數學分析 · MATH2050

Mathematical Analysis · MAT2050

學院 數據科學學院（SDS）

2026 年 1 月 18 日

Abstract

这不是完整版，最新版本可以参考我的 github 网页

大概会在寒假结束完成更新

This is the ***lecture notes*** for the ***course MAT2050***: **course Mathematical Analysis** . Nearly all content is based on the course materials of Professor Fengyi Yuan at The Chinese University of Hong Kong, Shenzhen (CUHK (SZ))[**?**].

We primarily reference materials from course MAT2050, including *lecture notes, homework (HW), and exam papers* by Professor Fengyi Yuan, as well as *tutorial slides* from the course id teaching faculty[**?**].

這是**課程 *MAT2050*：數學分析（Mathematical Analysis）**的課程講義与整理歸納。幾乎所有内容都是基於香港中文大學（深圳）Fengyi Yuan 教授的課程材料 [**?**]。

我們主要參考了 CSC3001 的材料，包括 Fengyi Yuan 教授的**講義、作業和試卷**，以及 CSC3001 教學團隊的**輔導幻燈片**。[**?**]。

目录

0.1 Sup and Inf 3
0.2 Density in R 5
0.3 Limit, Sequence & Monotone Convergence Theorem 6
0.4 Liminf and Limsup 9
0.5 Cauchy Sequence 15
0.6 Convergence of Series 17
0.7 Rearrangements of series 27
0.8 Finite and Countable sets 28
0.9 Metric Space 31
0.10 Limit Isolated Boundary Points 35
0.11 Boundedness and Convergence 36
0.12 Compactness 38
0.13 Sequentially Compactness 44
0.14 Connectedness & Subspace Topology 49
0.15 Continuity 54
0.16 Continuity in Metric Space 56
0.17 Topological characterization of continuity 58
0.18 Uniform continuity 62
0.19 Basic Definition 64
0.20 Chain rule 66
0.21 Local extrema 67
0.22 Mean value theorems 68
0.23 Vector–valued functions 69
0.24 L'Hospital's rule 72
0.25 Darboux Property 73
0.26 Taylor's theorem 74
0.27 Basic Concepts 75
0.28 Integrals against step functions 84
0.29 Differentiable integrator 86
0.30 Integration and differentiation(usual Riemann case) 87

0.31 Convergence in sequence or series of functions 90
0.32 The properties of Uni-convergence 96
0.33 Arzelà–Ascoli theorem 101
0.34 Multivariable Functions. 107
0.35 Inverse Function Theorem 111
0.36 Second Order Derivatives 120

数学分析

0.1 Sup and Inf

Definition

Let $A \subset \mathbb{R}$. M is an upper bound if $\forall a \in A$, $a \leq M$. m is a lower bound if $\forall a \in A$, $a \geq m$. A is upper bounded if such M exists, lower bounded if such m exists.

Example

$(-\infty, 0]$ is upper bounded; $(2, \infty)$ is lower bounded.

If A is lower (upper) bounded, there exists a unique m such that: 1. m is a lower (upper) bound, 2. For any other lower (upper) bound m', $m' \leq m$ (resp. $m' \geq m$).

Denote $m = \inf A$ (resp. $m = \sup A$).

【Remark】:

This follows from $\mathbb{R}$'s construction, but is treated as an axiom in this course.

Definition

Definition of Infimum

Let $A \subseteq \mathbb{R}$ be a nonempty set that is **bounded below**. A real number $\inf A$ is the **infimum** (greatest lower bound) of A **if and only if:**

1. $\inf A$ is a lower bound for A (i.e., $a \geq \inf A$ for all $a \in A$);
2. For **every** $\varepsilon > 0$, there exists $a_\varepsilon \in A$ such that $a_\varepsilon < \inf A + \varepsilon$.

Definition of Supremum

Let $A \subseteq \mathbb{R}$ be a nonempty set that is **bounded above**. A real number $\sup A$ is the **supremum** (least upper bound) of A if and only if:

1. $\sup A$ is an upper bound for A (i.e., $a \leq \sup A$ for all $a \in A$);
2. For **every** $\varepsilon > 0$, there exists $a_\varepsilon \in A$ such that $a_\varepsilon > \sup A - \varepsilon$.

【Remark】:

m^* may not belong to A, but it is arbitrarily close (continuity).

Fact and proposition

Since $\sup A$ is the smallest upper bound. Thus, there must exist $a_i \in A$ such that $\sup A - \epsilon < a_i \leq \sup A$.

If $a_i = \sup A + \epsilon$ (where $\epsilon > 0$), then $\sup A + \epsilon > \sup A$, contradicting the fact that $\sup A$ is an upper bound. Thus, such an a_i cannot exist for a valid $\sup A$

For example:

Example

Let $A = (1, 2]$. Then $\inf A = 1$ and $\sup A = 2$.

证明. ($\Rightarrow$) (1) is clear. For (2), suppose $\exists \varepsilon_0 > 0$ such that $\forall a \in A$, $a \geq m^* + \varepsilon_0$. Then $m^* + \varepsilon_0$ is a lower bound greater than m^*, contradiction.

($\Leftarrow$) For any lower bound m', if $m' > m^*$, let $\varepsilon = m' - m^*$. Then $\exists a_\varepsilon < m^* + \varepsilon = m'$, so m' is not a lower bound. Hence all lower bounds satisfy $m' \leq m^*$. □

Example

Let $A = \{x \in \mathbb{R} : x^2 < 2\}$. Then $\sup A = \sqrt{2}$.

证明. Claim: 2 is an upper bound of A. If $x \geq 2$, $x^2 \geq 4 > 2$, so $x \notin A$. By axiom, $m = \sup A$ exists. We need to prove $m^2 = 2$.

Case 1: $m^2 > 2$. Take $0 < \varepsilon < \frac{m^2-2}{2m}$. Then $(m-\varepsilon)^2 = m^2 - 2m\varepsilon + \varepsilon^2 > m^2 - 2m\varepsilon > 2$. By sup characterization, $\exists a_\varepsilon \in A$ with $a_\varepsilon > m - \varepsilon$, so $a_\varepsilon^2 > (m-\varepsilon)^2 > 2$, contradiction.

Case 2: $m^2 < 2$. Take $0 < \varepsilon < \min\left\{1, \frac{2-m^2}{2m+1}\right\}$. Then $\varepsilon^2 < \varepsilon$, and $(m+\varepsilon)^2 < m^2 + (2m+1)\varepsilon < 2$, so $m + \varepsilon \in A$, contradiction.

Thus $m^2 = 2$, so $m = \sqrt{2}$. □

Exercise

Question:Let $A, B \subset \mathbb{R}$ be nonempty and bounded. Prove: $\sup(A \cup B) = \max\{\sup A, \sup B\}$ and $\inf(A \cup B) = \min\{\inf A, \inf B\}$.

证明. proof:

Prove $\sup(A \cup B) = \max\{\sup A, \sup B\}$Let $M = \max\{\sup A, \sup B\}$.

M is an upper bound for $A \cup B$:For any $x \in A \cup B$, $x \in A$ (so $x \leq \sup A \leq M$) or $x \in B$ (so $x \leq \sup B \leq M$). Thus M bounds $A \cup B$ above.

M is the least upper bound: Suppose $L < M$ is an upper bound for $A \cup B$. Then $L < \sup A$ or $L < \sup B$ (whichever is the max). If $L < \sup A$, there exists $x \in A$ with $x > L$ (by definition of $\sup A$); since $x \in A \cup B$, L cannot be an upper bound. Contradiction. Thus $M = \sup(A \cup B)$. □

0.2 Density in R

Property

(Archimedean Property):If $x, y \in \mathbb{R}$ and $x > 0$, then $\exists n \in \mathbb{N}$ such that $nx > y$.

证明. Let $A = \{nx : n \in \mathbb{N}\}$. If it is false, A is upper bounded. Let $m^* = \sup A$. By sup characterization, $\exists a_\varepsilon \in A$ with $a_\varepsilon > m^* - \varepsilon$.

