infty,+\infty)$ are said to be open. The intervals $[a,b]$, $(-\infty,b]$, $[a,+\infty)$ and $(-\infty,+\infty)$ are said to be closed. Note that the interval $(-\infty,+\infty)$ is both open and closed.

The following four theorems are only true for functions that are continuous on **closed and bounded intervals** $[a,b]$.

Theorem. If $f:[a,b]\to\mathbb{R}$ is continuous function on the closed bounded interval $[a,b]$ then f is bounded.

Proof. If f were unbounded then there would exist a sequence $x_n\in[a,b]$ such that $|f(x_n)|\geqslant n$ for all n. Since the sequence $\{x_n\}$ is bounded, by the Bolzano–Weierstrass theorem it has a subsequence $\{x_{n_k}\}$ which converges to a limit $c\in[a,b]$ as $k\to\infty$. Since f is continuous, we must have $f(x_{n_k})\to f(c)$ as $k\to\infty$. However, $f(c)$ is a finite number and $|f(x_{n_k})|\geqslant n_k\to\infty$ as $k\to\infty$. The obtained contradiction proves the theorem.

Example. The function $f(x)=x$ is continuous on the closed interval $[1,\infty)$ but is not bounded. The function $f(x)=x^{-1}$ is continuous on the bounded half-open interval $(0,1]$ but is not bounded.

Maximum Theorem. If $f:[a,b]\to\mathbb{R}$ is continuous function on the closed bounded interval $[a,b]$ then it attains a maximum value.

Proof. Since f is bounded, its range has the least upper bound $M=\sup f$. Then, for each $n\in\mathbb{N}$, there exists $x_n\in[a,b]$ such that $|f(x_n)-M|<n^{-1}$ (otherwise M would be separated from the range of f and would not be its **least** upper bound). The sequence $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ which converges to a limit $c\in[a,b]$ as $k\to\infty$. Since f is continuous, we have $f(c)=\lim_{k\to\infty}f(x_{n_k})=M$.

Minimum Theorem. If $f:[a,b]\to\mathbb{R}$ is continuous function on the closed bounded interval $[a,b]$ then it attains a minimum value.

Proof. If $m=\inf f$ then $-m=\sup(-f)$. By the maximum theorem, the continuous function $-f$ attains the maximum value $-m$. This means that $-f(x)=-m$ and, consequently, $f(x)=m$ for some $x\in[a,b]$.

Example. The function $f(x)=(1-x)\sin(1/x)$ is continuous on the bounded half-open interval $(0,1]$ but does not attain the minimum and maximum values ±1. The function $f(x)=x^{-1}$ is continuous on the closed unbounded interval $[1,+\infty)$ but does not attain the minimum value 0.

Intermediate Value Theorem. Let $f:[a,b]\to\mathbb{R}$ be a continuous function on the closed bounded interval $[a,b]$, and let d be a number lying between $f(a)$ and $f(b)$. Then there exists $c\in[a,b]$ such that $f(c)=d$.

Proof. Assume that $f(a) \leqslant f(b)$, so that $f(a) \leqslant d \leqslant f(b)$. Let S be the set of numbers $x \in [a,b]$ such that $f(x) \leqslant d$. The set S is bounded from above by b and is not empty because $a \in S$. By the completeness axiom, it has the least upper bound $c = \sup S$. In other words, c is the smallest number in the interval $[a,b]$ such that $f(x) > d$ for all $x > c$.

If $f(c) < d$ then $c < b$ and $f(c) < d - \varepsilon$ for some $\varepsilon > 0$. By continuity, there exists $\delta > 0$ such that $c + \delta < b$ and $f(x) < f(c) + \varepsilon < d$ for all $x \in [c, c+\delta]$. In this case c is not an upper bound for S (since $c + \delta \in S$).

If $f(c) > d$ then $c > a$ and $f(c) > d + \varepsilon$ for some $\varepsilon > 0$. By continuity, there exists $\delta > 0$ such that $c - \delta > a$ and $f(x) > f(c) - \varepsilon > d$ for all $x \in [c-\delta, c]$. In this case c is not the **least** upper bound for S (since $c - \delta$ is also an upper bound).

Assume that $f(a) \geqslant f(b)$, so that $f(b) \leqslant d \leqslant f(a)$. Denote $g = -f$. Then $g(a) \leqslant g(b)$ and $g(a) \leqslant -d \leqslant g(b)$. Therefore, by the above, there exists a point $c \in [a,b]$ such that $g(c) = -d$ and $f(c) = d$.

Corollary. If $f : [a,b] \to \mathbb{R}$ is continuous function on the closed bounded interval $[a,b]$ then its range $f([a,b])$ is the closed interval $[m,M]$ where $m = \inf f$ and $M = \sup f$.

Proof. By the maximum and minimum theorems, there exist points $x_1 \in [a,b]$ and $x_2 \in [a,b]$ such that $f(x_1) = m$ and $f(x_2) = M$. Applying the intermediate value theorem to the interval $[x_1, x_2]$ we see that f takes all possible values between m and M.

Example. The equation $(x^2+1)\sin x = 1$ has a solution in the interval $[0, \pi/2]$. Indeed, $(x^2+1)(\sin x) - 1 = -1$ at the point $x = 0$ and $(x^2+1)(\sin x) - 1 = \pi^2/4$ at the point $x = \frac{\pi}{2}$. Since the function $(x^2+1)(\sin x) - 1$ is continuous, it takes all the values (including 0) from the interval $[-1, \pi^2/4]$.

We know that sums, products and quotients of continuous functions are continuous functions. It turns out that the composition of two continuous functions is also continuous.

Theorem. Let f be a continuous function on a closed interval $[a,b]$, and let $[m,M]$ be its range. If g is a continuous function on $[m,M]$ then the composition $g \circ f(x) = g(f(x))$ is a continuous function on $[a,b]$.

Proof. Let $c \in [a,b]$ and $x_n \in [a,b]$ be an arbitrary sequence which converges to c. Then, since f is continuous, $f(x_n)$ converges to $f(c) \in [m,M]$. Now, since g is continuous, $g(f(x_n))$ converges to $g(f(c))$. Thus we have $g \circ f(x_n) \to g \circ f(c)$ for every sequence $x_n \to c$. In other words, $g \circ f(x) \to g \circ f(c)$ as $x \to c$, which means that the function $g \circ f(x)$ is continuous.

Example. $\sin(1+x^3)$ is continuous at every point $x \in \mathbb{R}$ because $1 + x^3$ is continuous at all points x and $\sin y$ is continuous at all points y.

DIFFERENTIATION

Let f be a function defined on an interval I. We say that $f'(x) = k$ if

$$\forall \varepsilon > 0 \ \exists \, \delta_{x,\varepsilon} > 0 : \ 0 < |x - y| < \delta_{x,\varepsilon} \Rightarrow \left| \frac{f(x) - f(y)}{x - y} - k \right| < \varepsilon \, .$$

Equivalently, putting $y = x + \delta$,

$$\forall \varepsilon > 0 \ \exists \delta_{x,\varepsilon} > 0 : \ 0 < |\delta| < \delta_{x,\varepsilon} \Rightarrow \left| \frac{f(x + \delta) - f(x)}{\delta} - k \right| < \varepsilon,$$

or

$$\lim_{y \to x} \frac{f(x) - f(y)}{x - y} = k.$$

The number k is called the *derivative of* f *at the point* x. If $f'(x)$ exists for every point x from the interval, we can consider f' as a function of the variable x. This function is called *the derivative of* f. Another notation for the derivative is $\frac{\mathrm{d}}{\mathrm{d}x} f(x)$.

