5CCM221A and 5CCM225A Real Analysis

Alexander Pushnitski

19 April 2021
King's College London

Contents

1 Preliminaries

This section serves as a warm-up. We shall review the necessary topics from Sequences and Series (Sets on the Real Line and Sequences) and then discuss the simplest aspects of the analysis of functions of real variable: natural domain, boundedness, monotonicity. Then we shall review the concept of the limit of a function, which should be familiar from Calculus I. Finally, we shall discuss in detail the use of the $O(\cdot)$ and $o(\cdot)$ notation.

1.1 Recommended literature

These lecture notes should be sufficient for most purposes. If you wish to go a little beyond the lecture material, you can use the excellent book [Rudin] (see the bibliography at the end); in fact, our presentation will follow this book quite closely in Sections 6–10. Two other very useful books are [Haggarty] and [Brannan] (the latter is an Open University textbook designed for self-study). For background reading it suffices to consult lecture notes on Calculus I and Sequences and Series.

1.2 Brief Revision

Here is a brief reminder of relevant topics from *Sequences and Series*. For details, please check the *Sequences and Series* lecture notes.

1.2.1 Sets on the real line

Intervals We will deal with various sets on the real line $\mathbb{R}$. With $a < b$ we can form the following intervals:

- $[a, b] = \{x : a \leqslant x \leqslant b\}$ and $(a, b) = \{x : a < x < b\}$;
- $[a, b) = \{x : a \leqslant x < b\}$ and $(a, b] = \{x : a < x \leqslant b\}$;
- $(a, \infty) = \{x : a < x\}$ and $[a, \infty) = \{x : a \leqslant x\}$;
- $(-\infty, b) = \{x : x < b\}$ and $(-\infty, b] = \{x : x \leqslant b\}$;
- $(-\infty, \infty) = \mathbb{R}$.

Bounded sets A set $A \subset \mathbb{R}$ is said to be *bounded above*, if $A \subset (-\infty, M]$ for some $M \in \mathbb{R}$; any such M is called an *upper bound* of A. Similarly, A is said to be *bounded below*, if $A \subset [m, \infty)$ for some $m \in \mathbb{R}$, and any such m is called a *lower bound* of A. If A is both bounded above and bounded below, it is called *bounded*; otherwise it is called *unbounded*.

Open and closed sets We shall not discuss *general* open or closed sets on the real line in any detail, but here are some definitions for your information. If $x \in \mathbb{R}$, then a *neighbourhood* of x is any open interval (a, b) such that $x \in (a, b)$. Most commonly, one takes this interval of the form $(x - \varepsilon, x + \varepsilon)$ with $\varepsilon > 0$; in this case it is called an ε-neighbourhood of x. A point $x \in A$ is called an *interior point* of A, if A contains a neighbourhood of x. A set $A \subset \mathbb{R}$ is *open*, if it contains a neighbourhood of its every point (i.e. if all points of A are its interior points). A set $B \subset \mathbb{R}$ is *closed*, if its complement is open. These concepts will be discussed in more details in the *Metric Spaces and Topology* module.

Supremum and infimum

A number M is called a *least upper bound*, or a *supremum* of A, if M is an upper bound of A and if there is no $M' < M$ which is also an upper bound of A. Similarly, m is a *greatest lower bound* or an *infimum* of A, if m is a lower bound of A and no $m' > m$ is a lower bound.

The *completeness axiom* of real numbers asserts that every set that is bounded above has a supremum, and every set bounded below has an infimum.

Countable sets

Definition. A set A on the real line is called *countable*, if it is either finite or there is a one-to-one map between A and $\mathbb{N}$. If A is not countable, it is called *uncountable*.

Example. $\mathbb{Z}$, $\mathbb{Q}$ are countable; $\mathbb{R}$ is uncountable.

1.2.2 Sequences

A sequence of real numbers $\{a_n\}_{n=1}^{\infty}$ is said to *converge* to a *limit* $a \in \mathbb{R}$, if

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

Theorem (The Algebra of Limits for sequences)**.** *If* $\lim_{n\to\infty} a_n = a$ *and* $\lim_{n\to\infty} b_n = b$, *then*

- $\lim_{n\to\infty}(a_n \pm b_n) = a \pm b$;
- $\lim_{n\to\infty} a_n b_n = ab$;
- $\lim_{n\to\infty} a_n/b_n = a/b$, *if* $b \neq 0$.

A sequence $\{a_n\}_{n=1}^{\infty}$ is said to be a *Cauchy sequence*, if

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

Theorem (Cauchy's criterion)**.** *A sequence converges if and only if it is a Cauchy sequence.*

A sequence of real numbers $\{a_n\}_{n=1}^{\infty}$ is called *non-decreasing* (resp. *non-increasing*), if $a_{n+1} \geqslant a_n$ (resp. $a_{n+1} \leqslant a_n$) for all n.

Theorem. *If the sequence* $\{a_n\}_{n=1}^{\infty}$ *is non-decreasing and bounded above (or non-increasing and bounded below), then it converges.*

If $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers, and if $\{n_k\}_{k=1}^{\infty}$ is a strictly increasing (i.e. $n_{k+1} > n_k$ for all k) sequence of natural numbers, then $\{a_{n_k}\}_{k=1}^{\infty}$ is said to be a *subsequence* of $\{a_n\}_{n=1}^{\infty}$. A point $a \in \mathbb{R}$ is said to be a *limit point* of a sequence $\{a_n\}_{n=1}^{\infty}$, if there exists a subsequence $\{a_{n_k}\}_{k=1}^{\infty}$ which converges to a as $k \to \infty$.

Theorem (Bolzano-Weierstrass)**.** *Let* $\{x_n\}_{n=1}^{\infty}$ *be a sequence of real numbers such that* $x_n \in [a, b]$ *for all* $n \in \mathbb{N}$. *Then* $\{x_n\}_{n=1}^{\infty}$ *has a limit point in* $[a, b]$.

1.3 Functions of a real variable

Let $\Delta \subseteq \mathbb{R}$ be a nonempty subset of the real line (typically, it is an interval or a union of several intervals). A (real-valued) function f on Δ is a mapping $\Delta \to \mathbb{R}$, that is, an association $x \mapsto f(x)$ of each element x of Δ to some real number $f(x)$ which is called the value of the function f at the point x. The set Δ is called the *domain of definition* of the function. The set of all its values $f(x)$ (when x runs over Δ) is said to be the *range* of f; it is usually denoted by $\operatorname{ran} f$. Finally, the set of points $(x, y) \in \mathbb{R}^2$ such that $y = f(x)$ is called the *graph* of the function f. In symbols:

$$\Delta \subset \mathbb{R}, \quad f : \Delta \to \operatorname{ran} f \subset \mathbb{R}.$$

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

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

is well defined on the real line (do not try to sketch its graph, it is impossible!). More generally, if A is an arbitrary subset of $\mathbb{R}$, then the function

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

is called the *characteristic function* of the set A and is often denoted by χ_A or by 1_A. Dirichlet function corresponds to the case $A = \mathbb{Q}$.

Another nice example is *Thomae's function*, defined by

$$f(x) = \begin{cases} \frac{1}{q} & \text{if } x = \frac{p}{q},\ p \in \mathbb{Z},\ q \in \mathbb{N},\ p \text{ and } q \text{ coprime} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

One can also define a function as the sum of an infinite series. For example, the exponential function can be defined by the power series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

which converges on the whole real line, and the Riemann Zeta function is defined by the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

which converges for $s > 1$.

If a function $f(x)$ is given by an explicit formula, its *natural domain* is the set of values $x \in \mathbb{R}$ such that the formula makes sense; that is, there is no division by zero, all square roots and logarithm are taken of non-negative numbers, etc. *Natural domain* is not a very rigorous mathematical notion but it is convenient to use in concrete examples.

Example. The natural domain of $f(x) = \sqrt{4 - x^2}$ is $[-2, 2]$.

The natural domain of

$$f(x) = \frac{\log(x-2)}{x-5}$$

is $(2, 5) \cup (5, \infty)$.

The natural domain of $f(x) = \sqrt{\sin \sqrt{x}}$ is

$$[0, \pi^2] \cup [(2\pi)^2, (3\pi)^2] \cup [(4\pi)^2, (5\pi)^2] \cup \dots.$$

A function f defined on the whole of $\mathbb{R}$ (or on a symmetric interval $(-R, R)$) is called *even* if $f(x) = f(-x)$ and *odd* if $f(x) = -f(-x)$ for all x. A function f defined on the whole of $\mathbb{R}$ is called *periodic with period* T, if $f(x + T) = f(x)$ for all $x \in \mathbb{R}$.

Example. $\sin$ is odd 2π-periodic, $\cos$ is even 2π-periodic and $\tan$ is odd π-periodic.

1.4 Boundedness

Definition. A function is said to be *bounded* if its range is a bounded subset of $\mathbb{R}$. In other words, f is bounded if there exists a number $R > 0$ such that $|f(x)| \leqslant R$ for all x in the domain of f. If f is not bounded, it is said to be *unbounded*.

Example. The functions $\sin x$, $1/(1 + x^2)$, $\tan^{-1}(x)$, $x - \lfloor x \rfloor$ are bounded, and x, x^2, $1/(x - 1)$, $\tan x$, e^x are unbounded.

The notion of boundedness is often applied to a function restricted onto a set:

Example. The function $f(x) = \sqrt{x}$ is bounded on each interval of the form $[0, R]$ for $R > 0$, but it is unbounded on $[0, \infty)$. The function $\tan x$ is bounded on any interval of the form $(-\frac{\pi}{2} + \varepsilon, \frac{\pi}{2} - \varepsilon)$ for $\varepsilon > 0$; but it is unbounded on $(-\frac{\pi}{2}, \frac{\pi}{2})$. The function $f(x) = \frac{1}{x} - \frac{1}{x+x^2}$ is bounded on the interval $[-1/2, 1/2]$, even though each of the two terms $1/x$, $1/(x + x^2)$ separately is unbounded.

1.5 Monotonicity

Definition. Let f be a function on an interval $\Delta \subset \mathbb{R}$. f is called *increasing* (resp. *non-decreasing*) on Δ, if for any $x_1 < x_2$ on Δ we have $f(x_1) < f(x_2)$ (resp. $f(x_1) \leqslant f(x_2)$). Similarly, f is called *decreasing* (resp. *non-increasing*) on Δ, if for any $x_1 < x_2$ on Δ we have $f(x_1) > f(x_2)$ (resp. $f(x_1) \geqslant f(x_2)$).

Example. The function x^2 is decreasing on $(-\infty, 0]$ and increasing on $[0, \infty)$. The functions $\sin x$ and $\tan x$ are increasing on $(-\pi/2, \pi/2)$. The function e^{ax} is increasing on $\mathbb{R}$ for $a > 0$ and decreasing for $a < 0$. The function $\lfloor x \rfloor$ is non-decreasing on $\mathbb{R}$.

1.6 The limit of a function

Let $x_0 \in \mathbb{R}$ and let f be a function defined on a *punctured neighbourhood* of x_0, i.e. on $(x_0 - a, x_0 + a) \setminus \{0\} = (x_0 - a, x_0) \cup (x_0, x_0 + a)$ for some $a > 0$. (The function f may or may not be defined at the point x_0; this is not important for the following definition.)

Definition. We say that $f(x)$ converges to y_0 as $x \to x_0$ and write $\lim_{x \to x_0} f(x) = y_0$, if

$$\boxed{\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x : \; |x - x_0| < \delta \; \Rightarrow |f(x) - y_0| < \varepsilon.}$$

Example. The function $\frac{\sin x}{x}$ is defined for all $x \neq 0$, and $\lim_{x \to 0} \frac{\sin x}{x} = 1$.

The function $x \log |x|$ is defined for all $x \neq 0$, and $\lim_{x \to 0} x \log |x| = 0$.

Remark. We can rewrite this definition in the following way:

For any neighbourhood $U = (y_0 - \varepsilon, y_0 + \varepsilon)$ of y_0 there exists a neighbourhood $V = (x_0 - \delta, x_0 + \delta)$ of x_0 such that if $x \in V$, then $f(x) \in U$.

At a first glance, this seems more complicated; however, it turns out that the terminology of neighbourhoods is the one that allows the most far-reaching generalisations. This approach will be developed further in the *Metric Spaces and Topology* course.

In a similar spirit, one defines infinite limits and the limits at infinity. Here one should think of ∞ as special "singular point" on the real line, and its neighbourhoods are sets of the form (R, ∞); similarly, the neighbourhoods of $-\infty$ are sets of the form $(-\infty, -R)$. As an example, we give

Definition. Let f be a function defined on (a, ∞) for some $a \in \mathbb{R}$. We write $\lim_{x \to \infty} f(x) = y_0$, if for any $\varepsilon > 0$ there exists $R > 0$ such that for all $x > R$ we have $|f(x) - y_0| < \varepsilon$.

We can rewrite this as:

For any neighbourhood $U = (y_0 - \varepsilon, y_0 + \varepsilon)$ of y_0 there exists a neighbourhood $V = (R, \infty)$ of ∞ such that if $x \in V$, then $f(x) \in U$.

In the same spirit, one defines the relations $\lim_{x \to -\infty} f(x) = y_0$, $\lim_{x \to x_0} f(x) = \infty$, $\lim_{x \to \infty} f(x) = \infty$, etc – see exercises.

Example. Let us prove the relation $\lim_{x \to \infty} \dfrac{x + a}{x + b} = 1$ for any $a, b \in \mathbb{R}$. Given $\varepsilon > 0$, we must find a neighbourhood of infinity such that for all x in this neighbourhood, we have

$$\left| \frac{x + a}{x + b} - 1 \right| < \varepsilon.$$

Let us assume $x > 0$ and estimate the left hand side as follows:

$$\left| \frac{x + a}{x + b} - 1 \right| = \left| \frac{a - b}{x + b} \right| \leqslant \frac{|a - b|}{x - |b|}.$$

The desired estimate will be proven if we require that

$$\frac{1}{x - |b|} < \frac{\varepsilon}{|a - b|}$$

(we may assume that $a \neq b$, because if $a = b$, then our function is identically equal to 1 and there is nothing to prove). So we must take

$$x > |b| + \frac{|a - b|}{\varepsilon},$$

i.e. our neighbourhood of infinity is (R, ∞) with $R = |b| + |a - b|/\varepsilon$.

1.7 The O and o notation

Let f and g be two functions defined on an open interval $\Delta \subset \mathbb{R}$, and let $x_0 \in \Delta$. Suppose $g(x) \neq 0$ for small $x - x_0 \neq 0$. We write

$$f(x) = O(g(x)) \text{ as } x \to x_0, \text{ if } \frac{f(x)}{g(x)} \text{ is bounded for small } x - x_0;$$

$$f(x) = o(g(x)) \text{ as } x \to x_0, \text{ if } \lim_{x \to x_0} \frac{f(x)}{g(x)} = 0.$$

The same notation can be used if $x \to \infty$ or $x \to -\infty$.

Example. $f(x) = O(x)$ as $x \to 0$ means that $|f(x)| < C|x|$ for small $x \neq 0$; $f(x) = o(1)$ as $x \to \infty$ means that $f(x) \to 0$ as $x \to \infty$.

Example. The following statements hold true as $x \to 0$:

$$\begin{aligned} x^n &= o(x^m), \quad n > m; \\ O(x^n) + O(x^m) &= O(x^m), \quad n > m; \\ O(x^n)O(x^m) &= O(x^{n+m}); \\ O(x^n) &= o(1), \quad n > 0; \\ \log|x| &= O(x^{-1}). \end{aligned}$$

Example. The following statements hold true as $x \to \infty$:

$$\begin{aligned} x^m &= o(x^n), \quad n > m; \\ O(x^n) + O(x^m) &= O(x^n), \quad n > m; \\ O(x^n)O(x^m) &= O(x^{n+m}); \\ O(x^{-n}) &= o(1), \quad n > 0; \\ x &= O(e^{ax}), \quad a > 0. \end{aligned}$$

2 Continuity I: continuity at a point

Here we discuss the simplest aspects of the notion of continuity of a function. More advanced aspects are discussed in the next section.

2.1 Definition of continuity

Informally speaking, *continuous functions are those whose graph you can draw without lifting your pencil from the paper.*

Definition. Let f be a function on an open interval (a, b) and let $x_0 \in (a, b)$. We say that f is *continuous* at x_0, if

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

Equivalently:

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad |x - x_0| < \delta \quad \Rightarrow \quad |f(x) - f(x_0)| < \varepsilon$$

(here $\delta > 0$ in general depends both on ε and on x_0). One more equivalent way of writing this definition, in terms of neighbourhoods, is:

For any neighbourhood U of $f(x_0)$ there exists a neighbourhood V of x_0 such that if $x \in V$, then $f(x) \in U$.

Finally, we say that f is *continuous on* (a, b) and write $f \in C(a, b)$, if f is continuous at every point of (a, b).

Example. $f(x) = \sqrt{x}$ is a continuous function on the nonnegative half-line $[0, +\infty)$. Indeed, if $x_0 = 0$ then $|f(x) - f(x_0)| = \sqrt{x} < \varepsilon$ whenever $|x - x_0| = x < \varepsilon^2$, that is, we can take $\delta = \varepsilon^2$. If $x_0 > 0$ then

$$\sqrt{x} - \sqrt{x_0} = \frac{x - x_0}{\sqrt{x} + \sqrt{x_0}},$$

and so

$$|\sqrt{x} - \sqrt{x_0}| = \frac{|x - x_0|}{\sqrt{x} + \sqrt{x_0}} \leqslant \frac{|x - x_0|}{\sqrt{x_0}}.$$

In this case $|\sqrt{x} - \sqrt{x_0}| < \varepsilon$ whenever $|x - x_0| < \sqrt{x_0}\varepsilon$, that is, we can take $\delta = \sqrt{x_0}\varepsilon$. Observe that δ depends both on ε and on x_0.

2.2 Left and right limits; types of discontinuity

Definition. Let f be a function on an interval (a, b), and let $x_0 \in (a, b)$.

$\lim\limits_{x \to x_0+} f(x) = y_+$ means that
$\forall \varepsilon > 0 \, \exists \delta > 0 \, : \, x_0 < x < x_0 + \delta \Rightarrow |f(x) - y_+| < \varepsilon$.

$\lim\limits_{x \to x_0-} f(x) = y_-$ means that
$\forall \varepsilon > 0 \, \exists \delta > 0 \, : \, x_0 - \delta < x < x_0 \Rightarrow |f(x) - y_-| < \varepsilon$.