Take $\varepsilon = x$, then $\exists n$ such that $nx > m^* - x$, so $(n+1)x > m^*$ and $(n+1)x \in A$, contradiction. □

let x as 1, then we have:

Corollary & Secondary Conclusion

For all $y \in \mathbb{R}$, there exists $n \in \mathbb{N}$ such that $n > y$.

Definition

Integer part: For all $a \in \mathbb{R}$, there exists a unique $m \in \mathbb{Z}$ such that $m \leq a < m + 1$. Denote $m = \lfloor a \rfloor$.

证明. Use Archimedean property with $x = 1$ and $y = a$: $\exists n'$ such that $n' - 1 \leq a < n'$. From the finite set $\{-n', ..., 0, ..., n'\}$, pick $m = \max\{k \leq a\}$. Then $m \leq a < m + 1$, and m is unique. □

Theorem

Density of $\mathbb{Q}$: For all $a < b$, there exists $p \in \mathbb{Q}$ such that $a < p < b$.

证明. We need $a < \frac{m}{n} < b$, i.e., $na < m < nb = na + n(b - a)$. By Archimedean property, choose n such that $n(b - a) > 1$.

Let $m = \lfloor na \rfloor + 1$, then $na < m < nb$, so $a < \frac{m}{n} < b$. □

Definition

0.3 Limit, Sequence & Monotone Convergence Theorem

Numerical sequence:A numerical sequence is a function $f : \mathbb{N} \to \mathbb{R}$, denoted by $f(n) = a_n$ (or c_n, u_n, etc.). For each $n \in \mathbb{N}$, $a_n \in \mathbb{R}$.
The formula for a_n is the general term. Sequences can also be defined recursively (e.g., $u_{n+1} = \frac{1}{1+u_n}$, $u_1 = 1$).

Example

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

And now, we need to make a stage for the $\epsilon - \delta$Language:

Lemma

If $a \in \mathbb{R}$ and $\forall \varepsilon > 0$, $|a| < \varepsilon$, then $a = 0$.

证明. If $|a| > 0$, choose $\varepsilon = \frac{|a|}{2}$, which gives $|a| < \frac{|a|}{2}$, contradiction. □

Corollary & Secondary Conclusion

If $\forall \varepsilon > 0$, $|u_k - u| < \varepsilon$, then $u_k = u$.

Definition

Convergence of sequence:A sequence $(u_n)_{n=1}^{\infty}$ converges to $u^* \in \mathbb{R}$ if $\forall \varepsilon > 0$, $\exists N_\varepsilon \in \mathbb{N}$ such that $\forall n \geq N_\varepsilon$, $|u_n - u^*| < \varepsilon$. We write $u^* = \lim_{n\to\infty} u_n$.

Example

$\lim_{n\to\infty} \frac{1}{n} = 0$

证明. $\forall \varepsilon > 0$, let $N_\varepsilon = \lceil \frac{1}{\varepsilon} \rceil + 1$. For $n \geq N_\varepsilon$, $\frac{1}{n} < \varepsilon$, so $|\frac{1}{n} - 0| < \varepsilon$. □

Example

$u_n = (-1)^n$ diverges.

证明. Suppose $u_n \to u^* \in \mathbb{R}$. Then $\forall \varepsilon > 0$, $\exists N_\varepsilon$ such that $\forall n \geq N_\varepsilon$, $|u_n - u^*| < \varepsilon$. For even $n = 2k \geq N_\varepsilon$, $u_n = 1$, so $|1 - u^*| < \varepsilon$. For odd $n = 2k+1 \geq N_\varepsilon$, $u_n = -1$, so $|-1-u^*| < \varepsilon$. This implies $u^* = 1$ and $u^* = -1$, contradiction. □

we will show 4 important rules about the limit of sequence:

Theorem

- **Addition:** If $\lim u_n = u$ and $\lim v_n = v$, then $\lim(u_n + v_n) = u + v$.
- **Multiplication:** If $\lim u_n = u$ and $\lim v_n = v$, then $\lim(u_n v_n) = (\lim u_n)(\lim v_n)$.
- **Reciprocal:** If $\lim u_n = u$ and $u \neq 0$, then $\lim \frac{1}{u_n} = \frac{1}{\lim u_n}$.

Definition

Bounded sequence:A sequence $(u_n)_{n=1}^{\infty}$ is bounded if $\exists C > 0$ such that $|u_n| \leq C$ for all $n \in \mathbb{N}$.

Fact and proposition

If (u_n) converges, then it is bounded.

证明. Let $\lim u_n = u$. Choose $\varepsilon = 1$, then $\exists N$ such that $\forall n \geq N$, $|u_n - u| < 1$, so $|u_n| \leq |u| + 1$. Let $C = \max\{|u_1|, ..., |u_{N-1}|, |u| + 1\}$, then $|u_n| \leq C$ for all n. □

【Remark】:

A **bounded sequence** does not *necessarily* converge (e.g., $u_n = (-1)^n$).

Definition

Monotone sequence:A sequence is monotone if it is increasing ($u_n \leq u_{n+1}$ for all n) or decreasing ($u_n \geq u_{n+1}$ for all n).

And now we arrive the last Theorem in this lecture:

Theorem

Monotone Convergence Theorem: If (u_n) is **increasing and bounded from above (so bounded)**, then

$$\lim_{n\to\infty} u_n = \sup\{u_n : n \in \mathbb{N}\}.$$

If (u_n) is **decreasing and bounded from below**, then

$$\lim_{n\to\infty} u_n = \inf\{u_n : n \in \mathbb{N}\}.$$

证明. Only prove the increasing case. Let $A = \{u_n : n \in \mathbb{N}\}$. Since A is increasing and bounded above, $\sup A = u^*$ exists. $\forall \varepsilon > 0$, by sup characterization, $\exists N$ such that $u_N > u^* - \varepsilon$. By monotonicity, $\forall n \geq N$, $u^* - \varepsilon < u_N \leq u_n \leq u^*$, so $|u_n - u^*| < \varepsilon$. Hence $\lim u_n = u^*$. □

【Remark】:

A monotone sequence converges *if and only if* it is bounded.

Definition

Infinite limits: We say $\lim_{n\to\infty} u_n = +\infty$ *if* $\forall A > 0$, $\exists N$ such that $\forall n \geq N$, $u_n > A$. Similarly, $\lim_{n\to\infty} u_n = -\infty$ if $\forall A > 0$, $\exists N$ such that $\forall n \geq N$, $u_n < -A$.

Exercise

Question: Given $a_1 \in \mathbb{R}$, define

$$a_{n+1} = \frac{a_n + 4}{5} \quad (n \geq 1).$$

Show that (a_n) converges and find its limit (hint: solve $L = \frac{L+4}{5}$ and prove $|a_{n+1} - L| = \frac{1}{5}|a_n - L|$).

证明. proof:

□

Order preserving:

Lemma

If for two sequences (u_n) and (v_n) we have $u_n \le v_n$ for all n, then

$$\lim_{n\to\infty} u_n \le \lim_{n\to\infty} v_n$$

whenever both limits exist.

we will argue by contradiction

证明. Argue by contradiction. If $\lim u_n = u^* > v^* = \lim v_n$, then take $\varepsilon = \frac{u^*-v^*}{2} > 0$. By definition, there exist N_1, N_2 such that $n \ge N_1 \Rightarrow u_n > u^* - \varepsilon$ and $n \ge N_2 \Rightarrow v_n < v^* + \varepsilon$. For $n \ge \max\{N_1, N_2\}$ we get

$$u_n > u^* - \varepsilon = v^* + \varepsilon > v_n,$$

contradiction. □

Theorem

(Nested Interval Theorem). Given closed bounded intervals $I_1 \supset I_2 \supset I_3 \supset \cdots$, $I_n = [a_n, b_n]$, then

$$\bigcap_{k=1}^{\infty} I_k \neq \varnothing.$$

证明. Because in I_1, $a_1 \le a_2 \le \cdots \le b_2 \le b_1$, thus (a_n) and (b_n) are bounded. Moreover, $a_{n+1} \ge a_n$ (increasing), $b_{n+1} \le b_n$ (decreasing), thus (a_n) and (b_n) both converge. Let $a := \lim_{n\to\infty} a_n$, $b := \lim_{n\to\infty} b_n$. By Lemma 1, $a \le b$. Then $a_n \le a \le b \le b_n$ for every n. Hence $[a,b] \subset \bigcap_{k=1}^{\infty} I_k$. □

0.4 Liminf and Limsup

Recall: A sequence (u_n) has $\lim_{n\to\infty} u_n = u^*$ if

$$\forall \varepsilon > 0,\ \exists N_\varepsilon \in \mathbb{N},\ \forall n \ge N_\varepsilon,\ |u_n - u^*| < \varepsilon.$$

We generalize this to discuss "*limits*" for *any* sequence (as not all sequences converge in the usual sense).

