Week 1

Continuity of the inverse function. Infinite series

Version 2025/02/02 To accessible online version of this chapter

These notes are being developed to reflect the content of the Real Analysis course as taught in the 2024/25 academic year. The first half of the course is lectured by Dr Yuri Bazlov. Questions and comments on these lecture notes should be directed to `Yuri.Bazlov@manchester.ac.uk`. The second half will be lectured by Dr Mark Coleman.

Pre-requisite: Mathematical Foundations and Analysis

We build upon what was achieved in the **Mathematical Foundations and Analysis** (**MFA**) course, taught in Semester 1. **Limit of a sequence, continuous function** and **limit of a function** remain key notions in Real Analysis, which develops the "analysis" part of MFA further. Important functions of real variable, used in MFA, will be formally defined, and their properties proved. This includes the power function x^α with arbitrary real α; exponential function a^x and the logarithm; trigonometric functions.

Hence, the Real Analysis course will **constantly require the students to call on their knowledge of definitions and results introduced in MFA.**

An informal introduction (not covered in lectures)

The **first part** of the course will introduce **infinite series.** Recall Euler's identity

$$e^{i\pi} + 1 = 0,$$

often brought up as an example of a beautiful mathematical result. But what exactly is there to prove? Let us try to understand this equation and its connection to Real Analysis.

First of all, e is the irrational number discovered by Jacob Bernoulli in 1683 (and initially written as "b") in his study of compound interest. It is defined as

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n, \text{ equivalently, } e = \sum_{n=0}^{\infty}\frac{1}{n!} = 1 + \frac{1}{1} + \frac{1}{1\cdot 2} + \frac{1}{1\cdot 2\cdot 3} + \dots .$$

Both formulae involve an infinite process, which in practice can only approximate e to a given precision. Note how the limit of a sequence, defined in MFA, appears next to something new: a "**sum of infinitely many numbers**".

What is $e^{i\pi}$, though? Raising e to an imaginary power is done via the rule

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots$$

This is where the meaning of Euler's identity starts to come across. The "infinite sum" for $e^{i\pi}$ breaks down into the real and imaginary parts, leading to two equations,

$$1 - \frac{\pi^2}{2!} + \frac{\pi^4}{4!} - \frac{\pi^6}{6!} + \dots = -1 \text{ and } \pi - \frac{\pi^3}{3!} + \frac{\pi^5}{5!} - \frac{\pi^7}{7!} + \dots = 0.$$

Adding up ten terms of each "infinite sum", we obtain approximately -1.000000004 and -5.3×10^{-10} (check the calculation here!) which suggests that if we **keep adding new terms** generated by the same rule, **in the limit** we will indeed get -1 and 0, respectively. This is mysterious: π is transcendental, so no **finite** sum like this can be a rational number. A proper way to prove that π satisfies these equations is to express $\cos(x)$ and $\sin(x)$, functions given by ratios of sides in a right-angled triangle where x is in radians, as

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots, \qquad \sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

These infinite expansions, apparently known already to Madhava (ca. 1400), are **enormously important** in mathematics and have many compelling applications in the sciences. To prove these formulas is the same as to prove the formula $e^{ix} = \cos(x) + i\sin(x)$ (used in MFA without proof). Complete and rigorous theory, leading to these expansions, is part of what is covered in the Real Analysis course. In the **first part** of MATH11112, we will:

- formally define **infinite series** and make sense of "sums of infinitely many numbers";
- learn about ways to tell whether a given series **converges**, i.e., has a sum;
- understand **power series** which consist of power of x with some coefficients, and see why they define "smooth" continuous functions of x if they converge.

Expansion of functions as power series is intimately connected with **differentiation,** a formal treatment of which begins the **second part** of the course. Higher derivatives of a function are key to approximating the function by polynomials, called **Taylor polynomials.** It is this theory that allows us to write a "good" function as sum of a **Taylor series.**

The course concludes with the **third part** devoted to rigorous treatment of **integration.** A key result is the **Fundamental Theorem of Calculus,** which demonstrates that integration is truly a reverse operation to differentiation.

Further study of series: a power series is a sum of infinitely many functions of the form ax^n. In 1807, Joseph Fourier publicised a class of scientific problems which require calculating infinite sums of more sophisticated functions, such as $a\sin(nx)$. The theory of series that we develop in Real Analysis serves as a foundation for the study of Fourier series and other advanced series in mathematics, science and engineering.

End of the informal introduction.

The Inverse Function Theorem

We begin with a result which could have been proved in Mathematical Foundations and Analysis (MFA), given more time. The Inverse Function Theorem will be used to define a function $x \mapsto x^r$ where r is rational. First, a definition.

Definition.

Let $A \subseteq \mathbb{R}$. A real-valued function f on A is

increasing,	if	$\forall x, x' \in A,\ x < x' \implies f(x) \le f(x')$;
strictly increasing,	if	$\forall x, x' \in A,\ x < x' \implies f(x) < f(x')$;
decreasing,	if	$\forall x, x' \in A,\ x < x' \implies f(x) \ge f(x')$;
strictly decreasing,	if	$\forall x, x' \in A,\ x < x' \implies f(x) > f(x')$.

A function satisfying one of the above conditions is called **(strictly) monotone.**

The above applies to sequences which are functions on $A = \mathbb{N}$.

We can use monotone sequences to calculate limits of functions. The following is an MFA-style result:

Lemma 1.1: limit of f from the left via strictly increasing sequences.

For a function f, defined on (a, b), the following are equivalent:

(i) $\lim_{x \to b-} f(x) = \ell$.

(ii) For all strictly increasing sequences (x_n) such that $x_n \to b$ as $n \to \infty$, one has $\lim_{n\to\infty} f(x_n) = \ell$.

Remark: the Lemma can be expressed in words as follows:

> The limit of f at b from the left is the common limit of all sequences $(f(x_n))_{n\ge1}$, where a sequence $(x_n)_{n\ge1}$ is strictly increasing and converges to b.

Proof of the Lemma (not given in class). (i) $\Rightarrow$ (ii): let $\varepsilon > 0$ be arbitrary. First, we use the definition of $\lim_{x\to b-} f(x) = \ell$ to generate $\delta > 0$ such that $|f(x) - \ell| < \varepsilon$ for all $x \in (b - \delta, b)$.

Now we let $(x_n)_{n\ge1}$ be a strictly increasing sequence, and use the above $\delta > 0$ in the definition of "$x_n \to b$ as $n \to \infty$" to generate $N \in \mathbb{N}$ such that $n \ge N$ implies $|x_n - b| < \delta$. That is, $x_N, x_{N+1}, \ldots \in (b - \delta, b + \delta)$. Since the sequence is strictly increasing with limit b, no term can exceed b, so in fact $x_N, x_{N+1}, \ldots \in (b - \delta, b)$.

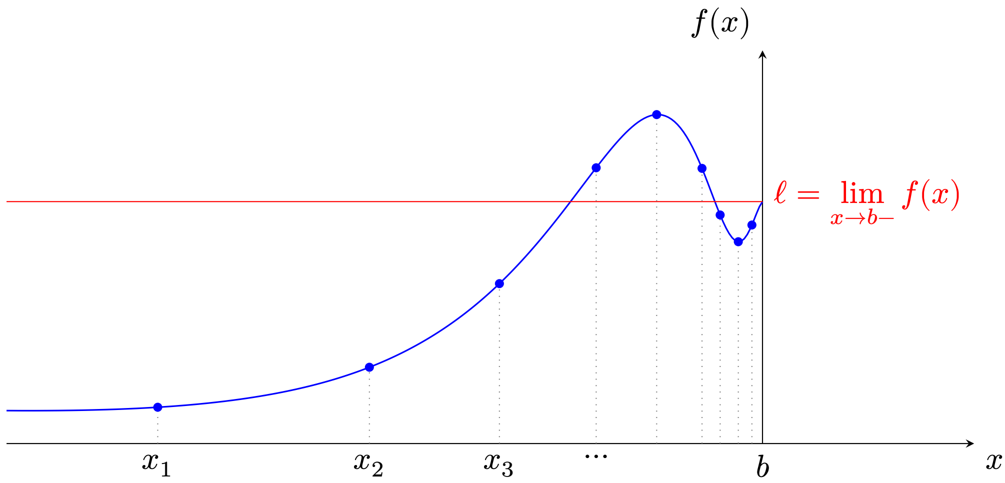

Figure 1.1: *Lemma 1.1 says that the limit of f at b from the left is the same as the common limit of all sequences $f(x_1), f(x_2), \ldots$ where $x_1, x_2, \ldots$ strictly increase and converge to b*

But then, by the choice of δ, $|f(x_n) - \ell| < \varepsilon$ for all $n \geq N$. We have shown that ℓ satisfies the definition of limit for the sequence $(f(x_n))_{n\geq 1}$, and so (ii) is proved.

(ii) $\Rightarrow$ (i): to prove the contrapositive of this implication, we assume that the statement "$\lim_{x\to b-} f(x) = \ell$" is false. This means that there exists some $\varepsilon_0 > 0$ such that for all $\delta > 0$, the interval $(b-\delta, b)$ contains a point, say $x(\delta)$, with $|f(x(\delta)) - \ell| \geq \varepsilon_0$.

Choose $\delta_1 = 1$ and construct $x(\delta_1) \in (b - \delta_1, b)$. We have $|f(x(\delta_1)) - \ell| \geq \varepsilon_0$.

Then, for each $n \geq 2$, choose $\delta_n = \min(\frac{1}{n}, b - x(\delta_{n-1}))$ and construct $x(\delta_n) \in (b - \delta_n, b)$.

Since $b - \frac{1}{n} < x(\delta_n) < b$, by the Sandwich Rule $x(\delta_n) \to b$ as $n \to \infty$. Also, since $x(\delta_n) > b - (b - x(\delta_{n-1})) = x(\delta_{n-1})$, the sequence $x(\delta_1), x(\delta_2), \ldots$ is strictly increasing.

We still have, by construction, that $|f(x(\delta_n)) - \ell| \geq \varepsilon_0$ for all n. Therefore, ℓ fails to satisfy the definition of the limit of the sequence $(f(x(\delta_n)))_{n\geq 1}$, i.e. (ii) is false. $\square$

The Lemma is illustrated by Figure 1.1. The next result mirrors the Lemma to deal with a limit from the right:

Corollary: limit of f from the right via strictly decreasing sequences.

For a function f, defined on (a, b), the following are equivalent:

(i) $\lim_{x\to a+} f(x) = \ell$.

(ii) For all strictly decreasing sequences (x_n) such that $x_n \to a$ as $n \to \infty$, one has $\lim_{n\to\infty} f(x_n) = \ell$. □

In other words,

> the limit of f at a from the right is the common limit of all sequences $(f(x_n))_{n\geq 1}$, where a sequence $(x_n)_{n\geq 1}$ is strictly decreasing and converges to a.

We are now ready to prove the first theorem of the course.

Theorem 1.2: the Inverse Function Theorem for strictly increasing functions.

A strictly increasing continuous function $f\colon [a, b] \to [f(a), f(b)]$ has an inverse $g\colon [f(a), f(b)] \to [a, b]$ which is strictly increasing and continuous.

Proof. $f\colon [a, b] \to [f(a), f(b)]$ is **surjective,** because for every $d \in [f(a), f(b)]$ the Intermediate Value Theorem (and its corollary in MFA) gives c in $[a, b]$ such that $f(c) = d$.

A strictly increasing f is **injective:** indeed, if $x_1 \neq x_2$, then either $x_1 < x_2$ and so $f(x_1) < f(x_2)$, or $x_1 > x_2$ and so $f(x_1) > f(x_2)$. In either case $f(x_1) \neq f(x_2)$.

We have shown that f is **bijective**, hence it **has an inverse** $g = f^{-1}\colon [f(a), f(b)] \to [a, b]$.

We prove that g **is strictly increasing** by contradiction. Assume not, then there exist y_1, y_2 such that $y_1 < y_2$ and $g(y_1) \geq g(y_2)$. Since f is increasing, $f(g(y_1)) \geq f(g(y_2))$. Since $f = g^{-1}$, this reads $y_1 \geq y_2$, but at the same time $y_1 < y_2$, a contradiction.

We prove that g **is continuous** at an arbitrary point d of its domain by verifying the **criterion of continuity**, seen in MFA:

$$\lim_{y\to d} g(y) = g(d).$$

By results from MFA, this is equivalent to

$$\lim_{y\to d-} g(y) = \lim_{y\to d+} g(y) = g(d).$$

We first show that $\lim_{y\to d-} g(y) = g(d)$. We would like to use Lemma 1.1 for this, so we let (y_n) be a strictly increasing sequence in $[f(a), f(b)]$ which converges to d. Since g is a strictly increasing function, $(g(y_n))$ is a strictly increasing sequence; it is also bounded (lies in $[a, b]$), hence by a result from MFA, $(g(y_n))$ has a limit, say c, in $[a, b]$.

Then by Lemma 1.1, $\lim_{n\to\infty} f(g(y_n)) = \lim_{x\to c-} f(x)$ which is $f(c)$ as f is continuous. Since $f = g^{-1}$, this says that $\lim_{n\to\infty} y_n = f(c)$. Thus, $d = f(c)$, hence $g(d) = c$.

We have proved that the common limit of all sequences $(g(y_n))$, where y_n strictly increases and converges to d, is $g(d)$. By Lemma 1.1, this means that $\lim_{y\to d-} g(y) = g(d)$.

The proof that $\lim_{y\to d+} g(y) = g(d)$ is completely similar, based on the Corollary to Lemma 1.1, and we omit it. Continuity of g at d is proved. □

Example: the p^{th} root function.

Let $p \in \mathbb{N}$. Show that the p^{th} power function $[0, +\infty) \to [0, +\infty)$, $x \mapsto x^p$, has a continuous inverse (denoted $y \mapsto \sqrt[p]{y}$ and called **the pth root function**).