Here y_+ and y_- are real numbers, called the *right and the left limit*, respectively. Of course, the left and right limits don't have to exist.

In terms of left and right limits, we can restate the definition of continuity of f at x_0 as follows: f is continiuous at x_0, if the left and right limits of $f(x)$ as $x \to x_0$ exist and

$$\boxed{\lim_{x \to x_{0+}} f(x) = \lim_{x \to x_{0-}} f(x) = f(x_0).}$$

If a function f is not continuous at a given point x, we say that x is a point of *discontinuity* for f. It is useful to distinguish different types of points of discontinuity. The following terminology is not completely standard, but the notions themselves are well-known and useful. We assume that f is defined on an interval (a, b) and $x_0 \in (a, b)$.

Removable discontinuity: The point x_0 is a *removable discontinuity* of f, if both limits

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

exist, are finite and are equal to each other, but $f(x_0)$ is either NOT equal to these limits or not defined. In brief:

$$\lim_{x \to x_{0-}} f(x) = \lim_{x \to x_{0+}} f(x) \neq f(x_0).$$

If f has a removable discontinuity at x_0, then by redefining the value of f at x_0, we can remove this discontinuity.

Example. Let $f(x) = \frac{\sin x}{x}$; then f is not defined at zero. However, as we know from Calculus I,

$$\lim_{x \to 0_+} \frac{\sin x}{x} = \lim_{x \to 0_-} \frac{\sin x}{x} = 1,$$

and so f has a removable discontinuity at the origin. Defining $f(0) = 1$, we make f continuous.

Jump discontinuity: The point x_0 is a *jump discontinuity* of f, if both limits

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

exist and are finite, but are not equal to each other. In brief:

$$\lim_{x \to x_{0-}} f(x) \neq \lim_{x \to x_{0+}} f(x).$$

However you define $f(x_0)$, this will not make f continuous.

Example. Let $f(x) = \lfloor x \rfloor$ be the floor function on $\mathbb{R}$, i.e. $\lfloor x \rfloor$ is the greatest integer n such that $n \leqslant x$. Then f has a jump discontinuity at every integer.

Infinite discontinuity: The point x_0 is an *infinite discontinuity* of f, if at least one of the limits

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

is infinite.

Example. The function $f(x) = 1/x$ has an infinite discontinuity at the origin.

Oscillatory discontinuity: The point x_0 is an *oscillatory discontinuity* of f, if at least one of the limits

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

does not exist.

Example. The function $f(x) = \sin(1/x)$ has an oscillatory discontinuity at the origin. The Dirichlet function has an oscillatory discontinuity at every real point.

Of course, one can also have mixed cases, e.g. one of the two-sided limits may be infinite and another one may not exist.

2.3 Continuity and convergent sequences

Revise the definition of convergence for sequences.

As above, we assume that f is defined on (a, b) and $x_0 \in (a, b)$.

Theorem. $\lim\limits_{x\to x_0} f(x) = y_0$ *if and only if for every sequence of points* $x_n \in (a, b)$ *such that* $\lim\limits_{n\to\infty} x_n = x_0$ *we have* $\lim\limits_{n\to\infty} f(x_n) = y_0$.

Proof. Assume first that $\lim\limits_{x\to x_0} f(x) = y_0$:

$$\forall \varepsilon > 0\ \exists \delta_\varepsilon > 0 :\ |x - x_0| < \delta_\varepsilon \ \Rightarrow\ |f(x) - y_0| < \varepsilon.$$

Let $\{x_n\}_{n=1}^{\infty}$ be a sequence which converges to x_0:

$$\forall \delta > 0\ \exists N_\delta :\ n \geqslant N_\delta \ \Rightarrow\ |x_n - x_0| < \delta.$$

Now for given $\varepsilon > 0$, take $\delta = \delta_\varepsilon$; then for $n \geqslant N_\delta$ we get $|x_n - x_0| < \delta$ and therefore $|f(x_n) - y_0| < \varepsilon$. This proves that $\lim\limits_{n\to\infty} f(x_n) = y_0$.

Assume now that $f(x)$ fails to converge to y_0 as $x \to x_0$; then there exists $\varepsilon > 0$ such that for all $\delta > 0$ we can find x'_δ such that $|x'_\delta - x_0| < \delta$ but $|f(x'_\delta) - y_0| \geqslant \varepsilon$. For every $n \in \mathbb{N}$, let us take $\delta = 1/n$ and denote by x_n the corresponding x'_δ. Then $x_n \to x_0$ because $|x_n - x_0| < 1/n \to 0$. On the other hand, $|f(x_n) - y_0| \geqslant \varepsilon$ for all n, which shows that the sequence $f(x_n)$ does not converge to y_0. ∎

Corollary. *A function f is continuous at x_0 if and only if for every sequence of points* $x_n \in (a, b)$ *such that* $\lim\limits_{n\to\infty} x_n = x_0$ *we have* $\lim\limits_{n\to\infty} f(x_n) = f(x_0)$.

Example. Thomae's function f is discontinuous at all rational points and is continuous at all irrational points. Indeed, if x is rational then $f(x) \neq 0$ but $f(x_n) = 0$ for any sequence of irrational numbers $x_n \to x$ (explain why such a sequence exists!) that is, $f(x_n) \not\to f(x)$. On the other hand, if x is irrational and $\{x_n\}$ is a sequence of rational numbers $x_n = p_n/q_n \to x$ 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 x). Therefore $f(x_n) \to 0 = f(x)$.

2.4 The algebra of continuous functions

The above theorem and the algebra of limits for sequences immediately imply

Theorem. *Let f and g be functions continuous at x_0. Then the functions $f+g$ and $f\cdot g$ are continuous at x_0. If $g(x_0)\neq 0$ then $\frac{f}{g}$ is also continuous at x_0.*

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 $f(x)=x$ is continuous. This is immediate from the definition of continuity, as you can take $\delta_\varepsilon=\varepsilon$.

In a similar way, one proves that the composition of two continuous functions is also continuous. One only needs to take care about the domains and ranges of the functions.

Theorem. *Let f be a function continuous at x_0, and let g be a function, defined on $\operatorname{ran} f$ and continuous at $f(x_0)$. Then the composition $g\circ f(x)=g(f(x))$ is continuous at x_0.*

Proof. Let $\{x_n\}$ be a sequence convergent to x_0. Since f is continuous, $f(x_n)$ converges to $f(x_0)$. Since g is continuous, we find that $g(f(x_n))$ converges to $g(f(x_0))$. Thus we have $g\circ f(x_n)\to g\circ f(x_0)$ for every sequence $x_n\to x_0$, i.e. $g\circ f$ is continuous at x_0. ■

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.

2.5 Monotonic functions

Theorem. *Let f be a monotonic function on (a,b); then all discontinuities of f are jump discontinuities.*

Proof. Consider the case of non-decreasing f; the case of non-increasing f is considered similarly (or by swapping f to $-f$). It suffices to prove that the left and right limits of $f(x)$ exist at every point $x_0\in(a,b)$. In fact, we have

$$\lim_{x\to x_{0+}} f(x)=\inf_{x_0<x<b} f(x),\qquad \lim_{x\to x_{0-}} f(x)=\sup_{a<x<x_0} f(x).$$

The proof of this is an exercise. ■

Thus, f is continuous if $\lim_{x\to x_{0+}} f(x)=\lim_{x\to x_{0-}} f(x)$ and discontinuous otherwise. Revise countable and uncountable sets.

Corollary. *Let f be a monotonic function on (a,b). Then the set of discontinuities of f is either finite or countably infinite.*

Sketch of proof. For every point x of discontinuity of f, select a rational number in the interval between $\lim\limits_{x\to x_{0+}} f(x)$ and $\lim\limits_{x\to x_{0-}} f(x)$. This establishes a bijection between a subset of rational numbers and the set of discontinuities of f. Since the set of rational numbers is countable, any of its subsets is either finite or countable. ■

3 Continuity II: continuity on an interval

Here we establish some nice properties of functions continuous on closed and bounded intervals and discuss the important notion of *uniform continuity*.

3.1 Preliminaries

Recall that f is called continous on (a, b), if it is continuous at every point of (a, b).

Definition. A function f is said to be continuous on $[a, b]$ if it is continuous at every point of (a, b), the limits $\lim\limits_{x\to a_+} f(x)$, $\lim\limits_{x\to b_-} f(x)$ exist and the relations

$$f(a) = \lim_{x\to a_+} f(x) \quad \text{and} \quad f(b) = \lim_{x\to b_-} f(x) \tag{3.1}$$

hold true. f is said to be continuous on $[a, b)$, if it is continuous at every point of (a, b), the limit $\lim_{x\to a_+} f(x)$ exists and the first relation in (3.1) holds. f is said to be continuous on $(a, b]$, if it is continuous at every point of (a, b), the limit $\lim_{x\to b_-} f(x)$ exists and the second relation in (3.1) holds.

Example. The functions $1/\sqrt{x}$ and $\sin(1/\sqrt{x})$ are continuous on $(0, \infty)$. The function

$$f(x) = \begin{cases} e^{-1/x}, & x > 0 \\ 0, & x = 0 \end{cases}$$

is continuous on $[0, \infty)$.

Notation: We denote by $C(a, b)$ the set of all functions continuous on (a, b). Similarly, $C[a, b]$ denotes the set of all functions continuous on $[a, b]$. Finally, $C(\mathbb{R})$ denotes the set of all functions continuous on the whole real line.

3.2 Three important theorems

The following three theorems are true for functions that are continuous on **closed and bounded intervals** $[a, b]$. These theorems are fundamental for the whole of analysis.

Revise the Bolzano-Weierstrass theorem.

Theorem (Boundedness Theorem). *If f is a continuous function on a closed bounded interval, then f is bounded.*

Proof. Denote the interval by $[a, b]$. Suppose, to get a contradiction, that f is unbounded. Then there exists a sequence $x_n \in [a, b]$ such that $|f(x_n)| \to \infty$ as $n \to \infty$. By the Bolzano–Weierstrass theorem, there exists 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 $|f(x_{n_k})| \to \infty$ as $k \to \infty$. The obtained contradiction proves the theorem. ■

As a simple example of application of this theorem, let us prove the following

Corollary. *Let $a \in \mathbb{R}$ and let f be a continuous function on $[a, \infty)$ such that the limit $\lim_{x\to\infty} f(x)$ exists. Then f is bounded on $[a, \infty)$.*

Proof. Denote the limit of f at infinity by A. Let us write the definition of convergence of f to A at infinity:

$$\forall \varepsilon > 0 \; \exists a_0 \geqslant a : \quad x \geqslant a_0 \;\Rightarrow\; |f(x) - A| < \varepsilon.$$

Let us take, for example, $\varepsilon = 1$ in this definition; we get:

$$\exists a_0 \geqslant a : \; x \geqslant a_0 \;\Rightarrow\; |f(x) - A| < 1$$

It follows that for $x \geqslant a_0$, we have

$$|f(x)| = |f(x) - A + A| \leqslant |f(x) - A| + |A| \leqslant 1 + |A|.$$

On the other hand, f is continuous on $[a, a_0]$ and so by the Boundedness Theorem, it is bounded there; let us denote by M some bound for f on $[a, a_0]$. Putting this together, we see that

$$|f(x)| \leqslant M, \quad a \leqslant x \leqslant a_0;$$
$$|f(x)| \leqslant 1 + |A|, \quad a_0 \leqslant x,$$

and so f is bounded on $[a, \infty)$ with the bound $\max\{M, 1 + |A|\}$. ■

The following two examples show that it is crucial that the interval in the Boundedness Theorem is both bounded and closed.

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.

Revise supremum and infimum.

Theorem (Maximum/Minimum Theorem)**.** *If f is a continuous function on a closed bounded interval, then f attains its maximum and minimum values.*

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 so would not be its *least* upper bound). By the Bolzano-Weierstrass theorem, 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 $f(c) = \lim_{k\to\infty} f(x_{n_k}) = M$. The case of minimum is considered in the same way, starting from $\inf f$. ■

The following two examples show that the boundedness and closedness of the interval are crucial for this theorem.

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 (sketch the graph!). The function $f(x) = x^{-1}$ is continuous on the closed unbounded interval $[1, +\infty)$ but does not attain the minimum value 0.

Theorem (Intermediate Value Theorem)**.** *Let f be a continuous function on a closed bounded interval $[a, b]$. Then f attains every value between $f(a)$ and $f(b)$.*

Proof. Let d be a point between $f(a)$ and $f(b)$; we need to prove that there exists $c \in [a, b]$ such that $f(c) = d$. If $f(a) = f(b)$, then $d = f(a)$ and so we can take $c = a$. Suppose $f(a) \neq f(b)$; let us assume for the sake of definiteness that $f(a) < f(b)$ (otherwise swap $f(a)$ and $f(b)$ in the rest of the proof).

We give a proof by bisection method, which should be familiar to you from the proof of the Bolzano-Weierstrass theorem. Denote $a_0 = a$, $b_0 = b$. We define inductively the sequence of intervals

$$[a_0, b_0] \supset [a_1, b_1] \supset [a_2, b_2] \supset \dots$$

as follows. If $[a_j, b_j]$ has already been defined, we consider the value $f(\frac{a_j+b_j}{2})$ and choose the next interval as follows:

- if $f(\frac{a_j+b_j}{2}) = d$, then we take $c = \frac{a_j+b_j}{2}$, the process terminates and the proof is finished;
- if $f(\frac{a_j+b_j}{2}) > d$, set $a_{j+1} = a_j$, $b_{j+1} = \frac{a_j+b_j}{2}$;
- if $f(\frac{a_j+b_j}{2}) < d$, set $a_{j+1} = \frac{a_j+b_j}{2}$, $b_{j+1} = b_j$.

Then at every step, if the process has not terminated, we have

$$f(a_j) < d \text{ and } f(b_j) > d. \tag{3.2}$$

The sequence a_j is monotone increasing and bounded above (for example, by b_0); similarly, b_j is monotone decreasing and bounded below. It follows that both sequences converge. Passing to the limit in the identity

$$b_j - a_j = \frac{b-a}{2^j},$$

we find that a_j and b_j converge to the same limit; denote this limit by c_*. It is evident that $c_* \in (a, b)$. Passing to the limit in (3.2) and using the continuity of f at c_*, we find $f(c_*) \leqslant d$ and $f(c_*) \geqslant d$; thus $f(c_*) = d$ and we can take $c = c_*$. ∎

Corollary. *If f is a continuous function on the closed bounded interval $[a, b]$, then its range is the closed interval $[m, M]$ where $m = \inf f$ and $M = \sup f$.*

Proof. By the maximum/minimum theorem, there exist points $a_1, b_1 \in [a, b]$ such that $f(a_1) = m$ and $f(b_1) = M$. By the intermediate value theorem, f attains all values between $[m, M]$. ∎

The intermediate value theorem is a powerful tool which allows us to assert that a solution to some equation exists even though there is no explicit formula for it. For example, it allows us to define inverse functions to all elementary functions, on suitable intervals.

Example. What is $a = \tan^{-1}(x)$? It is a solution $a \in (-\pi/2, \pi/2)$ to the equation $\tan a = x$. How do we know that such a solution exists for every $x \in \mathbb{R}$? Because for the function $f(t) = \tan t$, we know that

$$\lim_{t \to -\pi/2_+} f(t) = -\infty \quad \text{and} \quad \lim_{t \to \pi/2_-} f(t) = \infty,$$

and therefore one can choose an interval $[-\pi/2+\varepsilon, \pi/2-\varepsilon]$ with $\varepsilon > 0$ such that $f(-\pi/2+\varepsilon) < x$ and $f(\pi/2-\varepsilon) > x$. By the intermediate value theorem, there exists $a \in (-\pi/2+\varepsilon, \pi/2-\varepsilon)$ such that $f(a) = x$, i.e. $\tan a = x$. (The *uniqueness* of a requires a separate argument based on the monotonicity of $\tan$.)

3.3 Uniform continuity and Cantor's theorem

Definition. Let f be a function defined on an interval Δ, which may be bounded or unbounded. Then f is called *uniformly continuous on* Δ, if for any $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that

$$\boxed{\forall x, x' \in \Delta : \quad |x - x'| < \delta_\varepsilon \Rightarrow |f(x) - f(x')| < \varepsilon.} \tag{3.3}$$

Clearly, if function is uniformly continuous on Δ, then it is continuous at every point in Δ. The converse is false! Before discussing examples, let us make the following general point.

Question: How to prove that f is NOT uniformly continuous on Δ?

Answer: Construct two sequences $x_n, x'_n \in \Delta$ such that $|x_n - x'_n| \to 0$ as $n \to \infty$, yet $|f(x_n) - f(x'_n)| \geqslant \varepsilon$ for some $\varepsilon > 0$.

Example. The function $f(x) = 1/x$ on $\Delta = (0, 1)$ is continuous at every point of Δ. However, it is NOT uniformly continuous on Δ. Indeed, take $\varepsilon = 1/2$ and let $x_n = 1/2n$, $x'_n = 1/n$ for every $n \in \mathbb{N}$. Then $x_n - x'_n \to 0$, yet $f(x_n) - f(x'_n) = n > 1/2$ for all n.

Example. The function $f(x) = x^2$ on $\Delta = \mathbb{R}$ is continuous at every point of Δ. However, it is NOT uniformly continuous on Δ. Indeed, take $x_n = n + \frac{1}{n}$, $x'_n = n$ for every $n \in \mathbb{N}$. Then, clearly, $x_n - x'_n = 1/n \to 0$, yet

$$f(x) - f(x') = \left(n + \frac{1}{n}\right)^2 - n^2 = 2 + \frac{1}{n^2} > 2.$$

Example. The function $f(x) = \sin(1/x)$ on $\Delta = (0, 1)$ is continuous at every point of Δ. However, it is NOT uniformly continuous on Δ. Indeed, take $x_n = 1/(2\pi n)$, $x'_n = 1/(2\pi n + \frac{\pi}{2})$ for every $n \in \mathbb{N}$. Then $x_n \to 0$ and $x'_n \to 0$, and so $|x_n - x'_n| \to 0$ as $n \to \infty$. However, $f(x_n) = 0$ and $f(x'_n) = 1$, and so so $|f(x_n) - f(x'_n)| = 1$.