There are three types of divergence: divergence to $+\infty$, divergence to $-\infty$, or fluctuation/oscillation (no single limit, but bounded variation). For oscillating sequences, we use "*upper/lower bounds*": choose $A \le B$ and require $A - \varepsilon < u_n < B + \varepsilon$ for large n, with A, B chosen "optimally".

Definition

For any sequence (u_n).define:

$$\liminf_{n\to\infty} u_n = \sup_{k\geq 1} \inf_{n\geq k} u_n, \qquad \limsup_{n\to\infty} u_n = \inf_{k\geq 1} \sup_{n\geq k} u_n.$$

- Inner inf / sup control lower/upper bounds for u_n when n is sufficiently large.
- Outer sup / inf select the "optimal" A, B (tightest bounds).
- lim sup and lim inf always exist (may be $\pm\infty$) by ***supremum / infimum axioms***.

Example

- Let $u_n = (-1)^n$. For any n, $\{u_k : k \geq n\} = \{-1, 1\}$, so $\sup\{u_k : k \geq n\} = 1$ and $\inf\{u_k : k \geq n\} = -1$. Thus $\limsup_{n\to\infty} u_n = 1$, $\liminf_{n\to\infty} u_n = -1$.
- Let $u_n = n$. Then $\limsup_{n\to\infty} u_n = \liminf_{n\to\infty} u_n = +\infty$.

We can see that: $\limsup u_n$ is the *largest* limit of any convergent subsequence; $\liminf u_n$ is the *smallest* limit of any convergent subsequence.

Now let us check Liminf/Limsup for Bounded Sequences:

Let (u_n) be bounded. Define:

Definition

$a_n := \inf_{k\geq n} u_k$ $b_n := \sup_{k\geq n} u_k$.

Then we have these **propositions:**

Fact and proposition

- $a_{n+1} \geq a_n$ (so (a_n) is *increasing*),
- $b_{n+1} \leq b_n$ (so (b_n) is *decreasing*),
- Thus (a_n) and (b_n) converge, with $\liminf_{n\to\infty} u_n = \lim_{n\to\infty} a_n$ and $\limsup_{n\to\infty} u_n = \lim_{n\to\infty} b_n$.

证明. Let $A_n = \{u_k : k \geq n\}$. Since $k \geq n+1$ implies $k \geq n$, we have $A_{n+1} \subset A_n$. By the property of infimum: the infimum of a subset is no smaller than the infimum of the original set, so $a_{n+1} = \inf A_{n+1} \geq \inf A_n = a_n$. Thus (a_n) is increasing.

Similarly, by the property of supremum: the supremum of a subset is no larger than the supremum of the original set, so $b_{n+1} = \sup A_{n+1} \leq \sup A_n = b_n$. Thus (b_n) is decreasing.

Since (u_n) is bounded, there exists $C > 0$ such that $|u_k| \leq C$ for all k. Thus $a_n = \inf A_n \geq -C$ (lower bound for (a_n)) and $b_n = \sup A_n \leq C$ (upper bound for (b_n)). By the Monotone Convergence Theorem (MCT), (a_n) and (b_n) both converge.

By definition of $\liminf u_n = \sup_{k\geq 1} \inf_{n\geq k} u_k = \sup_{k\geq 1} a_k$. Since (a_n) is increasing, ***its limit equals its supremum***, so $\lim_{n\to\infty} a_n = \sup_{k\geq 1} a_k = \liminf u_n$.

Similarly, $\limsup u_n = \inf_{k\geq 1} \sup_{n\geq k} u_k = \inf_{k\geq 1} b_k$. Since (b_n) is decreasing, ***its limit equals its infimum***, so $\lim_{n\to\infty} b_n = \inf_{k\geq 1} b_k = \limsup u_n$. □

So we can have another version of the definition of *limsup*(or *liminf*):

Definition

For a sequence (a_n):

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \left(\sup_{k\geq n} a_k \right), \quad \liminf_{n\to\infty} a_n = \lim_{n\to\infty} \left(\inf_{k\geq n} a_k \right).$$

Now let's check the *Characterization of liminf / limsup* in bounded sequences

Property

Let $A = \liminf_{n\to\infty} u_n$ and $B = \limsup_{n\to\infty} u_n$. Then:

1. $A \leq B$.
2. $\forall \varepsilon > 0,\ \exists N_1$ such that $\forall n \geq N_1,\ u_n > A - \varepsilon$.
3. $\forall \varepsilon > 0,\ \exists N_2$ such that $\forall n \geq N_2,\ u_n < B + \varepsilon$.
4. If $A = B$, then (u_n) converges (to $A = B$).

证明. 1. For each n, $a_n = \inf_{k\geq n} u_k \leq \sup_{k\geq n} u_k = b_n$. Because limit preserves order, taking limits on both sides gives $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$, i.e., $A \leq B$.

2. Since $A = \lim_{n\to\infty} a_n$, by the definition of sequence convergence: $\forall \varepsilon > 0$, $\exists N_1$ such that $\forall n \geq N_1$, $|a_n - A| < \varepsilon$. This implies $a_n > A - \varepsilon$. But $a_n = \inf_{k\geq n} u_k$, so $u_k \geq a_n$ for all $k \geq n$. Thus $\forall n \geq N_1$, $u_n \geq a_n > A - \varepsilon$.

3. Similarly, $B = \lim_{n\to\infty} b_n$, so $\forall \varepsilon > 0$, $\exists N_2$ such that $\forall n \geq N_2$, $|b_n - B| < \varepsilon$, i.e., $b_n < B + \varepsilon$. Since $b_n = \sup_{k\geq n} u_k$, we have $u_k \leq b_n$ for all $k \geq n$. Thus $\forall n \geq N_2$, $u_n \leq b_n < B + \varepsilon$.

4. If $A = B = L$, then by (2) and (3): $\forall \varepsilon > 0$, $\exists N = \max\{N_1, N_2\}$ such that $\forall n \geq N$, $L - \varepsilon < u_n < L + \varepsilon$. By the definition of sequence convergence, $\lim_{n\to\infty} u_n = L = A = B$. □

Compared to the *proposition of inf and sup*, we can find the **difference** that:
In the proposition of inf and sup instead of liminf or limsup, we notice that:

- For every $\varepsilon > 0$, there exists $a_\varepsilon \in A$ such that $a_\varepsilon > \sup A - \varepsilon$;
- For every $\varepsilon > 0$, there exists $a_\varepsilon \in A$ such that $a_\varepsilon < \inf A + \varepsilon$.

The idea of "Constant": try to Construct a constant with respect to k:

Fact and proposition

$\sup_{k\geq n}(x_k + y_k) \leq \sup_{k\geq n} x_k + \sup_{k\geq n} y_k$.

So we can have these properties:

Property

Consider two sequences (x_n) and (y_n):

$$\limsup_{n\to\infty}(x_n + y_n) \leq \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n,$$

$$\liminf_{n\to\infty}(x_n + y_n) \geq \liminf_{n\to\infty} x_n + \liminf_{n\to\infty} y_n.$$

Hint: Think about the definition of the limsup and liminf to construct these structure.