Solution: define $f\colon [0, +\infty) \to [0, +\infty)$ by $f(x) = x^p$. Then f is strictly increasing on $[0, +\infty)$. Apply Inverse Function Theorem 1.2 to the restriction $[0, b] \xrightarrow{f} [0, b^p]$ to get a continuous inverse $\sqrt[p]{}\colon [0, b^p] \to [0, b]$. Since $b > 0$ can be made arbitrarily large, this defines the continuous function $\sqrt[p]{}$ on all of $[0, +\infty)$.

We compose continuous, strictly increasing functions to define a rational power function:

Example: raising to rational power $\frac{p}{q}$ where $p \in \mathbb{Z}$, $q \in \mathbb{N}$.

Define $x^{\frac{p}{q}} = (\sqrt[q]{x})^p$. This is a continuous function of x where $x \in (0, +\infty)$.

Remark: one can deduce from the definition of a rational power that

$$x > 1,\ r, s \in \mathbb{Q},\ r < s \quad \implies \quad x^r < x^s.$$

This allows us to formally define arbitrary real powers of x:

> **Definition: x^α where $x > 0$ and $\alpha \in \mathbb{R}$.**
>
> If $x \geq 1$, define $x^\alpha = \sup\{x^r : r \in \mathbb{Q}, r \leq \alpha\}$.
> If $0 < x < 1$, define $x^\alpha = (1/x)^{-\alpha}$.

A disadvantage of this definition is that proving the expected properties of powers such as $x^\alpha x^\beta = x^{\alpha+\beta}$ requires work. We will soon obtain a more useful expression for powers via the exponential function.

Infinite series: definition

> **Definition: infinite series, convergent series, sum.**
>
> For real numbers a_n, an **infinite series** is an expression of the form $\sum\limits_{n=1}^{\infty} a_n$ (also written as $a_1 + a_2 + \cdots + a_n + \ldots$, $\sum_{n\geq 1} a_n$ or just $\sum a_n$).
> The n**th partial sum** of this series is the finite sum of terms up to and including a_n: $s_n = a_1 + \cdots + a_n = \sum\limits_{i=1}^{n} a_i$.
> If the sequence of partial sums converges: $\lim_{n\to\infty} s_n = s$, we say that the series $\sum\limits_{n=1}^{\infty} a_n$ is **convergent with sum** s, and write $\sum\limits_{n=1}^{\infty} a_n = s$.

Remarks on the definition: (i) $\sum_{n=1}^{\infty} a_n = s$ is a actually a shorthand which means "the series $\sum_{n=1}^{\infty} a_n$ is convergent with sum s".

(ii) Any series that is not convergent is said to be a **divergent** series.

(iii) A series can start from $n = N$ (any integer) instead of $n = 1$: $a_N + a_{N+1} + \cdots = \sum\limits_{n=N}^{\infty} a_n$. For example, it is common to start from $n = 0$. The nth partial sum will still be the sum which ends with a_n: e.g., for the series $a_0 + a_1 + a_2 + a_3 + \ldots$,

- a_0 is the 0^{th} partial sum,
- $a_0 + a_1$ is the 1^{st} partial sum, $a_0 + a_1 + a_2$ is the 2^{nd} partial sum, and so on.

Basic examples of convergent/divergent series are discussed in week 1 supervision classes.

Alert: a strict definition of convergence.

The definition of convergence of the series $\sum_{n=1}^{\infty} a_n$, used in Real Analysis, is very strict: if the number sequence a_1, $a_1 + a_2$, $a_1 + a_2 + a_3, \ldots$ does not have a limit, then the series has **no sum.**

Weaker definitions can assign a "sum" to some particular types of series which we consider divergent: Cesàro sum, Abel sum etc. They are used in specialist applications which are beyond this course.

The geometric series

The next example is simple yet important: we will see that more complicated series can be studied by comparing them to a geometric series. We revisit a result seen in MFA.

Proposition 1.3: convergence and sum of geometric series.

Let $a, r \in \mathbb{R}$. The **geometric series with initial term** a **and ratio** r**,**

$$a + ar + ar^2 + ar^3 + \cdots = \sum_{n=0}^{\infty} ar^n,$$

is convergent if $|r| < 1$, with sum $\dfrac{a}{1-r}$.

Proof. The nth partial sum of the series is $s_n = a(1 + r + r^2 + \cdots + r^n)$. The calculation

$$\begin{aligned}
&(1 + r + r^2 + \cdots + r^n)(1 - r) \\
&= 1 - r + r - r^2 + r^2 - r^3 + \cdots + r^n - r^{n+1} \\
&= 1 - r^{n+1}
\end{aligned}$$

where the intermediate terms cancel, gives us the formula

$$s_n = a\frac{1 - r^{n+1}}{1 - r}.$$

If $|r| < 1$, we recall from MFA that r^{n+1} tends to 0 as $n \to \infty$, so by Algebra of Limits of convergent sequences,

$$s = \lim_{n\to\infty} s_n = a\frac{1-0}{1-r} = \frac{a}{1-r}.$$

The sum of the series, is, by definition, the limit of partial sums if it exists. Hence the sum of the geometric series is $a/(1-r)$ as claimed. □

Convergence of series with non-negative terms

Unlike the geometric series, usually there is **no nice formula** for the n^{th} partial sum s_n. We still want to decide if a series is convergent, so we prove theorems known as "convergence tests". Our first few tests work for series where all terms are non-negative.

Theorem 1.4: boundedness test for non-negative series.

Let a series $a_1 + a_2 + \ldots$ have $a_n \geq 0$ for all n. The following are equivalent:

(i) the partial sums $s_1, s_2, \ldots$ are bounded above;

(ii) the series is convergent.

If (i) and (ii) hold, the sum of the non-negative series is the least upper bound, $\sup\{s_n : n \geq 1\}$, of its partial sums.

Proof. The partial sums of a non-negative series form an increasing sequence, because $s_{n+1} = s_n + a_{n+1} \geq s_n$ for all n. We know from MFA that an increasing sequence $(s_n)_{n\geq 1}$ of real numbers has a limit iff it is bounded above, and then the limit is the supremum of the terms of the sequence. □

Corollary: only two convergence types for non-negative series.

A non-negative series $a_1 + a_2 + \ldots$ is either

- convergent with a non-negative finite sum: $\sum\limits_{n=1}^{\infty} a_n = s$, $0 \leq s < +\infty$, or
- divergent if $\sup\{s_n : n \geq 1\} = +\infty$.

In the latter case we use the symbolic notation "$\sum\limits_{n=1}^{\infty} a_n = +\infty$" and say that the non-negative series **diverges to** $+\infty$.

Alert.

Notation $\sum\limits_{n=1}^{\infty} a_n = +\infty$ (to mean that the series is divergent) and $\sum\limits_{n=1}^{\infty} a_n < +\infty$ (to mean that the series is convergent) is used **only** for series with non-negative terms!

The next test is used very often.

Theorem 1.5: the comparison test for non-negative series.

Assume that $0 \le a_n \le b_n$ for all n. If the series $b_1 + b_2 + \ldots$ is convergent (with sum T), then the series $a_1 + a_2 + \ldots$ is convergent (with sum at most T).
If $a_1 + a_2 + \ldots$ is divergent, then $b_1 + b_2 + \ldots$ is also divergent.

Proof. Write $s_n = a_1 + \cdots + a_n$ and $t_n = b_1 + \cdots + b_n$. As $a_1 \le b_1$, $a_2 \le b_2$ etc, we have $s_n \le t_n$, where $t_n \le T$ by Theorem 1.4. Hence $s_1, s_2, \ldots$ have an upper bound T, so by boundedness test, Theorem 1.4, the series $a_1 + a_2 + \ldots$ is convergent.

The sum $\sum_{n=1}^{\infty} a_n$ is the least upper bound of $(s_n)_{n\ge 1}$, and T is an upper bound. Hence $\sum_{n=1}^{\infty} a_n \le T$.

Now "$a_1 + a_2 + \ldots$ is divergent $\Rightarrow b_1 + b_2 + \ldots$ is divergent" follows by contrapositive. □

To use the Comparison Test, we need to compare with some easy series $\sum b_n$, yet the Test does not tell us how to find it. Hence we develop further convergence tests.

Theorem 1.6: the Ratio Test for positive series.

For a **positive** series $\sum_{n\ge 1} a_n$, suppose that the limit $\ell = \lim\limits_{n\to\infty} \frac{a_{n+1}}{a_n}$ exists. Then if $0 \le \ell < 1$, the series is convergent, and if $\ell > 1$, the series is divergent.

Proof. **The case** $0 \le \ell < 1$**.** Choose a positive ε such that $\ell + \varepsilon < 1$. For example, $\varepsilon = (1-\ell)/2$ works.

Since $\ell = \lim_{n\to\infty} \frac{a_{n+1}}{a_n}$, there exists N such that $\frac{a_{n+1}}{a_n} < \ell + \varepsilon$ for all $n \geq N$. Write this as $a_{n+1} < ra_n$, where $r = \ell + \varepsilon$. Then $a_{N+2} < ra_{N+1} < r^2 a_N$, and, repeating this, we obtain $a_{N+k} < r^k a_N$. We have

$$\begin{aligned} a_1 + a_2 + \dots + a_{N+k} &\leq a_1 + \dots + a_{N-1} + a_N + ra_N + \dots + r^k a_N \\ &\leq (a_1 + \dots + a_{N-1}) + \frac{a_N}{1-r}. \end{aligned}$$

The upper bound that we have obtained is a finite constant which does not depend on k. Thus, partial sums of the series $a_1 + a_2 + \dots$ are bounded, so by Theorem 1.4, the series is convergent.

The case $\ell > 1$. Put $\varepsilon = \ell - 1$. There is N such that $\ell - \varepsilon < \frac{a_{n+1}}{a_n}$ for $n \geq N$. But $\ell - \varepsilon = 1$, so $1 < \frac{a_{n+1}}{a_n}$, equivalently $a_n < a_{n+1}$, for $n \geq N$. In particular, all a_n for $n > N$ are greater than the positive constant a_N. Hence $s_n \geq (n - N)a_N$ which is unbounded. □

Alert: the Ratio test may be inconclusive.

If $\ell = 1$ or the limit does not exist, this test does not tell us anything: the harmonic series $\sum_{n\geq 1} \frac{1}{n}$ is divergent (see the next Chapter), and the series of inverse squares $\sum_{n\geq 1} \frac{1}{n^2}$ is convergent (see the first exercise sheet). Both have $\ell = 1$.

In the following test (*not taught in lectures, not examinable*), we use the nth root function $\sqrt[n]{}$, defined earlier.

Theorem 1.7: The nth Root Test for non-negative series.

For a non-negative series $\sum_{n\geq 1} a_n$, suppose that the limit $\ell = \lim_{n\to\infty} \sqrt[n]{a_n}$ exists. Then if $0 \leq \ell < 1$, the series is convergent, and if $\ell > 1$, the series is divergent.

Remark: Again, if $\ell = 1$ or the limit does not exist, this test does not tell us anything.

Proof. (*not given in class: very similar to the proof of the Ratio Test; not examinable.*)
The case $0 \leq \ell < 1$. Choose a positive ε so that $r = \ell + \varepsilon$ is still less than 1; for example, $\varepsilon = (1 - \ell)/2$ works.

By definition of limit, there is N such that $\sqrt[n]{a_n} < \ell + \varepsilon = r$ for all $n \geq N$. Then $a_n < r^n$ for $n \geq N$, so partial sums of the series are bounded by $a_1 + \cdots + a_N + \frac{1}{1-r}$, implying convergence.

The case $\ell > 1$. Put $\varepsilon = \ell - 1$. Since ℓ is the limit of $\sqrt[n]{a_n}$, there is N such that $\ell - \varepsilon < \sqrt[n]{a_n}$ for $n \geq N$. But $\ell - \varepsilon = 1$, so $1 < \sqrt[n]{a_n}$, equivalently $1 < a_n$, for $n \geq N$. We therefore have $s_n \geq n - N$ which is unbounded. □

Week 2

The Harmonic Series. Rearrangements. Series with positive and negative terms

Version 2025/02/04 To accessible online version of this chapter

We begin the chapter with a series which is an example to many results in this course. In particular, it shows that a positive series with $\lim_{n\to\infty}\frac{a_{n+1}}{a_n}=1$ may be divergent:

Proposition 2.1: the Harmonic Series is divergent.

The following series, called **the harmonic series,** is divergent:

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots=\sum_{n=1}^{\infty}\frac{1}{n}=+\infty.$$

Proof. We have

$$s_{2n}-s_n=\frac{1}{n+1}+\frac{1}{n+2}+\cdots+\frac{1}{2n}\geq n\times\frac{1}{2n}=\frac{1}{2}.$$

Hence $s_2\geq s_1+\frac{1}{2}$, $s_4\geq s_2+\frac{1}{2}\geq s_1+\frac{2}{2}$, and, continuing, we get $s_{2^n}\geq s_1+\frac{n}{2}$. This shows that the sequence $(s_n)_{n\geq 1}$ is unbounded. □

Comment: we can see that the partial sums s_n of the harmonic series diverge to $+\infty$ but "slowly". How large must n be so that $s_n\geq 10$? Do you need to add up more than a thousand terms of the harmonic series to get a sum of at least 10? Open the spreadsheet which tabulates partial sums of the harmonic series and scroll down to find out!

Rearrangements of a series with non-negative terms

When we add up **finitely many** numbers, the answer does not depend on the order of summands. Yet for **infinite series** this is more intricate: putting terms in a different order gives a very different sequence of partial sums. Let us formally define a rearrangement.

Definition: rearrangement.

A series $\sum_{n=1}^{\infty} b_n$ is called a **rearrangement** of a series $\sum_{n=1}^{\infty} a_n$ if there exists a bijective function $\sigma\colon \mathbb{N} \to \mathbb{N}$ such that $b_n = a_{\sigma(n)}$ for all n.

We now prove that rearranging a **non-negative** series does not change the sum.

Theorem 2.2: the rearrangement theorem for non-negative series.