It may well happen that the limit does not exist, in which case we say that f is not differentiable at x.

The right and left limits

$$\lim_{y \to x+0} \frac{f(x) - f(y)}{x - y} \quad \text{and} \quad \lim_{y \to x-0} \frac{f(x) - f(y)}{x - y}$$

(if they exist) are said to be the right and left derivatives of f at the point x. The function f is differentiable if both the right and left derivative exist and have the same value. If x is an end point of the interval I then one can speak only about one of these derivatives (the other limit does not make sense).

Theorem. If the derivative $f'(x)$ exists then f is continuous at x.

Proof. If f is differentiable at x then there exists a number $\delta_{x,1} > 0$ such that $\left| \frac{f(x) - f(y)}{x - y} - k \right| < 1$ whenever $0 < |x - y| < \delta_{x,1}$ (because we can take $\varepsilon = 1$ in the definition). Since $\left| \frac{f(x) - f(y)}{x - y} - k \right| \leqslant \left| \frac{f(x) - f(y)}{x - y} \right| + |k|$, this implies that $\left| \frac{f(x) - f(y)}{x - y} \right| < |k| + 1$. Multiplying both parts of this inequality by $|x - y|$, we see that $|f(x) - f(y)| < (|k| + 1)|x - y|$ whenever $0 < |x - y| < \delta_{x,1}$.
Let us fix an arbitrary $\varepsilon > 0$ and choose $\delta > 0$ so small that $\delta < \delta_{x,1}$ and $\delta < ((|k| + 1))^{-1} \varepsilon$. Then, by the above, we have $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$. This means that f is continuous at x.

The converse is not true; a continuous function may not be differentiable.

Example. The function $f(x) = |x|$ is continuous at the point $x = 0$ because $|f(0) - f(y)| = |y| \to 0$ as $y \to 0$. However,

$$\left| \frac{f(0+\delta) - f(0)}{\delta} - k \right| = \left| \frac{|\delta|}{\delta} - k \right| .$$

If $\delta < 0$ then the right hand side is equal to $|k+1|$. If $\delta > 0$ then the right hand side is equal to $|1-k|$. Clearly, for all sufficiently small ε at least one of these numbers is greater than ε. Therefore the derivative at $x = 0$ does not exist.

Theorem. Let f and g be differentiable functions. Then

(a) $\frac{\mathrm{d}}{\mathrm{d}x}(f(x) + g(x)) = f'(x) + g'(x)$;

(b) $\frac{\mathrm{d}}{\mathrm{d}x}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x)$;

(c) $\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$ provided that $g(x) \neq 0$.

Proofs.

(a) $\lim_{y\to x} \frac{f(x)+g(x)-(f(y)+g(y))}{x-y} = \lim_{y\to x}\left(\frac{f(x)-f(y)}{x-y} + \frac{f(x)-f(y)}{x-y}\right)$
$= \lim_{y\to x}\frac{f(x)-f(y)}{x-y} + \lim_{y\to x}\frac{g(x)-g(y)}{x-y} = f'(x) + g'(x)$
because the limit of the sum is equal to the sum of limits.

(b) $\lim_{y\to x} \frac{f(x)g(x)-f(y)g(y)}{x-y} = \lim_{y\to x}\left(\frac{(f(x)-f(y))g(x)}{x-y} + \frac{f(y)(g(x)-g(y)}{x-y}\right)$
$\lim_{y\to x}\frac{(f(x)-f(y))g(x)}{x-y} + \lim_{y\to x}\frac{f(y)(g(x)-g(y))}{x-y}$
$= \left(\lim_{y\to x} g(x)\right)\left(\lim_{y\to x}\frac{f(x)-f(y)}{x-y}\right) + \left(\lim_{y\to x} f(y)\right)\left(\lim_{y\to x}\frac{g(x)-g(y)}{x-y}\right)$
$= f'(x)g(x) + f(x)g'(x)$
because the limits of the products are equal to the products of limits and $\lim_{y\to x} f(y) = f(x)$ as f is continuous.

(c) $\lim_{y\to x}\frac{f(x)g^{-1}(x)-f(y)g^{-1}(y)}{x-y} = \lim_{y\to x}\frac{f(x)g(y)-f(y)g(x)}{(x-y)g(x)g(y)}$
$= \lim_{y\to x}\frac{(f(x)-f(y))g(x)-f(x)(g(x)-g(y))}{(x-y)g(x)g(y)}$
$\lim_{y\to x}\left(\frac{1}{g(x)g(y)}\right)\left(\lim_{y\to x}\frac{(f(x)-f(y))g(x)}{(x-y)} - \lim_{y\to x}\frac{f(x)(g(x)-g(y))}{(x-y)}\right)$
$= \frac{f'(x)g(x)-f(x)g'(x)}{g^2(x)}$
because $\lim_{y\to x}\left(\frac{1}{g(x)g(y)}\right) = \frac{1}{\lim_{y\to x} g(x)g(y)} = \frac{1}{g^2(x)}$ as g is continuous.

Example. The derivative of a constant function is identically equal to zero because in this case $f(x) - f(y) = 0$ for all x and y. This fact and the part (b) imply that $\frac{\mathrm{d}}{\mathrm{d}x}(c\,g(x)) = c\,g'(x)$ for any constant c.

Example. If n is a positive integer then $\frac{\mathrm{d}}{\mathrm{d}x}x^n = \lim_{x\to y} \frac{x^n - y^n}{x-y}$. Substituting the expansion $x^n - y^n = (x-y)(x^{n-1} + x^{n-2}y + \cdots + xy^{n-2} + y^{n-1})$ (see Sheet 1) and passing to the limit, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}x^n = \lim_{y\to x}(x^{n-1} + x^{n-2}y + \cdots + xy^{n-2} + y^{n-1})$$
$$= (x^{n-1} + x^{n-2}x + \cdots + xx^{n-2} + x^{n-1}) = nx^{n-1}\,.$$

Example. $\frac{\mathrm{d}}{\mathrm{d}x}e^x = e^x$.

Proof. We have $\frac{\mathrm{d}}{\mathrm{d}x}e^x = \lim_{y\to x}\frac{e^x - e^y}{x-y} = e^x \lim_{y\to x}\frac{e^{y-x}-1}{y-x} = e^x \lim_{z\to 0}\frac{e^z - 1}{z}$. Recall that the exponential function is given by the series $e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \ldots$ It follows that

$$\frac{e^z - 1}{z} = 1 + \frac{z}{2!} + \frac{z^2}{3!} + \cdots = 1 + z\,F(z)$$

where $F(z) = \frac{1}{2!} + \frac{z}{3!} + \frac{z^2}{4!} + \ldots$ The ratio test shows that the series in the right hand side is absolutely convergent for all z. If $|z| < 1$ then $|F(z)| \leqslant \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots$ because the modulus of each partial sum of the series defining F is estimated by the corresponding partial sum of the series on the right hand side. Thus F is estimated by the constant $C = \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots$ whenever $|z| < 1$. This implies that $\left|\frac{e^z-1}{z} - 1\right| < C\,|z|$ for all z with $|z| < 1$. It follows that $\lim_{z\to 0}\left|\frac{e^z-1}{z} - 1\right| = 0$ and, consequently, $\lim_{z\to 0}\frac{e^z-1}{z} = 1$.