证明. For the limsup inequality: Let $c_n = \sup_{k\geq n}(x_k+y_k)$, $a_n = \sup_{k\geq n} x_k$, $b_n = \sup_{k\geq n} y_k$. By the given fact, $c_n \leq a_n + b_n$ for all n.

Since (a_n) is decreasing and (b_n) is decreasing, their limits exist (may be $+\infty$): $\lim_{n\to\infty} a_n = \limsup x_n$, $\lim_{n\to\infty} b_n = \limsup y_n$. Taking limits on both sides of $c_n \leq a_n + b_n$, by limit preserves order, we get $\lim_{n\to\infty} c_n \leq \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$, i.e., $\limsup(x_n + y_n) \leq \limsup x_n + \limsup y_n$.

□

Exercise

(limsup preserve the order): If there exists N_1 such that $u_n \leq B'$ for all $n \geq N_1$, then $\limsup u_n \leq B'$. If there exists N_2 such that $u_n \geq A'$ for all $n \geq N_2$, then $\liminf u_n \geq A'$.

证明. Since$\sup_{k\geq n} u_k \leq B'$,and $\limsup_{n\to\infty} u_n = \inf_{k\geq 1} \sup_{n\geq k} u_n$,easy to prove this conclusion □

Exercise

Question: Consider two sequences (x_n) and (y_n). Show that, if $\lim_{n\to\infty} y_n = y$, then

$$\limsup_{n\to\infty}(x_n + y_n) = \limsup_{n\to\infty} x_n + y.$$

Subsequences

Definition

For a sequence (u_n), a *subsequence* is a sequence (u_{k_n}) where $k_1 < k_2 < \cdots$ (indices are strictly increasing).

Easy to get that: If $n_1 < n_2 < \cdots$ are positive integers, then $n_k \geq k$ for all k.

Lemma

(Convergent Subsequences from Liminf/Limsup): Let (u_n) be bounded, with $A = \liminf_{n\to\infty} u_n$ and $B = \limsup_{n\to\infty} u_n$. Then:

- There exist subsequences $(u_{k_j}) \to A$ and $(u_{p_j}) \to B$.
- For each $j \in \mathbb{N}$, these subsequences satisfy:

$$A - \frac{1}{j} < u_{k_j} < A + \frac{1}{j}, \quad B - \frac{1}{j} < u_{p_j} < B + \frac{1}{j},$$

  with $k_1 < k_2 < \cdots$ and $p_1 < p_2 < \cdots$.

【Remark】:

To solve this problem, we must completely understand the concept of *limsup* and *liminf* and their def. in two forms.

证明. We construct the subsequence $(u_{k_j}) \to A$ (the construction for $(u_{p_j}) \to B$ is analogous).

By definition, $A = \lim_{n\to\infty} a_n$ where $a_n = \inf_{k\geq n} u_k$ and (a_n) is increasing. We use induction to construct k_j:

- Step 1 ($j = 1$): Let $\varepsilon = 1$. Since $a_n \to A$, there exists N_1 such that $a_{N_1} > A-1$(the property of common limit). By definition of $a_{N_1} = \inf_{k\geq N_1} u_k$, there exists $k_1 \geq N_1$ such that $u_{k_1} < a_{N_1}+1 \leq A+1$. Also, $u_{k_1} \geq a_{N_1} > A-1$, so $A-1 < u_{k_1} < A+1$(the property of *inf.*).

- Step 2 ($j = 2$): Let $\varepsilon = 1/2$. Since $a_n \to A$, there exists $N_2 > \max\{N_1, k_1\}$ such that $a_{N_2} > A-1/2$. By definition of a_{N_2}, there exists $k_2 \geq N_2$ such that $u_{k_2} < a_{N_2}+1/2 < A+1/2$. Also, $u_{k_2} \geq a_{N_2} > A-1/2$, so $A-1/2 < u_{k_2} < A+1/2$. Note $k_2 > k_1$ because $k_2 \geq N_2 > k_1$.

- Inductive Step: Suppose $k_1 < k_2 < \cdots < k_{j-1}$ are constructed such that $A - 1/(j-1) < u_{k_{j-1}} < A+1/(j-1)$. Let $\varepsilon = 1/j$. There exists $N_j > \max\{N_{j-1}, k_{j-1}\}$ such that $a_{N_j} > A-1/j$. Choose $k_j \geq N_j$ such that $u_{k_j} < a_{N_j} + 1/j < A+1/j$, and $u_{k_j} \geq a_{N_j} > A-1/j$. Thus $A-1/j < u_{k_j} < A+1/j$, and $k_j > k_{j-1}$.

By induction, we get a strictly increasing index sequence (k_j) satisfying $A-1/j < u_{k_j} < A+1/j$ for all j. By the squeeze principle, $\lim_{j\to\infty} u_{k_j} = A$.

For $B = \limsup u_n = \lim_{n\to\infty} b_n$ (where $b_n = \sup_{k\geq n} u_k$ is decreasing), repeat the above construction with b_n instead of a_n to get a subsequence (u_{p_j}) such that $B-1/j < u_{p_j} < B+1/j$ and $p_1 < p_2 < \cdots$, so $\lim_{j\to\infty} u_{p_j} = B$. □

Now we can get our Main Theorem:

Theorem

Bolzano–Weierstrass Theorem: For any **bounded** numerical sequence (u_n), there exists a ***convergent subsequence***.

证明. Since (u_n) is bounded, by the earlier lemma, its liminf $A = \liminf_{n\to\infty} u_n$ exists (as a real number, not $\pm\infty$).

The lemma also guarantees the existence of a subsequence (u_{k_j}) such that $\lim_{j\to\infty} u_{k_j} = A$. Thus (u_{k_j}) is a convergent subsequence of (u_n). □

Exercise

Question: Consider a bounded sequence (x_n) and let $A = \liminf_{n\to\infty} x_n$, $B = \limsup_{n\to\infty} x_n$. Show that, if (x_{k_n}) is a convergent subsequence of (x_n), then

$$A \leq \lim_{n\to\infty} x_{k_n} \leq B.$$

证明. To prove $A \leq \lim x_{k_n} \leq B$ (where $L = \lim x_{k_n}$):Show $L \leq B$:By definition, $B = \lim_{n\to\infty} \sup_{m\geq n} x_m$ (and $\sup_{m\geq n} x_m$ is decreasing).

For the subsequence, $k_n \geq n$ (indices are strictly increasing), so $x_{k_n} \leq \sup_{m\geq n} x_m$. Taking limits:$L = \lim_{n\to\infty} x_{k_n} \leq \lim_{n\to\infty} \sup_{m\geq n} x_m = B$.

Show $A \leq L$:By definition, $A = \lim_{n\to\infty} \inf_{m\geq n} x_m$ (and $\inf_{m\geq n} x_m$ is increasing).

For the subsequence, $k_n \geq n$, so $x_{k_n} \geq \inf_{m\geq n} x_m$. Taking limits:$A = \lim_{n\to\infty} \inf_{m\geq n} x_m \leq \lim_{n\to\infty} x_{k_n} = L$.

Combining these gives $A \leq L \leq B$. □

0.5 Cauchy Sequence

Definition

Cauchy sequence. (u_n) is said to be a Cauchy sequence if

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} \text{ s.t. } \forall m, n \geq N,\ |u_m - u_n| < \varepsilon.$$

Lemma

A Cauchy sequence is bounded.

证明. proof:

Let (u_n) be a Cauchy sequence.

By the definition of a Cauchy sequence, for $\varepsilon = 1$, there exists $N \in \mathbb{N}$ such that for all $m, n \geq N$, $|u_m - u_n| < 1$.Fix $n = N$.