Remark. In simple examples, it is often easy to heuristically identify the behaviour of f that is "responsible" for the breakdown of uniform continuity. As we saw in the examples above, this can be either the growth or the oscillations of $f(x)$, as x approaches some "singular point". This singular point can be an interior point of Δ, a boundary point of Δ or the infinity, if Δ is unbounded.

Theorem (Cantor's theorem)**.** *Let f be a continuous function on a closed bounded interval Δ. Then f is uniformly continuous on Δ.*

Proof. We give a proof by contradiction. Assume that there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exist $x, x' \in \Delta$ such that

$$|x - x'| < \delta, \quad \text{yet} \quad |f(x) - f(x')| \geqslant \varepsilon.$$

Let us take a sequence $\delta_n \to 0$; to be definite, let us put $\delta_n = 1/n$. Then, by assumption, for every n there exist x_n and x'_n such that

$$|x_n - x'_n| \leqslant \tfrac{1}{n} \to 0 \quad \text{as } n \to \infty, \quad \text{yet} \quad |f(x_n) - f(x'_n)| \geqslant \varepsilon \tag{3.4}$$

for some $\varepsilon > 0$. By the Bolzano-Weierstrass Theorem (remember that Δ is closed and bounded!) the sequence $\{x_n\}_{n=1}^{\infty}$ has a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ which converges to some $x_* \in \Delta$. By (3.4), we find that x_* is also the limit of the $\{x'_{n_k}\}_{k=1}^{\infty}$. Then, in view of the continuity of f at x_*, we must have

$$f(x_{n_k}) \to f(x_*) \quad \text{and} \quad f(x'_{n_k}) \to f(x_*)$$

as $k \to \infty$, and therefore

$$f(x_{n_k}) - f(x'_{n_k}) \to f(x_*) - f(x_*) = 0.$$

This contradicts (3.4). ∎

Examples at the start of this subsection show that neither boundedness nor closedness of Δ can be dropped from the hypothesis of the theorem.

3.4 Lipschitz continuous functions

A useful subclass of uniformly continuous functions is the set of Lipschitz continuous functions.

Definition. A function f defined on an interval Δ is called *Lipschitz continuous* on Δ, if there exists a constant $C > 0$ (called Lipschitz constant) such that

$$\boxed{\forall x_1, x_2 \in \Delta : \quad |f(x_1) - f(x_2)| \leqslant C|x_1 - x_2|.}$$

Theorem. *If f is Lipschitz continuous on Δ, then f is uniformly continuous on Δ.*

Proof. Take $\delta_\varepsilon = \varepsilon/C$ in the definition of uniform continuity. ∎

Example. The function $f(x) = |x|$ is Lipschitz continuous on $\mathbb{R}$ with the Lipschitz constant $C = 1$.

Example. Consider the function $f(x) = \sqrt{x}$ for $x \geqslant 0$. Then f is Lipschitz continuous on $[\varepsilon, \infty)$ for any $\varepsilon > 0$. Indeed, for $\varepsilon \leqslant x_1 \leqslant x_2$

$$\sqrt{x_2} - \sqrt{x_1} = \frac{x_2 - x_1}{\sqrt{x_2} + \sqrt{x_1}} \leqslant \frac{x_2 - x_1}{2\varepsilon},$$

and so f is Lipschitz continuous with the Lipschitz constant $C_\varepsilon = 1/2\varepsilon$. However, f is NOT Lipschitz continuous on $[0, \infty)$. (One can "guess" this from the fact that the Lipschitz constant blows up: $C_\varepsilon \to \infty$ as $\varepsilon \to 0$.) Indeed, taking $x_2 = 0$, Lipschitz continuity condition says

$$\sqrt{x_1} \leqslant C|x_1|,$$

i.e. $1 \leqslant C\sqrt{x_1}$, which of course cannot be true for small x_1.

Example. The function $f(x) = x^2$ is Lipschitz continuous on $[-R, R]$ with the Lipschitz constant $2R$:

$$|x_1^2 - x_2^2| = |x_1 + x_2||x_1 - x_2| \leqslant (|x_1| + |x_2|)|x_1 - x_2| \leqslant 2R|x_1 - x_2|.$$

But f is NOT Lipschitz continuous on the whole of $\mathbb{R}$; you can "guess" this from the fact that the Lipschitz constant $C_R = 2R$ blows up as $R \to \infty$.

Remark. Every differentiable function whose derivative is bounded on Δ is Lipschitz continuous. Indeed, assuming that $|f'(x)| \leqslant B$ on Δ and taking $x_1 < x_2$, we have, by the Fundamental Theorem of Calculus (which we will only prove properly towards the end of this course!),

$$|f(x_2) - f(x_1)| = \left|\int_{x_1}^{x_2} f'(t)dt\right| \leqslant \int_{x_1}^{x_2} |f'(t)|dt \leqslant B\int_{x_1}^{x_2} dt = B|x_1 - x_2|.$$

But the class of Lipschitz continuous functions is wider than the class of differentiable functions: $f(x) = |x|$ is not differentiable at zero. This will be properly discussed later in the course.

Warning: there are uniformly continuous functions with unbounded derivative!

3.5 Concluding remarks

How to check whether f is uniformly continuous on Δ?

1. Is f differentiable on Δ? If it is, can you check that its derivative is bounded on Δ? If it is, then f is Lipschitz continuous and hence uniformly continuous.

2. Can you prove directly that f is Lipschitz continuous on Δ? If you can, then f is uniformly continuous.

3. Is it true that Δ is closed and bounded and f continuous on Δ? If it is, you are in luck – Cantor's theorem guarantees that f is uniformly continuous on Δ.

4. If all the above fails, try to prove that f is NOT uniformly continuous. Can you identify a singular point (finite or infinite) in the domain of f, such that f grows or oscillates near that point?

5. Using the previous step, try to construct sequences x_n, x_n' converging to this singular point, such that $|x_n - x_n'| \to 0$, yet $|f(x_n) - f(x_n')| \geqslant \varepsilon > 0$.

Example. Let us prove that $f(x) = xe^{-x}$ is uniformly continuous on $[0, 1]$. This is easy: f is continuous on $[0, 1]$, hence uniformly continuous by Cantor's theorem.

Now let us prove that the same function is uniformly continuous on $[0, \infty)$. We have $f'(x) = (1 - x)e^{-x}$; this function is bounded on $[0, \infty)$ and therefore f is Lipschitz continuous, and therefore uniformly continuous on $[0, \infty)$.

Example. Consider the function $f(x) = e^{x^2}$ on $[0, \infty)$. Inspecting the derivative, we see that it is unbounded. Furthermore, we see that f grows fast at infinity. Let us check that f is not uniformly continuous. Take $x_n = n + \frac{1}{n}$ and $x_n' = n$ for $n \in \mathbb{N}$, then we get

$$f(x_n) - f(x_n') = e^{n^2+2+n^{-2}} - e^{n^2} = e^{n^2}(e^{2+n^{-2}} - 1) \geqslant e(e^2 - 1)$$

for all n.

Example. Consider the function $f(x) = \sqrt{x}$ on $[0, 1]$. Its derivative is unbounded on $[0, 1]$, and the function is not Lipschitz continuous. However, $[0, 1]$ is closed and bounded, and so f is uniformly continuous there by Cantor's theorem.

4 Differentiation I: theory

Here we establish the theoretical basis for the notion of the derivative. We review the definition and basic algebraic properties of the derivative (these should be familiar to you from Calculus I); further, we establish the Mean Value Theorem and Taylor's formula.

4.1 Basics

4.1.1 Definition of derivative

Let f be a function defined on an open interval Δ, and let x be a point in Δ. We say that f is differentiable at a point x if there exists a real number denoted by $f'(x)$, such that

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

The number $f'(x)$ is called the *derivative of f at the point x.*

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

We say that f is differentiable on the interval Δ if $f'(x)$ exists at every point $x \in \Delta$. In this case we can consider f' as a function of x defined on Δ. This function is called *the derivative of f.* Another notation for the derivative is $\frac{\mathrm{d}}{\mathrm{d}x} f(x)$.

Revise the $O(\cdot)$ and $o(\cdot)$ notation.

Using the $o(\cdot)$ notation, one can write the definition of derivative as follows:

$$f(x+h) = f(x) + f'(x)h + o(h), \quad h \to 0.$$

It follows that

$$f(x+h) = f(x) + o(1), \quad h \to 0,$$

which is the definition of continuity of f at x. Thus, if a function is differentiable at x, it must be continuous at x. (This was also discussed in Calculus I.) We will see very soon that the converse is not true.

4.1.2 Derivative and the slope of the graph of f

Geometrically one can think of the derivative as the slope of the graph of f at the point $(x, f(x))$. If $x' \neq x$ is another point in the interval Δ, then the slope of the line passing through the two points $(x, f(x))$ and $(x', f(x'))$ is given by the ratio

$$\frac{f(x') - f(x)}{x' - x}.$$

The limit of this slope as $x \to x'$ (if it exists) is exactly the slope of the tangent line to the graph of f.

4.1.3 Left and right derivatives

The following discussion should be compared to the discussion of types of discontinuity of a function.

As above, let Δ be an open interval and $x \in \Delta$. The right and left limits

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

(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 derivatives exist and have the same value.

It may well happen that the right and left derivatives of f at a point exist but are not equal to each other; example: $f(x) = |x|$ at the point $x = 0$.

Remark. One of the turning points in the history of mathematics was the construction by Weierstrass in 1872 of a continuous *nowhere differentiable* function. We shall discuss this example at the end of this course - see Section 10.5.

4.1.4 Functions differentiable on an interval

We say that f is *differentiable* on (a, b), if the derivative of f exists at every point $x \in (a, b)$. We say that f is *continuously differentiable* on (a, b), if f is differentiable on (a, b) and the derivative f' is continuous on (a, b). The set of all continuously differentiable functions on (a, b) is denoted by $C^1(a, b)$.

In other words, $f \in C^1(a, b)$ means $f' \in C(a, b)$.

4.2 The algebra of differentiation

Here we recall the familiar "rules" of differentiation: sum rule, product rule, quotient rule and the chain rule.

Theorem (Sum, product, quotient rules)**.** *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* $g(x) \neq 0$.

The proofs are standard and were provided in Calculus I; we shall not repeat them here. Starting from the fact that the derivative of a constant function is zero and the derivative of $f(x) = x$ is one, it is easy to build up formulas for the derivatives of any rational functions.

Theorem (Chain rule)**.** *If $f(x)$ is differentiable at $x = x_0$ and $g(y)$ is differentiable at $y_0 = f(x_0)$ then $g(f(x))$ is differentiable at $x = x_0$ and*

$$\frac{\mathrm{d}}{\mathrm{d}x} g(f(x))|_{x=x_0} = g'(f(x_0))f'(x_0).$$

For the proof, we refer to Calculus I notes.

4.3 The Mean Value Theorem

Definition (minimum/maximum)**.** Let f be defined on an interval Δ. We say that $c \in \Delta$ is a *minimum* (resp. *maximum*) of f on Δ, if $f(c) \leqslant f(x)$ (resp. $f(c) \geqslant f(x)$) for all $x \in \Delta$. Maxima and minima are collectively known as *extrema*.

Definition (local minimum/maximum)**.** Let f be a function defined on an interval Δ. The point $c \in \Delta$ is called a *local minimum* (resp. *local maximum*) of f, if there exists $\varepsilon > 0$ such that c is a maximum (resp. minimum) of f on $\Delta \cap (c-\varepsilon, c+\varepsilon)$. In detail, this means that $f(c) \leqslant f(x)$ (resp. $f(c) \geqslant f(x)$) for all $x \in \Delta \cap (c-\varepsilon, c+\varepsilon)$.

The following theorem is basic in the whole of analysis. It is sometimes called Fermat's theorem (even though differential calculus has not yet been rigorously developed during Pierre de Fermat's time).

> **Theorem** (Fermat's 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 $g(x) = -f(x)$). ■

Warning: Despite the previous theorem, in general f does not have to be differentiable at the point where it has a local minimum or maximum. For instance, $f(x) = |x|$ attains its (global) minimum at $x = 0$, but $f'(0)$ does not exist.

Theorem (Rolle's Theorem)**.** *Let f be a continuous function on $[a, b]$ such that $f(a) = f(b)$. If f is differentiable on (a, b) then there exists a point c in (a, b) at which $f'(c) = 0$.*

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

> **Theorem** (Mean Value Theorem (MVT))**.** *Let f be a continuous function on $[a, b]$. If f is differentiable on (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. ■

Example. Using the Mean Value Theorem, let us prove the inequality

$$|\sin x - \sin y| \leqslant |x - y|.$$

Assume for definiteness that $x < y$ and apply the MVT to $[x, y]$; we get

$$\frac{\sin x - \sin y}{x - y} = \cos c$$

for some $c \in (a, b)$. Since $|\cos c| \leqslant 1$, this gives the required inequality.

Remark. Another way of writing the MVT is

$$f(x+h) - f(x) = f'(c)h,$$

where c is a point between x and $x + h$. This expression is useful because it allows us to link the increment $f(x + h) - f(x)$ with the derivative of f. It is important to note that there is, perhaps, an even more useful way of linking the increment with the derivative: the Fundamental Theorem of Calculus:

$$f(x+h) - f(x) = \int_x^{x+h} f'(t)dt, \quad h > 0.$$

However, at this point in the course we have not yet rigorously developed the theory of integration, and therefore we prefer not to use the latter formula in order to avoid circular arguments.

4.4 Higher derivatives and Taylor's formula

Notation: 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, and so on). We say that f is n times continuously differentiable if the final n'th derivative is continuous on (a, b) (the function f itself and its first $(n - 1)$ derivatives are automatically continuous). The usual notation for the derivative of order n is $f^{(n)}$, so that $f^{(n)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} f^{(n-1)}(x)$. The set of all n times continuously differentiable functions on (a, b) is denoted by $C^n(a, b)$. In other words, $f \in C^n(a, b)$ means that $f, f', \ldots, f^{(n)} \in C(a, b)$.

Theorem (Taylor's formula)**.** *If $f \in C^n(a, b)$ and $x_0 \in (a, b)$, then for each $x \in (a, b)$ there exists a point c between x_0 and x such that*

$$\begin{aligned} f(x) =& \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(n)}(c)}{n!}(x - x_0)^n \\ =& f(x_0) + f'(x_0)\,(x - x_0) + \cdots + \frac{f^{(n-1)}(x_0)}{(n-1)!}(x - x_0)^{n-1} \\ &+ \frac{f^{(n)}(c)}{n!}(x - x_0)^n. \end{aligned}$$

Note that $n = 1$ gives the Mean Value Theorem.

Proof. For simplicity of notation, let us take $x_0 = 0$ (the general case is recovered by considering $f(x - x_0)$ in place of $f(x)$). Let us denote by P the Taylor polynomial,

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!} t^k,$$

and let M be a real number defined by the condition

$$f(x) = P(x) + \frac{M}{n!} x^n.$$

We need to prove that $M = f^{(n)}(c)$ for some c between 0 and x. Denote

$$g(t) = f(t) - P(t) - \frac{M}{n!} t^n;$$

observe that after differentiating g n times, all powers of t less than n vanish, and we obtain

$$g^{(n)}(t) = f^{(n)}(t) - M.$$

Thus, we need to prove that $g^{(n)}(c) = 0$ for some c between 0 and x. Now, computing the derivatives of g at zero, we obtain

$$g(0) = g'(0) = g''(0) = \cdots = g^{(n-1)}(0) = 0.$$

Further, by our definition of the number M, we have $g(x) = 0$. Now applying Rolle's theorem to g on the interval between 0 and x, we find that $g'(c_1) = 0$ for some c_1 between 0 and x. Next, similarly applying Rolle's theorem to g' on the interval between 0 and c_1, we find that $g''(c_2) = 0$ for some c_2 between 0 and c_1. Continuing in the same way, after n steps we arrive at the conclusion that $g^{(n)}(c_n) = 0$ for some c_n between 0 and c_{n-1}, i.e. between 0 and x. ■

Remark. Since by assumption the derivative $f^{(n)}$ is continuous on (a, b), it is bounded on any neighbourhood of x_0. It follows that for the remainder term in Taylor's formula we have

$$\frac{f^{(n)}(c)}{n!}(x - x_0)^n = O((x - x_0)^n), \quad x \to x_0.$$

Example. For a polynomial $p(x) = c_n x^n + \cdots + c_1 x + c_0$, Taylor's formula reduces to an algebraic operation of re-expanding $p(x)$ in terms of the powers of $(x - x_0)$. For example, Taylor's formula for $p(x) = x^2$ at $x_0 = 3$ is

$$x^2 = (3 + x - 3)^2 = 9 + 6(x - 3) + (x - 3)^2.$$

Example. Let $\alpha \in \mathbb{R}$, $\alpha \neq 0$; consider the function $f(x) = (1 + x)^\alpha$ for $|x| < 1$. Computing the derivatives, we find

$$\begin{aligned} f'(x) &= \alpha(1 + x)^{\alpha-1}, \\ f''(x) &= \alpha(\alpha - 1)(1 + x)^{\alpha-2}, \\ &\cdots \\ f^{(n)}(x) &= \alpha(\alpha - 1)\cdots(\alpha - n + 1)(1 + x)^{\alpha-n}. \end{aligned}$$

Thus, Taylor's formula for $f(x)$ at $x_0 = 0$ is

$$(1 + x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha - 1)}{2!}x^2 + \cdots \frac{\alpha(\alpha - 1)\cdots(\alpha - n + 1)}{n!}x^n + O(x^{n+1}), \quad x \to 0.$$

We conclude by displaying Taylor's formula for some elementary functions:

$$
\begin{aligned}
e^x &= 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + O(x^{n+1}), \\
\sin x &= x - \frac{x^3}{3!} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + O(x^{2n+1}), \\
\cos x &= 1 - \frac{x^2}{2!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + O(x^{2n+2}), \\
\log(1+x) &= x - \frac{x^2}{2} + \cdots + (-1)^{n-1}\frac{x^n}{n} + O(x^{n+1}).
\end{aligned}
$$

4.5 Asymptotic expansions

Let $c_0, c_1, c_2, \ldots$ be real numbers and let f be a function defined in a neighbourhood of $x_0 \in \mathbb{R}$. We will say that f has an *asymptotic expansion near* x_0,

$$
f(x) \sim \sum_{n=0}^{\infty} c_n (x - x_0)^n, \quad x \to x_0, \tag{4.1}
$$

if for every $N \in \mathbb{N}$ we have the asymptotic formula

$$
f(x) = \sum_{n=0}^{N} c_n (x - x_0)^n + O(x^{N+1}), \quad x \to x_0.
$$

NB: this does not mean that the series $\sum c_n (x - x_0)^n$ is necessarily convergent!