Example. $\frac{\mathrm{d}}{\mathrm{d}x}\sin x = \cos x$ and $\frac{\mathrm{d}}{\mathrm{d}x}\cos x = -\sin x$.

Proof. The trigonometric formulae

$$\sin x - \sin y = 2\,\cos\left(\frac{x+y}{2}\right)\,\sin\left(\frac{x-y}{2}\right),$$

$$\cos x - \cos y = -2\,\sin\left(\frac{x+y}{2}\right)\,\sin\left(\frac{x-y}{2}\right)$$

imply that

$$\frac{\mathrm{d}}{\mathrm{d}x}\sin x = \lim_{y\to x}\frac{\sin x - \sin y}{x-y} = \lim_{y\to x}\cos\left(\frac{x+y}{2}\right)\,\lim_{y\to x}\left(\frac{2}{x-y}\sin\left(\frac{x-y}{2}\right)\right)$$
$$= \cos x\,\lim_{z\to 0}\left(\frac{\sin z}{z}\right)$$

and, similarly,

$$\frac{\mathrm{d}}{\mathrm{d}x}\cos x = -\sin x\,\lim_{z\to 0}\left(\frac{\sin z}{z}\right).$$

Since $\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \ldots$, we have $\sin z = z + z^2\,F(z)$, where

$$F(z) = -\frac{z}{3!} + \frac{z^3}{5!} - \frac{z^5}{7!} + \ldots$$

By the ratio test, the series in the right hand side is absolutely convergent for all z and y. If $|z| < 1$ then

$$|F(z)| \leqslant C = \frac{1}{3!} + \frac{1}{5!} + \frac{1}{5!} + \dots$$

It follows that $|\sin z - z| = |z^2 F(z)| \leqslant C \frac{|z|^2}{2}$. Therefore

$$\lim_{z \to 0} \left| \frac{\sin z}{z} - 1 \right| = \lim_{z \to 0} z^{-1} |\sin z - z| = 0$$

and, consequently, $\lim_{z \to 0} \frac{\sin z}{z} = 1$.

Chain rule (differentiating a function of a function). If $f(x)$ is differentiable at $x = c$ and $g(y)$ is differentiable at $y = f(c)$ then $g(f(x))$ is differentiable at $x = c$ and

$$\frac{\mathrm{d}}{\mathrm{d}x} g(f(x))|_{x=c} = g'(f(c)) f'(c).$$

Proof. By definition,

$$\lim_{x \to c} \frac{g(f(x)) - g(f(c))}{x - c} = \lim_{x \to c} \left(\frac{g(f(x)) - g(f(c))}{f(x) - f(c)} \, \frac{f(x) - f(c)}{x - c} \right)$$
$$\left(\lim_{x \to c} \frac{g(f(x)) - g(f(c))}{f(x) - f(c)} \right) \left(\lim_{x \to c} \frac{f(x) - f(c)}{x - c} \right) = \left(\lim_{x \to c} \frac{g(f(x)) - g(f(c))}{f(x) - f(c)} \right) f'(c) \, .$$

Since f is continuous at c, we have $f(x) \to f(c)$ as $x \to c$. Therefore

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

Example. Let $f(x) = \ln x$ where $x > 0$. By definition, $g(f(x)) = x$ for all x where $g(y) = e^y$. Since $(e^y)' = e^y$, differentiation this identity and applying the previous theorem, we obtain $\frac{\mathrm{d}}{\mathrm{d}x} e^{\ln x} = (\ln x)' \, e^{\ln x} = (\ln x)' x = 1$. Thus we have $(\ln x)' = x^{-1}$.

FINDING MAXIMAL AND MINIMAL VALUES

Definition. Let f be a function defined on an interval (a, b). We say that f has a local maximum at a point $c \in (a, b)$ if there exists $\varepsilon > 0$ such that $f(c) \geqslant f(x)$ for all $x \in (c - \varepsilon, c + \varepsilon)$. Similarly, f has a local minimum at $c \in (a, b)$ if there exists $\varepsilon > 0$ such that $f(c) \leqslant f(x)$ for all $x \in (c - \varepsilon, c + \varepsilon)$.

Theorem. Let f be a differentiable function on the interval (a, b). If f has a local maximum or a local minimum at $c \in (a, b)$ then $f'(c) = 0$.

Proof. If f has a local maximum at c then there exists $\varepsilon > 0$ such that $\frac{f(x)-f(c)}{x-c} \geqslant 0$ for all $x \in (c - \varepsilon, c)$ and $\frac{f(x)-f(c)}{x-c} \leqslant 0$ for all $x \in (c, c + \varepsilon)$. Therefore, in the definition of the derivative, the left limit is nonnegative and the right limit is nonpositive. Since f is differentiable, both these limits exist and coincide. This implies that they are equal to zero. The corresponding result for a local minimum is obtained in a similar way (or by applying the local maximum result to the function $g(x) = -f(x)$).

A function may have several local maxima and minima. For example, $f(x) = \cos x$ has local maxima at the points $x = 2\pi n$ and local minima at the points $\pi + 2\pi n$, where $n = 0, \pm 1, \pm 2, \ldots$

It may well happen that $f'(c) = 0$ but c is not a local mimimum or local maximum of f. Such a point c is called a *saddle* point.

Example. Let $f(x) = x^3$ then $f'(x) = 3x^2 = 0$ at $x = 0$. However, the function f does not have a local minimum or a local maximum at this point.

Finding the maximum value of a differentiable function on an interval. The maximal value of a function f on an interval I either coincides with a local maxima or is attained at an end point of the interval. In order to find it, one has to do the following:
(a) to find all points $c_1, c_2, \ldots \in I$ at which $f'(c_k) = 0$;
(b) to evaluate $f(c_k)$;
(c) to evaluate f at the end points of the interval **if they are included in I;**
(d) to select the point at which f takes the maximal value. This may be either one of the points c_k, or an end point of the interval.

In a similar way one can find the minimum value. **Do not forget (c) and (d)!**

Warning: the equality $f'(c) = 0$ does not imply that $f(c)$ is the maximal (or minimal) value of f. It may well happen that c is a local maximum (or minimum), or that the value of f at an end point is greater (or smaller) than $f(c)$.

MEAN VALUE THEOREMS

Rolle's Theorem. If $f : [a, b] \to \mathbb{R}$ is continuous and it is differentiable at every $x \in (a, b)$ and $f(a) = f(b) = 0$ then there exists a point c in (a, b) at which $f'(c) = 0$.

Proof. A continuous function on a bounded closed interval attains its maximum and minimum values. If both these values are zero, the function is identically equal to zero and $f' = 0$ everywhere. If one of these values is not zero and is attained at the point c then $c \in (a, b)$ and, by the previous theorem $f'(c) = 0$.

Mean Value Theorem. If $f : [a, b] \to \mathbb{R}$ is continuous and it is differentiable at every $x \in (a, b)$ then there exists a point c in (a, b) at which $f'(c) = \frac{f(b)-f(a)}{b-a}$.