Suppose that all terms in a series $a_1 + a_2 + \dots$ are non-negative. Then all rearrangements of this series have the same sum, and if the series $a_1 + a_2 + \dots$ is divergent, then all rearrangements are divergent.

Proof. Let $\sigma\colon \mathbb{N} \to \mathbb{N}$ be a bijection so that $a_{\sigma(1)} + a_{\sigma(2)} + \dots$ is a rearrangement of the original series. We let

$$s_n = a_1 + a_2 + \dots + a_n, \quad t_n = a_{\sigma(1)} + a_{\sigma(2)} + \dots + a_{\sigma(n)}$$

be the nth partial sum of the series, respectively, rearranged series. Denote by $M(n)$ the largest among the indices $\sigma(1), \dots, \sigma(n)$. Then $a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(n)}$ is a sublist of the list $a_1, \dots, a_{M(n)}$ of non-negative real numbers, and so

$$t_n \le s_{M(n)}.$$

If $\sum_{n=1}^{\infty} a_n$ is convergent with sum $S < +\infty$, then $s_{M(n)} \leq S$, and so $t_n \leq S$, for all n. By the boundedness test, Theorem 1.4, the rearranged series is **convergent** with

$$\sum_{n=1}^{\infty} a_{\sigma(n)} \leq \sum_{n=1}^{\infty} a_n,$$

which proves that **rearranging a non-negative series cannot increase its sum.**

But the series $\sum_{n=1}^{\infty} a_n$ can also be viewed as a rearrangement of $\sum_{n=1}^{\infty} a_{\sigma(n)}$, via the bijective function σ^{-1}. Since rearranging cannot increase the sum, we must conclude that

$$\sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} a_{\sigma(n)}.$$

From the two inequalities we conclude that the series and its rearrangement have the same sum. Finally, our observation that $\sum_{n=1}^{\infty} a_n$ is a rearrangement of $\sum_{n=1}^{\infty} a_{\sigma(n)}$ proves, by contrapositive, the implication "$\sum_{n=1}^{\infty} a_n$ is divergent $\Rightarrow \sum_{n=1}^{\infty} a_{\sigma(n)}$ is divergent". □

Summation of double series

For theoretical and practical reasons, we would like to be able to calculate a sum of all numbers in a **double series,** which is defined as an array **infinite in two directions,**

$$\begin{matrix} a_{00} & a_{01} & a_{02} & a_{03} & \cdots \\ a_{10} & a_{11} & a_{12} & a_{13} & \cdots \\ a_{20} & a_{21} & a_{22} & a_{23} & \cdots \\ a_{30} & a_{31} & a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

where $a_{m,n} \in \mathbb{R}$ for all $m, n \geq 0$. Double series allow several methods of summation. First, we can **enumerate** all terms of a double series by non-negative integers, which gives a (single) series. (Two such enumerations will be shown later in Figure 2.2.) Secondly, let

- RowSum_m = the sum of the infinite series $a_{m0} + a_{m1} + a_{m2} + \ldots$ if it exists;
- $\mathrm{ColumnSum}_n$ = the sum of the infinite series $a_{0n} + a_{1n} + a_{2n} + \ldots$ if it exists;

- $\mathrm{DiagSum}_d =$ the finite sum $a_{d0} + a_{d-1,1} + \cdots + a_{0d}$, equivalently $\sum\limits_{m+n=d} a_{m,n}$;
- $\mathrm{SquareSum}_{b\times b} = \sum\{a_{m,n} : 0 \le m \le b,\, 0 \le n \le b\}$.

Figure 2.1 illustrates how some of these sums are calculated. The next result considers

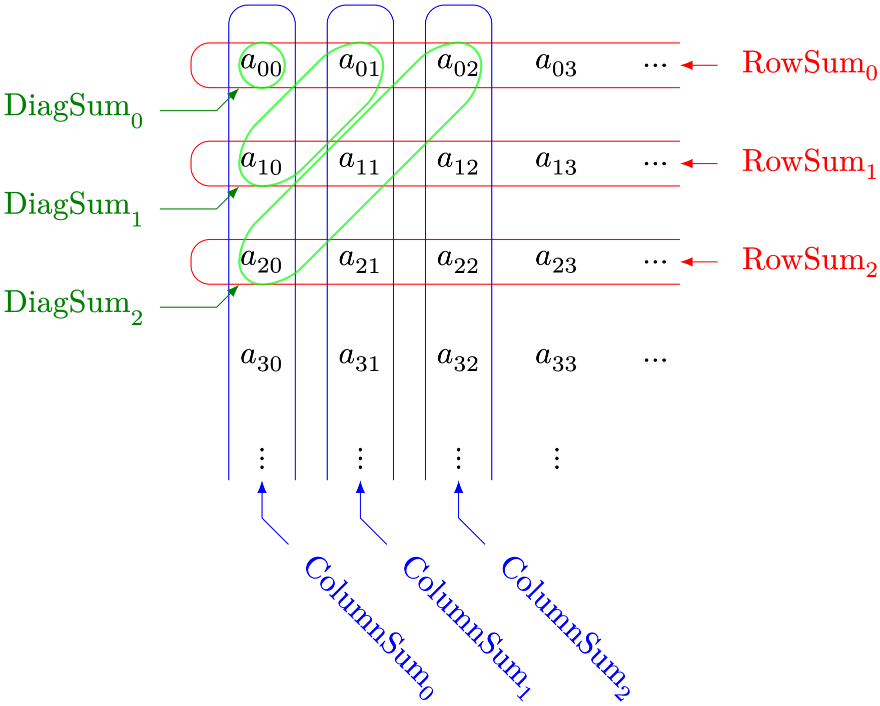

Figure 2.1: *Row sums, column sums and diagonal sums in a double series*

summing a double series by enumeration, by squares, by diagonals, by rows and by columns. If the terms are all non-negative, the methods will return the same answer:

Proposition 2.3: summation of double series with non-negative terms.

Suppose that a double series $(a_{m,n})$ has **non-negative** terms, and a way to enumerate all these terms gives a single series with sum S. Then:

1. all ways to enumerate the terms $a_{m,n}$ will result in series with sum S;
2. $\lim\limits_{b\to\infty} \mathrm{SquareSum}_{b\times b} = S$ and $\sum\limits_{d=0}^{\infty} \mathrm{DiagSum}_d = S$;
3. $\sum\limits_{m=0}^{\infty} \mathrm{RowSum}_m = S$ and $\sum\limits_{n=0}^{\infty} \mathrm{ColumnSum}_n = S$.

Proof. 1. All the different ways of arranging the terms $a_{m,n}$ in a single series are rearrangements of one another, hence must have the same sum, S, by Theorem 2.2.

2. To show that $\sum_{d=0}^{\infty} \operatorname{DiagSum}_d = S$ and $\lim_{b\to\infty} \operatorname{SquareSum}_{b\times b} = S$, we consider two special ways to enumerate the $a_{m,n}$, shown in Figure 2.2.

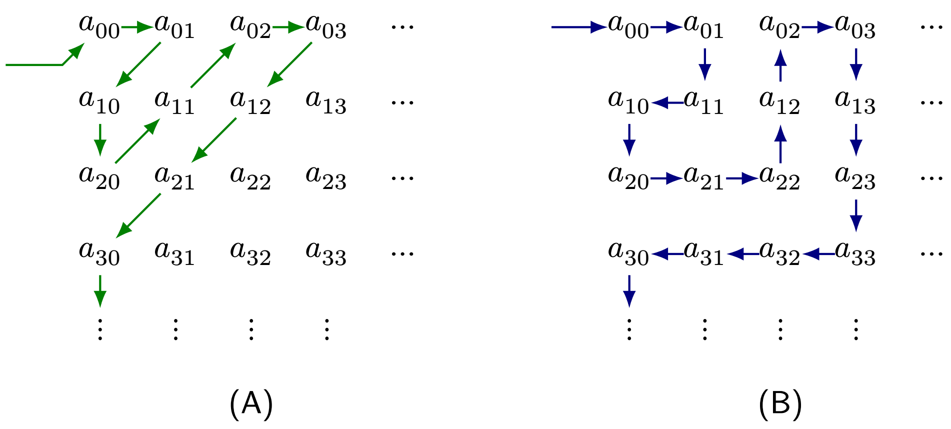

Figure 2.2: *Two examples of enumerating terms of a double series*

In Figure 2.2(A), partial sums of the resulting single series (which, by part 1., converge to S) contain a subsequence of sums of the form $\operatorname{DiagSum}_0 + \operatorname{DiagSum}_1 + \cdots + \operatorname{DiagSum}_d$. All subsequences of a convergent sequence have the same limit, so this subsequence must also converge to S.

Figure 2.2(B) similarly shows that $\lim_{b\to\infty} \operatorname{SquareSum}_{b\times b}$ must be S.

3. The top row of the $b \times b$ square sum is $a_{00} + a_{01} + \cdots + a_{0b}$. This is a partial sum which is less than or equal to the infinite sum RowSum_0. In the same way, the remaining rows of $\operatorname{SquareSum}_{b\times b}$ are bounded above by $\operatorname{RowSum}_1, \ldots, \operatorname{RowSum}_b$. Therefore,

$$\operatorname{SquareSum}_{b\times b} \leq \sum_{m=0}^{b} \operatorname{RowSum}_m.$$

Taking the limit as $b \to \infty$, we have

$$S = \lim_{b\to\infty} \operatorname{SquareSum}_{b\times b} \leq \sum_{m=0}^{\infty} \operatorname{RowSum}_m.$$

It remains to show that the opposite inequality, $\sum_{m=0}^{\infty} \mathrm{RowSum}_m \leq S$, also holds. Since the infinite sum $\sum_{m=0}^{\infty} \mathrm{RowSum}_m$ is the limit of partial sums, it is enough to show that for every fixed M, the sum $\sum_{m=0}^{M} \mathrm{RowSum}_m$ is at most S.

By definition of the row sums, we have

$$\begin{aligned} \mathrm{RowSum}_0 &= \lim_{n\to\infty} a_{00} + a_{01} + \cdots + a_{0n}, \\ \mathrm{RowSum}_1 &= \lim_{n\to\infty} a_{10} + a_{11} + \cdots + a_{1n}, \\ &\vdots \\ \mathrm{RowSum}_M &= \lim_{n\to\infty} a_{M,0} + a_{M,1} + \cdots + a_{M,n}. \end{aligned}$$

Adding together these limits, we have

$$\sum_{m=0}^{M} \mathrm{RowSum}_m = \lim_{n\to\infty} (\text{sum over the } M \times n \text{ rectangle}).$$

Yet any rectangle is contained in a $b \times b$ square for large enough b, and by part 2., sums over all squares are $\leq S$. Hence $\sum_{m=0}^{M} \mathrm{RowSum}_m$ is a limit of a sequence bounded above by S, which means that $\sum_{m=0}^{M} \mathrm{RowSum}_m \leq S$.

We have proved that $S \leq \sum_{m=0}^{\infty} \mathrm{RowSum}_m \leq S$, so $\sum_{m=0}^{\infty} \mathrm{RowSum}_m = S$. Finally, the argument for the column sums, $\sum_{n=0}^{\infty} \mathrm{ColumnSum}_n = S$, is the same as for row sums, and we omit it. The Proposition is proved. □

Alert: changing the order of summation.

In analysis, passing from sum by rows to sum by columns is known as **changing the order of summation.** We have proved that

$$\sum_{m=0}^{\infty}\Big(\sum_{n=0}^{\infty} a_{m,n}\Big) = \sum_{n=0}^{\infty}\Big(\sum_{m=0}^{\infty} a_{m,n}\Big),$$

assuming that **all** $a_{m,n}$ **are non-negative.** The equality **does not hold** in general, see example below. Finding conditions (such as non-negativity) which allow changing the order of infinite summation is an important problem in Analysis.

Example: if we drop the assumption that all $a_{m,n}$ are non-negative, the "sum of the double series" may depend on the method of summation. Consider the double series

$$\begin{array}{cccccc} 1 & -1 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 0 & 1 & -1 & \dots \\ 0 & 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \qquad a_{m,n} = \begin{cases} 1, & \text{if } m = n, \\ -1, & \text{if } m = n-1, \\ 0, & \text{otherwise.} \end{cases}$$

We have $\mathrm{RowSum}_m = 0$ for all m. Yet $\mathrm{ColumnSum}_0 = 1$ and $\mathrm{ColumnSum}_n = 0$ for all $n \geq 1$, and so

$$\sum_{m=0}^{\infty}\Big(\sum_{n=0}^{\infty} a_{m,n}\Big) = 0, \qquad \sum_{n=0}^{\infty}\Big(\sum_{m=0}^{\infty} a_{m,n}\Big) = 1.$$

Series with terms of different signs. The Nullity Test

We will now develop several convergence tests for infinite series without the assumption that all terms are non-negative. Our first test can only show divergence:

Theorem 2.4: the nullity test for divergence.

If $a_n \not\to 0$ as $n \to \infty$, then the series $\sum\limits_{n=1}^{\infty} a_n$ is divergent.

Proof. Write s_n for the nth partial sum. Assume that the series is convergent, i.e., $\lim\limits_{n\to\infty} s_n = s$ for some real number s. Then $\lim\limits_{n\to\infty} s_{n-1} = s$ as well. By AoL of convergent sequences, $\lim_{n\to\infty}(s_n - s_{n-1}) = s - s = 0$. But $s_n - s_{n-1} = a_n$ and so $\lim_{n\to\infty} a_n = 0$. We proved:

$$\text{the series } \sum_{n=1}^{\infty} a_n \text{ is convergent} \quad \Rightarrow \quad \lim_{n\to\infty} a_n = 0,$$

which is the contrapositive of, hence is equivalent to, the statement of the Theorem. □

Example: application of the nullity test.

Show that the series (a) $\sum_{n\geq 1} \frac{n}{n+1}$ and (b) $\sum_{n\geq 0}(-1)^n$ are divergent.