Taylor's formula is one of the key mechanisms (but not the only one) that produces asymptotic expansions. This formula gives us an excellent way of computing the coefficients c_n in (4.1), if f is infinitely differentiable in a neighbourhood of x_0. However, it relies on our ability to compute higher derivatives. In practice, it is often more convenient to proceed directly. This is illustrated by the following examples.

Example. We start from the asymptotic expansion

$$
\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n, \quad x \to 0.
$$

Here all coefficients c_n are equal to one. This asymptotic expansion can either be justified by applying Taylor's formula to $f(x) = 1/(1-x)$ or by more elementary means, using the formula for the sum of the geometric progression:

$$
\frac{1}{1-x} - \sum_{n=0}^{N} x^n = \frac{1}{1-x} - \frac{1-x^{N+1}}{1-x} = \frac{x^{N+1}}{1-x} = O(x^{N+1}), \quad x \to 0.
$$

Remark. In this example, the formal series that gives our asymptotic expansion is actually *convergent* for $|x| < 1$. But at this point we are not overly interested in this aspect; it will be discussed in Section 10.

Example. Compute the first three non-trivial terms of Taylor's expansion of

$$f(x) = \frac{1+x+x^2}{1-x+x^2} \quad \text{as } x \to 0.$$

Solution: Instead of using Taylor's formula, we use the previous example and some elementary algebra:

$$\begin{aligned}
\frac{1+x+x^2}{1-x+x^2} =&(1+x+x^2)\frac{1}{1-x(1-x)} \\
=&(1+x+x^2)(1+x(1-x)+x^2(1-x)^2 \\
&+O(x^3(1-x)^3)) \\
=&(1+x+x^2)(1+x(1-x)+x^2(1-x)^2)+O(x^3) \\
=&(1+x+x^2)(1+x(1-x)+x^2)+O(x^3) \\
=&(1+x+x^2)(1+x)+O(x^3) \\
=&1+2x+2x^2+O(x^3), \quad x \to 0.
\end{aligned}$$

Example. Compute the first three non-trivial terms of the Taylor's expansion for

$$f(x) = \frac{x}{e^x-1} \quad \text{as } x \to 0.$$

Solution: Using the Taylor's formula for exponential, we obtain

$$\begin{aligned}
\frac{x}{e^x-1} &= \frac{x}{x+\frac{x^2}{2!}+\frac{x^3}{3!}+O(x^4)} \\
&= \frac{1}{1+\frac{x}{2}+\frac{x^2}{6}+O(x^3)} \\
&= 1-\left(\frac{x}{2}+\frac{x^2}{6}+O(x^3)\right)+\left(\frac{x}{2}+\frac{x^2}{6}+O(x^3)\right)^2+O(x^3) \\
&= 1-\frac{1}{2}x-\frac{1}{6}x^2+\frac{1}{4}x^2+O(x^3) \\
&= 1-\frac{1}{2}x+\frac{1}{12}x^2+O(x^3).
\end{aligned}$$

5 Differentiation II: applications

This section has a slightly more "practical" flavour than the previous one. We look at how the notion of derivative can be applied to the study of functions: monotonicity, extrema, convexity, behaviour at infinity. All this information is put together when one sketches the graph of a function; this is discussed at the end of the section.

5.1 Monotonicity and Extrema

Revise the definition of increasing and decreasing functions.

Theorem. *Let f be continuous and differentiable on (a, b).*

(i) *If $f'(x) > 0$ (resp. $f'(x) \geqslant 0$) for all $x \in (a, b)$ then the function f is strictly increasing (resp. non-decreasing) on (a, b).*

(ii) *If $f'(x) < 0$ (resp. $f'(x) \leqslant 0$) for all $x \in (a, b)$ then the function f is strictly decreasing (resp. non-increasing) on (a, b).*

Proof. (i) Assume that $f'(x) > 0$ for all x, and let $x_1 < x_2$ be two points in (a, b). Applying Mean Value Theorem to the interval $[x_1, x_2]$, we obtain $\frac{f(x_2)-f(x_1)}{x_2-x_1} > 0$. Since $x_2 - x_1 > 0$, this implies that $f(x_2) > f(x_1)$.

In cases $f'(x) \geqslant 0$ and (ii), the proof is the same up to trivial changes. ■

Corollary 5.1. *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. By the previous statement, f is both non-increasing and non-decreasing; hence it must be constant. ■

We already know a *necessary* condition for the extremum of a differentiable function in terms of its derivative, given by Fermat's theorem. The following theorem gives a sufficient condition in terms of the second derivative.

Theorem 5.2. *Let f be continuous and twice differentiable on (a, b). Let $c \in (a, b)$;*

(i) *if $f'(c) = 0$ and $f''(c) > 0$, then f has a local minimum at c;*

(ii) *if $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum at c.*

Proof. We will only consider the case (i); the other case is considered in the same way (or by applying (i) to $g(x) = -f(x)$).

Assume $f'(c) = 0$ and $f''(c) > 0$. Consider the definition of the right derivative at the point c, applied to the function f'. Taking $\varepsilon = f''(c)/2$ in this definition, we find

$$\left| \frac{f'(x) - f'(c)}{x - c} - f''(c) \right| \leqslant f''(c)/2, \quad \forall x \in (c, c + \delta]$$

for some $\delta > 0$. This implies that

$$\frac{f'(x) - f'(c)}{x - c} \geqslant f''(c)/2 > 0, \quad \forall x \in (c, c + \delta]$$

and so, recalling that $f'(c) = 0$ and $x - c > 0$, we get

$$f'(x) > 0, \quad x \in (c, c + \delta].$$

By the previous theorem, it follows that f is increasing on $[c, c + \delta]$. A similar argument shows that f is decreasing on $[c - \delta', c]$ for some $\delta' > 0$. The conclusion is that c is the minimum point of f on $[c - \delta', c + \delta]$, as required. ■

Remark. If $f'(c) = 0$ and $f''(c) = 0$, then c may be a local minimum (e.g. $f(x) = x^4$ at $x = 0$) or local maximum (e.g. $f(x) = -x^4$) or neither (e.g. $f(x) = x^3$).

5.2 Convexity and the second derivative

Definition. Let f be a continuous function on an interval Δ. We say that f is *convex* on Δ (or *convex downward*), if for any $x_1, x_2 \in \Delta$ and for any $\theta \in [0, 1]$ we have

$$f(\theta x_1 + (1 - \theta)x_2) \leqslant \theta f(x_1) + (1 - \theta)f(x_2).$$

We say that f is *concave* on Δ, if $(-f)$ is convex on Δ.

The geometric meaning of convexity is that the graph of f between the points x_1 and x_2 lies below the line segment which links the points $(x_1, f(x_1))$ and $(x_2, f(x_2))$. The notion of convexity is very natural and underlies much of geometry and analysis.

Let $x_0 \in \mathbb{R}$ and $\delta > 0$. If f is convex on $(x_0 - \delta, x_0)$ and concave on $(x_0, x_0 + \delta)$ (or concave on $(x_0 - \delta, x_0)$ and convex on $(x_0, x_0 + \delta)$), then x_0 is called a *point of inflection*.

Theorem. *Let f be twice differentiable on (a, b) and suppose that $f'' > 0$ on (a, b). Then f is convex on (a, b).*

Proof. Let $x_1, x_2 \in (a, b)$ with $x_1 \neq x_2$. Define a continuous function g on $[0, 1]$ by

$$g(\theta) = f(\theta x_1 + (1 - \theta)x_2) - \theta f(x_1) - (1 - \theta)f(x_2), \quad \theta \in [0, 1].$$

A direct calculation shows that $g(0) = g(1) = 0$ and

$$g''(\theta) = (x_1 - x_2)^2 f''(\theta x_1 + (1 - \theta)x_2) > 0.$$

We need to prove that $g \leqslant 0$ on $[0, 1]$. Assume, to get a contradiction, that g is positive at some points of the interval $[0, 1]$. Then, by the maximum theorem, it attains its (positive) maximum at some point $\theta_0 \in (0, 1)$. By Fermat's theorem, we have $g'(\theta_0) = 0$ at this point. Also by our assumption $g''(\theta_0) > 0$. Then, by Theorem 5.2, we see that g has a local *minimum* at θ_0. It follows that θ_0 is both a local maximum and a local minimum of g. This is only possible if g is constant in a neighbourhood of θ_0. But if g is constant, its first and second derivatives must be zero in that neighbourhood, which contradicts the above strict inequality. ■

Example. The function $f(x) = x^2$ is convex on $\mathbb{R}$. The function $f(x) = 1/x$ is concave on $(-\infty, 0)$ and convex on $(0, \infty)$.

5.3 Global extrema

How to find the maximum value of a differentiable function f on $[a, b]$?

(i) find all points $c_1, c_2, \ldots \in [a, b]$ at which $f'(c_k) = 0$;

(ii) evaluate $f(c_k)$ for all k;

(iii) evaluate f at the endpoints a, b of the interval;

(iv) from the list $f(a)$, $f(b)$, $f(c_1)$, $f(c_2)$, $\ldots$, select the maximal value.

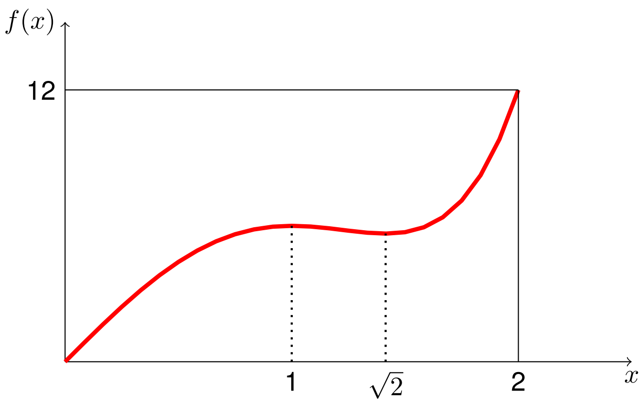
Remark.

1. Do not forget (iii) and (iv)! The function may take its maximal value at an endpoint!
2. The equality $f'(c) = 0$ does not imply that $f(c)$ is the maximum (or minimum) value of f. It may happen that f has only a local maximum (or minimum) at c, or that the value of f at an end point is greater (or smaller) than $f(c)$.
3. At step (i), you can compute $f''(c_k)$ and discard the points where the second derivative has a "wrong sign"; but this is not necessary.
4. Note that f may have several global maxima on $[a, b]$.
5. Of course, the minimum of f is investigated in the same way.

Example. Consider the function $f(x) = x^5 - 5x^3 + 10x$ on the interval $[0, 2]$. We have $f'(x) = 5x^4 - 15x^2 + 10 = 5(x^2 - 2)(x^2 - 1)$; the equation $f'(x) = 0$ has solutions $x = 1$ and $x = \sqrt{2}$ in the interval $[0, 2]$. We have

$$f(0) = 0, \; f(1) = 6, \; f(\sqrt{2}) = 4\sqrt{2} \approx 5.65, \; f(2) = 12$$

and so the minimum of f is attained at 0 and the maximum of f is attained at 2. The point 1 is a local maximum and $\sqrt{2}$ is a local minimum, but none of them are global. See the plot below.



5.4 Behaviour at infinity

Revise asymptotic expansions, see Section 4.5, and the limits at infinity.

Let f be a function on the real line such that $\lim_{x\to\infty} f(x) = c_0$. One can enquire about the rate of convergence of $f(x) \to c_0$ as $x \to \infty$. The precise way of expressing this idea is by looking at asymptotic expansions at infinity:

$$f(x) = \sum_{n=0}^{N} \frac{c_n}{x^n} + o(x^{-N}), \quad x \to \infty.$$

Example. Let us compute the first four terms of the asymptotic expansion of the function $f(x) = \dfrac{x^2+2x+1}{2x^2+1}$ in negative powers of x as $x \to \infty$. We expand $f(x)$ as a function of $1/x$:

$$\begin{aligned}
\frac{x^2+2x+1}{2x^2+1} &= \frac{1}{2}\frac{1+2/x+1/x^2}{1+1/2x^2} \\
&= \frac{1}{2}\left(1+\frac{2}{x}+\frac{1}{x^2}\right)\left(1-\frac{1}{2x^2}+O(x^{-4})\right) \\
&= \frac{1}{2}\left(1+\frac{2}{x}+\frac{1}{x^2}\right)\left(1-\frac{1}{2x^2}\right)+O(x^{-4}) \\
&= \frac{1}{2}+\frac{1}{x}+\frac{1}{4x^2}-\frac{1}{2x^3}+O(x^{-4}), \quad x \to \infty.
\end{aligned}$$

Example. Let us find the first three terms of the expansion of $f(x) = \sqrt{x^2+x}$ as $x \to \infty$. Using the expansion

$$\sqrt{1+\varepsilon} = 1 + \frac{1}{2}\varepsilon - \frac{1}{8}\varepsilon^2 + O(\varepsilon^3), \quad \varepsilon \to 0,$$

we find

$$\begin{aligned}
\sqrt{x^2+x} = x\sqrt{1+\frac{1}{x}} &= x\left(1+\frac{1}{2x}-\frac{1}{8x^2}+O(x^{-3})\right) \\
&= x+\frac{1}{2}-\frac{1}{8x}+O(x^{-2})
\end{aligned}$$

as $x \to \infty$.

If $f(x) = ax+b+o(1)$ as $x \to \infty$, we say that the straight line $y = ax+b$ is an *asymptote* of the graph of f. This means that the graph of $f(x)$ "gets close" to this line at infinity. In the previous example, we saw that $y = x + \frac{1}{2}$ is an asymptote for $f(x) = \sqrt{x^2+x}$.

5.5 Graph sketching

To sketch the graph of a concrete function $f(x)$, given by an explicit formula, you need to follow these steps:

1. Find the **natural domain** of the function (the subset of $\mathbb{R}$ where f is well defined);

2. Investigate the limiting behaviour at the boundary of the natural domain; if this includes infinity, determine the **asymptotes** at infinity, if they exist;
3. Find intervals of continuity, determine the **discontinuities and their type** (removable, jump, infinite or oscillatory);
4. Find **zeros** and intervals of constant sign ($f > 0$ or $f < 0$);
5. Find **extrema (max/min)** and intervals of monotonicity;
6. Find the intervals of **convexity/concavity** (this step is more difficult and can sometimes be skipped).

Example. Sketch the graph of the function $f(x) = \dfrac{x^4}{(1+x)^3}$ on its natural domain.

Solution: The function has an infinite discontinuity at $x = -1$, with $\lim\limits_{x\to -1_\pm} f(x) = \pm\infty$. As $x \to \infty$, we have

$$\begin{aligned} f(x) &= x^4(1+x)^{-3} = x(1 + \tfrac{1}{x} + O(x^{-2}))^{-3} \\ &= x(1 - \tfrac{3}{x} + O(x^{-2})) = x - 3 + O(x^{-1}), \end{aligned}$$

and so the line $y = x - 3$ is an asymptote. The function is non-negative for $x > -1$ and negative for $x < -1$, with a zero at $x = 0$. Computing the derivatives,

$$f'(x) = \frac{x^3(x+4)}{(1+x)^4}, \qquad f''(x) = \frac{12x^2}{(1+x)^5},$$

we find that the sign of the first derivative is

$(-\infty,-4)$	$(-4,-1)$	$(-1,0)$	$(0,\infty)$
$f' > 0$	$f' < 0$	$f' < 0$	$f' > 0$

Thus, there is a local maximum at $x = -4$ and a local minimum at $x = 0$. Similarly, $f'' < 0$ for $x < -1$ and $f'' > 0$ for $x > -1$; thus, the graph is concave on $(-\infty, -1)$ and convex on $(-1, \infty)$. Putting this information together, we sketch the graph as in Figure 1. Compare this with the graph given by your favourite graphic calculator. Observe that the graph given by a calculator, while being more accurate, may be less suggestive than the plot we have constructed by hand.

Example. Sketch the graph of the function $f(x) = \dfrac{\log x}{\sqrt{x}}$ on its natural domain.

Solution: The natural domain is $(0, \infty)$. We have

$$\lim_{x\to 0_+} f(x) = -\infty, \qquad \lim_{x\to\infty} f(x) = 0.$$

The function is negative for $x < 1$ and positive for $x > 1$, with a zero at $x = 1$. Computing the derivatives,

$$f'(x) = \frac{2 - \log x}{2x\sqrt{x}}, \qquad f''(x) = \frac{3\log x - 8}{4x^2\sqrt{x}},$$

we find that the function is increasing on $(0, e^2)$ and decreasing on (e^2, ∞), with a local maximum at e^2. The function is concave on $(0, e^{8/3})$ and convex on $(e^{8/3}, \infty)$, with an inflection point at $e^{8/3}$. Putting this information together, we can sketch the graph as shown in Figure 2.

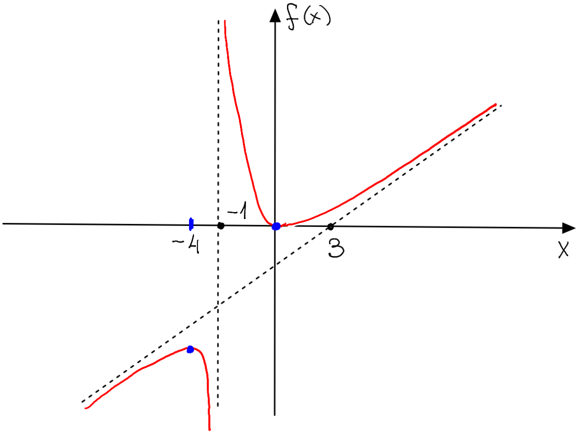


Figure 1: The graph sketch of $f(x) = \dfrac{x^4}{(1+x)^3}$

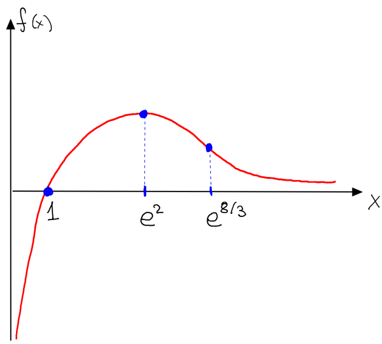


Figure 2: The graph sketch of $f(x) = (\log x)/\sqrt{x}$

$*\,*\,*\,*\,*$ READING WEEK $*\,*\,*\,*\,*$

---

6 Integration I: Integrability

6.1 Definitions

If f is a non-negative 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. One cannot hope to integrate every conceivable function, obtaining a well-defined real number as its integral. In some cases the function may be so irregular that it is not clear how to start to define its integral.