Proof. Consider the function

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

Since f is continuous on $[a, b]$ and differentiable on (a, b), the same is true about g. Also, $g(a) = g(b) = 0$. Applying Rolle's Theorem to g, we obtain the required result.

Corollary. If f is differentiable on an interval (a, b) and $f'(x) = 0$ for all $x \in (a, b)$ then f is constant on (a, b).

Proof. Let $a_1, b_1 \in (a, b)$ and $a_1 < b_1$. Applying Mean Value Theorem to the interval $[a_1, b_1]$, we obtain $\frac{f(b_1)-f(a_1)}{b_1-a_1} = 0$, that is, $f(b_1) = f(a_1)$. Since this is true for all $a_1, b_1 \in (a, b)$, the function f is constant.

Cauchy's Mean Value Theorem. Suppose that f and g are continuous on $[a, b]$ and differentiable on (a, b). Suppose further that g' is never zero on (a, b). Then there is some $c \in (a, b)$ such that

$$\frac{f(a) - f(b)}{g(a) - g(b)} = \frac{f'(c)}{g'(c)}.$$

Proof. Note first of all that $g(a) - g(b) \neq 0$. Indeed, if $g(a) = g(b)$ then, by Rolle's theorem, $g'(x) = 0$ at some point $x \in (a, b)$.

Let $\varphi(x) = (g(b) - g(a))\, f(x) - (f(b) - f(a))\, g(x) + f(b)g(a) - f(a)g(b)$. Then $\varphi(a) = \varphi(b) = 0$ and φ satisfies the conditions of Rolle's theorem. Thus there exists $c \in (a, b)$ such that

$$\varphi'(c) = (g(b) - g(a))\, f'(c) - (f(b) - f(a))\, g'(c) = 0.$$

This implies the required result.

TAYLOR'S THEOREM

Definition. We say that f is n times differentiable on (a, b) if each derivative of order up to n exists at every point of the interval (the derivative of order two is the derivative of the first derivative, the derivative of order three is obtained by differentiation the derivative of order three and so on). We say that it is n times continuously differentiable if the final derivative is continuous (the function and its first $(n - 1)$ derivatives are automatically continuous). If the interval is $[a, b]$ then we require the one-sided derivatives of all orders up to n to exist at the end-points if the interval. The usual notation for the derivative of order n is $f^{(n)}$, so that $f^{(n)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} f^{(n-1)}(x)$.

Theorem (Taylor's formula). If f is n times continuously differentiable on the interval $(a-\varepsilon, a+\varepsilon)$ then, for each $x \in (a-\varepsilon, a+\varepsilon)$, there exists a point c lying between a and x (that is, $c \in (a,x)$ if $x > a$ and $c \in (x,a)$ if $x < a$), such that

$$f(x) = f(a) + f'(a)\,(x-a) + f^{(2)}(a)\,\frac{(x-a)^2}{2!} + f^{(3)}(a)\,\frac{(x-a)^3}{3!} + \cdots + f^{(n-1)}(a)\,\frac{(x-a)^{n-1}}{(n-1)!} + R_n\,,$$

where $R_n = f^{(n)}(c)\,\dfrac{(x-a)^n}{n!}$.

Remark. Sometimes it is more convenient to write down Taylor's formula as

$$f(a+h) = f(a) + f'(a)\,h + f^{(2)}(a)\,\frac{h^2}{2!} + f^{(3)}(a)\,\frac{h^3}{3!} + \cdots + f^{(n-1)}(a)\,\frac{h^{n-1}}{(n-1)!} + R_n\,,$$

where $R_n = f^{(n)}(c)\,\dfrac{h^n}{n!}$ and c is a point lying between a and $a+h$.

Proof of Taylor's formula. Let $g(x) = (x-a)^m$ and

$$\tilde f(x) = f(x) - \sum_{k=0}^{n-1} f^{(k)}(a)\,\frac{(x-a)^k}{k!}\,.$$

Note that the functions $\tilde f$ and g and their first $(n-1)$ derivatives are equal to zero at the point $x = a$. Therefore, applying Cauchy's Mean Value Theorem, we obtain

$$\frac{\tilde f(x)}{g(x)} = \frac{\tilde f(x) - \tilde f(a)}{g(x) - g(a)} = \frac{\tilde f'(c_1)}{g'(c_1)} = \frac{\tilde f'(c_1) - f'(a)}{g'(c_1) - g'(a)} = \frac{\tilde f^{(2)}(c_2)}{g^{(2)}(c_2)}$$
$$= \frac{\tilde f^{(2)}(c_2) - \tilde f^{(2)}(a)}{g^{(2)}(c_2) - g^{(2)}(a)} = \frac{\tilde f^{(3)}(c_3)}{g^{(3)}(c_3)} = \cdots = \frac{\tilde f^{(n-1)}(c_{n-1}) - \tilde f^{(n-1)}(a)}{g^{(n-1)}(c_{n-1}) - g^{(n-1)}(a)} = \frac{\tilde f^{(n)}(c_n)}{g^{(n)}(c_n)}\,,$$

where $c_1 \in (a,x)$, $c_2 \in (a,c_1)$, $c_3 \in (a,c_2)$, ..., $c_n \in (a,c_{n-1}) \subset (a,x)$ if $x > a$, and $c_1 \in (x,a)$, $c_2 \in (c_1,a)$, $c_3 \in (c_2,a)$, ..., $c_n \in (c_{n-1},a) \subset (x,a)$ if $x < a$.
Since $g^{(n)}(x) = n!$ and $\tilde f^{(n)}(x) = f^{(n)}(x)$ for all x, we get $\tilde f(x) = \frac{g(x)}{n!}\,f^{(n)}(c_n)$. This is equivalent to Taylor's formula with $c = c_n$.

Remark. Note that the mean value theorem is a particular case of Taylor's theorem corresponding to $n = 1$.

ANALYTIC FUNCTIONS

Definition. If $f(x) = \sum_{k=0}^{\infty} f^{(n)}(a)\,\frac{(x-a)^n}{n!}$ for all $x \in (a-\varepsilon, a+\varepsilon)$ (or, equivalently, $R_m \to 0$ as $n \to \infty$) for all $x \in (a-\varepsilon, a+\varepsilon)$ then the function f is said to be *analytic* on the interval $(a-\varepsilon, a+\varepsilon)$. This series is often called Taylor's expansion of f at the point a.

In other words, the function f is analytic on $(a-\varepsilon, a+\varepsilon)$ if $f(a+h) = \sum_{n=0}^{\infty} f^{(n)}(a)\,\frac{h^n}{n!}$ for all $h \in (-\varepsilon, +\varepsilon)$.

Remark. It may well happen that a function f has infinitely many derivatives, the Taylor series $\sum_{n=0}^{\infty} f^{(n)}(a)\,\frac{(x-a)^n}{n!}$ converges but the function f is not analytic. The function is analytic if the series converges AND its sum is equal to $f(x)$.

Example. Let $f(x) = \begin{cases} e^{-1/x^2}, & x \neq 0\,, \\ 0\,, & x = 0\,. \end{cases}$ For this function, all the derivatives $f^{(n)}$ are equal to zero at the origin. Therefore all terms in Taylor's series with $a = 0$ are equal to zero, and the series does not define the function f.