Solution: (a) $\lim_{n\to\infty} \frac{n}{n+1} = \lim_{n\to\infty} \frac{1}{1+1/n} = 1 \neq 0$ so by the Nullity Test, the series is divergent. (b) $\lim_{n\to\infty}(-1)^n$ does not exist and in particular is not 0, so $\sum_{n\geq 0}(-1)^n$ is also a divergent series.

Remark: if $a_n \to 0$ as $n \to \infty$, the Nullity Test is **inconclusive:** the series $\sum\limits_{n=1}^{\infty} a_n$ may still be divergent. The Harmonic Series is an example.

Algebra of infinite sums

The next result allows us to construct new convergent series out of existing examples.

Proposition 2.5: Algebra of Infinite Sums (AoIS).

If $\sum_{n=1}^{\infty} a_n$ is convergent with sum S and $\sum_{n=1}^{\infty} b_n$ is convergent with sum T, then for any real numbers λ, μ the series $\sum_{n=1}^{\infty}(\lambda a_n + \mu b_n)$ is convergent with sum $\lambda S + \mu T$.

Proof. The nth partial sum of the series $\sum_{n=1}^{\infty}(\lambda a_n+\mu b_n)$ is $(\lambda a_1+\mu b_1)+\cdots+(\lambda a_n+\mu b_n)$. This is a **finite** sum, so we can rearrange to get $\lambda s_n + \mu t_n$, where $s_n = a_1 + \cdots + a_n$ and $t_n = b_1 + \cdots + b_n$ (partial sums). We are given that $s_n \to S$ and $t_n \to T$ as $n \to \infty$, so by AoL of convergent sequences, $\lambda s_n + \mu t_n \to \lambda S + \mu T$ as claimed. □

Alert: no $\times$ or $/$ for series.

Unlike the Algebra of Limits of convergent **sequences**, AoIS does not allow multiplication or division of **series.**

Absolute convergence

The following definition comes from "absolute value", the old name for the modulus $|a|$.

Definition: absolutely convergent.

The series $\sum_{n=1}^{\infty} a_n$ is **absolutely convergent** if $\sum_{n=1}^{\infty} |a_n| < +\infty$.

The next very strong test can establish convergence of many series.

Theorem 2.6: Absolute Convergence Theorem; the Infinite Triangle Inequality.

Suppose the series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent. Then:

1. there are non-negative p_n, q_n such that $\sum_{n=1}^{\infty} p_n < +\infty$, $\sum_{n=1}^{\infty} q_n < +\infty$ and $a_n = p_n - q_n$ for all n;
2. $\sum_{n=1}^{\infty} a_n$ is convergent;
3. $\left|\sum_{n=1}^{\infty} a_n\right| \le \sum_{n=1}^{\infty} |a_n|$ **(infinite triangle inequality)**.

Proof. For a real number a, denote

$$a^+ = \begin{cases} a, & \text{if } a \ge 0, \\ 0, & \text{if } a < 0, \end{cases} \qquad a^- = \begin{cases} 0, & \text{if } a \ge 0, \\ -a, & \text{if } a < 0. \end{cases}$$

1. Assume that the series $\sum a_n$ is absolutely convergent, meaning that $\sum_{n=1}^{\infty} |a_n|$ has finite sum $M < +\infty$. Put $p_n = a_n^+$ and $q_n = a_n^-$ for all n. Then we have

$$0 \le p_n, q_n \le |a_n|, \quad a_n = p_n - q_n, \quad |a_n| = p_n + q_n.$$

By Comparison Test, Theorem 1.5, the series $\sum_{n=1}^{\infty} p_n$ is convergent, with some finite sum $P \le M$, and likewise $\sum_{n=1}^{\infty} q_n = Q \le M$.

2. By AoIS, the series $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} (p_n - q_n)$ is **convergent** with sum $P - Q$.

3. Since $0 \le P, Q \le M$, we have $|P - Q| \le M$, which reads $\left|\sum_{n=1}^{\infty} a_n\right| \le \sum_{n=1}^{\infty} |a_n|$. $\square$

Remark. How to test the series $\sum_{n=1}^{\infty} |a_n|$ for convergence? For example, we can use the Comparison Test, the Ratio Test or the Root Test.

The next result shows that rearrangements of absolutely convergent series are as well-behaved as rearrangements of non-negative series.

Claim 2.7: rearrangements of absolutely convergent series and double series.

(a) All rearrangements of absolutely convergent $\sum_n a_n$ converge with the same sum.
(b) If $\sum_{m,n} |a_{m,n}| < +\infty$, summing the double series $(a_{m,n})$ by rows, by columns, by diagonals and by squares gives the same answer. In particular, changing the order of summation is allowed: $\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{m,n} = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{m,n}$.

Explanation: (a) If $\sum a_n$ is absolutely convergent, Theorem 2.6 gives us non-negative p_n and q_n such that $a_n = p_n - q_n$ for all n and $\sum p_n = P < +\infty$, $\sum q_n = Q < +\infty$. Then $a_{\sigma(n)} = p_{\sigma(n)} - q_{\sigma(n)}$ and so by AoIS, $\sum a_{\sigma(n)} = (\sum p_{\sigma(n)}) - (\sum q_{\sigma(n)})$. Yet by Rearrangement Theorem 2.2 for non-negative series, $\sum p_{\sigma(n)} = P$ and $\sum q_{\sigma(n)} = Q$, and so $\sum a_{\sigma(n)} = P - Q$, regardless of the rearrangement.

(b) Similarly to (a), write $a_{m,n} = p_{m,n} - q_{m,n}$ with non-negative $p_{m,n}$ and $q_{m,n}$ and $\sum\limits_{m,n} p_{m,n} = P < +\infty$, $\sum\limits_{m,n} q_{m,n} = Q < +\infty$. Then by Proposition 2.3 and AoIS, all methods of summation of the double series $a_{m,n}$ will return the answer $P - Q$.

The Alternating Series Test

If a series with positive and negative terms is not absolutely convergent, it might still satisfy the assumptions of the next test which shows convergence.

Theorem 2.8: The Alternating Series Test.

Let $a_1, a_2, \dots \geq 0$ be a **decreasing** sequence which **tends to zero.** Then the series $a_1 - a_2 + a_3 - a_4 + \dots = \sum_{n=1}^{\infty} (-1)^{n+1} a_n$ is convergent.

Proof. Similarly to the Absolute Convergence Theorem, the idea is to write the given series as a linear combination of two convergent series; yet in this case, we cannot use two series with non-negative terms. Consider two series:

Series 1:	$(a_1 - a_2)$	$+0$	$+(a_3 - a_4)$	$+0$	$+(a_5 - a_6)$	$+0$	$+\dots$
Series 2:	a_2	$-a_2$	$+a_4$	$-a_4$	$+a_6$	$-a_6$	$+\dots$

To obtain a partial sum of Series 1, we start with a_1, subtract a_2, add a_3, subtract a_4, add a_5 etc. By assumption, $a_2, a_3, \dots$ decrease, so each time we subtract more than we add; hence all partial sums are bounded above by a_1. Series 1 has non-negative terms, hence is convergent by Boundedness Test, Theorem 1.4.

Series 2 has partial sums a_2, 0, a_4, 0, a_6, 0 and so on. The nth partial sum is between 0 and a_n, and $a_n \to 0$ as $n \to \infty$. By Sandwich Rule the partial sums have limit 0, hence Series 2 is convergent.

The required series $a_1 - a_2 + a_3 - a_4 + \dots$ is obtained by adding Series 2 to Series 1, hence is convergent by Algebra of Infinite Sums, Proposition 2.5. □

Example: Alternating Harmonic Series.

Show that $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ is a convergent series.

Solution. $(\frac{1}{n})_{n\geq 1}$ is a decreasing sequence which has limit 0. So by the Alternating Series Test, Theorem 2.8, the series is convergent.

Remark: we are not ready to calculate the sum of the Alternating Harmonic Series just yet. Methods from this course will allow us to prove that its sum is $\ln(2)$.

Conditional convergence. Rearrangements

We proved that absolute convergence implies convergence. Is the converse true? That is, if a series converges, does it have to be absolutely convergent? Here is a counterexample:

Example: a convergent series which is not absolutely convergent.

Show that the (convergent) alternating harmonic series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$ is not absolutely convergent.

Solution. The series is convergent by the Alternating Series Test. But the series

$$1 + \left|-\tfrac{1}{2}\right| + \left|\tfrac{1}{3}\right| + \left|-\tfrac{1}{4}\right| + \cdots = 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \ldots,$$

made up of absolute values, is the Harmonic Series which, as we proved, is divergent.

Series with this property have a special name:

Definition: conditionally convergent.

A series is **conditionally convergent** if it is convergent but not absolutely convergent.

Rearrangements of conditionally convergent series do not have the same sum:

Theorem 2.9: Riemann's rearrangement theorem.

Suppose $\sum_{n=1}^{\infty} a_n$ is a conditionally convergent series. Then for any real number α there is a rearrangement of $\sum_{n=1}^{\infty} a_n$ which converges with sum α. There is also a rearrangement which diverges. □

We do not go through the proof of this striking result in class, but it may later be added as an appendix to this week's notes (not examinable).

Week 3

Power series

Version 2025/02/12 To accessible online version of this chapter

Notice (Feb 2025): Week 3 lectures begin with the Alternating Series Test, Theorem 2.8.

We now consider series where the nth term depends on a variable x and we ask for which x does the series converge. We study the simplest case of a series with x: power series.

Definition: power series.

Let $c_0, c_1, c_2, \ldots$ be real numbers. A series of the form $\sum_{n=0}^{\infty} c_n x^n$, which we also write as $C(x)$, is called a **power series** in the variable x.

A power series can be convergent for some values of x and divergent for others:

Example: geometric series as a power series.

Let $G(x)$ denote the power series $1 + x + x^2 + x^3 + \ldots$ Then $G(x)$ is convergent when $|x| < 1$ and divergent when $|x| \geq 1$. Indeed, for any given value of x, $G(x)$ becomes a geometric series with ratio x, and we apply the known results.

The set of x where $G(x)$ is convergent is the interval $(-1, 1)$. We will show that the picture for other power series is broadly similar, although their **interval of convergence** can be closed, half open or open, or be the whole real line $\mathbb{R}$. We begin with a lemma.

Lemma 3.1: absolute convergence for smaller modulus.

If, for a given $x_0 \in \mathbb{R}$, the power series $C(x_0)$ is convergent, then for all y with $|y| < |x_0|$, $C(y)$ is absolutely convergent.

Proof. Assume $C(x_0) = \sum_{n=0}^{\infty} c_n x_0^n$ is a convergent series. Then by the Nullity Test, the sequence $(c_n x_0^n)_{n\geq 0}$ has limit 0. Sequences that have a limit are bounded, so there exists $M \geq 0$ such that $|c_n x_0^n| \leq M$ for all n.

Given y: $|y| < |x_0|$, set $r = \frac{|y|}{|x_0|}$. As $|r| < 1$, the geometric series $\sum_n Mr^n$ converges, and

$$0 \leq |c_n y^n| = |c_n x_0^n| r^n \leq M r^n \text{ for all } n.$$

Hence by the Comparison Test, $\sum_{n=0}^{\infty} |c_n y^n|$ is convergent. This means that the series $C(y) = \sum_{n=0}^{\infty} c_n y^n$ is absolutely convergent. □

Corollary 3.2: the set of points where the power series is convergent.

The set of $t \in \mathbb{R}$ such that the series $C(t)$ is convergent is one of the following:

i. interval $(-R, R)$, $[-R, R]$, $(-R, R]$ or $[-R, R)$ where $R > 0$;

ii. $\{0\}$;

iii. $\mathbb{R}$.

We formally put $R = 0$ in ii. and $R = \infty$ in iii., so that there is a unique value $R \in [0, \infty]$ such that $C(t)$ is absolutely convergent when $|t| < R$ and divergent when $|t| > R$.

Proof. These are all the possible subsets I of $\mathbb{R}$ which contain 0 and have the property: *if* $x_0 \in I$ *then* I *contains the interval* $(-|x_0|, |x_0|)$, required by Lemma 3.1. See the illustration in Figure 3.1. □

Definition: interval of convergence, radius of convergence.

The set I given in Corollary 3.2 is called the **interval of convergence** of the power series $C(x)$, and $R \in [0, \infty]$ is the **radius of convergence.**

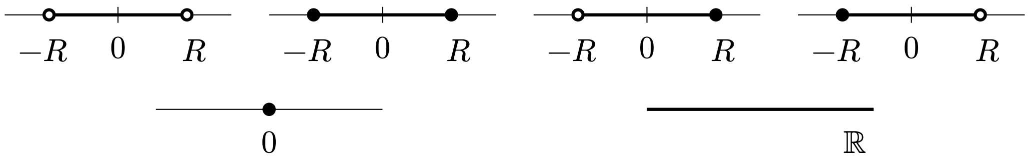

Figure 3.1: The possible intervals I of convergence of a power series

Remark: suppose the radius of convergence R of $C(x)$ is strictly between 0 and ∞. At $x = R$, the series $C(x)$ may

- be absolutely convergent; or
- be conditionally convergent; or
- be divergent.

At $x = -R$ one of the three options will also hold. We note that the series is absolutely convergent at $x = R$ iff it is absolutely convergent at $x = -R$, because $\sum_{n\geq 0} |a_n R^n|$ and $\sum_{n\geq 0} |a_n (-R)^n|$ are the same series.

We have already obtained the following

Result: interval of convergence of $G(x) = \sum_{n=0}^{\infty} x^n$.

$G(x) = 1 + x + x^2 + \ldots$ has radius of convergence 1, interval of convergence $(-1, 1)$.

We will use a practical method to find the interval of convergence of a power series:

Method: finding the interval of convergence.

1. Apply the Ratio Test to the **non-negative** series $\sum_{n=0}^{\infty} |a_n||x|^n$.
2. The answer will show absolute convergence of $\sum_{n=0}^{\infty} a_n x^n$ when $|x| < R$, where R is the radius of convergence.
3. Use further tests to check convergence of $\sum_{n=0}^{\infty} a_n x^n$ at $x = -R$ and $x = R$.