In this module we explain the basics of the theory of integration over a bounded interval $a < b$, due to the 19th Century German mathematician Bernhard Riemann. Functions that can be integrated according to this construction are said to be Riemann integrable and we denote the space of all such functions by $\mathcal{R}[a, b]$ ($\mathcal{R}$ for Riemann). We mostly follow [Rudin, Chapter 6].

Definition (Partition)**.** Let $[a, b]$ be a given interval. By a *partition* P of $[a, b]$ we mean a finite set of points $x_0, x_1, \ldots, x_n$, where

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

We write $\Delta x_i = x_i - x_{i-1}$, $i = 1, \ldots, n$.

Definition (Upper and lower Riemann sums)**.** Let f be a bounded real-valued function on $[a, b]$ and let P be a partition of $[a, b]$. We set

$$M_i = \sup_{x_{i-1} \leqslant x \leqslant x_i} f(x), \qquad U(P, f) = \sum_{i=1}^{n} M_i \Delta x_i,$$

$$m_i = \inf_{x_{i-1} \leqslant x \leqslant x_i} f(x), \qquad L(P, f) = \sum_{i=1}^{n} m_i \Delta x_i.$$

The quantities $U(P, f)$ and $L(P, f)$ are called the *upper and lower Riemann sums* of f, corresponding to the partition P.

Definition (Riemann integral)**.** Let f be a bounded real-valued function on $[a, b]$. We set

$$\overline{\int_a^b} f(x)dx = \inf_P U(P, f), \tag{6.1}$$

$$\underline{\int_a^b} f(x)dx = \sup_P L(P, f), \tag{6.2}$$

where the inf and sup are taken over all partitions P of $[a, b]$. The left hand sides of (6.1) and (6.2) are called the upper and lower Riemann integrals of f over $[a, b]$. If the upper and lower integrals are equal, we denote their common value by

$$\int_a^b f(x)dx,$$

we say that f is *Riemann integrable* on $[a, b]$ and we write $f \in \mathcal{R}[a, b]$.

Remark. It is immediate from the definition of upper and lower sums that

$$L(P, f) \leqslant U(P, f)$$

for any partition P. Very soon we will see that a stronger statement is true:

$$L(P_1, f) \leqslant U(P_2, f) \tag{6.3}$$

for any two partitions P_1, P_2.

Our immediate plans are as follows:

- prove that in all cases
$$\underline{\int_a^b} f(x)dx \leqslant \overline{\int_a^b} f(x)dx \tag{6.4}$$
(this is a consequence of (6.3));
- prove that functions from a wide class (including continuous functions) are Riemann integrable;
- establish simple standard properties of the Riemann integral (such as linearity with respect to f), most of them are being familiar to you from Calculus I.

Example. It is straightforward to see that a constant function is Riemann integrable, so our set $\mathcal{R}[a, b]$ is non-empty! On the other hand, for the Dirichlet function (on any interval $[a, b]$) we have $M_i = 1$ and $m_i = 0$ for all i, and therefore it is not Riemann integrable.

6.2 Riemann's criterion

Here we prove the estimates (6.3) and (6.4).

Definition (Refinement)**.** We say that the partition P^* is a *refinement* of P if $P \subset P^*$; that is, every point of P is a point of P^*. Given two partitions, P_1 and P_2, we say that $P^* = P_1 \cup P_2$ is the *common refinement* of P_1 and P_2.

Below f is a bounded real-valued function on $[a, b]$ and P is a partition of $[a, b]$.

Lemma (Refinement Lemma)**.** *If P^* is a refinement of P, then*

$$L(P, f) \leqslant L(P^*, f), \tag{6.5}$$
$$U(P, f) \geqslant U(P^*, f). \tag{6.6}$$

Proof. To prove (6.5), suppose first that P^* contains just one point more than P. Let this extra point be x_*, and suppose $x_{i-1} < x_* < x_i$, where x_{i-1} and x_i are two consecutive points of P. Put

$$w_1 = \inf_{x_{i-1} \leqslant x \leqslant x_*} f(x), \quad w_2 = \inf_{x_* \leqslant x \leqslant x_i} f(x),$$

and let, as before,

$$m_i = \inf_{x_{i-1} \leqslant x \leqslant x_i} f(x).$$

Since $[x_{i-1}, x_*] \subset [x_{i-1}, x_i]$, we have $w_1 \geqslant m_i$ and similarly $w_2 \geqslant m_i$. Hence

$$\begin{aligned} L(P^*, f) - L(P, f) &= w_1(x_* - x_{i-1}) + w_2(x_i - x_*) - m_i(x_i - x_{i-1}) \\ &= (w_1 - m_i)(x_* - x_{i-1}) + (w_2 - m_i)(x_i - x_*) \geqslant 0. \end{aligned}$$

If P^* contains k points more than P, we repeat this argument k times, and arrive at (6.5). The proof of (6.6) is analogous. ∎

Theorem. *We have*

$$\underline{\int_a^b} f(x)dx \leqslant \overline{\int_a^b} f(x)dx.$$

Proof. Let P^* be the common refinement of two partitions P_1 and P_2. By the Refinement Lemma,

$$L(P_1, f) \leqslant L(P^*, f) \quad \text{and} \quad U(P^*, f) \leqslant U(P_2, f).$$

We also have

$$L(P^*, f) \leqslant U(P^*, f),$$

which follows from the definition of the upper and lower Riemann sums. It follows that

$$L(P_1, f) \leqslant U(P_2, f)$$

for any partitions P_1 and P_2. If P_2 is fixed and the sup is taken over all P_1, this gives

$$\underline{\int_a^b} f(x)dx \leqslant U(P_2, f).$$

The theorem follows by taking the inf over all P_2. ∎

From here we derive Riemann's criterion for integrability of f.

Theorem. $f \in \mathcal{R}[a, b]$ *if and only if for every* $\varepsilon > 0$ *there exists a partition* P *such that*

$$U(P, f) - L(P, f) < \varepsilon. \tag{6.7}$$

Proof. The "if" part: For every P we have

$$L(P, f) \leqslant \underline{\int_a^b} f(x)dx \leqslant \overline{\int_a^b} f(x)dx \leqslant U(P, f).$$

Thus our assumption (6.7) implies that

$$0 \leqslant \overline{\int_a^b} f(x)dx - \underline{\int_a^b} f(x)dx < \varepsilon.$$

Hence if (6.7) can be satisfied for every $\varepsilon > 0$, then

$$\underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx,$$

which means $f \in \mathcal{R}[a,b]$.

The "only if" part: Suppose $f \in \mathcal{R}[a,b]$ and let $\varepsilon > 0$ be given. Then there exist partitions P_1 and P_2 such that

$$U(P_2, f) - \int_a^b f(x)dx < \varepsilon/2,$$

$$\int_a^b f(x)dx - L(P_1, f) < \varepsilon/2.$$

Let P be the common refinement of P_1 and P_2. Then the above two inequalities, together with the Refinement Lemma (see (6.5) and (6.6)), show that

$$U(P,f) \leqslant U(P_2,f) < \int_a^b f(x)dx + \frac{\varepsilon}{2} < L(P_1,f) + \varepsilon \leqslant L(P,f) + \varepsilon,$$

and so (6.7) holds for the partition P. ■

6.3 Monotonic functions are Riemann integrable

Revise the definition of monotonic function.

Theorem. *Let f be bounded and monotonic on $[a,b]$; then $f \in \mathcal{R}[a,b]$.*

Proof. Let $\varepsilon > 0$ be given; choose a partition P of $[a,b]$ such that $\Delta x_i < \varepsilon$ for all i. We suppose that f is monotonically increasing (the proof is analogous if f is decreasing). Then

$$M_i = f(x_i), \quad m_i = f(x_{i-1}), \quad i = 1, \dots, n,$$

so that

$$\begin{aligned} U(P,f) - L(P,f) &= \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\Delta x_i \\ &< \varepsilon \sum_{i=1}^{n} (f(x_i) - f(x_{i-1})) = \varepsilon(f(b) - f(a)). \end{aligned}$$

Since ε is arbitrary, by Riemann's criterion we find that $f \in \mathcal{R}[a,b]$. ■

6.4 Continuous functions are Riemann integrable

Revise the definition of uniformly continuous function and Cantor's theorem.

Theorem. *If f is continuous on $[a,b]$, then f is Riemann integrable on $[a,b]$. In symbols: $C[a,b] \subset \mathcal{R}[a,b]$.*

Proof. First we note the identity

$$\sup_{x,x' \in \Delta} |f(x) - f(x')| = \sup_{x \in \Delta} f(x) - \inf_{x \in \Delta} f(x) \tag{6.8}$$

for any interval Δ; we leave the proof as an exercise.

Let $\varepsilon > 0$ be given. By Cantor's theorem, f is uniformly continuous on $[a, b]$ and therefore there exists $\delta > 0$ such that

$$|f(x) - f(x')| < \varepsilon/(b - a) \tag{6.9}$$

if $x, x' \in [a, b]$ and $|x - x'| < \delta$.

Let P be any partition of $[a, b]$ such that $\Delta x_i < \delta$ for all i. Taking max over $x \in [x_{i-1}, x_i]$ and min over $x' \in [x_{i-1}, x_i]$ in (6.9) and using (6.8), we find that

$$M_i - m_i < \varepsilon/(b - a)$$

for all $i = 1, \dots, n$. It follows that

$$U(P, f) - L(P, f) = \sum_{i=1}^{n} (M_i - m_i)\Delta x_i < \frac{\varepsilon}{b - a} \sum_{i=1}^{n} \Delta x_i = \varepsilon.$$

By Riemann's criterion, we conclude that $f \in \mathcal{R}[a, b]$. ■

6.5 Piecewise integrable functions are integrable

Here we prove that if the interval $[a, b]$ is split into several pieces and a function is integrable on each piece, then it is integrable on the whole of $[a, b]$.

Lemma 6.1. *Let $a < c < b$ and let f be a bounded function on $[a, b]$ such that $f \in \mathcal{R}[a, c]$ and $f \in \mathcal{R}[c, b]$. Then $f \in \mathcal{R}[a, b]$.*

Proof. In this proof, we shall temporarily change the notation for upper and lower Riemann sums, indicating the dependance of the interval which is being partitioned. Let $\varepsilon > 0$ be given. By Riemann's criterion, there exist partitions P_1 of $[a, c]$ and P_2 of $[c, b]$ such that

$$U(P_1, f; [a, c]) - L(P_1, f; [a, c]) < \varepsilon \quad \text{and} \quad U(P_2, f; [c, b]) - L(P_2, f; [c, b]) < \varepsilon.$$

Let P be the partition of $[a, b]$ obtained as the union of P_1 and P_2. Then

$$\begin{aligned} U(P, f; [a, b]) &= U(P_1, f; [a, c]) + U(P_2, f; [c, b]), \\ L(P, f; [a, b]) &= L(P_1, f; [a, c]) + L(P_2, f; [c, b]). \end{aligned} \tag{6.10}$$

Combining these relations, we get

$$U(P, f; [a, b]) - L(P, f; [a, b]) < 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, by Riemann's criterion we obtain that $f \in \mathcal{R}[a, b]$. ■

From this statement it follows that if f is bounded and has only finitely many jump discontinuities on $[a, b]$ (and no other discontinuities), then it is Riemann integrable on $[a, b]$. It will be convenient to postpone the proof until Section 6.7, where we prove a slightly more general statement, allowing for oscillatory discontinuities.

6.6 Compositions of integrable functions

Revise the definition of Lipschitz continuity.

Theorem 6.2. *Let $f \in \mathcal{R}[a,b]$ and let φ be a Lipschitz continuous function on $\mathbb{R}$. Then $\varphi \circ f \in \mathcal{R}[a,b]$.*

Proof. Let P be a partition $x_0, \ldots, x_n$ of $[a,b]$ such that

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

By the definition of Lipschitz continuity, for every $x, x' \in [a,b]$ we have

$$|\varphi(f(x)) - \varphi(f(x'))| \leqslant A|f(x) - f(x')|$$

for some constant A. Taking a supremum over $x, x' \in [x_{i-1}, x_i]$ and applying (6.8), we find

$$\sup_{[x_{i-1},x_i]} \varphi(f(x)) - \inf_{[x_{i-1},x_i]} \varphi(f(x)) \leqslant A\big(\sup_{[x_{i-1},x_i]} f(x) - \inf_{[x_{i-1},x_i]} f(x)\big).$$

Multiplying by Δx_i and summing over i, we find

$$U(P, \varphi \circ f) - L(P, \varphi \circ f) \leqslant A(U(P,f) - L(P,f)) < A\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we find that $\varphi \circ f \in \mathcal{R}[a,b]$. ■

The conclusion of this theorem is still true if φ is continuous (but not necessarily Lipschitz continuous), but the proof of this fact is slightly more delicate; see [Rudin, Theorem 6.11].

6.7 Oscillatory discontinuities

Consider the function $f(x) = \sin(1/x)$ on $(0,1)$. Is it Riemann integrable on $[0,1]$? The following statement gives an affirmative answer.

Lemma. *Let f be bounded on $[a,b]$ and suppose that $f \in \mathcal{R}[a+\varepsilon, b-\varepsilon]$ for all sufficiently small $\varepsilon > 0$. Then $f \in \mathcal{R}[a,b]$.*

The proof is outlined in the exercises to this section.

Example. Define $f(x) = \sin(1/x)$ for $x \in (0,1]$ and $f(0) = 0$. Then f is bounded on $[0,1]$ and $f \in C[\varepsilon, 1]$ for any $0 < \varepsilon < 1$. Thus, $f \in \mathcal{R}[\varepsilon, 1]$. It follows that f is Riemann integrable on $[0,1]$.

Theorem. *Let f be bounded on $[a,b]$ and suppose that f has finitely many points of discontinuity on $[a,b]$. Then $f \in \mathcal{R}[a,b]$.*

Proof. Let the discontinuities of f be located at the points

$$a < x_1^* < x_2^* < \cdots < x_k^* < b.$$

By the previous Lemma, f is Riemann integrable on all intervals $[a, x_1^*]$, $[x_1^*, x_2^*]$, ..., $[x_k^*, b]$. Applying Lemma 6.1 k times, we find consecutively that $f \in \mathcal{R}[a, x_2^*]$, $\mathcal{R}[a, x_3^*]$, etc. and eventually $f \in \mathcal{R}[a,b]$. ■

Remark. One can prove a much more general statement, completely characterising Riemann integrable functions. Let f be a bounded function on $[a, b]$; then $f \in \mathcal{R}[a, b]$ if and only if the set of discontinuities of f is a *set of measure zero*. A set $E \subset \mathbb{R}$ is said to be of measure zero, if for any $\varepsilon > 0$ one can cover the set E by intervals (a_j, b_j), $j = 1, 2, 3, \ldots$ such that the total length

$$\sum_{j=1}^{\infty} (b_j - a_j) < \varepsilon.$$

We shall not prove this; for details, see [Rudin, Theorem 11.33].

7 Integration II: Properties of the Riemann integral

7.1 Identities for the Riemann integral

We start with a trivial remark that a constant function is integrable and

$$\int_a^b C\,dx = C(b-a).$$

In particular, the integral of a zero function is zero.
Revise Lemma 6.1 and its proof.

Theorem 7.1. *Let* $a < c < b$ *and suppose* $f \in \mathcal{R}[a,b]$. *Then*

$$\boxed{\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.}$$

Proof. We come back to the proof of Lemma 6.1. Let P_1, P_2, P be as in that proof. We have

$$\int_a^c f(x)dx \leqslant U(P_1, f; [a,c]) < L(P_1, f; [a,c]) + \varepsilon \leqslant \int_a^c f(x)dx + \varepsilon,$$
$$\int_c^b f(x)dx \leqslant U(P_2, f; [c,b]) < L(P_2, f; [c,b]) + \varepsilon \leqslant \int_c^b f(x)dx + \varepsilon.$$

Using this and (6.10), we find

$$\int_a^b f(x)dx \leqslant U(P, f; [a,b]) = U(P_1, f; [a,c]) + U(P_2, f; [c,b])$$
$$< \int_a^c f(x)dx + \int_c^b f(x)dx + 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we obtain the inequality

$$\int_a^b f(x)dx \leqslant \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Quite similarly, working with the lower Riemann sums instead of the upper ones, we obtain the opposite inequality

$$\int_a^b f(x)dx \geqslant \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Putting this together gives the required statement. ■

Theorem. $\mathcal{R}[a,b]$ *is a linear space and the map*

$$f \to \int_a^b f(x)\,dx$$

is linear on $\mathcal{R}[a,b]$. *In other words, if* $f, g \in \mathcal{R}[a,b]$ *and* $\lambda, \mu \in \mathbb{R}$, *then* $\lambda f + \mu g \in \mathcal{R}[a,b]$ *and*

$$\boxed{\int_a^b (\lambda f(x) + \mu g(x))\,dx = \lambda \int_a^b f(x)\,dx + \mu \int_a^b g(x)\,dx\,.}$$

Proof. We shall only prove the statement for $\lambda = \mu = 1$; the rest is covered in the exercises. Since $f, g \in \mathcal{R}[a, b]$, there exist partitions P_1 and P_2 such that

$$U(P_1, f) - L(P_1, f) < \varepsilon, \quad U(P_2, g) - L(P_2, g) < \varepsilon.$$

Let P^* be the common refinement of P_1 and P_2. By the Refinement Lemma (see (6.5) and (6.6)), we have

$$U(P^*, f) - L(P^*, f) < \varepsilon, \quad U(P^*, g) - L(P^*, g) < \varepsilon. \tag{7.1}$$

If P^* consists of the points $x_0, \ldots, x_n$, we find for every interval $[x_{i-1}, x_i]$

$$\sup_{[x_{i-1},x_i]} (f(x) + g(x)) \leqslant \sup_{[x_{i-1},x_i]} f(x) + \sup_{[x_{i-1},x_i]} g(x)$$

and so, summing over i,

$$U(P^*, f + g) \leqslant U(P^*, f) + U(P^*, g). \tag{7.2}$$

In a similar way we find

$$L(P^*, f + g) \geqslant L(P^*, f) + L(P^*, g).$$

Putting this together with (7.2), we find

$$U(P^*, f + g) - L(P^*, f + g) \leqslant (U(P^*, f) - L(P^*, f)) + (U(P^*, g) - L(P^*, g)).$$

Combining with (7.1), we arrive at

$$U(P^*, f + g) - L(P^*, f + g) < 2\varepsilon$$

and so $f + g \in \mathcal{R}[a, b]$.