For any $m \geq N$, the triangle inequality implies $|u_m| = |u_m - u_N + u_N| \leq |u_m - u_N| + |u_N| < 1 + |u_N|$.Define $C = \max\{|u_1|, |u_2|, \dots, |u_{N-1}|, 1 + |u_N|\}$. For every $n \in \mathbb{N}$, if $n < N$, $|u_n| \leq C$ by the definition of C; if $n \geq N$, $|u_n| < 1 + |u_N| \leq C$.Thus, $|u_n| \leq C$ for all $n \in \mathbb{N}$, meaning (u_n) is bounded.

□

Theorem

The sequence (u_n) converges, ***iff*** it is Cauchy.

【Remark】:

For the converse, there are two ways to prove it!

证明. Sufficiency ((u_n) converges $\implies$ (u_n) is Cauchy)

Suppose $\lim_{n\to\infty} u_n = u$. By the definition of limit, for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $|u_n - u| < \frac{\varepsilon}{2}$.

For all $m, n \geq N$, the triangle inequality gives:$|u_m - u_n| \leq |u_m - u| + |u_n - u| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$Thus, (u_n) is a Cauchy sequence.

Necessity ((u_n) is Cauchy $\implies$ (u_n) converges)

proof 1:Bolzano-Weierstrass:

By the lemma above, a Cauchy sequence is bounded. By the Bolzano-Weierstrass Theorem, every bounded sequence has a convergent subsequence. Let (u_{k_n}) be a convergent subsequence of (u_n), and let $\lim_{n\to\infty} u_{k_n} = u$.To prove $\lim_{n\to\infty} u_n = u$: For any $\varepsilon > 0$, by the Cauchy property, there exists $N_1 \in \mathbb{N}$ such that for all $m, n \geq N_1$, $|u_m - u_n| < \frac{\varepsilon}{2}$.

By the convergence of (u_{k_n}), there exists $N_2 \in \mathbb{N}$ such that for all $n \geq N_2$, $|u_{k_n} - u| < \frac{\varepsilon}{2}$.

Let $N = \max\{N_1, N_2\}$. For any $n \geq N$, we have $k_n \geq n \geq N$ (since subsequence indices satisfy $k_n \geq n$). Thus:$|u_n - u| \leq |u_n - u_{k_n}| + |u_{k_n} - u| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$Hence, $\lim_{n\to\infty} u_n = u$.

proof 2:Constant thinking, fixed technique:

A Cauchy sequence is bounded, so its liminf $A = \liminf_{n\to\infty} u_n$ and limsup $B = \limsup_{n\to\infty} u_n$ exist and are finite.

For any $\varepsilon > 0$, by the Cauchy property, there exists $N \in \mathbb{N}$ such that for all $m, n \geq N$, $|u_m - u_n| \leq \varepsilon$, which implies $u_m \leq u_n + \varepsilon$ for all $m \geq N$.Fix $n \geq N$.

Taking limsup on both sides of $u_m \leq u_n + \varepsilon$ (as $m \to \infty$), we get $B \leq u_n + \varepsilon$.Taking liminf on both sides as $n \to \infty$, we obtain $B \leq A + \varepsilon$.

Since $\varepsilon > 0$ is arbitrary, $B \leq A$.By definition, $A \leq B$ always holds.

Thus, $A = B$, which implies (u_n) converges (the limit is $A = B$).

□

Iteration for $\sqrt{2}$

Consider $g(t) = \frac{t}{2} + \frac{1}{t}$. This gives a sequence: $x_1 = 1$, $x_{n+1} = g(x_n)$. Clearly $x_n \in \mathbb{Q}$ for all n.

> ***Claim***
>
> (x_n) is Cauchy; the limit is $\sqrt{2}$.

证明. proof:

First, we establish two key lemmas, then prove the claim.

Lemma 1: For all $n \in \mathbb{N}$, $x_n \in [1, 2]$We use mathematical induction:

Base case: $x_1 = 1$, which is clearly in $[1, 2]$.

Inductive step: Assume $x_n \in [1, 2]$. Then $x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}$. By the AM-GM inequality, $\frac{x_n}{2} + \frac{1}{x_n} \geq 2\sqrt{\frac{x_n}{2} \cdot \frac{1}{x_n}} = \sqrt{2} \geq 1$. Also, since $x_n \leq 2$, $\frac{x_n}{2} \leq 1$ and $\frac{1}{x_n} \leq 1$, so $x_{n+1} \leq 1 + 1 = 2$. Thus, $x_{n+1} \in [1, 2]$.

By induction, $x_n \in [1, 2]$ for all $n \in \mathbb{N}$.

Lemma 2: For all $p, n \in \mathbb{N}$, $|x_{n+p} - x_n| \leq \frac{1}{2^{n-1}}|x_{p+1} - x_1|$Note that $x_{n+1} - x_n = \frac{x_n}{2} + \frac{1}{x_n} - x_n = \frac{1}{x_n} - \frac{x_n}{2} = \frac{2-x_n^2}{2x_n}$.

For $x_{n+p}-x_n$, we iterate this relation:$|x_{n+p}-x_n| = \left|\frac{2-x_{n+p-1}^2}{2x_{n+p-1}} - \frac{2-x_{n-1}^2}{2x_{n-1}}\right| = \frac{1}{2}\left|\frac{(2-x_{n+p-1}^2)x_{n-1}-(2-x_{n-1}^2)x_n}{x_{n+p-1}x_{n-1}}\right.$ the numerator and using $x_k \in [1, 2]$ (so $\frac{1}{x_{n+p-1}x_{n-1}} \in [\frac{1}{4}, 1]$), we get $|x_{n+p} - x_n| \leq \frac{1}{2}|x_{n+p-1}-x_{n-1}|$.Iterating this inequality $n-1$ times:$|x_{n+p}-x_n| \leq \frac{1}{2^{n-1}}|x_{p+1}-x_1|$Since $x_1 = 1$ and $x_{p+1} \in [1, 2]$, $|x_{p+1}-x_1| \leq 1$, so $|x_{n+p}-x_n| \leq \frac{4}{2^n}$ (tightened bound from the PDF).

Proof of the Claim

(x_n) is Cauchy: For any $\varepsilon > 0$, choose $N \in \mathbb{N}$ such that $\frac{4}{2^N} \leq \varepsilon$. For all $n \geq N$ and $p \in \mathbb{N}$:$|x_{n+p} - x_n| \leq \frac{4}{2^n} \leq \frac{4}{2^N} \leq \varepsilon$

By the definition of a Cauchy sequence, (x_n) is Cauchy.The limit is $\sqrt{2}$: Let $x = \lim_{n\to\infty} x_n$. Since $x_n \in [1, 2]$ for all n, $x \in [1, 2]$ (limits preserve closed intervals) and $x > 0$.Taking the limit on both sides of $x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}$ (limits preserve arithmetic operations):$x = \frac{x}{2} + \frac{1}{x}$Multiply both sides by 2x (since $x > 0$):$2x^2 = x^2 + 2 \implies x^2 = 2$Since $x > 0$, $x = \sqrt{2}$.

□

0.6 Convergence of Series

Today's topic: series — infinite summation, a special type of sequence. In other words, we compute

$$\sum_{n=1}^{\infty} a_n \equiv a_1 + a_2 + a_3 + \cdots$$

Two possible scenarios for $(*)$:

1. We can assign a real value to $(*) \Rightarrow$ **convergence**.

2. We cannot assign a real value to $(*)$; sometimes we set it to $\pm\infty$. But in other cases, we cannot meaningfully define any value to $(*) \Rightarrow$ **divergence**.

Consider a numerical sequence (a_n).

Definition

Consider the **partial-sum sequence**

$$S_n := \sum_{k=1}^{n} a_k.$$

(S_n) is called the *partial sums* of the series $\sum_{k=1}^{\infty} a_k$.

If $S_n \to S$ (as $n \to \infty$), the series is **convergent**, and we write $S = \sum_{k=1}^{\infty} a_k$. If S_n does not converge, the series is **divergent**.