Example 1: find the interval of convergence of the series $\sum_{n=1}^{\infty} \frac{x^n}{n} = x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots$ and determine the type of convergence at the endpoints of the interval.

Solution. We apply the Ratio Test to the series $\sum_{n=1}^{\infty} \frac{|x^n|}{n}$:

$$\ell = \lim_{n\to\infty} \frac{|x^{n+1}|/(n+1)}{|x^n|/n} = \lim_{n\to\infty} |x|\frac{n}{n+1} = |x| \lim_{n\to\infty} \frac{1}{1+\frac{1}{n}} = |x|\frac{1}{1+0} = |x|.$$

If $\ell < 1$, that is $|x| < 1$, the series $\sum_{n=1}^{\infty} \frac{x^n}{n}$ is absolutely convergent. If $|x| > 1$, the series is not absolutely convergent. We conclude that the radius of convergence is $R = 1$.

What happens when $x = \pm 1$?

At $x = -1$ the series $\sum_{n=1}^{\infty} \frac{x^n}{n}$ becomes $-1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \ldots$ This is -1 times the alternating harmonic series, **convergent (conditionally)** by the Alternating Series Test.

At $x = 1$ the series is the harmonic series $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$, **divergent**. We have:

> **Result: the interval of convergence of $\sum_{n=1}^{\infty} \frac{x^n}{n}$.**
>
> The interval of convergence is $[-1, 1)$. At $x = -1$ convergence is conditional.

Example 2: find the interval of convergence of the series $\sum_{n=1}^{\infty} nx^n = x + 2x^2 + 3x^3 + \ldots$ and determine the type of convergence at the endpoints of the interval.

Solution. We apply the Ratio Test to the series $\sum_{n=1}^{\infty} nx^n$:

$$\ell = \lim_{n\to\infty} \frac{(n+1)|x^{n+1}|}{n|x^n|} = \lim_{n\to\infty} |x|\frac{n+1}{n} = |x|.$$

If $\ell < 1$, that is $|x| < 1$, the series $\sum_{n=1}^{\infty} \frac{x^n}{n}$ is absolutely convergent. If $|x| > 1$, the series is not absolutely convergent. We conclude that the radius of convergence is $R = 1$.

At $x = \pm 1$ the series is $\sum_{n=1}^{\infty} n(\pm 1)^n$, divergent by Nullity Test.

> **Result: the interval of convergence of $\sum_{n=1}^{\infty} nx^n$.**
>
> The interval of convergence is $(-1, 1)$.

Remark. A limitation of power series that it can only define a function on an interval centred at zero. To overcome this, one can consider, for $a \in \mathbb{R}$, **power series at a:**

$$\sum_{n=0}^{\infty} c_n (x-a)^n.$$

Such a series defines a function on an interval I such that $(a-R, a+R) \subseteq I \subseteq [a-R, a+R]$ where R is the radius of convergence.

The function defined by a power series is continuous

Every power series $C(x)$ can be viewed as a real-valued function on I, its interval of convergence. We claim that, at least on the open interval $(-R, R)$, this function is continuous. As $C(x)$ is an infinite sum of terms of the form $c_n x^n$ which are continuous functions of x, it makes sense to study the continuity of a sum of a series of functions. We begin with the following

> **Proposition 3.3: infinite sum of increasing continuous functions.**
>
> Suppose each of the functions $f_1(x), f_2(x), \ldots$ is **increasing** and continuous on $[c, d]$, and for each $x \in [c, d]$ the series $\sum_{m=1}^{\infty} f_m(x)$ is convergent with sum $F(x)$. Then $F(x)$ is a continuous function on $[c, d]$.

Proof. By criterion of continuity, we need to prove, for each $b \in (c, d)$, that $\lim_{x\to b-} F(x) = \lim_{x\to b+} F(x) = F(b)$. We will only prove that $\lim_{x\to b-} F(x)$ is $F(b)$; the limit $\lim_{x\to b+}$ is similar and is left to the student. By Lemma 1.1, we need to show that

$$\lim_{n\to\infty} F(x_n) = F(b), \tag{*}$$

given any strictly increasing sequence $x_1, x_2, \ldots$ with limit b. To prove (*), we sum the double series $a_{m,n} = f_m(x_n) - f_m(x_{n-1})$,

$$\begin{array}{cccc}
f_1(x_2) - f_1(x_1) & f_1(x_3) - f_1(x_2) & f_1(x_4) - f_1(x_3) & \ldots \\
f_2(x_2) - f_2(x_1) & f_2(x_3) - f_2(x_2) & f_2(x_4) - f_2(x_3) & \ldots \\
f_3(x_2) - f_3(x_1) & f_3(x_3) - f_3(x_2) & f_3(x_4) - f_3(x_3) & \ldots \\
\vdots & \vdots & \vdots & \ddots
\end{array}$$

by two methods. The first method is **summation by columns,** where we calculate

$$\text{ColumnSum}_n = \sum_{m=1}^{\infty} f_m(x_n) - f_m(x_{n-1}) = F(x_n) - F(x_{n-1}),$$

so that $\sum_{n=2}^{\infty} \text{ColumnSum}_n = \sum_{n=2}^{\infty} F(x_n) - F(x_{n-1})$. A partial sum $(F(x_2) - F(x_1)) + (F(x_3) - F(x_2)) + \cdots + (F(x_n) - F(x_{n-1}))$ of this series is **telescoping:** the intermediate terms cancel, leaving $F(x_n) - F(x_1)$. The sum of the series is the limit of partial sums, so

$$\sum_{n=2}^{\infty} \text{ColumnSum}_n = \lim_{n\to\infty} F(x_n) - F(x_1).$$

The second method is **summation by rows:** using telescoping sums again, we calculate

$$\text{RowSum}_m = \sum_{n=2}^{\infty} f_m(x_n) - f_m(x_{n-1}) = \lim_{n\to\infty} f_m(x_n) - f_m(x_1).$$

Here f_m is continuous, so $x_n \to b$ as $n \to \infty$ implies $\lim_{n\to\infty} f_m(x_n) = f_m(b)$. Therefore

$$\sum_{m=1}^{\infty} \text{RowSum}_m = \sum_{m=1}^{\infty} f_m(b) - f_m(x_1) = F(b) - F(x_1).$$

Note that all the $a_{m,n}$ are **non-negative:** $x_n > x_{n-1}$, f_m is increasing, so $f_m(x_n) \geq f_m(x_{n-1})$. Hence by Proposition 2.3, the sum of all the $a_{m,n}$ does not depend on the method of summation, which means that $\lim_{n\to\infty} F(x_n) - F(x_1) = F(b) - F(x_1)$. This proves (*) and the Proposition. □

If we consider the function $f_m(x) = c_m x^m$ on, say, an interval $[0, d]$, it can be increasing or decreasing, depending on the sign of c_m. Recall that a function which is increasing on $[c, d]$ or is decreasing on $[c, d]$ is called a **monotone** function. With this in mind, we give a modified version of the previous result.

> **Proposition 3.4: absolutely convergent sum of monotone continuous functions.**
>
> Suppose each of the functions $f_1(x), f_2(x), \ldots$ is **monotone** and continuous on $[c, d]$, and for each $x \in [c, d]$ the series $\sum_{m=1}^{\infty} f_m(x)$ is **absolutely convergent** and has sum $F(x)$. Then $F(x)$ is a continuous function on $[c, d]$.

Proof (not given in class). We literally repeat the proof of Proposition 3.3, **not** including the last paragraph which begins with the words "Note that all the $a_{m,n}$ are **non-negative.**": $a_{m,n} = f_m(x_n) - f_m(x_{n-1})$ may be negative, so we cannot use Proposition 2.3 to say that the sum of the double series does not depend on the method of summation.

Instead, we will use Claim 2.7. For that, we need to check that $\sum_{m,n} |a_{m,n}| < +\infty$. Note that in the mth row, all numbers $a_{m,2}, a_{m,3}, \ldots$ are either all non-negative if f_m is an increasing function, or all non-positive if f_m is decreasing. That is,

$$\sum_{n=2}^{\infty} |a_{m,n}| = |f_m(x_2) - f_m(x_1)| + |f_m(x_3) - f_m(x_2)| + \ldots$$
$$= \begin{cases} (f_m(x_2) - f_m(x_1)) + (f_m(x_3) - f_m(x_2)) + \ldots, \\ \quad \text{if } f_m \text{ is increasing,} \\ -(f_m(x_2) - f_m(x_1)) - (f_m(x_3) - f_m(x_2)) - \ldots, \\ \quad \text{if } f_m \text{ is decreasing,} \end{cases}$$

which is a telescopic series with sum $|f_m(b) - f_m(x_1)|$. By the triangle inequality, $|f_m(b) - f_m(x_1)|$ is bounded above by $|f_m(b)| + |f_m(x_1)|$.

Therefore, $\sum_{m,n} |a_{m,n}| \le \sum_{m=1}^{\infty} |f_m(b)| + \sum_{m=1}^{\infty} |f_m(x_1)|$. This is finite by the assumption about absolute convergence. So Claim 2.7 guarantees that summation of the $a_{m,n}$ by colums gives the same answer as summation by rows, leading to the same conclusion as in Proposition 3.3. □

Theorem 3.5: a function defined by a power series is continuous.

Let $C(x)$ denote the sum of a power series with radius of convergence R. Then $C(x)$ is a continuous function on the interval $(-R, R)$.

Proof. Let $C(x) = c_0 + c_1 x + c_2 x^2 + \ldots$ Each term $c_m x^m$ is a continuous and monotone function on $[0, R)$, the convergence for $|x| < R$ is absolute, so by Proposition 3.4, $C(x)$ is continuous on $[0, R)$.

On $(-R, 0]$, $C(x)$ is continuous for exactly the same reason. We have to split the interval $(-R, R)$ into $(-R, 0]$ and $[0, R)$ because if m is even, x^m is not monotone on the whole $(-R, R)$ but is monotone on $(-R, 0]$ and on $[0, R)$.

It is easy to check (using one-sided limits at 0) that a function, continuous on $(-R, 0]$ and on $[0, R)$, is continuous on $(-R, R)$. □

Week 4

e^x, $\ln$, differentiation

Version 2025/02/23 To accessible online version of this chapter

Remark: we will use **the Binomial Theorem** which says that for $x, y \in \mathbb{R}$ and $n \geq 0$, $(x+y)^n$ expands as $\binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n}y^n$ where $\binom{n}{i} = \frac{n!}{(n-i)!\,i!}$. The Binomial Theorem is taught in Probability I, and the standard proof is by induction.

The exponential function

Theorem 3.5 allows us to define new continuous functions by power series (with non-zero radius of convergence). Here is the most important example.

Definition: the exponential function and the number e.

$$\exp(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \qquad e = \exp(1).$$

Theorem 4.1: continuity and the law of the exponential.

$\exp$ is a continuous function on $\mathbb{R}$. One has $\exp(x)\exp(y) = \exp(x+y)$ for all x, y.

Proof. Apply the Ratio Test to the series $\sum_{n=0}^{\infty} \frac{|x|^n}{n!}$ to find $\ell = \lim_{n\to\infty} \frac{x^{n+1}/(n+1)!}{|x|^n/n!} = \lim_{n\to\infty} |x|/(n+1) = 0$ for all $|x|$. Since $0 < 1$, the power series $\exp(x)$ is absolutely convergent for all $x \in \mathbb{R}$ (the radius of convergence is $R = \infty$). By Theorem 3.5, it follows that $\exp(x)$ is a continuous function on all of $\mathbb{R}$.

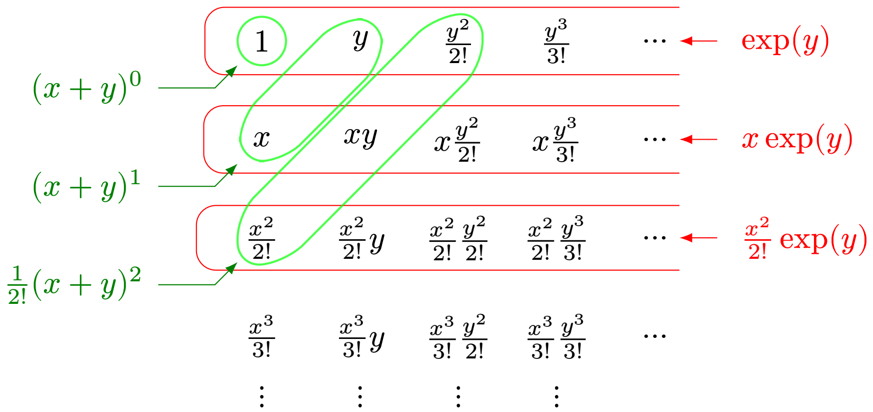

Figure 4.1: *The double series used to prove* $\exp(x)\exp(y) = \exp(x+y)$

To prove $\exp(x)\exp(y) = \exp(x+y)$, we compare two methods of summation of the double series $a_{m,n} = \dfrac{x^m}{m!}\dfrac{y^n}{n!}$, see Figure 4.1. We have

$$\mathrm{RowSum}_m = \frac{x^m}{m!}\Big(1 + y + \frac{y^2}{2!} + \dots\Big) = \frac{x^m}{m!}\exp(y),$$

hence **summation by rows** gives

$$\sum_{m=0}^{\infty} \mathrm{RowSum}_m = \sum_{m=0}^{\infty} \frac{x^m}{m!}\exp(y) = \exp(x)\exp(y).$$

We now calculate the dth diagonal sum (multiplying and dividing by $d!$ for emphasis):

$$\mathrm{DiagSum}_d = \frac{1}{d!}\Big(x^d + \frac{d!}{(d-1)!\,1!}x^{d-1}y^1 + \frac{d!}{(d-2)!\,2!}x^{d-2}y^2 + \dots + y^d\Big).$$

By **the Binomial Theorem,** the expression in brackets is the expansion of $(x+y)^d$. Thus, **summation by diagonals** gives

$$\sum_{d=0}^{\infty} \mathrm{DiagSum}_d = \sum_{d=0}^{\infty} \frac{1}{d!}(x+y)^d = \exp(x+y).$$

We claim that the sum of all numbers in this double series does not depend on the method of summation, and so $\exp(x)\exp(y) = \exp(x+y)$. We need to **justify** this claim.