Assume that $|f^{(n)}(x)| \leqslant C_n$ for all $x \in (a-\varepsilon, a+\varepsilon)$, where C_n are some constants. Then, obviously, $|R_n| \leqslant C_n\,\frac{|x-a|^n}{n!} \leqslant C_n\,\frac{\varepsilon^n}{n!}$. If $C_n\,\frac{\varepsilon^n}{n!} \to 0$ as $n \to \infty$ then

$$\left| f(x) - \sum_{k=0}^{m-1} f^{(n)}(a)\,\frac{(x-a)^n}{n!} \right| \;=\; |R_m| \;\to\; 0$$

as $m \to \infty$ for all $x \in (a-\varepsilon, a+\varepsilon)$. This implies that f is analytic on $(a-\varepsilon, a+\varepsilon)$.

Example. The functions $f(x) = \sin x$ and $f(x) = \cos x$ are analytic in any open interval $(a-\varepsilon, a+\varepsilon)$. Indeed, in either case $|f^{(n)}(x)| \leqslant 1$ for all x, so that $|R_n| \leqslant \frac{\varepsilon^n}{n!} \to 0$ for all h. The function $f(x) = e^x$ is also analytic on any interval because $|f^{(n)}(x)| \leqslant e^{|a|+\varepsilon}$ for all $x \in (a-\varepsilon, a+\varepsilon)$. One can easily check that
$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} \dots$
$\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + ...$
$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + ...$

are their Taylor's expansions at the point 0.

Example. $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$ is Taylor's expansion of the function $f(x) = \log(1+x)$ at $x = 0$. Indeed, we have $f'(x) = (1+x)^{-1}$ and

$$f^{(n)}(x) \;=\; (-1)^{n-1}(n-1)!\,(1+x)^{-n},$$

so that $f'(0) = 1$ and $f^{(n)}(0) = (-1)^{n-1}(n-1)!$.

In this case

$$|f^{(n)}(x)| \leqslant C_n = (n-1)!(1-\varepsilon)^{-n}$$

for all $x \in (-\varepsilon, \varepsilon)$. If $\varepsilon < \frac{1}{2}$ then $\frac{\varepsilon}{1-\varepsilon} < 1$ and

$$C_n \frac{\varepsilon^n}{n!} = \frac{1}{n}\left(\frac{\varepsilon}{1-\varepsilon}\right)^n \underset{n\to\infty}{\to} 0\,.$$

It follows that $\log(1+x)$ is analytic on the interval $(-\frac{1}{2}, \frac{1}{2})$. We shall see later that $\log(1+x)$ is analytic on the bigger interval $(-1,1)$, but the proof of this fact is more complicated.

FUNCTIONS DEFINED BY POWER SERIES

From now on, for the sake of simplicity, we shall be assuming that $a = 0$. All further results remain valid for series of the form $\sum_{n=0}^{\infty} c_n(x-a)^n$ and obtained by substituting $y = x - a$.

Consider a power series $\sum_{n=0}^{\infty} c_n x^n$, and denote be its radius of convergence by $\hat{R}$. Let $f(x) = \sum_{n=0}^{\infty} c_n x^n$. Since the series converges for each $x \in (-\hat{R}, \hat{R})$, one can consider $f(x)$ as a function on the open interval $(-\hat{R}, \hat{R})$.

If $f_m(x) = \sum_{n=0}^{m} c_n x^n$ then $f_m(x) \underset{m\to\infty}{\to} f(x)$ for every $x \in (-\hat{R}, \hat{R})$. Also, $f_m(x)$ are continuous functions. Since $f_m(x) \underset{x\to c}{\to} f_m(c)$ for every $c \in (-\hat{R}, \hat{R})$ and $f_m(x) \underset{m\to\infty}{\to} f(x)$ for each x, it is tempting to deduce that $f(x)$ is also a continuous function on $(-\hat{R}, \hat{R})$. However, this argument is flawed: for general continuous functions f_m the conclusion may well be wrong.

Example. Consider the functions $f_m(x) = \begin{cases} 1 & \text{if } |x| > 1/m, \\ m|x| & \text{if } |x| \leqslant 1/m, \end{cases}$ on the interval $(-1,1)$. Obviously, the functions f_m are continuous. If $x \neq 0$ then $f_m(x) = 1$ for all sufficiently large m. If $x = 0$ then $f(x) = 0$ for all m. Thus the functions $f_m(x)$ converge to the discontinuous function $f(x) = \begin{cases} 1 & \text{if } |x| \neq 0, \\ 0 & \text{if } |x| = 0, \end{cases}$.

Definition. The series converges *uniformly* on an interval $[a,b]$ if there exists a sequence of numbers ε_m such that $\varepsilon_m \to 0$ as $m \to \infty$ and $|f(x) - f_m(x)| \leqslant \varepsilon_m$ for all $x \in [a,b]$.

Theorem. The power series $\sum_{n=0}^{\infty} c_n x^n$ converges uniformly on every closed subinterval $[a,b] \subset (-\hat{R}, \hat{R})$.

Proof. Recall that the power series is absolutely convergent for each $x \in (-\hat{R}, \hat{R})$, that is the series $\sum_{n=0}^{\infty} |c_n|\,|x|^n$ converges. This means that $\sum_{n=0}^{m} |c_n|\,|x|^n \to C(x)$ as $m \to \infty$, where $C(x)$ is some finite number. Therefore

$$\sum_{n=m+1}^{\infty} |c_n|\,|x|^n = C(x) - \sum_{n=0}^{m} |c_n|\,|x|^n \underset{m\to\infty}{\to} 0\,, \qquad \forall x \in (-\hat{R}, \hat{R})\,.$$

Since the interval $[a,b]$ is closed, its end points are are separated from $\pm\hat{R}$. This implies that there exists $r \in (0,\hat{R})$ such that $[a,b] \subset [-r,r]$.

Denote $\varepsilon_m = \sum_{n=m+1}^{\infty} |c_n|\, r^n$. By the above, $\varepsilon_m \to 0$ as $m \to \infty$. Also, since $|x| \leqslant r$ for all $x \in [a,b]$, we have

$$\left|\sum_{n=m+1}^{k} c_n\, x^n\right| \;\leqslant\; \sum_{n=m+1}^{k} |c_n|\, |x|^n \;\leqslant\; \sum_{n=m+1}^{\infty} |c_n|\, r^n \;=\; \varepsilon_m\,, \qquad \forall x \in [a,b]\,,$$

for all $k = m+1, m+2, \ldots$ Passing to the limit as $k \to \infty$, we see that

$$|f(x) - f_m(x)| \;=\; \left|\sum_{n=m+1}^{\infty} c_n\, x^n\right| \;=\; \lim_{k\to\infty}\left|\sum_{n=m+1}^{k} c_n\, x^n\right| \;\leqslant\; \varepsilon_m\,.$$

This proves the theorem.

Theorem. The function $f(x) = \sum_{n=0}^{\infty} c_n x^n$ is continuous on the open interval $(\hat{R},\hat{R})$.

Proof. Let us fix a point $c \in (-\hat{R},\hat{R})$ and a positive number ε. We have to show that there exists $\delta > 0$ such that $|f(x) - f(c)| < \varepsilon$ whenever $|x - c| < \delta$.