Next, from (7.1) and

$$L(P^*, f) \leqslant \int_a^b f(x)dx \leqslant U(P^*, f)$$

we find

$$U(P^*, f) < L(P^*, f) + \varepsilon \leqslant \int_a^b f(x)dx + \varepsilon$$

and similarly for g,

$$U(P^*, g) < L(P^*, g) + \varepsilon \leqslant \int_a^b g(x)dx + \varepsilon.$$

Using these inequalities and (7.2), we get

$$\begin{aligned}\int_a^b (f(x) + g(x))dx &\leqslant U(f + g, P^*) \leqslant U(f, P^*) + U(g, P^*) \\ &\leqslant \int_a^b f(x)dx + \int_a^b g(x)dx + 2\varepsilon.\end{aligned}$$

Since ε is arbitrary, we get

$$\int_a^b (f(x)+g(x))dx \leqslant \int_a^b f(x)dx + \int_a^b g(x)dx.$$

Quite similarly we obtain the opposite inequality

$$\int_a^b (f(x)+g(x))dx \geqslant \int_a^b f(x)dx + \int_a^b g(x)dx.$$

This gives the required statement with $\lambda = \mu = 1$. ■

7.2 Inequalities for the Riemann integral

Theorem. *If $f, g \in \mathcal{R}[a,b]$ are such that $f \leqslant g$ on $[a,b]$, then*

$$\int_a^b f(x)dx \leqslant \int_a^b g(x)dx.$$

In particular (by taking $f=0$), the integral of a non-negative function is non-negative.

Proof. From $f \leqslant g$ we find, for any partition P,

$$U(P,f) \leqslant U(P,g).$$

Taking infimum over all partitions, we get the required inequality. ■

Theorem 7.2. *If $f \in \mathcal{R}[a,b]$ then $|f| \in \mathcal{R}[a,b]$ and*

$$\boxed{\left|\int_a^b f(x)\,dx\right| \leqslant \int_a^b |f(x)|\,dx \leqslant (b-a) \sup_{x\in[a,b]} |f(x)|.} \tag{7.3}$$

We recall that all functions $f \in \mathcal{R}[a,b]$ are bounded by definition, and so the supremum in the right hand side is finite.

Proof. The inclusion $|f| \in \mathcal{R}[a,b]$ follows from Theorem 6.2, because the function $\varphi(t) = |t|$ is Lipschitz continuous. Since $f \leqslant |f| \leqslant M := \sup_{x\in[a,b]} |f(x)|$, by monotonicity of the integral we have

$$\int_a^b f(x)dx \leqslant \int_a^b |f(x)|dx \leqslant \int_a^b M\,dx = M(b-a).$$

In the same way from $f \geqslant -|f|$ we get the lower bound

$$\int_a^b f(x)dx \geqslant -\int_a^b |f(x)|dx.$$

Putting these estimates together, we obtain the desired relation (7.3). ■

Revise convex functions.

Theorem (Jensen's inequality). *Let $f \in \mathcal{R}[a,b]$ and let φ be a <u>convex</u> function defined on the range of f. Then $\varphi \circ f \in \mathcal{R}[a,b]$ and*

$$\boxed{\varphi\left(\frac{1}{b-a}\int_a^b f(x)dx\right) \leqslant \frac{1}{b-a}\int_a^b \varphi \circ f(x)dx.}$$

Sketch of proof. We shall leave out the proof of the inclusion $\varphi \circ f \in \mathcal{R}[a,b]$. Assuming this inclusion, let $P = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a,b]$, and let $M_i = \sup_{[x_{i-1},x_i]} f$; then, by the convexity of φ

$$\varphi\left(\frac{1}{b-a}U(P,f)\right) = \varphi\left(\sum_{i=1}^n \frac{\Delta x_i}{b-a}M_i\right) \leqslant \sum_{i=1}^n \frac{\Delta x_i}{b-a}\varphi(M_i) = \frac{1}{b-a}\sum_{i=1}^n \Delta x_i \varphi(M_i);$$

here, crucially, we have used that

$$\sum_{i=1}^n \Delta x_i = b-a.$$

Further, it is not difficult to see that

$$\varphi(M_i) = \varphi\big(\sup_{[x_{i-1},x_i]} f\big) \leqslant \sup_{[x_{i-1},x_i]} \varphi \circ f(x_i),$$

and so we obtain

$$\varphi\left(\frac{1}{b-a}U(P,f)\right) \leqslant \frac{1}{b-a}U(P,\varphi \circ f).$$

Taking the infimum over all partitions P, we arrive at the desired inequality. ∎

Example. We have, for $f \in \mathcal{R}[0,1]$,

$$\left(\int_0^1 f(x)dx\right)^2 \leqslant \int_0^1 f(x)^2 dx,$$
$$\exp\left(\int_0^1 f(x)dx\right) \leqslant \int_0^1 \exp(f(x))dx.$$

7.3 The fundamental theorem of calculus

Revise Lipschitz continuity and left and right derivatives.

The first half of the fundamental theorem of calculus is:

Theorem (FTC part 1). *Let $f \in \mathcal{R}[a,b]$. For $x \in [a,b]$, put*

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

Then F is continuous on $[a,b]$. Furthermore, if f is continuous at a point $x_0 \in (a,b)$, then F is differentiable at x_0, and

$$F'(x_0) = f(x_0).$$

Proof. Since $f \in \mathcal{R}[a,b]$, f is bounded by definition. Let $M > 0$ be such that $|f(t)| \leqslant M$ for all $t \in [a,b]$. If $a \leqslant x_1 \leqslant x_2 \leqslant b$, then, using Theorem 7.1 at the first step and Theorem 7.2 at the last step, we find

$$|F(x_2) - F(x_1)| = \left| \int_a^{x_1} f(t)\mathrm{d}t - \int_a^{x_2} f(t)\mathrm{d}t \right| = \left| \int_{x_1}^{x_2} f(t)\mathrm{d}t \right| \leqslant M(x_2 - x_1).$$

It follows that F is Lipschitz continuous and therefore continuous on $[a,b]$.

Now suppose f is continuous at x_0. Given $\varepsilon > 0$, choose $\delta > 0$ such that

$$|f(t) - f(x_0)| < \varepsilon$$

if $|t - x_0| < \delta$ and $t \in [a,b]$. Hence, if $t \in (x_0, x_0 + \delta)$, we find

$$\begin{aligned} \left| \frac{F(t) - F(x_0)}{t - x_0} - f(x_0) \right| &= \left| \frac{1}{t - x_0} \int_{x_0}^{t} f(s)\mathrm{d}s - \frac{1}{t - x_0} \int_{x_0}^{t} f(x_0)\mathrm{d}s \right| \\ &= \left| \frac{1}{t - x_0} \int_{x_0}^{t} (f(s) - f(x_0))\mathrm{d}s \right| < \varepsilon, \end{aligned}$$

where we have used Theorem 7.2 (with $[x_0, t]$ in place of $[a,b]$) at the last step. This argument shows that

$$\lim_{t \to x_0+} \frac{F(t) - F(x_0)}{t - x_0} = f(x_0),$$

i.e. F has the right derivative $f(x_0)$ at the point x_0. A similar argument proves that F has the left derivative $f(x_0)$ at the same point. ■

Revise Corollary 5.1.

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 (FTC part 2). *Let $f \in C[a,b]$ with $f \in C^1(a,b)$ and $f' \in \mathcal{R}[a,b]$. Then for all $x \in [a,b]$,*

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

Proof. Let

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

It follows from the previous Theorem that $G'(x) = 0$ for all $x \in (a,b)$. By Corollary 5.1, G is constant on (a,b). Furthermore, since f is continuous at $x = a$ and $x = b$, we see that G is also continuous at a and b (here we also use the previous Theorem). Since $G(a) = 0$, we find that G is identically zero on $[a,b]$, as required. ■

7.4 Integration by parts and the change of variable

From the fundamental theorem of Calculus we easily obtain the familiar “rules” of integration: integration by parts and the change of variable.

Theorem (Integration by parts)**.** *Let* $f, g \in C[a,b]$ *with* $f', g' \in C[a,b]$*; then*

$$\int_a^b f'(x)g(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b f(x)g'(x)dx.$$

Note that by assumption, both $f'g$ and fg' are continuous on $[a,b]$ and therefore the two integrals are well-defined.

Sketch of proof. Apply the Fundamental Theorem of Calculus, part 2 to the product $f(x)g(x)$ and use the product rule. ■

Theorem (Change of variable)**.** *Let* $\varphi \in C[a,b]$ *and* $\varphi' \in C[a,b]$*, with* $\varphi(a) = A$ *and* $\varphi(b) = B$*, and let* $f \in C[A,B]$*. Then*

$$\int_A^B f(y)dy = \int_a^b f(\varphi(x))\varphi'(x)dx.$$

In the language of Calculus, here we are making the change of variable $y = \varphi(x)$. The proof is outlined in the exercises.

8 Integration III: Improper integrals

8.1 Unbounded intervals

Revise the limit of a function at infinity

Definition (Improper integrals 1).

- Let $f : [a, \infty) \to \mathbb{R}$, $f \in \mathcal{R}[a, b]$ for all $b > a$; we define
$$\int_a^\infty f(x)dx = \lim_{b\to\infty} \int_a^b f(x)dx,$$
if the limit exists.

- Let $f : (-\infty, b] \to \mathbb{R}$, $f \in \mathcal{R}[a, b]$ for all $a < b$; we define
$$\int_{-\infty}^b f(x)dx = \lim_{a\to-\infty} \int_a^b f(x)dx,$$
if the limit exists.

- Let $f : \mathbb{R} \to \mathbb{R}$, $f \in \mathcal{R}[a, b]$ for all $a < b$; we define
$$\int_{-\infty}^\infty f(x)dx = \lim_{b\to\infty} \int_0^b f(x)dx + \lim_{a\to-\infty} \int_a^0 f(x)dx,$$
if both limits exist.

Remark.

1. Of course, the existence of the limits above is not automatic; the limits may or may not exist. If the limit exists, we say that the corresponding integral "converges", or "is well defined", or "exists".
2. It is important that in the definition of $\int_{-\infty}^\infty f(x)dx$ we take *separate* limits; the limit
$$\lim_{b\to\infty} \int_{-b}^b f(x)dx$$
may exist even if the two separate limits above do not exist (example: $f(x) = x$).
3. In the definition of $\int_{-\infty}^\infty f(x)dx$, instead of integrating over $(-\infty, 0)$ and $(0, \infty)$, we could have integrated over $(-\infty, c)$ and (c, ∞) for any "reference point" $c \in \mathbb{R}$. The result is independent of the choice of the reference point c (the proof of this is an exercise).

Example. Let $\alpha \in \mathbb{R}$; then

$$\boxed{\int_1^\infty \frac{1}{x^\alpha}dx \quad \text{converges iff } \alpha > 1.}$$

This can be seen from the explicit formula for the integral

$$\int_1^b \frac{dx}{x^\alpha} = \frac{b^{1-\alpha} - 1}{1 - \alpha}, \quad \alpha \neq 1;$$

the special case $\alpha = 1$ must be considered separately.

Example. Let $\beta \in \mathbb{R}$; then the integral

$$\boxed{\int_2^\infty \frac{1}{x(\log x)^\beta}dx \quad \text{converges iff } \beta > 1.}$$

To see this, it suffices to make a change of variable $y = \log x$ and use the previous example.

8.2 Unbounded functions

Definition (Improper integrals 2)**.** Let $[a, b]$ be a bounded interval. If $f \in \mathcal{R}[a + \varepsilon, b]$ for all $\varepsilon > 0$ (but possibly $f \notin \mathcal{R}[a, b]$), we define

$$\int_a^b f(x)dx = \lim_{\varepsilon \to 0_+} \int_{a+\varepsilon}^b f(x)dx.$$

If this limit exists, we will say that the integral $\int_a^b f(x)dx$ converges. In a similar way, if $f \in \mathcal{R}[a, b - \varepsilon]$ for all $\varepsilon > 0$ (but possibly $f \notin \mathcal{R}[a, b]$), we define

$$\int_a^b f(x)dx = \lim_{\varepsilon \to 0_+} \int_a^{b-\varepsilon} f(x)dx,$$

and say that the integral $\int_a^b f(x)dx$ converges if this limit exists.

Example. Let $\alpha \in \mathbb{R}$; then the integral

$$\boxed{\int_0^1 \frac{dx}{x^\alpha} \quad \text{converges iff } \alpha < 1.}$$

Further, let $\beta \in \mathbb{R}$; then the integral

$$\boxed{\int_0^{1/2} \frac{dx}{x|\log x|^\beta} \text{ converges iff } \beta > 1.}$$

Similarly, suppose $[a, b]$ is a bounded interval and $c \in (a, b)$. Suppose $f \notin \mathcal{R}[a, b]$, but $f \in \mathcal{R}[a, c - \varepsilon]$ and $f \in \mathcal{R}[c + \varepsilon, b]$ for all $\varepsilon > 0$. Then we define

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx,$$

where each of the integrals in the right hand side is understood according to the previous definition.

Of course, the above definitions can be combined.

Example. The integral

$$\int_0^\infty \frac{dx}{x^\alpha} \quad \text{does NOT converge for any } \alpha \in \mathbb{R}.$$

Indeed, for the convergence of $\int_0^1$ we must have $\alpha < 1$ and for the convergence of $\int_1^\infty$ we must have $\alpha > 1$.

A rigorous proof of the fact that an improper integral *diverges* can often be a little laborious. The following example illustrates this point.

Example. Does the integral
$$\int_0^\infty \frac{dx}{e^x - 1}$$
converge? Denote $f(x) = 1/(e^x - 1)$. Clearly, we have two points to inspect: $x = 0$ and $x = \infty$. As $x \to \infty$ we have $f(x) = O(e^{-x})$, and so the integral clearly converges at infinity. As $x \to 0$ we have
$$f(x) = \frac{1}{x + \frac{1}{2}x^2 + O(x^3)};$$
thus, the integrand behaves roughly speaking as $1/x$ and we suspect that the integral diverges. Let us prove this. We consider

$$g(x) = \frac{1}{x} - f(x) = \frac{e^x - 1 - x}{x(e^x - 1)} = 1/2 + O(x), \quad x \to 0_+.$$

It follows that g extends to a continuous function on $[0, \infty)$, and so the integral $\int_0^1 g(x)dx$ converges. Now we use the following fact (proof is a very easy exercise): if the improper integrals
$$\int_0^1 f(x)dx \quad \text{and} \quad \int_0^1 g(x)dx$$
converge, then the improper integral $\int_0^1 (f(x) + g(x))dx$ also converges. Using this, we argue by contradiction: suppose $\int_0^1 f(x)dx$ converges, then $\int_0^1 (1/x)dx = \int_0^1 (f(x) + g(x))dx$ must also converge, which is a contradiction.

8.3 Revision: series

Revise Cauchy's criterion for convergence.

We recall (see lecture notes on Sequences and Series for the details) that a series of real (or complex) numbers $\sum_{n=1}^\infty a_n$ is called *convergent* to the *sum* A, if the sequence of partial sums
$$A_N = \sum_{n=1}^N a_n$$
converges to A. If the series $\sum_{n=1}^\infty a_n$ converges, then $a_n \to 0$ as $n \to \infty$, but the converse is false (example: $a_n = 1/\sqrt{n}$). A useful fact is that if $\sum_{n=1}^\infty a_n$ converges, then

$$\sum_{n=N}^\infty a_n \to 0, \quad N \to \infty$$

(proof is a simple exercise).

A series of complex numbers $\sum_{n=1}^\infty a_n$ is called *absolutely convergent*, if the series $\sum_{n=1}^\infty |a_n|$ is convergent; the latter fact is usually written as

$$\sum_{n=1}^\infty |a_n| < \infty.$$

If a series is absolutely convergent, then it is convergent. Below we recall the proof of this statement because our discussion of the integrals to follow is completely parallel to this.

Lemma 8.1. *Let $|a_n| \leqslant b_n$ for all $n = 1, 2, \dots$, and assume that the series $\sum_{n=1}^{\infty} b_n$ converges. Then the series $\sum_{n=1}^{\infty} a_n$ converges and, denoting its sum by A, we have the remainder estimate*

$$\left| A - \sum_{n=1}^{N} a_n \right| \leqslant \sum_{n=N+1}^{\infty} b_n, \quad N = 1, 2, \dots.$$

Proof. Denote

$$A_N = \sum_{n=1}^{N} a_n, \quad B_N = \sum_{n=1}^{N} b_n.$$

For any $K > N$ we have

$$|A_K - A_N| = \left| \sum_{n=1}^{K} a_n - \sum_{n=1}^{N} a_n \right| = \left| \sum_{n=N+1}^{K} a_n \right| \leqslant \sum_{n=N+1}^{K} |a_n| \leqslant \sum_{n=N+1}^{K} b_n = B_K - B_N.$$

As $\sum_n b_n$ converges, the sequence B_N is Cauchy and so, by the above inequality, A_N is also Cauchy; hence $\sum_n a_n$ converges. Finally, taking $K \to \infty$ in the above inequality, we obtain the required estimate. ∎

8.4 Absolute and conditional convergence of integrals

To make the following discussion more concrete, we focus on integrals of bounded functions over $[a, \infty)$; similar considerations apply to other unbounded intervals and also to unbounded functions integrated over finite or infinite intervals. The following discussion is completely parallel to that of the absolute and conditional convergence of series.

Let a be fixed and let $f \in \mathcal{R}[a, b]$ for any $b > a$. We say that the integral

$$\int_a^{\infty} f(x)dx$$

converges absolutely, if the integral

$$\int_a^{\infty} |f(x)|dx$$

converges (as an improper integral, i.e. in the sense of the definition at the start of this section). If the integral of f converges but does not converge absolutely, we will say that it *converges conditionally*.

If an integral converges absolutely, it converges. In fact, we prove a slightly more general statement.