If both x and y are non-negative, then all the numbers $\dfrac{x^m}{m!}\dfrac{y^n}{n!}$ are non-negative. In this case, Proposition 2.3 guarantees that the sum, $\sum_{m,n} \dfrac{x^m}{m!}\dfrac{y^n}{n!}$, is independent of the method of summation, and so $\exp(x)\exp(y) = \exp(x+y)$.

Without the assumption that x, y are non-negative, we can show that the sum of all the absolute values in the table is finite:

$$\left|\frac{x^m}{m!}\frac{y^n}{n!}\right| = \frac{|x|^m}{m!}\frac{|y|^n}{n!} \quad \Rightarrow \quad \sum_{m,n}\left|\frac{x^m}{m!}\frac{y^n}{n!}\right| = \exp(|x|+|y|) < +\infty,$$

so by Claim 2.7, the sum $\sum_{m,n} \dfrac{x^m}{m!}\dfrac{y^n}{n!}$ is still independent of the method of summation, and we still have $\exp(x)\exp(y) = \exp(x+y)$. □

Discussion of the e^x notation. The law of the exponential tells us that, for all $n \in \mathbb{N}$,

$$\exp(n) = \exp(\underbrace{1+1+\cdots+1}_{n}) = \exp(1)\exp(1)\ldots\exp(1) = e^n.$$

It also follows that, for $p, q \in \mathbb{N}$, $\left(\exp(\frac{p}{q})\right)^q = \exp(q\frac{p}{q}) = \exp(p)$ which is e^p, and so by definition of the qth root and the $(p/q)^{\text{th}}$ power,

$$\exp(\tfrac{p}{q}) = \sqrt[q]{e^p} = e^{\frac{p}{q}}.$$

The law of the exponential also tells us that $\exp(-x)\exp(x) = \exp(0) = 1$, hence

$$\exp(-x) = \frac{1}{\exp(x)} \quad \Rightarrow \quad \exp(-\tfrac{p}{q}) = 1/e^{\frac{p}{q}} = e^{-\frac{p}{q}}.$$

Therefore, $\exp(x) = e^x$ for all rational numbers x. Motivated by this, we extend the notation to all real x:

> **Notation:** e^x.
>
> $\exp(x)$ is written as e^x for all $x \in \mathbb{R}$.

Definition of $\ln$, the natural logarithm function

We are going to introduce the inverse function to e^x. Let us show that e^x is bijective.

Proposition 4.2: properties of e^x.

The function $f(x) = e^x$ is a strictly increasing bijection $\mathbb{R} \to (0, +\infty)$.

Proof. Observe that $x > 0 \implies e^x = 1 + x + \dfrac{x^2}{2} + \cdots > 1 + x$. In particular, e^x is positive for positive x. Then $e^{-x} = 1/e^x$ implies that e^x is positive for all x, and is indeed a function from $\mathbb{R}$ to $(0, +\infty)$.

For all $x, y \in \mathbb{R}$ we have $e^y - e^x = e^x(e^{y-x} - 1)$. If $x < y$, then $e^{y-x} > 1$ as observed above, so $e^y > e^x$. We have shown that e^x is strictly increasing, hence injective.

To show that e^x is surjective, let $d \in (0, +\infty)$ be arbitrary. If $d > 1$, note that $e^d > 1 + d > d$ as shown above. Also $e^0 = 1 < d$. The function is continuous, so by the Intermediate Value Theorem there exists $c \in [0, d]$ such that $e^c = d$.

If $d < 1$ then $\frac{1}{d} > 1$ and by the above, $\frac{1}{d} = e^c$ for some c. We then have $d = e^{-c}$ by the law of the exponential. Finally, if $d = 1$ then $d = e^0$. We have proved that e^x is surjective, and so it is bijective. □

We immediately deduce

Theorem 4.3: natural logarithm $\ln$.

There is a strictly increasing continuous bijection $\ln\colon (0, +\infty) \to \mathbb{R}$ such that $\ln e^x = x$ for all $x \in \mathbb{R}$, $e^{\ln y} = y$ for all $y > 0$ and $\ln(yz) = \ln y + \ln z$ for all $y, z > 0$.

Sketch of proof. e^x is a bijection from $\mathbb{R}$ to $(0, +\infty)$ so it must have an inverse $(0, +\infty) \to \mathbb{R}$, which we denote $\ln$ and call the **natural logarithm** function. Inverse means that $\ln e^x = x$ and $e^{\ln y} = y$.

Using the Inverse Function Theorem 1.2, we conclude that $\ln$ is strictly increasing and continuous.

By definition of $\ln$, $x = \ln e^x$ for all x. Set $x = \ln y + \ln z$ to get $\ln y + \ln z = \ln e^{(\ln y + \ln z)}$. By the law of the exponential, this equals $\ln(e^{\ln y} e^{\ln z})$. Yet $e^{\ln y} = y$ and $e^{\ln z} = z$, so the answer simplifies to $\ln(yz)$. We proved the logarithm law, $\ln y + \ln z = \ln(yz)$. $\square$

Differentiation of functions: an informal introduction

We begin the second part of the course: the theory of **differentiation.**

To differentiate a "smooth" function f at point $a \in \mathbb{R}$ means to calculate the **derivative**, $f'(a)$, of f at a. The derivative, if it exists, shows "how fast" the function f grows (or decreases) at the point a. It is impossible to measure growth by looking just at the value of f at a. Rather, the derivative is defined via taking the **limit**; we illustrate this in Fig. 4.2.

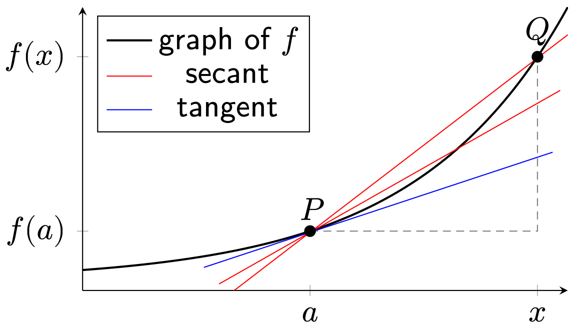

Figure 4.2: The secant passing through the points $(a, f(a))$ and $(x, f(x))$ on the graph is $m = \frac{f(x)-f(a)}{x-a}$. As $x \to a$, we expect the secant to get closer to the tangent at $(a, f(a))$.

We first present the idea **informally** (rigorous definitions are below). Fix a point $P = (a, f(a))$ on the graph of a function f. The slope, or gradient, of the secant passing through P and another point $Q = (x, f(x))$ on the graph is

$$m_{PQ} = \frac{f(x) - f(a)}{x - a}.$$

As Q "gets closer" to P, the secants "seem" to approach a fixed line, the **tangent** to the graph at P. The gradient of the tangent at P, if it exists, is the derivative of f at a:

$$m_{\text{tangent at } P} = f'(a).$$

Why differentiate functions? It turns out that derivatives appear in powerful results which allow us to approximate functions by extremely good functions — **polynomials** — and to represent some functions as sums of infinite **power series.** But first, we build up theory to

- differentiate basic functions, such as polynomials, rational functions, exponential, logarithm, trigonometric and inverse trigonometric functions;
- use rules of differentiation, to find derivatives of new functions constructed from basic functions.

Definition of the derivative of f at a

We now start our rigorous treatment of differentiation.

Definition: open neighbourhood of the point $a \in \mathbb{R}$.

An **open neighbourhood** of a is an open interval $(a - \delta, a + \delta)$ for some $\delta > 0$.

Definition: differentiable at a**, derivative at** a.

Let $A \subseteq \mathbb{R}$, and let $f\colon A \to \mathbb{R}$ be a function. **Suppose that** $a \in A$ **and** A **contains an open neighbourhood of the point** a**.** We say that f is **differentiable** at a if

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists. The value of this limit is the **derivative** of f at a, and is denoted $f'(a)$.

Remark: for f to be differentiable at a, $f'(a)$ must be a real number, not infinity.

Definition: differentiable on an open interval.

f is **differentiable on an open interval** I if it is differentiable at every point of I.

Remark: if f is defined on a closed interval $[a, b]$, we will not try to differentiate f at a or at b. Though possible via one-sided limits, we will not need this.

Notation: $\frac{d}{dx}f(x)$.

If a function $f(x)$ is differentiable on an open interval, taking the derivative of f **at each point of the interval** defines a new function. We will write $f'(x)$, or $\frac{d}{dx}f(x)$, to denote the **derivative** of $f(x)$ as a function of x.

There are functions whose derivatives can be computed **by definition,** i.e., by calculating the limit given in the definition of $f'(a)$ without using any further theorems.

Example: derivative of a constant function.

Given $c \in \mathbb{R}$, define a **constant function** on $\mathbb{R}$ by the formula $f(x) = c$ for all x. This function has derivative 0 at all points of $\mathbb{R}$.

Justification: by definition, the derivative at a is $\lim_{x\to a} \frac{c-c}{x-a} = \lim_{x\to a} 0 = 0$.

Remark: Remember that the limit, $\lim_{x\to a} g(x)$, of $g(x)$ as x tends to a, **does not require** $g(x)$ **to be defined at** a**.** Indeed, the MFA definition of limit (revisit it!) looks only at points x such that $0 < |x - a| < \delta$, and this **excludes** the case $x = a$.

For example, the expression $\frac{c-c}{x-a}$ above is **undefined** when $x = a$. But it is of no concern to us: $\frac{c-c}{x-a}$ has value 0 for all x such that $x \neq a$, and so we can write $\lim_{x\to a} \frac{c-c}{x-a} = \lim_{x\to a} 0$.

To conclude: when calculating a limit $\lim_{x\to a}$, we can always assume $x \neq a$.

Example: derivative of the function x.

$\frac{d}{dx}x = 1$ on $\mathbb{R}$.

Justification: by definition, the derivative of x at a is $\lim_{x\to a} \frac{x-a}{x-a} = \lim_{x\to a} 1 = 1$.

Theorem 4.4: differentiable implies continuous.

If f is differentiable at a, then f is continuous at a.

Proof. The criterion of continuity says that f is continuous at a iff $\lim_{x\to a} f(x) = f(a)$. Rearranging, we obtain: f is continuous at $a \iff \lim_{x\to a}(f(x) - f(a)) = 0$.

Assume f is differentiable at a, so that the limit $\lim_{x\to a} \frac{f(x)-f(a)}{x-a} = L$ exists. Then

$$\begin{aligned}\lim_{x\to a}(f(x) - f(a)) &= \lim_{x\to a} \frac{f(x) - f(a)}{x - a}(x - a) && \text{(can assume } x \neq a\text{)}\\ &= \lim_{x\to a} \frac{f(x) - f(a)}{x - a} \lim_{x\to a}(x - a) && \text{(by AoL for functions)}\\ &= L \cdot 0 = 0.\end{aligned}$$

Thus, f verifies the (rearranged) criterion of continuity above, so is continuous at a. □

Alert: continuous at $a \not\Rightarrow$ differentiable at a.

The converse to Theorem 4.4 does not hold. For example, $f(x) = |x|$ is continuous but not differentiable at 0.

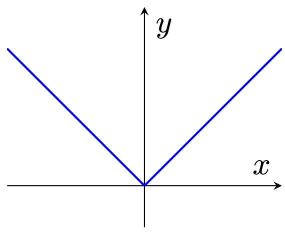

Figure 4.3: *Visibly, the graph of $f(x) = |x|$ is "not smooth" at $x = 0$.*

Justification. "Differentiable at 0" requires the limit $\lim_{x\to 0} \frac{|x|-|0|}{x-0} = \lim_{x\to 0} \frac{|x|}{x}$ to exist. Yet the function is defined by $|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0, \end{cases}$ see the graph in Fig. 4.3. Hence

$$\lim_{x\to 0+} \frac{|x|}{x} = \lim_{x\to 0+} \frac{x}{x} = 1, \quad \lim_{x\to 0-} \frac{|x|}{x} = \lim_{x\to 0-} \frac{-x}{x} = -1.$$

The one-sided limits are **not** equal, so the limit $\lim_{x\to 0}$ does not exist.

Rules of differentiation: sums and products

We can obtain new differentiable functions from known ones by addition and multiplication.

Theorem 4.5: sum and product rules of differentiation.

Suppose that the functions f, g are differentiable at a. Then

- the function $f + g$ is differentiable at a, and $(f + g)'(a) = f'(a) + g'(a)$;
- the function fg is differentiable at a, and $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$.

Proof. **The sum rule (proof not given in class)**: by definition of the function $f + g$, $\frac{(f+g)(x) - (f+g)(a)}{x-a}$ is the same as $\frac{f(x) + g(x) - (f(a) + g(a))}{x-a}$ which rearranges as $\frac{f(x) - f(a)}{x-a} + \frac{g(x) - g(a)}{x-a}$. Taking the limit as $x \to a$ and using AoL for functions, we obtain $(f + g)'(a) = f'(a) + g'(a)$ as claimed.

The product rule: by definition, $(fg)(x) = f(x)g(x)$. Start with

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

where we subtract then add $f(a)g(x)$ in the numerator. The RHS rearranges as

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

We are given that g is differentiable at a. Differentiable implies continuous, so g is continuous at a. Hence $\lim_{x \to a} g(x) = g(a)$. Taking $\lim_{x \to a}$ in the last displayed formula and using AoL, we get $f'(a)g(a) + f(a)g'(a)$, as claimed. $\square$

Now, using only $+$ and $\times$, we can construct all **polynomials in** x from constants and the function x. If we apply the rules of differentiation, we obtain

Corollary.

A polynomial in x is differentiable for all $x \in \mathbb{R}$.