In the last lecture we proved that the series converges uniformly on every closed subinterval $[a,b] \subset (-\hat{R},\hat{R})$. In other words, there exist constants ε_m such that $\varepsilon_m \to 0$ as $m \to \infty$ and $|f(x) - f_m(x)| \leqslant \varepsilon_m$, where $f_m(x) = \sum_{n=0}^{m} c_n x^n$.

Let us choose $\delta' > 0$ so small that $[c-\delta', c+\delta'] \subset (-\hat{R},\hat{R})$. Since f_m converge to f uniformly on $[c-\delta', c+\delta']$, we have $|f(x) - f_m(x)| \leqslant \varepsilon/3$ for all $x \in [c-\delta', c+\delta']$ and all sufficiently large m. Let us fixed an arbitrary m, for which this estimate holds.

The function f_m is a polynomial and, therefore, is continuous. It follows that there exist $\delta'' > 0$ such that $|f_m(x) - f_m(c)| < \varepsilon/3$ whenever $|x - c| < \delta''$.

Define $\delta = \min\{\delta', \delta''\}$. If $|x - c| \leqslant \delta$ then, by the above,

$$\begin{aligned} |f(x) - f(c)| &\leqslant |f(x) - f_m(x)| + |f_m(x) - f_m(c)| + |f_m(c) - f(c)| \\ &\leqslant\; \varepsilon/3 + \varepsilon/3 + \varepsilon/3 \;=\; \varepsilon\,. \end{aligned}$$

This completes the proof.

Theorem. The series $\sum_{n=0}^{\infty} c_n (n+1)^{-1} x^{n+1}$ and $\sum_{n=1}^{\infty} c_n n x^{n-1}$, obtained by formal integration and differentiation of the series $\sum_{n=0}^{\infty} c_n x^n$, have the same radius of convergence $\hat{R}$ as $\sum_{n=0}^{\infty} c_n x^n$. If $f(x) = \sum_{n=0}^{\infty} c_n x^n$, $g(x) = \sum_{n=0}^{\infty} c_n (n+1)^{-1} x^{n+1}$ and $h(x) = \sum_{n=1}^{\infty} c_n n x^{n-1}$ then $g' = f$ and $f' = h$.

The function $f(x) = \sum_{n=0}^{\infty} c_n x^n$ is differentiable on the interval $(-\hat{R},\hat{R})$. Its derivative coincides with the sum of the series $\sum_{n=1}^{\infty} c_n n x^{n-1}$ which has the same radius of convergence $\hat{R}$.

Proof uses integrals. It will be given in the last week of the course.

Corollary. The function $f(x) = \sum_{n=0}^{\infty} c_n x^n$ is infinitely differentiable on the open interval $(-\hat{R}, \hat{R})$, and its derivatives are obtained by formal differentiation of the series $\sum_{n=0}^{\infty} c_n x^n$.

Proof is obtained by repeatedly applying the above theorem.

Corollary. If $f(x) = \sum_{n=0}^{\infty} c_n x^n$ then $c_n = \frac{1}{n!} f^{(n)}(0)$, that is, $\sum_{n=0}^{\infty} c_n x^n$ is Taylor's expansion of the function f at the point 0.

Proof. Differentiating the series m times, we see that

$$f^{(m)}(0) \;=\; \sum_{n=0}^{\infty} c_n \frac{d^m}{dx^m}\, x^n \bigg|_{x=0} \;=\; c_n \frac{d^m}{dx^m}\, x^m \bigg|_{x=0} \;=\; m!\, a_m\,.$$

Example. If $f(x) = \log(1+x)$ then $f'(x) = (1+x)^{-1}$. By the geometric progression formula, $(1+x)^{-1} = 1 - x + x^2 - x^3 + \ldots$ This series converges for $|x| < 1$ and diverges for $|x| > 1$, so its radius of convergence is 1 (this can also be shown by applying the ratio test). Integrating, we obtain $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$ It follows that $\log(1+x)$ is an analytic function on $(-1, 1)$.

INTEGRATION LECTURE NOTES

by E B Davies

1. Introduction

If f is a real-valued function on a bounded interval $[a,b]$ the integral $\int_a^b f(x)\,\mathrm{d}x$ is intended to be a measure of the area under the graph of the function. If the function takes both positive and negative values then one defines the integral to be the difference of the areas of A and B where

$$A = \{(x,y) :\ a \leqslant x \leqslant b \ \text{ and } 0 < y < f(x)\}$$
$$B = \{(x,y) :\ a \leqslant x \leqslant b \ \text{ and } 0 > y > f(x)\}.$$

The above idea of integral transfers the problem of giving a precise definition of integral to that of giving a precise definition of area, which turns out to be no easier. We shall not attempt to define integration from first axioms in these notes, but concentrate on the properties that a successful integration procedure should have, and then develop the theory from there.

The first issue is that one cannot hope to integrate every conceivable function, obtaining a well-defined real number as its integral. In some cases the integral is infinite, in others both sets A and B have infinite areas, and there is no sensible meaning to $\infty - \infty$, while in yet others the function may be so irregular that it is not clear how to start to define its integral.

In these notes we only consider the integrals of piecewise continuous functions on a bounded interval $[a,b]$. Piecewise continuous functions $f : [a,b] \to \mathbb{R}$ are functions for which there exist numbers

$$a = a_0 < a_1 < \cdots < a_n = b$$

such that f is continuous on each interval (a_{k-1}, a_k) and the limits

$$\lim_{x\to a_k - 0} f(x)\,, \qquad \lim_{x\to a_k + 0} f(x)$$

exist for all relevant k. In other words, this means that f has a jump discontinuity of size $f(a_k+0) - f(a_k-0)$ at each point a_k, but is otherwise continuous. The space PC of such functions contains the space $\mathcal{S}$ of step functions and the space C of continuous functions on $[a,b]$ (recall that a step function f is a piecewise continuous functions in the sense of our definition, which takes constant values on each interval (a_{k-1}, a_k)).

It is evident from the definition that if f and g are piecewise continuous functions on $[a,b]$ and $\alpha, \beta \in \mathbb{R}$ then the combination $(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x)$ is also piecewise continuous. The set of jump points of $\alpha f + \beta g$ is just the union of the jump points of f and of g, or possibly a smaller set if the new function has a jump of size zero between two consecutive intervals. We summarize this by saying that PC is a vector space of functions. It is not finite-dimensional because it contains the polynomials of all orders.

If we write

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

to abbreviate the notation, then $I : PC \to \mathbb{R}$ is a real valued function on PC, assigning the value $I(f) \in \mathbb{R}$ to each piecewise continuous function $f \in PC$. Three properties of this integral are of importance.

Proposition 1.

(1) The map $I :\ PC \to \mathbb{R}$ is linear, that is, $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ for all $\alpha, \beta \in \mathbb{R}$ and $f, g \in PC$.

(2) If $f \in PC$ and $f \geqslant 0$ in the sense that $f(x) \geqslant 0$ for all $x \in [a, b]$ then $I(f) \geqslant 0$.

(3) If $f(x) = c$ for all $x \in [a, b]$ then $I(f) = c(b - a)$.

(4) If $a \leqslant b \leqslant c$ then $I_{a,c}(f) = I_{a,b}(f) + I_{b,c}(f)$, where the subscripts denote the interval of integration.

All of the other properties of integrals can be proved from the above five statements, which we take as **axioms**, in the same way as we assumed the least upper bound property for real numbers, rather than proving it by constructing the reals from first principles. We need to know that $I(f) = 0$ if $a = b$, but this follows from hypothesis (3).