Theorem 8.2. *Let $a \in \mathbb{R}$ and let f, g be functions on $[a, \infty)$; assume that $f, g \in \mathcal{R}[a, b]$ for any $b > a$ and $|f(x)| \leqslant g(x)$ for all $x \geqslant a$. If the integral $\int_a^{\infty} g(x)dx$ converges, then the integral $\int_a^{\infty} f(x)dx$ converges as well.*

Before coming to the proof, we need to translate Cauchy's criterion for convergence of sequences into the language of functions.

Lemma. *Let $a \in \mathbb{R}$ and let f be a function on $[a, \infty)$. Then the limit $\lim_{x\to\infty} f(x)$ exists if and only if for any sequence of points $\{x_n\}_{n=1}^{\infty}$ such that $x_n \to \infty$, the limit $\lim_{n\to\infty} f(x_n)$ exists.*

Proof. The "only if" part is an easy exercise. Let us prove the "if" part. Assume that for any $x_n \to \infty$, the limit $\lim_{n\to\infty} f(x_n)$ exists.

1) Let us check that the limit $\lim_{n\to\infty} f(x_n)$ is independent of the choice of the sequence $x_n \to \infty$. Suppose we have two sequences $x_n \to \infty$, $x'_n \to \infty$ such that $\lim_{n\to\infty} f(x_n) \neq \lim_{n\to\infty} f(x'_n)$. Consider the sequence

$$\{\tilde{x}_n\} = \{x_1, x'_1, x_2, x'_2, x_3, x'_3, \dots\};$$

then $\tilde{x}_n \to \infty$. But $f(\tilde{x}_n)$ tends to different limits over the subsequences with even and odd indices, so the limit $\lim_{n\to\infty} f(\tilde{x}_n)$ cannot exist - contradiction!

2) Denote the common value of all limits $\lim_{n\to\infty} f(x_n)$ by A. Let us prove (by contradiction) that $\lim_{x\to\infty} f(x) = A$. Suppose this is false; then there exists $\varepsilon > 0$ such that for all $R > 0$ there exists $x_R > R$ with $|f(x_R) - A| \geqslant \varepsilon$. Take $R = 1, 2, \dots$; we obtain a sequence $x_n \to \infty$ such that $f(x_n) \not\to A$ – contradiction! ■

Now we can come back to the proof of the theorem.

Proof of Theorem 8.2. Denote

$$F(b) = \int_a^b f(x)dx, \quad G(b) = \int_a^b g(x)dx.$$

By definition, convergence of the integral of g means that $G(b)$ converges to a finite limit as $b \to \infty$. By the previous lemma, this means that for any sequence $b_n \to \infty$, the limit of $G(b_n)$ exists. Take two terms b_n and b_m of this sequence; suppose for definiteness that $b_n < b_m$. Then

$$|F(b_m) - F(b_n)| = \left| \int_{b_n}^{b_m} f(x)dx \right| \leqslant \int_{b_n}^{b_m} |f(x)|dx \leqslant \int_{b_n}^{b_m} g(x)dx = G(b_m) - G(b_n).$$

Since $\{G(b_n)\}_{n=1}^{\infty}$ is Cauchy, it follows that $\{F(b_n)\}_{n=1}^{\infty}$ is Cauchy, and therefore it converges. Using the previous lemma again, we find that $F(b)$ converges to a finite limit as $b \to \infty$, as required. ■

Example. For $\alpha > 0$, let us examine the convergence of the integral

$$\int_1^{\infty} \frac{\sin x}{x^\alpha} dx. \tag{8.1}$$

The estimate $|\sin x| \leqslant 1$ shows that our integral converges absolutely when $\alpha > 1$.

What happens when $0 < \alpha \leqslant 1$? Let us show that in this case our integral converges conditionally. Integrating by parts, we find

$$\int_1^b \frac{\sin x}{x^\alpha} dx = -\int_1^b \frac{(\cos x)'}{x^\alpha} dx = -\frac{\cos b}{b^\alpha} + \cos 1 - \alpha \int_1^b \frac{\cos x}{x^{1+\alpha}} dx.$$

Here $\cos b/b^\alpha$ converges to zero as $b \to \infty$, while the integral in the right hand side converges absolutely because $1+\alpha > 1$. It follows that the integral (8.1) converges. Finally, let us show that in this case it does NOT converge absolutely. Observe that

$$|\sin x| \geqslant (\sin x)^2$$

for all x. Using this and the standard trigonometric identity for $(\sin x)^2$, we obtain

$$\int_1^b \frac{|\sin x|}{x^\alpha} dx \geqslant \int_1^b \frac{(\sin x)^2}{x^\alpha} dx = \frac{1}{2}\int_1^b \frac{1-\cos 2x}{x^\alpha} dx = \frac{1}{2}\int_1^b \frac{dx}{x^\alpha} - \frac{1}{2}\int_1^b \frac{\cos 2x}{x^\alpha} dx.$$

As above, integrating by parts, it is easy to see that the second integral in the right hand side converges to a finite limit as $b \to \infty$. On the other hand, since $\alpha \leqslant 1$, the first integral tends to infinity as $b \to \infty$. It follows that the left hand side tends to infinity.

Here is a very useful statement generalising the previous example.

Theorem 8.3. *Let f and g be continuous functions on $[a,\infty)$, such that (i) the integral*

$$\int_a^x f(t)dt$$

is bounded and (ii) $g(x)$ is continuously differentiable, goes to zero as $x \to \infty$ and is monotone. Then the integral

$$\int_a^\infty f(x)g(x)dx$$

converges.

The proof uses integration by parts and is left as an exercise.

Finally, we mention the integral test for convergence of series:

Theorem (Integral test)**.** *Let $f(x) \geqslant 0$ be a bounded non-increasing function on $[0,\infty)$. Then the series*

$$\sum_{n=1}^\infty f(n)$$

converges if and only if the integral

$$\int_0^\infty f(x)dx$$

converges. Furthermore, one has the estimate

$$\int_1^\infty f(x)dx \leqslant \sum_{n=1}^\infty f(n) \leqslant \int_0^\infty f(x)dx.$$

The proof is an exercise.

9 Sequences and series of functions I

In this section and the next one, we study sequences and series of continuous functions. We mainly focus on real-valued functions, although all of the results below are also valid for complex-valued functions. We follow [Rudin, Section 7] almost verbatim.

9.1 Interchanging limits: discussion of the main problem

Let f_1, f_2, f_3, ...be functions defined on an interval $\Delta \subset \mathbb{R}$. One says that f_n converge to a function f on Δ *pointwise* if $f_n(x) \to f(x)$ for each $x \in \Delta$ as $n \to \infty$. Similarly, if the series

$$\sum_{n=1}^{\infty} f_n(x)$$

converges for every $x \in \Delta$, we say that it converges *pointwise* and denote the sum by $f(x)$. The set of x for which the sequence or series converges is called its *domain of convergence*.

The main problem discussed here is whether important properties of the functions f_n are preserved under the limit operation. For instance, if the functions f_n are continuous, or differentiable, or integrable, is the same true for the limit function? What are the relations between f_n' and f', or between the integrals of f_n and f?

To say that f is continuous at a point $x \in \Delta$ means

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

Hence, to ask whether the limit of a sequence of continuous functions is continuous is the same as to ask whether

$$\lim_{t \to x} \lim_{n \to \infty} f_n(t) = \lim_{n \to \infty} \lim_{t \to x} f_n(t),$$

i.e. whether the two limit operations can be interchanged. This lies at the heart of the matter. Indeed, differentiation and integration are also limiting operations (for differentiation this is obvious, and for integration we take a limit of Riemann sums), and we are asking whether these limiting operations can be interchanged with $\lim\limits_{n \to \infty}$.

We shall now show by means of several examples that limit operations cannot in general be interchanged without affecting the result. Afterward, we shall define a new type of convergence, which is called *uniform convergence*, and prove that uniform limits can be interchanged with other limits.

First we illustrate the problem in the simplest possible setting.

Example (Two limits are not interchangeable)**.** For $n, m \in \mathbb{N}$, we set

$$s_{m,n} = \frac{m}{m+n}.$$

Then it is straightforward to see that

$$\lim_{n \to \infty} \lim_{m \to \infty} \frac{m}{m+n} = \lim_{n \to \infty} 1 = 1,$$

whereas

$$\lim_{m \to \infty} \lim_{n \to \infty} \frac{m}{m+n} = \lim_{m \to \infty} 0 = 0.$$

Example (Continuity is not interchangeable with limit)**.** Consider the functions $f_n(x) = x^n$ on the interval $[0, 1]$. Obviously, the functions f_n are continuous. It is clear that $f_n(x) \to 0$ as $n \to \infty$ for all $x < 1$ and $f_n(1) = 1$. Thus the functions f_n converge to the discontinuous function

$$f(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x = 1 \end{cases}$$

Example (Differentiation is not interchangeable with limit)**.** Let

$$f_n(x) = \frac{x}{1 + n^2x^2}, \quad x \in [-1, 1], \quad n \in \mathbb{N}.$$

Then it is clear that

$$f(x) = \lim_{n\to\infty} f_n(x) = 0,$$

and so $f'(x) = 0$. On the other hand,

$$f_n'(x) = \frac{1 - n^2x^2}{(1 + n^2x^2)^2}$$

which converges to the function

$$\lim_{n\to\infty} f_n'(x) = \begin{cases} 1, & x = 0, \\ 0, & x \neq 0. \end{cases}$$

Thus,

$$\frac{d}{dx} \lim_{n\to\infty} f_n(x) \neq \lim_{n\to\infty} \frac{d}{dx} f_n(x).$$

Example (Integration is not interchangeable with limit)**.** Let

$$f_n(x) = n^2xe^{-nx}, \quad x \geqslant 0, \quad n \in \mathbb{N}.$$

By the theorem "exponentials beat powers", we find

$$f(x) = \lim_{n\to\infty} f_n(x) = 0$$

for all $x > 0$; of course the same is true for $x = 0$. On the other hand, by a change of variable,

$$\int_0^\infty f_n(x)dx = \int_0^\infty xe^{-x}dx = 1,$$

and so

$$\lim_{n\to\infty} \int_0^\infty f_n(x)dx \neq \int_0^\infty \big(\lim_{n\to\infty} f_n(x)\big)dx.$$

9.2 Uniform convergence

Let f_n, $n = 1, 2, \ldots$, be a sequence of functions on an interval $\Delta \subset \mathbb{R}$, and let f be another function on Δ.

Definition. The sequence f_n converges to f *uniformly* on Δ if

$$\boxed{\forall \varepsilon > 0 \quad \exists n_\varepsilon \ : \ n > n_\varepsilon \ \Rightarrow \ |f(x) - f_n(x)| < \varepsilon \quad \forall x \in \Delta\,.}$$

Let us compare with the definition of pointwise convergence:

$$\forall x \in \Delta \quad \forall \varepsilon > 0 \quad \exists n_{\varepsilon,x} \ : \ n > n_{\varepsilon,x} \ \Rightarrow \ |f(x) - f_n(x)| < \varepsilon.$$

The crucial difference is that in pointwise convergence, the number $n_{\varepsilon,x}$ depends not only on ε but also on x. It is clear that uniform convergence implies pointwise convergence. We shall soon see that the converse is false. But first let us consider some examples of uniform convergence.

How to prove that $f_n(x) \to f(x)$ uniformly on Δ? One needs to find a sequence M_n such that

$$|f_n(x) - f(x)| \leqslant M_n \text{ for all } n \text{ and all } x \in \Delta$$

and $M_n \to 0$ as $n \to \infty$. The sequence M_n must be independent of x! This has a clear geometric interpretation: the graph of f_n must be confined to the strip between the lines $f(x) - M_n$ and $f(x) + M_n$, and the width $2M_n$ of this strip must tend to zero as $n \to \infty$.

Example. Let $f_n(x) = \frac{nx}{1+n+x}$, $x \in [0,1]$. It is easy to see that $f_n(x) \to x$ pointwise. In fact, the convergence is uniform. Indeed,

$$\left| x - \frac{nx}{1+n+x} \right| = \frac{x + x^2}{1+n+x} \leqslant \frac{2}{1+n}.$$

So we have found $M_n = 2/(1+n)$.

Next, let us give some examples showing that pointwise convergence does not imply uniform convergence. First let us make the following general point.

How to prove that $f_n(x)$ does NOT converge to $f(x)$ uniformly on Δ? One needs to find a sequence of points $x_n \in \Delta$ such that for some fixed $\varepsilon > 0$ and all n, we have $|f_n(x_n) - f(x_n)| > \varepsilon$.

Example. For $x \in \mathbb{R}$ and $n \in \mathbb{N}$, let

$$f_n(x) = \frac{1}{1 + |x - n|}.$$

It is clear that for every $x \in \mathbb{R}$, we have $f_n(x) \to 0$ as $n \to \infty$; thus we have the convergence $f_n \to 0$ pointwise. However, this convergence is NOT uniform! Indeed, take $x_n = n$; then

$$f_n(x_n) = f(n) = 1$$

for all n.

Example. Let $f_n(x) = nxe^{-nx}$ for $x \in [0,1]$. By the theorem “exponentials beat powers", we have $f_n \to 0$ pointwise. However, this convergence is not uniform. Indeed,

$$f_n(1/n) = e^{-1}$$

for all n.

Example. Let $f_n(x) = \sin(x/n)$, $x \in \mathbb{R}$. Then $f_n(x) \to 0$ pointwise. However, the convergence is NOT uniform; for example,

$$f_n(n) = \sin(1)$$

for all n.

Our immediate plans are as follows:

- In the rest of this section, we apply the notion of uniform convergence to series of functions;
- In the next section, we show that in terms of uniform convergence one can give sufficient conditions for interchanging limiting operations.

9.3 Uniform convergence of series

The most common way in which sequences of functions appear in mathematics (both pure and applied) is through series of functions. As we know, series is just a special kind of sequence. If $f_n(x)$, $n = 1, 2, \dots$ are functions defined on the same interval $\Delta \subset \mathbb{R}$, we discuss series of the form

$$\sum_{n=1}^{\infty} f_n(x). \tag{9.1}$$

Denote

$$F_N(x) = \sum_{n=1}^{N} f_n(x)$$

the Nth partial sum of the series. In accordance with our earlier definition of uniform convergence, we will say that the series (9.1) converges to $F(x)$ *uniformly* on Δ, if the sequence $F_N \to F$ uniformly on Δ.

The following sufficient condition is extremely useful:

Theorem (Weierstrass M-test)**.** *Let $\{f_n(x)\}_{n=1}^{\infty}$ be a sequence of functions on an interval Δ, and suppose that*

$$|f_n(x)| \leqslant M_n, \quad x \in \Delta, \quad n = 1, 2, \dots,$$

where the series $\sum_{n=1}^{\infty} M_n$ converges. Then the series (9.1) converges absolutely and uniformly on Δ.

Proof. By our assumptions, the series (9.1) converges absolutely for any $x \in \Delta$; denote its sum by $F(x)$:

$$F(x) = \lim_{N\to\infty} F_N(x) = \lim_{N\to\infty} \sum_{n=1}^{N} f_n(x).$$

By Lemma 8.1 we have

$$|F(x) - F_N(x)| \leqslant \alpha_N := \sum_{n=N+1}^{\infty} M_n.$$

Note that M_N is independent of x! Since $\sum_{n=1}^{\infty} M_n$ converges, we have $\alpha_N \to 0$ as $N \to \infty$. Thus, $F_N \to F$ uniformly on Δ. ∎

Example. It may happen that the series (9.1) converges absolutely, but not uniformly on an interval Δ. Indeed, for a set $A \subset \mathbb{R}$, let us denote by χ_A be the characteristic function of A:

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Let $f_n(x) = \chi_{[n-1,n)}(x)$; consider the series $\sum_{n=1}^{\infty} f_n(x)$ for $x > 0$. We have $F_N(x) = \chi_{[0,N)}(x) \to 1$ as $N \to \infty$, so the series converges pointwise to the function identically equal to 1. On the other hand, $F_N(N+1) = 0$, so convergence is not uniform.

Below we illustrate the above concepts by considering three common types of series: Power series, Fourier series and Dirichlet series.

9.4 Power series

Power series are series of the form

$$\sum_{n=0}^{\infty} a_n x^n, \tag{9.2}$$

where the coefficients a_n are real numbers and $x \in \mathbb{R}$. It is often more natural to regard x as complex variable, but here we confine ourselves to considering real numbers.

Theorem 9.1. *If the power series* (9.2) *converges for some* $x_0 \neq 0$*, then it converges absolutely on the interval* $(-|x_0|, |x_0|)$*; for any* $\varepsilon > 0$*, the convergence is uniform on the sub-interval* $[-|x_0| + \varepsilon, |x_0| - \varepsilon]$.

Proof. Denote $R = |x_0|$ and let $\varepsilon > 0$ be sufficiently small so that $R - \varepsilon > 0$. If (9.2) converges for $x = x_0$ then in particular the terms of the series are bounded:

$$|a_n||x_0|^n = |a_n|R^n \leqslant C$$

for all n. Then for $|x| \leqslant R - \varepsilon$ we have, denoting $\alpha = (R - \varepsilon)/R < 1$,

$$|a_n x^n| \leqslant |a_n| R^n \alpha^n \leqslant C\alpha^n,$$

and so we have the absolute and uniform convergence by the Weierstrass M-test with $M_n = C\alpha^n$. Since $\varepsilon > 0$ can be chosen arbitrary small, the series converges absolutely for all $|x| < R$. ■

Consider the domain of convergence of (9.2), i.e. the set

$$D = \{x : \text{ (9.2) is convergent at } x\}.$$

The previous theorem tells us that if $x \in D$, then the whole interval $(-|x|, |x|)$ belongs to D. It follows that D can be $\{0\}$, the whole real line, or an interval of the form $(-R, R)$, $[-R, R]$, $(-R, R]$ or $[-R, R)$ for some $R > 0$. This number R is called the *radius of convergence* of the series. In other words,

$$R = \sup\{|x| : \text{ (9.2) is convergent at } x\}$$

(with the understanding that R may also be 0 or ∞).

Remark. Without any changes, this theorem and the subsequent reasoning can be applied to the power series (9.2) considered as a function of a complex variable x. It then shows that the domain of convergence of a power series (9.2) in the complex plane is always a disk centered at the origin. (This explains the term "radius of convergence".) This reasoning doesn't tell us anything about convergence on the boundary of the disk (i.e. on the circle $|x| = R$). In fact, a power series may diverge at every point on the boundary, or diverge at some points and converge at other points, or converge at all the points on the boundary. We shall not discuss this in any detail. Power series of a complex variable are a central topic of the *Complex Analysis* module.