Differentiating infinite sums

The sum rule of differentiation **does not** extend to infinite sums. A function defined as a sum of series of differentiable functions may not be differentiable.

Yet one can show that a function defined as a sum of a power series is differentiable on $(-R, R)$, where R is the radius of convergence. We will not go through the proof of this in class. Interested students are invited to construct a proof as an exercise, along the following lines (**not done in class and not examinable**):

Let $f(x) = \sum_{n=0}^{\infty} c_k x^k$ where the radius of convergence is $R > 0$. Let $a \in (-R, R)$. By Algebra of Infinite Sums, we have $f(x) - f(a) = F_a(x)(x-a)$ where $F_a(x) = \sum_{n=1}^{\infty} c_k(x^{k-1} + ax^{k-2} + \cdots + a^{k-2}x + a^{k-1})$. By Proposition 4.6 below, $f(x)$ will be differentiable at a if $F_a(x)$ is shown to be continuous at a.

We note that $F_a(x)$ is obtained if the double series $a_{m,n} = c_{m+n+1}a^m x^n$, $m, n \geq 0$, is summed by diagonals. Yet summation by columns gives the same answer (*this needs to be justified by demonstrating that* $\sum m, n|a_{m,n}| < +\infty$ *when* $a, x \in (-R, R)$) and returns a power series in x. By Theorem 3.5, the sum of a power series is a continuous function, so F_a is continuous on $(-R, R)$, as required.

One concludes from the above that $\left(\sum_{n=0}^{\infty} c_n x^n\right)' = \sum_{n=0}^{\infty}(c_n x^n)' = \sum_{n=1}^{\infty} nc_n x^{n-1}$. So in particular, since $\left(\frac{x^n}{n!}\right)' = \frac{nx^{n-1}}{n!} = \frac{x^{n-1}}{(n-1)!}$, differentiating the exponential series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ term-by-term gives the same series, so $(e^x)' = e^x$.

Instructions for the exam: differentiating a power series term-by-term as above without giving full justification will not be accepted in the exam. If asked to justify differentiation of e^x, give a result obtained below, Proposition 4.7.

Proving "differentiable" by constructing slope function

Rather than showing directly that $\lim_{x\to a} \frac{f(x)-f(a)}{x-a}$ exists, we may use the following:

> **Proposition 4.6: differentiability means continuity of the slope function at a.**
>
> A function $f(x)$, defined in an open neighbourhood of $a \in \mathbb{R}$, is differentiable at a, if and only if there is a function $F_a(x)$ such that $f(x) - f(a) = F_a(x)(x - a)$ for all x, and $F_a(x)$ is continuous at $x = a$. If these conditions hold, $f'(a)$ equals $F_a(a)$.

Proof. If such F_a exists and is continuous at a, we have $\lim_{x\to a} \frac{f(x)-f(a)}{x-a} = \lim_{x\to a} F_a(x)$ which, by continuity, is $F_a(a)$. That is, $f'(a)$ exists and equals $F_a(a)$.

Now suppose that f is differentiable at a. Then, defining

$$F_a(x) = \begin{cases} \frac{f(x)-f(a)}{x-a}, & x \neq a, \\ f'(a), & x = a. \end{cases}$$

guarantees $\lim_{x\to a} F_a(x) = F_a(a)$, so by criterion of continuity F_a is continuous at a. □

We call F_a the **slope function** for f at a, because $F_a(x)$ is the slope (the gradient) of the secant through the points $(a, f(a))$ and $(x, f(x))$ on the graph of f. It is useful to note the slope function for the polynomial x^n:

$$f(x) = x^n \quad \Rightarrow \quad F_a(x) = \frac{x^n - a^n}{x - a} = x^{n-1} + x^{n-2}a + \cdots + a^n.$$

This formula defines a polynomial function of x which is continuous everywhere, including at $x = a$. One has $F_a(a) = na^{n-1}$ which is the derivative of x^n at $x = a$.

Differentiating e^x

We use the method of continuous slope function to differentiate e^x.

> **Proposition 4.7: derivative of e^x.**
>
> $\frac{d}{dx} e^x = e^x$.

Proof. To differentiate e^x at 0, write

$$e^x - e^0 = x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \underset{\text{AoIS}}{=} x\sum_{k=1}^{\infty}\frac{x^{k-1}}{k!} = (x-0)F_0(x).$$

The slope function $F_0(x)$ is the sum of a power series convergent for all x, hence is continuous by Theorem 3.5, and by Proposition 4.6 $\frac{d}{dx}(e^x)|_{x=0}$ exists and equals $F_0(0) = 1$. This proves the **Special Limit for e^x:**

$$\lim_{x\to 0}\frac{e^x - 1}{x} = 1.$$

Indeed, the left-hand side is exactly the derivative of e^x at $x = 0$ which we have just found to be 1. We now differentiate e^x at an arbitrary $x \in \mathbb{R}$:

$$\frac{d}{dx}e^x = \lim_{y\to x}\frac{e^y - e^x}{y - x} = \lim_{y\to x} e^x\frac{e^{y-x} - 1}{y - x} \underset{h=y-x}{=} e^x \lim_{h\to 0}\frac{e^h - 1}{h}.$$

By the Special Limit, this is $e^x \times 1 = e^x$. □

The Chain Rule and the Quotient Rule

We will work in the situation

$$\mathbb{R} \xrightarrow{g} \mathbb{R} \xrightarrow{f} \mathbb{R}$$

We will write g as a function of $y \in \mathbb{R}$ and f a function of $x \in \mathbb{R}$.

Theorem 4.8: The Chain Rule.

If $g(y)$ is differentiable at $y = k$ and $f(x)$ is differentiable at $x = g(k)$ then $(f \circ g)(y)$ is differentiable at $y = k$, and $(f \circ g)'(k) = f'(g(k))g'(k)$.

Proof. By Proposition 4.6, whenever f is differentiable at a point ℓ, one has

$$f(x) - f(\ell) = F_\ell(x)(x - \ell)$$

for all x, where the slope function F_ℓ is continuous at ℓ. In particular, this holds for $x = g(y)$ and $\ell = g(k)$:

$$f(g(y)) - f(g(k)) = F_\ell(g(y))(g(y) - g(k)) = F_\ell(g(y))G_k(y)(y - k),$$

where we assumed that g was differentiable at k and applied Proposition 4.6 to g.

The function $F_\ell(g(y))$ is continuous at $y = k$, because $g(y)$ is continuous (even differentiable!) at k, F_ℓ is continuous at $g(k) = \ell$, and a composition of continuous functions is continuous. The function $G_k(y)$ is continuous at k. Therefore, by Algebra of Continuous Functions, $F_\ell(g(y))G_k(y)$ is a continuous function of y. It immediately follows by Proposition 4.6 that the function $f(g(y))$ is differentiable at $y = k$, with

$$F_\ell(g(k))G_k(k) = f'(g(k))g'(k)$$

as its derivative at k, as claimed. □

Example.

Find $\frac{d}{dy}e^{-\frac{y^2}{2}}$.

Solution. Put $f(x) = e^x$ and $g(y) = -\frac{1}{2}y^2$ so that our required function is $f(g(y))$. To apply the Chain Rule, we must check that the assumptions of Theorem 4.8 are met:

- $g(y) = -\frac{1}{2}y^2$ is a polynomial, hence is differentiable for all y, with $g'(y) = -y$;
- $f(x) = e^x$ is differentiable for all x by Proposition 4.7, with $f'(x) = e^x$.

Hence we are allowed to use the Chain Rule: $\frac{d}{dy}e^{-\frac{y^2}{2}} = f'(g(y))g'(y) = e^{-\frac{y^2}{2}} \cdot (-y) = -ye^{-\frac{y^2}{2}}$.

Corollary: the Quotient Rule.

If $g(a) \neq 0$ and $f(y)$, $g(y)$ are differentiable at $y = a$, then

$$\left(\frac{1}{g}\right)'(a) = -\frac{g'(a)}{g(a)^2}, \quad \left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.$$

Proof. If $h(x) = \frac{1}{x}$ then, for any $\ell \neq 0$, $h'(\ell) = \lim_{x\to\ell} \frac{\frac{1}{x}-\frac{1}{\ell}}{x-\ell} = \lim_{x\to\ell} \frac{\ell-x}{(x-\ell)x\ell}$. When calculating $\lim_{x\to\ell}$, we may assume that $x \neq \ell$, so this simplifies to $\lim_{x\to\ell} \frac{-1}{x\ell} = -\frac{1}{\ell^2}$.

Writing $\frac{1}{g(y)}$ as $h(g(y))$ and applying the Chain Rule, we have $(\frac{1}{g})'(a) = h'(g(a))g'(a) = -\frac{1}{g(a)^2}g'(a)$ as claimed. Now, to obtain $(\frac{f}{g})'$, apply the Product Rule to $f \cdot \frac{1}{g}$. □

Week 5

Differentiating $\ln$, x^b, $\sin$ and $\cos$

Version 2025/02/26 To accessible online version of this chapter

We start this chapter by differentiating $\ln(x)$, the natural logarithm function. Since we introduced $\ln$ as the inverse function of e^x, we will need the following result.

The Inverse Rule of differentiation

Theorem 5.1: The Inverse Rule.

Let $f(x)$ be strictly monotonic and continuous on $[a, b]$, and let g be the inverse of f so that g is strictly monotonic and continuous by the Inverse Function Theorem. Suppose f is **differentiable** at $\ell \in (a, b)$ and that **the derivative** $f'(\ell)$ **is not zero**. Then g is differentiable at $k = f(\ell)$, and $g'(k) = \dfrac{1}{f'(\ell)}$.

Proof. We begin in the same way as in the proof of the Chain Rule: using Proposition 4.6, write $f(x) - f(\ell) = F_\ell(x)(x - \ell)$ for all x and in particular for $x = g(y)$, so

$$f(g(y)) - f(g(k)) = F_\ell(g(y))(g(y) - g(k)).$$

Since f and g are inverse to each other, $f(g(y)) = y$, so

$$y - k = F_\ell(g(y))(g(y) - g(k)).$$

When $y \neq k$, we have $y - k \neq 0$, and the equation shows that $F_\ell(g(y)) \neq 0$ (because the left-hand side is not 0). When $y = k$, we have $F_\ell(g(y)) = F_\ell(\ell) = f'(\ell)$ which is not 0 by assumption. Hence $F_\ell(g(y))$ is **never zero,** and we can divide by it:

$$\frac{1}{F_\ell(g(y))}(y - k) = g(y) - g(k).$$

Since $\frac{1}{F_\ell(g(y))}$ is continuous at $y = k$ by Algebra of Continuous Functions and continuity of composition, by Proposition 4.6 $g(y)$ is differentiable at k with

$$\frac{1}{F_\ell(g(k))} = \frac{1}{F_\ell(\ell)} = \frac{1}{f'(\ell)}$$

as derivative. □

We immediately deduce

Corollary 5.2: inverse of a function differentiable on an interval.

If f is strictly monotone and differentiable on an open interval, its inverse function f^{-1} is differentiable everywhere it is defined **except** the points $f(\ell)$ with $f'(\ell) = 0$, and

$$\frac{d(f^{-1})}{dy}(y) = \frac{1}{\dfrac{df}{dx}(f^{-1}(y))}. \qquad \square$$

Can the derivative of a strictly increasing function be zero at some points? Yes:

Example: a point where the inverse to a monotone function is not differentiable.

Construct a continuous strictly increasing function $f(x)$ such that $f'(\ell) = 0$ for some point ℓ. Check that f^{-1} is **not** differentiable at $f(\ell)$ in your example.

Solution: for example, $f(x) = x^3$ and $\ell = 0$. We have $f'(0) = 0$. For the inverse function $\sqrt[3]{y}$, the limit $\lim\limits_{y\to 0} \dfrac{\sqrt[3]{y} - \sqrt[3]{0}}{y - 0} = \lim\limits_{y\to 0} \dfrac{1}{(\sqrt[3]{y})^2}$ is ∞. This suggests "infinite derivative" at 0,

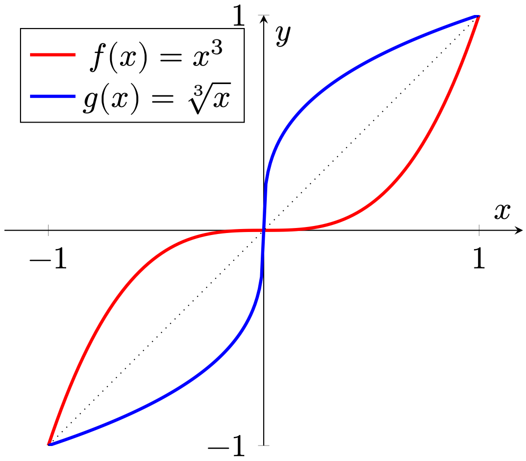

Figure 5.1: x^3 *has zero derivative at* 0*, the inverse function is not differentiable at* 0

yet formally means that $\sqrt[3]{y}$ is **not** differentiable at 0. Figure 5.1 shows horizontal tangent to the graph of x^3, and vertical tangent for $\sqrt[3]{x}$, at $(0, 0)$.

Corollary: derivative of $\ln$**.**
$\frac{d}{dy} \ln(y) = \frac{1}{y}$ for all $y \in (0, +\infty)$.

Indeed, $\ln$ is the inverse function to $\exp$, so, using the Inverse Rule and the Theorem which says that $\exp' = \exp$, we calculate

$$\ln'(y) = \frac{1}{\exp'(\ln(y))} = \frac{1}{\exp(\ln(y))} = \frac{1}{y}.$$

Functions x^b and a^x

Definition: a^b**.**

For positive real a and real b we define

$$a^b = e^{b \ln a}.$$

Remark: this satisfies the exponent laws $a^{b+c} = a^b a^c$, $(ac)^b = a^b c^b$, $a^0 = 1$ and $a^{-b} = 1/a^b$ (exercise: deduce these from the properties of the functions $\ln$ and $\exp$). In particular, a^n defined as $a \cdot a \cdot \ldots \cdot a$ (n factors) coincides with $e^{n \ln a}$.