Elementary but important consequences of Proposition 1 are:

(5) If $f \geqslant g$ in the natural sense then $I(f) \geqslant I(g)$.

(6) $|I(f)| \leqslant I(|f|)$ for all $f \in PC$.

(7) If $|f(x)| \leqslant c$ for all $x \in [a, b]$ then $|I(f)| \leqslant c(b - a)$.

(8) If $|f(x) - g(x)| \leqslant \varepsilon$ for all $x \in [a, b]$ then $|I(f) - I(g)| \leqslant \varepsilon(b - a)$.

Remark. If $\varepsilon > 0$ is small, then (8) can be rephrased as: if f and g are uniformly very close to each other then their integrals are almost equal.

Proofs.

(5) By (1) and (2), $I(f) - I(g) = I(f - g) \geqslant 0$.

(6) Since $f(x) \leqslant |f(x)|$, (5) implies that $I(f) \leqslant I(|f|)$. On the other hand, since $-|f| \leqslant f$, from (1) and (5) it follows that $-I(|f|) = I(-|f|) \leqslant I(f)$. Finally, the inequalities $-I(|f|) \leqslant I(f) \leqslant I(|f|)$ are equivalent to the estimate $|I(f)| \leqslant I(|f|)$.

(7) By the above, $|I(f)| \leqslant I(|f|) \leqslant I(c)$. According to (3), the right hand side coincides with $c\,(b - a)$.

(8) By (1), $I(f) - I(g) = I(f-g)$. Therefore (8) is a special case of (7).

From now on we will use the standard notation $\int_a^b f(x)\,\mathrm{d}x$ rather than our new notation $I(f)$, but we insist on basing our proofs on the properties listed above.

2. Theorems about integration

Our goal in this section is to show that all the standard theorems of integration theory follow from Proposition 1. The most important of these is the fundamental theorem of calculus, which we separate into two parts.

We define the indefinite integral F of a piecewise continuous function f on $[a,b]$ by $F(x) = \int_a^x f(s)\,\mathrm{d}s$, where $x \in [a,b]$ is considered as a variable, so that F is a function of x.

Theorem 2. The indefinite integral $F(x)$ is a continuous function of x on $[a,b]$.

Proof Let $c \in [a,b]$ and $\varepsilon > 0$. We need to show that there exists $\delta > 0$ such that $|F(x) - F(c)| < \varepsilon$ whenever $|x - c| < \delta$.

If $x > c$ then the axioms (4) implies that

$$|F(x) - F(c)| \;=\; \left|\int_a^x f(s)\,\mathrm{d}s - \int_a^c f(s)\,\mathrm{d}s\right| \;=\; \left|\int_c^x f(s)\,\mathrm{d}s\right|.$$

Now from (7) it follows that $|F(x) - F(c)| \leqslant (x - c)\sup|f|$. Thus we can take $\delta = \varepsilon\,(\sup|f|)^{-1}$. This proves the right continuity.

If $x < c$ then similar arguments show that $|F(x) - F(c)| = \left|\int_x^c f(s)\,\mathrm{d}s\right|$ and, consequently, $|F(x) - F(c)| \leqslant (c - x)\sup|f|$. This implies the left continuity at c.

The first half of the fundamental theorem of calculus is:

Theorem 3. If f is piecewise continuous on $[a,b]$ and continuous at a particular point $x \in (a,b)$ then F is differentiable at x and $F'(x) = f(x)$.

Proof If $\delta > 0$ then

$$\delta^{-1}\left(F(x+\delta) - F(x)\right) = \delta^{-1}\left(\int_a^{x+\delta} f(s)\,\mathrm{d}s - \int_a^x f(s)\,\mathrm{d}s\right) = \delta^{-1}\int_x^{x+\delta} f(s)\,\mathrm{d}s.$$

Therefore

$$\begin{aligned}|\delta^{-1}\left(F(x+\delta) - F(x)\right) - f(x)| \;&=\; \left|\delta^{-1}\int_x^{x+\delta} f(s)\,\mathrm{d}s - \delta^{-1}\int_x^{x+\delta} f(x)\,\mathrm{d}s\right| \\ &=\; \delta^{-1}\left|\int_x^{x+\delta}\{f(s) - f(x)\}\,\mathrm{d}s\right| \;\leqslant\; \delta^{-1}\int_x^{x+\delta}|f(s) - f(x)|\,\mathrm{d}s.\end{aligned}$$

Since f is continuous at x, for any $\varepsilon > 0$ there exists $\delta > 0$ such that if $|x - s| < \delta$ then $|f(s) - f(x)| < \varepsilon$. Hence

$$|\delta^{-1}\{F(x+\delta)-F(x)\}-f(x)| \leqslant \delta^{-1}\int_x^{x+\delta}|f(s)-f(x)|\,\mathrm{d}s \leqslant \delta^{-1}\int_x^{x+\delta}\varepsilon\,\mathrm{d}s = \varepsilon.$$

This proves that F is differentiable from the right with right derivative $f(x)$. The proof of the left differentiability is similar.

The proof of our next result uses the following consequence of the mean value theorem.

Lemma 4. If a continuous function $f : [a, b] \to \mathbb{R}$ is differentiable and $f'(x) = 0$ for all $x \in (a, b)$, then f is a constant function.

We are now able to prove the other half of the fundamental theorem of calculus, in which we differentiate before integrating, rather than after.

Theorem 5. Suppose $f : [a, b] \to \mathbb{R}$ is continuous and has a continuous derivative on (a, b) with finite limits $\lim_{x\to a+0} f'(x)$ and $\lim_{x\to b-0} f'(x)$ at the end points. Then

$$f(x) = f(a) + \int_a^x f'(s)\,\mathrm{d}s\,, \qquad \forall x \in [a, b].$$

<u>Proof</u> Let

$$G(x) = f(a) + \int_a^x f'(s)\,\mathrm{d}s - f(x)\,.$$

It follows from Theorem 3 that $G'(x) = 0$ for all $x \in (a, b)$. Since $G(a) = 0$, Lemma 4 now establishes that G is identically zero.

Other standard integration procedures can now be proved easily, for example the procedure for integrating by parts.

Theorem 6. If functions $f, g : [a, b] \to \mathbb{R}$ satisfy the conditions of Theorem 5 then

$$\int_a^b f(s)g'(s)\mathrm{d}s = f(b)g(b) - f(a)g(a) - \int_a^b f'(s)g(s)\,\mathrm{d}s.$$

<u>Proof</u> If we rewrite the formula as

$$f(b)g(b) = f(a)g(a) + \int_a^b (fg)'(s)\,\mathrm{d}s$$

then it becomes a special case of Theorem 5.

The next theorem justifies the process of integration by substitution.

Theorem 7. Let $f : [a, b] \to \mathbb{R}$ be a strictly increasing function satisfying the conditions of Theorem 5. If $f(a) = c$, $f(b) = d$ and g is a continuous function on the interval $[c, d]$ then

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

Proof Consider the function

$$G(s) = \int_c^{f(s)} g(y)\,\mathrm{d}y - \int_a^s g\left(f(x)\right) f\prime(x)\,\mathrm{d}x.$$

Direct calculations show that $G(a) = 0$ and $G\prime(s) = 0$ for all s. Therefore G is identically zero and $G(b) = 0$.