There is a very useful formula for the radius of convergence.

Theorem. *The radius of convergence R of a power series* (9.2) *is given by the formula*

$$R = 1/\alpha, \quad \alpha = \limsup_{n\to\infty} |a_n|^{1/n},$$

where $R = \infty$ if $\alpha = 0$ and $R = 0$ if $\alpha = \infty$.

The proof is outlined in the exercises.

9.5 Fourier series

Let $\{a_n\}_{n\in\mathbb{Z}}$ be a sequence of complex (or real) numbers; the index n takes all integer values from $-\infty$ to ∞. A *Fourier series* is a series of the form

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx}.$$

Convergence is understood as the convergence of partial sums

$$\sum_{n=-N}^{N} a_n e^{inx}$$

as $N \to \infty$. Fourier series are considered in much detail in the *Fourier Analysis* course. Since every term e^{inx} is 2π-periodic, the sum of the Fourier series is also a 2π-periodic function. (Alternatively, we may restrict x to an interval of the length 2π, e.g. $x \in [-\pi, \pi]$.) One of the central statements in Fourier Analysis is that *any* continuous 2π-periodic function can be represented as a sum of the Fourier series.

Assume that

$$\sum_{n=-\infty}^{\infty} |a_n| < \infty, \text{ i.e. } \sum_{n=0}^{\infty} |a_n| < \infty \text{ and } \sum_{n=1}^{\infty} |a_{-n}| < \infty.$$

Then, since

$$|a_n e^{inx}| = |a_n|,$$

by the Weierstrass M-test, the Fourier series converges absolutely and uniformly on $\mathbb{R}$. We will come back to Fourier series in the next section.

9.6 Dirichlet series

Let $\{a_n\}_{n=1}^{\infty}$ be real or complex numbers; the Dirichlet series is

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s}; \tag{9.3}$$

here we follow the tradition (going back to Riemann) to denote the variable by s rather than x. Usually s is regarded as a complex variable, but here we restrict ourselves to the real case. The simplest and the most famous example is the Riemann Zeta function, which corresponds to $a_n \equiv 1$. Dirichlet series are a central object in *Number Theory*.

The following theorem is a little more subtle than its analogue for power series (Theorem 9.1).

Theorem. *If the Dirichlet series converges for some $s = s_0 \in \mathbb{R}$, then it converges on (s_0, ∞); for any $\varepsilon > 0$, the convergence is uniform on $(s_0 + \varepsilon, \infty)$.*

Proof. We shall only prove convergence in the range (s_0, ∞); the proof of the uniform convergence in $(s_0 + \varepsilon, \infty)$ is outlined in the exercises.

Assume $s_0 = 0$ (otherwise repeat the proof below with $a_n n^{-s_0}$ in place of a_n). Convergence of the series for $s = 0$ means that $\sum a_n$ converges (and in particular, a_n are bounded). We observe that the convergence of (9.3) for $s > 1$ easily follows by comparison with $\sum n^{-s}$. The case of $s \leqslant 1$ is more subtle; we shall deal with it by directly considering the partial sum of the series (9.3) and using the trick called Abel summation (which is completely analogous to integration by parts). Denote

$$r_n = \sum_{j=n}^{\infty} a_j$$

and observe that

$$a_n = r_n - r_{n+1}, \quad n \geqslant 1.$$

We have

$$\begin{aligned}
\sum_{n=1}^{N} a_n n^{-s} &= \sum_{n=1}^{N} (r_n - r_{n+1}) n^{-s} = \sum_{n=1}^{N} r_n n^{-s} - \sum_{n=1}^{N} r_{n+1} n^{-s} \\
&= r_1 + \sum_{n=2}^{N} r_n n^{-s} - \sum_{n=2}^{N} r_n (n-1)^{-s} - r_{N+1} N^{-s} \\
&= r_1 - r_{N+1} N^{-s} - \sum_{n=2}^{N} r_n ((n-1)^{-s} - n^{-s}).
\end{aligned} \tag{9.4}$$

Next, by using the mean value theorem, is it easy to prove the estimate

$$(n-1)^{-s} - n^{-s} = O(n^{-s-1}), \quad n \to \infty.$$

Since r_n is bounded, it follows by comparison with $\sum n^{-s-1}$ that the series

$$\sum_{n=2}^{\infty} r_n ((n-1)^{-s} - n^{-s})$$

converge, i.e. the sum in the right hand side of (9.4) converges to a finite limit as $N \to \infty$. The term $r_{N+1}N^{-s}$ converges to zero as $N \to \infty$. Putting this together, we see that the partial sum of the Dirichlet series (9.3) converges to a finite limit as $N \to \infty$, i.e. our Dirichlet series converges, as claimed. ■

It follows that the domain of convergence of a Dirichlet series must be of one of the following forms: the whole of $\mathbb{R}$, empty set, or the half-line (σ, ∞) or $[\sigma, \infty)$. All of these possibilities can be realised. The number σ is called the *abscissa of convergence*.

If s is regarded as a complex variable, then, following the same logic, one proves that if (9.3) converges for some s_0, then it converges for all s with $\mathrm{Re}\, s > \mathrm{Re}\, s_0$. Thus, the *half-plane of convergence* $\{s \in \mathbb{C} : \mathrm{Re}\, s > \sigma\}$ for Dirichlet series is the analogue of the *disk of convergence* $\{x \in \mathbb{C} : |x| < R\}$ for power series.

Remark. By following the same logic (see the remark about $s > 1$ at the start of the proof of Theorem 9.6), one proves that if (9.3) converges for s_0, then it converges absolutely for $s > s_0 + 1$. This result is sharp (in the sense that 1 in $s > s_0 + 1$ cannot be replaced by any number less than one), as demonstrated by the example of the alternate zeta series

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^s},$$

which converges iff $s > 0$ and converges absolutely iff $s > 1$ (see exercises).

10 Sequences and series of functions II

10.1 Uniform convergence and continuity

Theorem 10.1. *If a sequence of continuous functions f_n converges uniformly on Δ to a function f, then f is also continuous.*

Proof. Let us fix a point $c \in \Delta$ and a positive number ε. We have to show that there exists $\delta > 0$ such that $|f(x) - f(c)| < \varepsilon$ whenever $x \in \Delta$ and $|x - c| < \delta$. Since f_n converge to f uniformly on Δ, we have

$$|f(c) - f_n(c)| < \varepsilon/3 \text{ and } |f(x) - f_n(x)| < \varepsilon/3$$

for all $x \in \Delta$ whenever n is sufficiently large. Let us fix an arbitrary n, for which this is true.

Since f_n is a continuous function on the interval Δ, there exists $\delta > 0$ such that

$$|f_n(x) - f_n(c)| < \varepsilon/3 \text{ whenever } x, c \in \Delta \text{ and } |x - c| < \delta.$$

By the above,

$$\begin{aligned} |f(x) - f(c)| &\leqslant |f(x) - f_n(x)| + |f_n(x) - f_n(c)| + |f_n(c) - f(c)| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon \end{aligned}$$

whenever $x, c \in \Delta$ and $|x - c| < \delta$. This completes the proof. ∎

Remark. The pointwise convergence is not sufficient to prove the theorem. Indeed, in this case n in the inequality $|f(x) - f_n(x)| < \varepsilon/3$ may depend on x and, if it does, we are not able to estimate $|f_n(x) - f_n(c)|$ for all $x \in (c - \delta, c + \delta)$.

Corollary 10.2. *If $\sum_{n=1}^{\infty} f_n(x)$ is a series of continuous functions on Δ which converges uniformly on Δ, then the sum of the series is a continuous function.*

Proof. Apply the theorem above to the sequence of partial sums. ∎

Example. Consider the power series

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

with the radius of convergence $R > 0$. Each term of the series is a continuous function on $(-R, R)$. As we already know, the convergence is uniform on the interval $(-R + \varepsilon, R - \varepsilon)$ for any $\varepsilon > 0$; it follows that $f(x)$ is continuous on this interval. Since $\varepsilon > 0$ is arbitrary, we conclude that $f(x)$ is continuous on $(-R, R)$. In fact, as we shall see soon, $f(x)$ is differentiable infinitely many times on $(-R, R)$.

Example. Consider the Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx}, \qquad \sum_{n=-\infty}^{\infty} |a_n| < \infty.$$

Every term of the series is a continuous function of $x \in \mathbb{R}$. As we already know, the series converges uniformly on $\mathbb{R}$. Thus, the sum of the series is a continuous function on $\mathbb{R}$.

10.2 Uniform convergence and integration

Revise the definition of Riemann integral.

Theorem 10.3. *Suppose $f_n \in \mathcal{R}[a,b]$ for $n = 1, 2, \ldots$, and suppose that $f_n \to f$ uniformly on $[a,b]$. Then $f \in \mathcal{R}[a,b]$ and*

$$\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

Proof. Put

$$\varepsilon_n = \sup_{x\in[a,b]} |f_n(x) - f(x)|, \quad n = 1, 2, \ldots;$$

by the definition of uniform convergence we have $\varepsilon_n \to 0$ as $n \to \infty$. We have

$$f_n - \varepsilon_n \leqslant f \leqslant f_n + \varepsilon_n,$$

so that the upper and lower integrals of f satisfy

$$\int_a^b (f_n(x) - \varepsilon_n)dx \leqslant \underline{\int_a^b} f(x)dx \leqslant \overline{\int_a^b} f(x)dx \leqslant \int_a^b (f_n(x) + \varepsilon_n)dx. \tag{10.1}$$

Hence

$$0 \leqslant \overline{\int_a^b} f(x)dx - \underline{\int_a^b} f(x)dx \leqslant 2\varepsilon_n(b-a).$$

Since $\varepsilon_n \to 0$, the upper and lower integrals of f are equal; thus $f \in \mathcal{R}[a,b]$. Another application of (10.1) now yields

$$\left| \int_a^b f(x)dx - \int_a^b f_n(x)dx \right| \leqslant \varepsilon_n(b-a).$$

Taking $n \to \infty$, we obtain the second claim. ∎

In Exercises, you are asked to prove an analogue of this theorem for improper integrals.

Corollary. *If $f_n \in \mathcal{R}[a,b]$ and if the series*

$$f(x) = \sum_{n=1}^{\infty} f_n(x),$$

converges uniformly on $[a,b]$, then $f \in \mathcal{R}[a,b]$ and

$$\int_a^b f(x)dx = \sum_{n=1}^{\infty} \int_a^b f_n(x)dx.$$

In other words, uniformly convergent series may be integrated term by term.

Example. Let us come back to the example of Fourier series

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx}, \qquad \sum_{n=-\infty}^{\infty} |a_n| < \infty.$$

Can we compute the *Fourier coefficients* a_n directly in terms of $f(x)$? It turns out that this question has a very simple answer! Let us multiply $f(x)$ by e^{-imx} and integrate over $[-\pi, \pi]$. By the above Corollary, we can integrate the series term by term, which gives

$$\int_{-\pi}^{\pi} f(x)e^{-imx} dx = \sum_{n\in\mathbb{Z}} a_n \int_{-\pi}^{\pi} e^{i(n-m)x} dx = 2\pi a_m,$$

where at the last step we have used the elementary formula

$$\int_{-\pi}^{\pi} e^{ikx} dx = \begin{cases} 2\pi, & k = 0, \\ 0, & k \neq 0. \end{cases}$$

To summarize: the n'th Fourier coefficient of f can be computed by the formula

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-inx} dx.$$

10.3 Uniform convergence and differentiation

Theorem. *For $n = 1, 2, \dots$, let $f_n \in C[a, b]$ be a sequence of differentiable functions such that $f_n' \in C[a, b]$ and both sequences f_n and f_n' converge uniformly on $[a, b]$. Then the limit of f_n is differentiable and*

$$\frac{d}{dx} \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \frac{d}{dx} f_n(x), \qquad x \in (a, b).$$

Proof. Denote

$$f(x) = \lim_{n\to\infty} f_n(x), \qquad g(x) = \lim_{n\to\infty} f_n'(x).$$

For each n, by the Fundamental Theorem of Calculus (part 2)

$$f_n(x) - f_n(a) = \int_a^x f_n'(t)dt.$$

Since $f_n \to f$ and $f_n' \to g$ uniformly we can pass to the limit in both parts of this equation, and by Theorem 10.3 we may interchange integration with the limit on the right hand side:

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

By Theorem 10.1, g is continuous. Thus, by the other part of the Fundamental Theorem of Calculus (part 1), the right hand side is differentiable in x, and

$$f'(x) = g(x),$$

as required. ■

Corollary 10.4. *For $n = 1, 2, \dots$, let $f_n \in C[a,b]$ be a sequence of differentiable functions with $f_n' \in C[a,b]$ such that both series*

$$\sum_{n=1}^{\infty} f_n, \quad \sum_{n=1}^{\infty} f_n'$$

converge uniformly on $[a,b]$. Then the sum of the first series is differentiable and

$$\frac{d}{dx}\sum_{n=1}^{\infty} f_n(x) = \sum_{n=1}^{\infty} \frac{d}{dx} f_n(x), \quad x \in (a,b).$$

10.4 Application to power series

Theorem. *Let $R > 0$ be the radius of convergence of a power series*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Then f is differentiable on $(-R,R)$ and its derivative is given by the power series

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}, \tag{10.2}$$

which converges in $(-R,R)$.

Proof. Let $|x| < R - \varepsilon$; denoting $\alpha = |x|/(R-\varepsilon)$, we have

$$|n a_n x^{n-1}| = n|a_n||x|^{n-1} = n\alpha^{n-1}(|a_n|(R-\varepsilon)^{n-1}).$$

We know that

- since $R - \varepsilon \in (-R,R)$, the series $\sum a_n (R-\varepsilon)^n$ converges and so $|a_n|(R-\varepsilon)^n$ is bounded;
- the series $\sum n\alpha^n$ converges for $0 \leqslant \alpha < 1$.

It follows by comparison test (Lemma 8.1) that the series (10.2) converges. Thus, the radius of convergence of (10.2) is at least $R - \varepsilon$. Since $\varepsilon > 0$ can be taken arbitrary small, we see that the radius of convergence of (10.2) is at least R. (In fact, it is exactly R, but we shall not dwell on this.)

Now let us apply Corollary 10.4. It follows that the identity (10.2) holds for $|x| < R - \varepsilon$. Since $\varepsilon > 0$ can be chosen arbitrarily small, we conclude that (10.2) holds for all $|x| < R$. ∎

Corollary. *Let*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

be a power series with the radius of convergence $R > 0$. Then f has derivatives of all orders in $(-R, R)$, which are given by

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)a_n x^{n-k}.$$

In particular,

$$f^{(k)}(0) = k!a_k, \quad k = 0, 1, \ldots. \tag{10.3}$$

Proof. Apply the previous theorem successively to f, f', f'', etc. To get (10.3), just put $x = 0$ in the previous formula. ■

Using (10.3), we can rewrite our original power series as

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n, \tag{10.4}$$

which of course, agrees with Taylor's formula. Most elementary functions can be represented as power series:

$$\begin{aligned}
e^x &= 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots, && x \in \mathbb{R} \\
\sin x &= x - \frac{x^3}{3!} + \cdots + (-1)^{n-1} \frac{x^{2n-1}}{(2n-1)!} + \cdots, && x \in \mathbb{R} \\
\cos x &= 1 - \frac{x^2}{2!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots, && x \in \mathbb{R} \\
\log(1+x) &= x - \frac{x^2}{2} + \cdots + (-1)^{n-1} \frac{x^n}{n} + \cdots, && |x| < 1 \\
(1+x)^\alpha &= 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \cdots + \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!} x^n + \cdots, && |x| < 1.
\end{aligned}$$

Remark. There exist functions f which have derivatives of all orders, yet they cannot be expanded in power series, i.e. formula (10.4) for such functions is WRONG! The standard example is

$$f(x) = e^{-1/x^2}, \quad x \neq 0$$

and $f(0) = 0$. It is easy to see that f is differentiable to any order and $f^{(n)}(0) = 0$ for all n (to prove this, you need to use some algebra and the theorem "exponentials beat powers"). Thus the power series with coefficients $a_n = f^{(n)}(0)/n!$ is simply identical zero, but our function is not zero for $x \neq 0$!

10.5 Continuous nowhere differentiable functions

Theorem. *There exists a continuous function on $\mathbb{R}$ which is nowhere differentiable.*

Sketch of proof. Define

$$\varphi(x) = |x|, \quad -1 \leqslant x \leqslant 1,$$

and extend φ to the real line as a 2-periodic function, i.e. by requiring that

$$\varphi(x+2) = \varphi, \quad x \in \mathbb{R}.$$

(Draw a graph of this function!) Then, evidently, φ is continuous on $\mathbb{R}$. Define

$$f(x) = \sum_{n=0}^{\infty} (3/4)^n \varphi(4^n x).$$

Since $0 \leqslant \varphi \leqslant 1$, the Weierstrass M-test shows that the series converges uniformly on $\mathbb{R}$. Corollary 10.2 shows that the sum $f(x)$ is continuous on $\mathbb{R}$.

One can show that f is nowhere differentiable. For the proof, we refer to [Rudin, Theorem 7.18]. Here we only give a brief heuristic comment explaining this fact. Formally, we have

$$f'(x) = \sum_{n=0}^{\infty} 3^n \varphi'(4^n x).$$

Note that $\varphi'(x)$ always takes values ± 1. Thus, the n'th term in the above series is $\pm 3^n$; this does not converge to zero as $n \to \infty$, and so the series diverges.

Of course, this “proof" is not rigorous; what it shows precisely is that the term by term differentiation in the formula for $f'(x)$ is not legitimate. ■

The first example of this kind was constructed by K.Weierstrass in 1872. Weierstrass considered the series

$$\sum_{n=1}^{\infty} a^n \sin(b^n x),$$

where $0 < a < 1$ and $b > 1$. Condition $0 < a < 1$ ensures that the series converges uniformly on the real line. Weierstrass proved that if b is a sufficiently large integer, then the sum of the series is nowhere differentiable. G.H.Hardy in 1916 has sharpened this result by showing that that the conclusion is true if and only if $ab \geqslant 1$.

References

[Rudin] W. Rudin, *Principles of Mathematical Analysis,* Third Edition, McGraw-Hill.

[Haggarty] R. Haggarty, *Fundamentals of Mathematical Analysis,* Second Edition, Addison-Wesley.

[Brannan] D. Brannan, *A First Course in Mathematical Analysis,* Cambridge University Press.