Example: derivative of x^b.

Fix $b \in \mathbb{R}$. The function x^b is defined for positive x. Show that

$$\frac{d}{dx}(x^b) = bx^{b-1}.$$

Solution. Indeed, by Chain Rule $\frac{d}{dx}(x^b) = \frac{d}{dx}(\exp(b\ln(x)) = \exp'(b\ln(x)) \cdot b\ln'(x)$. Using the derivatives of $\exp(x)$ and of $\ln(x)$ obtained earlier, this is $\exp(b\ln(x)) \cdot bx^{-1} = x^b \cdot bx^{-1} = bx^{b-1}$ where we use the exponent laws. □

Exercise. Let $a > 0$. Use the Chain Rule to show that $\frac{d}{dx}(a^x) = a^x \ln a$.

The limit definition of e^x (*optional material*)

The section "The limit definition of e^x" was not covered in class and is not examinable.

Although in this course we define e^x as the sum $\sum_{k\geq 0} x^k/k!$, originally it was defined via the following limit.

Proposition 5.3: the limit definition for e^x (not examinable).

For all $x \in \mathbb{R}$, $e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$. In particular, $e = \lim_{n\to\infty}(1 + \frac{1}{n})^n$.

Proof. The function $\ln(1+y)$ is differentiable (by Chain Rule), and its derivative at $y = 0$ is $\frac{1}{1+0} = 1$. By Proposition 4.6 we can write $\ln(1+y) = F_0(y)y$ where the slope function F_0 is continuous at 0 with $F_0(0) = 1$. Then for a fixed $x \in \mathbb{R}$,

$$\left(1 + \frac{x}{n}\right)^n = e^{n\ln(1+\frac{x}{n})} = e^{nF_0(\frac{x}{n})\frac{x}{n}} = e^{F_0(\frac{x}{n})x}.$$

The sequence $y_n = \frac{x}{n}$ is monotone with limit 0 so Lemma 1.1 and its Corollary tell us that $\lim_{n\to\infty} e^{F_0(y_n)x}$ is $\lim_{y\to 0} e^{F_0(y)x}$. Since $e^{F_0(y)x}$ is a continuous function of y, by criterion of continuity its limit as $y \to 0$ equals its value at 0, which is $e^{F_0(0)x} = e^{1\cdot x} = e^x$. □

End of the non-examinable section.

The functions $\sin$ and $\cos$

In this course, $\sin\alpha$ and $\cos\alpha$ are defined as the ratio of sides in a right-angled triangle. Thus, if P_α is the point on the unit circle centred at the origin O such that the angle between the x axis and OP_α is α, then P_α has coordinates $(\cos\alpha, \sin\alpha)$. This defines the sine and cosine of all $\alpha \in \mathbb{R}$, see Fig. 5.2. Note that α is in **radians:** the angle equal to the full revolution (full circle) is 2π. The definition implies that for all $\alpha \in \mathbb{R}$,

$$\cos(-\alpha) = \cos\alpha, \ \sin(-\alpha) = -\sin\alpha, \ \cos\alpha = \sin(\frac{\pi}{2} - \alpha).$$

We also define $\tan\alpha = \dfrac{\sin\alpha}{\cos\alpha}$ whenever $\cos\alpha \neq 0$. When $\alpha \in (-\frac{\pi}{2}, \frac{\pi}{2})$, the straight line OP_α intersects the tangent to the circle at P_0 at the point $T_\alpha(1, \tan\alpha)$, see Fig. 5.2.

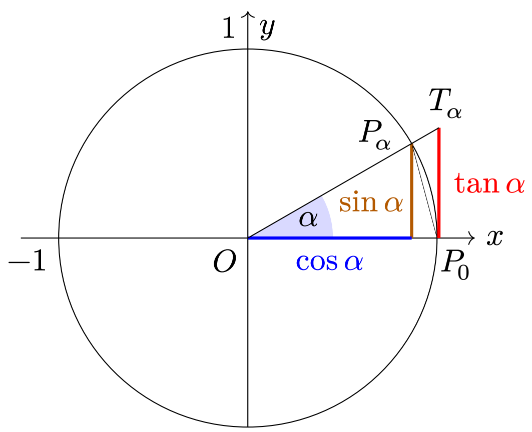

Figure 5.2: *definition of* $\sin\alpha$, $\cos\alpha$ *and* $\tan\alpha$

Lemma 5.4: the sine-angle-tangent inequality.

$0 \le \sin\alpha \le \alpha \le \tan\alpha$ for all $\alpha \in (0, \frac{\pi}{2})$.

Proof. In Fig. 5.2, area($\triangle OP_0P_\alpha$) $= \frac{1}{2} \times$ base $\times$ height $= \frac{1}{2} \times 1 \times \sin\alpha = \frac{1}{2}\sin\alpha$.

The area of the **sector** OP_0P_α with central angle α is $\frac{\alpha}{2\pi} \times$ area(circle) $= \frac{\alpha}{2\pi} \times \pi = \frac{1}{2}\alpha$.

Area($\triangle OP_0T_\alpha$) is $\frac{1}{2} \times 1 \times \tan\alpha$. We have $\triangle OP_0P_\alpha \subseteq$ sector $OP_0P_\alpha \subseteq \triangle OP_0T_\alpha$, therefore $\frac{1}{2}\sin\alpha \le \frac{1}{2}\alpha \le \frac{1}{2}\tan\alpha$. □

Remark. Graphs in Fig. 5.3 illustrate the behaviour of $\sin x$, x and $\tan x$ for small x.

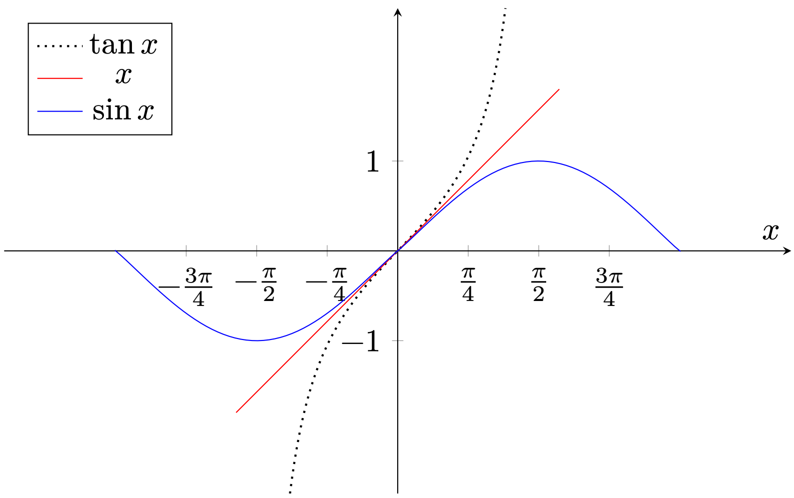

Figure 5.3: *graphs of* $\sin x$, x *and* $\tan x$ *for small* x

Corollary 5.5: limit of sine at 0.

$\lim_{x\to 0} |\sin x| = 0$.

Proof. The inequality in Lemma 5.4 can be written as $0 \le |\sin x| \le x$ when $x \in (0, \frac{\pi}{2})$, as $\sin x$ is positive for these x. Since $\lim_{x\to 0+} 0 = \lim_{x\to 0+} x = 0$, we have $\lim_{x\to 0+} |\sin x| = 0$ by Sandwich Rule.

Putting $t = -x$, we have $\lim_{x\to 0-} |\sin x| = \lim_{t\to 0+} |\sin(-t)| = \lim_{t\to 0+} |\sin t| = 0$.

The one-sided limits exist and are equal, so $\lim_{x\to 0} |\sin x|$ is their common value 0. □

We need another result about $\sin$ and $\cos$ with a geometric proof.

Lemma 5.6: sine and cosine subtraction formulas.

i. $\sin y - \sin x = 2\sin(\frac{y-x}{2})\cos(\frac{y+x}{2})$;

ii. $\cos y - \cos x = -2\sin(\frac{y-x}{2})\sin(\frac{y+x}{2})$.

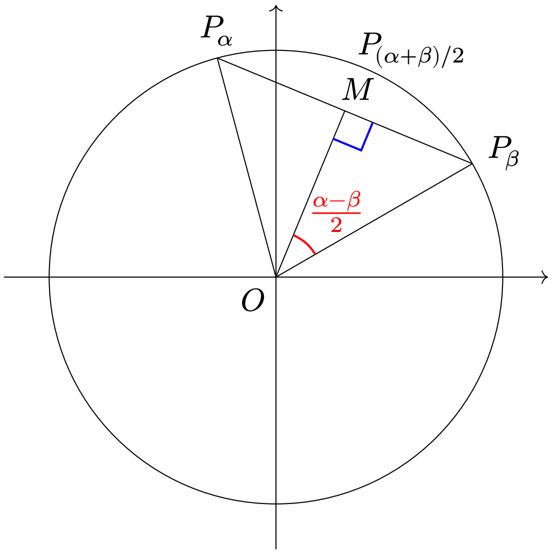

Figure 5.4: *proof of the sine addition formula for* $\sin\alpha + \sin\beta$

Proof of Lemma 5.6 — proof not examinable. If M is the midpoint of the segment $P_\alpha P_\beta$, the vector $\overrightarrow{OM}$ is $\frac{1}{2}(\overrightarrow{OP}_\alpha + \overrightarrow{OP}_\beta)$ and so the y-coordinate of M is $\frac{1}{2}(\sin\alpha + \sin\beta)$. On the other hand, see Fig. 5.4, $\overrightarrow{OM}$ is proportional to $\overrightarrow{OP}_{(\alpha+\beta)/2}$ and, from the right-angled triangle $\triangle OMP_\beta$ we can see that $\overrightarrow{OM} = \cos(\frac{\alpha-\beta}{2})\overrightarrow{OP}_{(\alpha+\beta)/2}$. This expresses the y-coordinate of M via the y-coordinate of $P_{(\alpha+\beta)/2}$:

$$\frac{1}{2}(\sin\alpha + \sin\beta) = \cos(\tfrac{\alpha-\beta}{2})\sin(\tfrac{\alpha+\beta}{2}).$$

This **sine addition formula** holds for all $\alpha, \beta \in \mathbb{R}$. Substitute $\alpha = y$, $\beta = -x$ to obtain the sine subtraction formula i. as claimed.

Substitute $\alpha = \frac{\pi}{2} - y$, $\beta = x - \frac{\pi}{2}$: the LHS becomes $\frac{1}{2}(\sin(\frac{\pi}{2} - y) - \sin(\frac{\pi}{2} - x))$ which is $\frac{1}{2}(\cos y - \cos x)$, the RHS becomes $\cos(\frac{\pi}{2} - \frac{y+x}{2})\sin(\frac{-y+x}{2}) = -\sin(\frac{y+x}{2})\sin(\frac{y-x}{2})$, equivalent to the cosine subtraction formula ii. ☐

Proposition 5.7: continuity of $\sin$ and $\cos$.

$\sin x$ and $\cos x$ are continuous functions on $\mathbb{R}$.

Proof. In the sine subtraction formula (Lemma 5.6), we bound the modulus of $\cos$ by 1:

$$|\sin y - \sin x| = 2\left|\sin \tfrac{y-x}{2}\right| \left|\cos \tfrac{y+x}{2}\right| \le 2|\sin h|,$$

where we put $h = (y-x)/2$. Opening out the modulus, we get

$$-2|\sin h| \le \sin y - \sin x \le 2|\sin h|.$$

Taking the limit as $y \to x$, equivalently $h \to 0$, we have $\lim_{h\to 0} 2|\sin h| = 0$ by Corollary 5.5. Hence by Sandwich Rule

$$\lim_{y\to x} \sin(y) - \sin(x) = 0 \quad \Leftrightarrow \quad \lim_{y\to x} \sin(y) = \sin(x),$$

so by the criterion of continuity, $\sin$ is continuous at x.

Continuity of $\cos$ is proved similarly and is left to the student. ☐

Comparing the graphs of $\sin x$ and x in Fig. 5.3, we suspect that these two functions have the same gradient at 0. Let us formalise this.

Proposition 5.8: special limit for sine.

$\lim_{x\to 0} \dfrac{\sin x}{x} = 1.$

Proof. We compute the one-sided limits. If $x \in (0, \frac{\pi}{2})$, we have

$$\cos x = \frac{\sin x}{\tan x} \leq \frac{\sin x}{x} \leq \frac{\sin x}{\sin x} = 1$$

from the sine-angle-tangent inequality in Lemma 5.4. As $x \to 0+$, $\cos x \to \cos 0 = 1$ because $\cos$ is a continuous function. Hence by Sandwich Rule $\lim_{x\to 0+} \frac{\sin x}{x} = 1$.

In $\lim_{x\to 0-} \frac{\sin x}{x}$, change the variable, $x = -y$. The limit becomes $\lim_{y\to 0+} \frac{\sin(-y)}{-y} = \lim_{y\to 0+} \frac{\sin y}{y} = 1$.

Both one-sided limits are equal to 1, so $\lim_{x\to 0} \frac{\sin x}{x}$ exists and equals 1, as claimed. □

Theorem 5.9: differentiation of $\sin$ and $\cos$.

Functions $\sin x$ and $\cos x$ are differentiable everywhere on $\mathbb{R}$, and

$$\sin' x = \cos x, \qquad \cos' x = -\sin x.$$

Proof. Use the sine subtraction formula to write $\sin y - \sin x = 2 \sin h \cos(x + h)$ with $h = (y - x)/2$. Then

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

and so

$$\sin' x = \lim_{h\to 0} \frac{\sin h}{h} \times \lim_{h\to 0} \cos(x + h) = 1 \times \cos(x + 0) = \cos x$$

by AoL, the Special Limit for sine and continuity of $\cos$.

To differentiate $\cos$, use $\cos y - \cos x = -2 \sin h \sin(x + h)$ and conclude similarly. □