Property

Cauchy's Criterion. The series $\sum_{n=1}^{\infty} a_n$ converges ***if and only if***

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N} \text{ such that } \forall m > n > N,\ \left| \sum_{k=n+1}^{m} a_k \right| < \varepsilon.$$

Proof. A series $\sum a_n$ converges $\Leftrightarrow$ its partial-sum sequence (S_n) converges $\Leftrightarrow$ (S_n) is a Cauchy sequence. But for $m < n$,

$$|S_n - S_m| = \left| \sum_{k=m+1}^{n} a_k \right|.$$

Property

If $\sum_{n=1}^{\infty} a_n$ converges, then $a_n \to 0$ (as $n \to \infty$).
Importantly, if a_n does not converge, or converges to a limit other than 0, then $\sum a_n$ diverges.

Exercise

Prove Harmonic series: $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges (even though $\lim_{n\to\infty} \frac{1}{n} = 0$).

证明. proof:

□

Definition

$\sum a_n$ is said to be **absolutely convergent** if $\sum |a_n|$ converges.

If $\sum_{n=1}^{\infty} |a_n|$ converges, then so does $\sum_{n=1}^{\infty} a_n$.

Absolute convergence usually induces better properties (rearrangement: next lecture).

Test

Method

(Comparison Test). Consider two numerical series $\sum a_n$ and $\sum b_n$ with $0 \le a_n \le b_n$. Then:

1. If $\sum b_n$ converges, so does $\sum a_n$.
2. If $\sum a_n$ diverges, so does $\sum b_n$.

Proof. Let (S_n) and (T_n) be sequences of partial sums. Then $\forall n$, $0 \le S_n \le T_n$. Thus, if (T_n) is bounded, (S_n) is bounded; if (S_n) is unbounded, (T_n) is unbounded. $\square$

Method

(Root Test). Consider (a_n) and let $\alpha = \limsup_{n\to\infty} \sqrt[n]{|a_n|}$.

1. If $0 \le \alpha < 1$, then $\sum |a_n|$ converges (hence $\sum a_n$ converges absolutely).
2. If $\alpha > 1$, then $a_n \nrightarrow 0$ and $\sum a_n$ diverges.

Proof. (1) If $0 \le \alpha < 1$, choose $\alpha < b < 1$. By lim sup characterization, $\exists N$ such that $\forall n \ge N$, $\sqrt[n]{|a_n|} < b < 1$. Thus $|a_n| \le b^n$, and $\sum a_n$ converges (by comparison with $\sum b^n$, a convergent geometric series).

(2) If $\alpha > 1$, lim sup implies a subsequence a_{n_k} with $\sqrt[n_k]{|a_{n_k}|} > 1$, so $|a_{n_k}| \ge 1$ for large k. Thus $a_n \nrightarrow 0$. $\square$

【Remark】:

The test is useless when $\alpha = 1$.

Method

(Ratio Test). Consider (a_n) and let $\alpha = \limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$.
If $0 \le \alpha < 1$, then $\sum |a_n|$ converges (so $\sum a_n$ converges absolutely).

Proof. If $\alpha < 1$, choose $\alpha < r < 1$ and let $\varepsilon = \frac{r-\alpha}{2}$. By lim sup characterization, $\exists N$

such that $\forall n \geq N$,

$$\left|\frac{a_{n+1}}{a_n}\right| < \alpha + \varepsilon < r.$$

By iteration, for $n \geq N$:

$$|a_{n+1}| < r|a_n| < r^2|a_{n-1}| < \cdots < r^{n-N+1}|a_N| \leq Cr^n$$

where $C = r^{-N+1}|a_N|$. Thus $\sum |a_n| \leq \sum_{n<N} |a_n| + C \sum_{n \geq N} r^n < \infty$ (since $\sum r^n$ converges). □

【Remark】:

The test is also useless when $\alpha = 1$.

Then, we will introduce a special series: Power series.

Definition

For a numerical sequence (c_n) and real number z, the series

$$\sum_{n=0}^{\infty} c_n z^n = c_0 + c_1 z + c_2 z^2 + \cdots$$

is a **power series** in z.

Example

Example: $\sum_{n \geq 1} x^n$ is a power series and converges if $|x| < 1$.

Definition

Let $\ell = \limsup_{n \to \infty} \sqrt[n]{|c_n|}$ and define

$$R = \begin{cases} \frac{1}{\ell}, & \ell > 0, \\ +\infty, & \ell = 0, \\ 0, & \ell = +\infty. \end{cases}$$

Then $\sum c_n z^n$ converges absolutely for $|z| < R$ and diverges for $|z| \geq R$.
And we denote R as the **radius of convergence**.

Proof. Let $a_n = c_n z^n$. Then

$$\alpha = \limsup_{n \to \infty} \sqrt[n]{|a_n|} = \limsup_{n \to \infty} \sqrt[n]{|c_n|}\, |z| = \ell\, |z|.$$

The conclusion follows from the Root Test. □

Example

Example 1 (Geometric series). $\sum_{n=1}^{\infty} z^n$ (where $c_n \equiv 1$) has radius of convergence $R = 1$.

Example 2 (Exponential series). $\sum_{n=1}^{\infty} \frac{z^n}{n!}$ has radius of convergence $R = +\infty$.

From the example above, we have:

Definition

(Exponential). We define

$$e^z := \sum_{n=0}^{\infty} \frac{z^n}{n!} \quad \text{(note we start from } n = 0\text{)}.$$

Lemma

(Bounding the tail of the series for e):Let $S_n = \sum_{k=0}^{n} \frac{1}{k!}$. Then $\forall n \geq 1$, then:
For all $n \geq 1$,

$$0 \leq e - S_n = \sum_{k=n+1}^{\infty} \frac{1}{k!} \leq \frac{1}{n \cdot n!}.$$

证明. For $k \geq n+1$, $\frac{1}{k!} \leq \frac{1}{(n+1)^{k-n} n!}$. Hence

$$\sum_{k=n+1}^{\infty} \frac{1}{k!} \leq \frac{1}{n!} \sum_{r=1}^{\infty} \frac{1}{(n+1)^r} = \frac{1}{n!} \cdot \frac{1}{n} = \frac{1}{n \cdot n!}.$$

□

And using the lemma above, we can gain:

Corollary & Secondary Conclusion

e is irrational.

证明. Suppose $e = \frac{p}{q}$ with $p, q \in \mathbb{N}$, $\gcd(p, q) = 1$. Then

$$q!\, e = p(q-1)! \in \mathbb{Z}, \quad q!\, S_q = \sum_{k=0}^{q} \frac{q!}{k!} \in \mathbb{Z}.$$

And Thus the sum $q! \cdot S_q$ is a sum of integers, so it is also an integer.
But $0 < q!\,(e - S_q) < \frac{q!}{q \cdot q!} = \frac{1}{q} < 1$, a contradiction. □

Theorem

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

证明. Let $t_n = \left(1+\frac{1}{n}\right)^n$. Using the binomial theorem,

$$\left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n}\binom{n}{k}\frac{1}{n^k} = \sum_{k=0}^{n}\frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{1}{k!}.$$

But $\frac{n(n-1)\cdots(n-k+1)}{n^k} < 1$, hence $t_n \le \sum_{k=0}^{n}\frac{1}{k!} \le e$. Thus $\limsup t_n \le e$.

On the other hand, fix $m \in \mathbb{N}$. Then

$$t_n = \sum_{k=0}^{n}\frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{1}{k!} \ge \sum_{k=0}^{m}\frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{1}{k!} \xrightarrow{n\to\infty} \sum_{k=0}^{m}\frac{1}{k!} = S_m.$$

Therefore $\liminf_{n\to\infty} t_n \ge S_m$. This is for any m. Thus $\liminf t_n \ge e$. Hence $\lim t_n = e$.

□

And then we will discuss a new topic: Abel’ spartial summation.