Example. The integral $\int_a^b f(x^2)\,x\,\mathrm{d}x$ can be rewritten in the form $\frac{1}{2}\int_a^b f(g(x))g\prime(x)\,\mathrm{d}x$ with $g(x) = x^2$. Applying Theorem 7, we obtain

$$\int_a^b f(x^2)\,x\,\mathrm{d}x \;=\; \frac{1}{2}\int_{a^2}^{b^2} f(x)\,\mathrm{d}x\,.$$

Theorem 8. Let $a(s)$ and $b(s)$ be differentiable functions of s taking values in an interval $[A, B]$ such that $a(s) \leqslant b(s)$, and let $g[A, B] \to \mathbb{R}$ be a continuous function of x. Then

$$\frac{\mathrm{d}}{\mathrm{d}s}\int_{a(s)}^{b(s)} g(x)\,\mathrm{d}x = g(b(s))b'(s) - g(a(s))a'(s)$$

for all s.

Proof If we define

$$F(s) := \int_A^s g(x)\,\mathrm{d}x$$

then the integral equals $F(b(s)) - F(a(s))$ and this is a restatement of the rule

$$\frac{\mathrm{d}}{\mathrm{d}s}[F(b(s)) - F(a(s))] = F'(b(s))b'(s) - F'(a(s))a'(s).$$

The differentiability of the LHS and the validity of the formula are standard results in the theory of differentiation.

Definition. If $x < y$ then we define $\int_y^x f(s)\,\mathrm{d}s = -\int_x^y f(s)\,\mathrm{d}s$. This is just a convenient agreement, under which all previous statements about integrals of the form $\int_y^x$ remain valid without the assumption $y < x$.

3. Integrating and differentiating power series

Let $f(x)=\sum_{n=0}^{\infty} c_n\, x^n$, and let $\hat{R}$ be the radius of convergence. Recall that that the function f defined by the power series is continuous on $(-\hat{R},\hat{R})$ and the power series converges uniformly on every closed subinterval interval $[a,b]\subset(-\hat{R},\hat{R})$, that is, there exists a sequence ε_m such that $\varepsilon_m\to 0$ as $m\to\infty$ and

$$\left|f(x)-\sum_{n=0}^{m} c_n\, x^n\right| \leqslant \varepsilon_n\,, \qquad \forall x\in[a,b]\,.$$

Theorem 9. We have $\int_0^x f(y)\,\mathrm{d}y=\sum_{n=0}^{\infty} c_n\,(n+1)^{-1}\,x^{n+1}$, where the series in the right hand side has the same radius of convergence $\hat{R}$.

Proof. The radius of convergence of the power series $\sum_{n=0}^{\infty} c_n\,(n+1)^{-1}\,x^{n+1}$ is equal to $\left(\limsup |c_n\,(n+1)^{-1}|^{1/n}\right)^{-1}$ (see Week 4). By definition of the upper limit, there exist positive integers $n_1<n_2<n_3\ldots$ such that

$$\limsup_{n\to\infty}\left|c_n\,(n+1)^{-1}\right|^{1/n} \;=\; \lim_{k\to\infty}\left|c_{n_k}\,(n_k+1)^{-1}\right|^{1/n_k} \;=\; \lim_{k\to\infty}|c_{n_k}|^{1/n_k}\;(n_k{+}1)^{-1/n_k}\,.$$

By the "powers beat logarithm" theorem (see Numbers and Functions lecture notes), $\frac{\log(n+1)}{n}\to 0$ and, consequently,

$$(n_k+1)^{-1/n_k} \;=\; e^{\frac{-\log(n_k+1)}{n_k}} \;\to\; 1 \quad \text{as } k\to\infty\,.$$

It follows that

$$\lim_{k\to\infty}|c_{n_k}|^{1/n_k}\;(n_k+1)^{-1/n_k}=\lim_{k\to\infty}|c_{n_k}|^{1/n_k}\,\lim_{k\to\infty}(n_k+1)^{-1/n_k}=\lim_{k\to\infty}|c_{n_k}|^{1/n_k}$$

for any sequence of positive integers $n_1<n_2<n_3\ldots$. Since $\hat{R}=\left(\limsup|c_n|^{1/n}\right)^{-1}$, it follows that the radius of convergence of the series $\sum_{n=0}^{\infty} c_n\,(n{+}1)^{-1}\,x^{n+1}$ is equal to $\hat{R}$.

It remains to show that $\int_0^x f(y)\,\mathrm{d}y=\sum_{n=0}^{\infty} c_n\,(n+1)^{-1}\,x^{n+1}$ for all $x\in(-\hat{R},\hat{R})$. Assume, for the sake of definiteness, that $x>0$. If $x<0$, the equality is proved by applying the same arguments to the interval $[x,0]$.

Let $f_m(y)=\sum_{n=0}^{m} c_n\,y^n$. Since the series is uniformly convergent on the closed subinterval $[0,x]\subset(-\hat{R},\hat{R})$, there exist constants $\varepsilon_m>0$ such that $|f(y)-f_m(y)|\leqslant\varepsilon_m$ for all $y\in[0,x]$ and $\varepsilon_m\to 0$ as $m\to\infty$.

Clearly, $\int_0^x f_m(y)\,\mathrm{d}y = \sum_{n=0}^m c_n\,\frac{x^{n+1}}{n+1}$. Also, from the basic properties of the integral it follows that

$$\left|\int_0^x f(y)\,\mathrm{d}y - \int_0^x f_m(y)\,\mathrm{d}y\right| \leqslant \left|\int_0^x (f(y) - f_m(y))\,\mathrm{d}y\right|$$
$$\leqslant \int_0^x |f(y) - f_m(y)|\,\mathrm{d}y \leqslant \int_0^x \varepsilon_m\,\mathrm{d}y \leqslant \varepsilon_m\, x\,.$$

The right hand side converges to zero as $m \to \infty$, which implies that

$$\sum_{n=0}^{\infty} c_n\,(n+1)^{-1}\,x^{n+1} = \lim_{m\to\infty}\sum_{n=0}^{m} c_n\,\frac{x^{n+1}}{n+1} = \lim_{m\to\infty}\int_0^x f_m(y)\,\mathrm{d}y = \int_0^x f(y)\,\mathrm{d}y\,,$$

that is, $\int_0^x f(y)\,\mathrm{d}y \;=\; \sum_{n=0}^{\infty} c_n\,(n+1)^{-1}\,x^{n+1}$.

Corollary 10. Let $f(x) = \sum_{n=1}^{\infty} c_n\,x^n$ and $g(x) = \sum_{n=1}^{\infty} c_n\,nx^{n-1}$, and let $\hat{R}$ be the radius of convergence of the second series. Then the first series also converges on $(-\hat{R},\hat{R})$ and $g = f'$ for all $x \in (-\hat{R},\hat{R})$.

Proof. Integrating the series $\sum_{n=1}^{\infty} c_n\,ny^{n-1}$ from 0 to x we obtain $\sum_{n=1}^{\infty} c_n\,x^n$. Therefore $\sum_{n=1}^{\infty} c_n\,x^n$ has the same radius of convergence, and the required result is a consequence of the previous theorem and the fundamental theorem of calculus.

4. Improper integrals

This section is not examinable and, therefore, has been removed from the notes.