Theorem

Consider two numerical sequences $(a_n)_{n\ge0}$ and $(b_n)_{n\ge0}$. Let $A_n = \sum_{k=0}^{n} a_k$. For integers $p \le q$,

$$\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q} A_n(b_n - b_{n+1}) + A_q b_{q+1} - A_{p-1}b_p. \tag{1}$$

use the Table below for further understanding:

Discrete Scenario (Series)	Continuous Scenario (Integral)	Core Role
Sum of product of sequences $\sum a_n b_n$	Integral of product of functions $\int a(x)b(x)dx$	“Product-type object” to be handled
Partial sum $A_n = \sum_{k=0}^{n} a_k$	Primitive function $A(x) = \int a(t)dt$	Intermediate object from “accumulation/integration”
Difference $b_n - b_{n+1}$	Differential $db(x) = b'(x)dx$	“Discrete differentiation/differential” of b
Boundary term $A_q b_{q+1} - A_{p-1}b_p$	Boundary value $A(x)b(x)\|_a^b$	“Endpoint term” left after splitting

表 1: Analogy between discrete and continuous scenarios

证明. To prove the summation by parts formula:

Recall $A_n = \sum_{k=0}^{n} a_k \implies a_n = A_n - A_{n-1}$ (define $A_{-1} = 0$). Substitute into the left-hand side (LHS):

$$\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q}(A_n - A_{n-1})b_n = \sum_{n=p}^{q} A_n b_n - \sum_{n=p}^{q} A_{n-1}b_n.$$

Reindex the second sum ($m = n - 1$):

$$\sum_{n=p}^{q} A_{n-1} b_n = \sum_{m=p-1}^{q-1} A_m b_{m+1} = A_{p-1} b_p + \sum_{n=p}^{q-1} A_n b_{n+1}.$$

Split the first sum:

$$\sum_{n=p}^{q} A_n b_n = \sum_{n=p}^{q-1} A_n b_n + A_q b_q.$$

Subtract the reindexed sum from the split sum:

$$\left(\sum_{n=p}^{q-1} A_n b_n + A_q b_q \right) - \left(A_{p-1} b_p + \sum_{n=p}^{q-1} A_n b_{n+1} \right) = -A_{p-1} b_p + A_q b_q + \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}).$$

Expand the right-hand side (RHS) sum:

$$\sum_{n=p}^{q} A_n (b_n - b_{n+1}) = \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) + A_q (b_q - b_{q+1}).$$

Substitute into the RHS and simplify (cross terms cancel):

$$\sum_{n=p}^{q} A_n (b_n - b_{n+1}) + A_q b_{q+1} - A_{p-1} b_p = \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p.$$

This matches the simplified LHS. Thus:

$$\boxed{\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q} A_n (b_n - b_{n+1}) + A_q b_{q+1} - A_{p-1} b_p}$$

□

Theorem

Dirichlet Test: Consider (a_n) and (b_n), and let $A_n = \sum_{k=0}^{n} a_k$. Suppose

1. A_n is bounded;
2. $b_0 \geq b_1 \geq b_2 \geq \cdots$ (i.e. (b_n) is decreasing);
3. $b_n \to 0$ $(n \to \infty)$.

Then $\sum_{n=0}^{\infty} a_n b_n$ converges.

证明. We use Cauchy's criterion. For $p \leq q$,

$$\left| \sum_{n=p}^{q} a_n b_n \right| = \left| \sum_{n=p}^{q} A_n (b_n - b_{n+1}) + A_q b_{q+1} - A_{p-1} b_p \right|$$

By the triangle inequality:

$$\leq \sum_{n=p}^{q} |A_n|(b_n - b_{n+1}) + |A_q|b_{q+1} + |A_{p-1}|b_p$$

Let $|A_n| \leq M$ (for some constant M). Then:

$$\leq M \left(\sum_{n=p}^{q} (b_n - b_{n+1}) + b_{q+1} + b_p \right)$$

Telescope the sum $\sum_{n=p}^{q}(b_n - b_{n+1}) = b_p - b_{q+1}$, so:

$$= M\left(b_p - b_{q+1} + b_{q+1} + b_p\right) = 2Mb_p.$$

Note $b_p \to 0$, so for any $\varepsilon > 0$, choose N such that $b_p < \frac{\varepsilon}{2M}$ for all $p \geq N$. Then for all $q \geq p \geq N$:

$$\left| \sum_{n=p}^{q} a_n b_n \right| \leq 2Mb_p < \varepsilon.$$

□

Example

Example 1: Violates Condition 1 (A_n is not bounded)

- Let $a_n = 1$ for all n (so $A_n = \sum_{k=0}^{n} 1 = n+1$, which is **unbounded**).

- Let $b_n = \frac{1}{n+1}$ (**decreasing** and $b_n \to 0$ as $n \to \infty$).

Series: $\sum_{n=0}^{\infty} a_n b_n = \sum_{n=0}^{\infty} \frac{1}{n+1}$.

Conclusion: This is the harmonic series, which **diverges**.

Example 2: Violates Condition 2 (b_n is not decreasing)

- Let $a_n = (-1)^n$ (so $A_n = \sum_{k=0}^{n}(-1)^k$, which is **bounded** by 1).

- Define b_n as:

$$b_n = \begin{cases} \frac{1}{n+1}, & \text{if } n \text{ is even,} \\ \frac{2}{n+1}, & \text{if } n \text{ is odd.} \end{cases}$$

Here, $b_n \to 0$ (as $n \to \infty$), but b_n is **not decreasing** (e.g., $b_1 = 1$, $b_2 = \frac{1}{3} < 1$, but $b_3 = \frac{1}{2} > \frac{1}{3}$).

Series: $\sum_{n=0}^{\infty} a_n b_n = 1 - 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} - \frac{1}{3} + \frac{1}{7} - \frac{1}{4} + \cdots$.

Conclusion: The series **diverges** (partial sums grow negatively without bound).

Example 3: Violates Condition 3 ($b_n \not\to 0$)

- Let $a_n = (-1)^n$ (so $A_n = \sum_{k=0}^{n}(-1)^k$, which is **bounded** by 1).

- Let $b_n = 1$ for all n (**decreasing**, since $1 \geq 1 \geq 1 \geq \cdots$, but $b_n \not\to 0$).

Series: $\sum_{n=0}^{\infty} a_n b_n = \sum_{n=0}^{\infty}(-1)^n$.

Conclusion: This alternating series oscillates ($1 - 1 + 1 - 1 + \cdots$) and **diverges**.

Corollary & Secondary Conclusion

Suppose $u_1 \geq u_2 \geq \cdots$ and $u_n \to 0$. Then $\sum_{n=1}^{\infty}(-1)^n u_n$ converges. In particular, $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}$ converges, but not absolutely.

证明. proof:

□

Definition

Consider two series $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$. Put

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0.$$

The series $\sum_{n=0}^{\infty} c_n$ is said to be the **product** (Cauchy product) of $\sum a_n$ and $\sum b_n$.

For finite sums, $\left(\sum_{k=0}^{N} a_k\right)\left(\sum_{k=0}^{N} b_k\right)$ expands into the diagonal sums that form $c_0, c_1, \ldots, c_{2N}$. If both series converge, one expects the product rule.

Theorem

Product rule. If $\sum a_n$ *converges absolutely* and $\sum b_n$ *converges*, then the product $\sum c_n$ converges, and

$$\sum_{n=0}^{\infty} c_n = \left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right).$$

Hint: We let $B_n = \sum_{k=0}^{n} b_k$ and $B = \sum_{k=0}^{\infty} b_k$ (so $B_n \to B$). Then

证明. Let $B_n = \sum_{k=0}^{n} b_k$ and $B = \sum_{k=0}^{\infty} b_k$ (so $B_n \to B$). Then

$$\sum_{k=0}^{n} c_k = \sum_{k=0}^{n} \sum_{j=0}^{k} a_j b_{k-j} = \sum_{j=0}^{n} a_j \sum_{k=j}^{n} b_{k-j} = \sum_{j=0}^{n} a_j B_{n-j}$$

Rewrite this as:

$$= \underbrace{\sum_{j=0}^{n} a_j (B_{n-j} - B)}_{=:\gamma_n} + \left(\sum_{j=0}^{n} a_j\right) B = \gamma_n + A_n B,$$

where $A_n = \sum_{j=0}^{n} a_j$.

Fix any $N \in \mathbb{N}$. Then

$$|\gamma_n| \le \sum_{j=0}^{N} |a_j||B_{n-j} - B| + \sum_{j>N} |a_j||B_{n-j} - B|.$$

Choose N so that $\sum_{j>N} |a_j| < \varepsilon$ (possible by absolute convergence). For this fixed N, since $B_n \to B$, we can pick n large so that $|B_{n-j} - B| < \varepsilon$ for all $0 \le j \le N$. On the other hand, $\{B_n\}$ is bounded (as it converges): suppose $|B_n| \le M$ for all n, and $|B| \le M$. Then

$$\limsup_{n\to\infty} |\gamma_n| \le \varepsilon \left(\sum_{j=0}^{N} |a_j|\right) + 2M \left(\sum_{j>N} |a_j|\right) \le \varepsilon \left(2M + \sum_{j=0}^{\infty} |a_j|\right).$$

Since $\varepsilon > 0$ is arbitrary, $\lim_{n\to\infty} \gamma_n = 0$. Therefore,

$$\sum_{k=0}^{n} c_k = \gamma_n + A_n B \xrightarrow{n\to\infty} AB.$$

$\square$

0.7 Rearrangements of series

Definition

Consider a bijection $\varphi : \mathbb{N} \to \mathbb{N}$. For $n \in \mathbb{N}$ define $\tilde{a}_n = a_{\varphi(n)}$. The series $\sum \tilde{a}_n$ is a **rearrangement** of $\sum a_n$.

Example

$$\varphi(n) = \begin{cases} n+1, & n \text{ odd}, \\ n-1, & n \text{ even}, \end{cases} \quad \text{maps } 1, 2, 3, 4, 5, \ldots \text{ to } 2, 1, 4, 3, 6, 5, \ldots.$$

Before we talk about our main theorem in this section, we first start a claim:

Claim

Absolute Convergence can control "*Tails*"

By definition, absolute convergence means $\sum |a_n|$ converges. For any $\varepsilon > 0$, the Cauchy criterion for convergent series guarantees there exists $N \in \mathbb{N}$ such that:$\sum_{k=N+1}^{\infty} |a_k| < \varepsilon$.

This says the "tail" of $\sum |a_n|$ (terms after N) is arbitrarily small.

Theorem

Suppose $\sum a_n$ *converges absolutely* and let $A = \sum a_n$. Then for any rearrangement $\sum \tilde{a}_n$, it converges to A.

证明. Given $\varepsilon > 0$, choose N such that $\sum_{k=N+1}^{\infty} |a_k| < \varepsilon$. Because φ is bijective, for this N, there exists n large with $\{\varphi(1), \ldots, \varphi(n)\} \supset \{1, \ldots, N\}$. In fact, take $n = \max\{\varphi^{-1}(1), \varphi^{-1}(2), \ldots, \varphi^{-1}(N)\} + 1$. With this n,

$$\tilde{a}_1 + \tilde{a}_2 + \cdots + \tilde{a}_n - (a_1 + a_2 + \cdots + a_n)$$

cancels all terms before N. Thus:

$$\left|\sum_{k=1}^{n} \tilde{a}_k - \sum_{k=1}^{n} a_k\right| \leq \sum_{k=N+1}^{\infty} |a_k| < \varepsilon.$$

By the triangle inequality:

$$\left|\sum_{k=1}^{n} \tilde{a}_k - A\right| \leq \left|A - \sum_{k=1}^{n} a_k\right| + \left|\sum_{k=1}^{n} \tilde{a}_k - \sum_{k=1}^{n} a_k\right| < 2\varepsilon.$$

Hence $\sum_{k=1}^{n} \tilde{a}_n \to A$. □

【Remark】:

If a series converges but not absolutely,then its rearrangements may give any numbers as the limit!

0.8 Finite and Countable sets

Definition

Finite sets: Given $n \in \mathbb{N}$, let $J_n = \{1, 2, \ldots, n\}$.
Consider a nonempty set X. Then X is said to be **finite** if there exists $n \in \mathbb{N}$ and a bijection $f : J_n \to X$.

By convention, $\varnothing$ is finite. If X is not finite, it is said to be infinite.

Definition

Countable sets: Consider a nonempty set X. X is said to be countable if either X is finite, or if there exists a bijection $f : \mathbb{N} \to X$.

Alternatively, X is countable if there exists a sequence (u_n) such that $X = \{u_1, u_2, \ldots\}$ (all elements of X can be listed).

Example

1. $\mathbb{N}$ is infinite and countable.

2. $\mathbb{Z}$ is infinite and countable. One enumeration is given by

$$f(n) = \begin{cases} \frac{n-1}{2}, & n \text{ odd}, \\ -\frac{n}{2}, & n \text{ even}, \end{cases}$$

   and $\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$.

3. $\mathbb{Q}$ is countable. One listing is

$$\mathbb{Q} = \left\{0, \frac{1}{1}, -\frac{1}{1}, \frac{1}{2}, -\frac{1}{2}, \frac{2}{1}, -\frac{2}{1}, \frac{1}{3}, -\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}, \frac{3}{1}, -\frac{3}{1}, \ldots\right\}.$$

4. The set $\mathbb{R}$ (real numbers) is uncountable.

and we have these 2 properties:

Property

- **Proposition 1.** If A is infinite and $A \subset X$, then X is infinite.
- **Proposition 2.** If X is countable and $A \subset X$, then A is countable.

Definition

Let X be a nonempty set. Suppose for each $n \in \mathbb{N}$ we have a subset $E_n \subset X$. Denote this family by $(E_n)_{n\in\mathbb{N}}$. Define

$$\bigcap_{n=1}^{\infty} E_n = \{x \in X : \forall n \in \mathbb{N},\ x \in E_n\}, \qquad \bigcup_{n=1}^{\infty} E_n = \{x \in X : \exists n \in \mathbb{N},\ x \in E_n\}.$$

Lemma

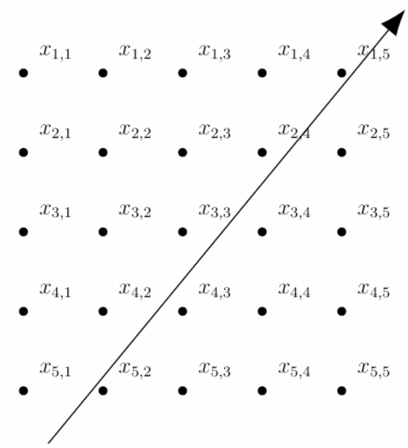
If for all n, E_n is countable, then $\bigcup_{n=1}^{\infty} E_n$ is countable.

证明. proof:

Traverse the Array Along Diagonals

To list all elements of $\bigcup_{n=1}^{\infty} E_n$ in a single sequence, traverse the array along diagonals defined by $n + k =$ constant (where n = row index, k = column index):

Diagonal 1 $(n + k = 2)$: Only $(n, k) = (1, 1)$, so element $x_{1,1}$.



Diagonal 2 $(n+k=3)$: $(n,k)=(1,2),(2,1)$, so elements $x_{1,2}, x_{2,1}$.
Diagonal 3 $(n+k=4)$: $(n,k)=(1,3),(2,2),(3,1)$, so elements $x_{1,3}, x_{2,2}, x_{3,1}$.
Diagonal m $(n+k=m+1)$: $(n,k)=(1,m),(2,m-1),\dots,(m,1)$, so m elements: $x_{1,m}, x_{2,m-1}, \dots$

□

The superposition of countably infinitely many "countable infinities" still results in a countable infinity.

Theorem

The set $\mathbb{R}$ (real numbers) is uncountable.

Hint: If $\mathbb{R}$ were countable, we could list all real numbers as a sequence:$\mathbb{R}=\{r_1,r_2,r_3,\dots\}=\{r_k\}_{k\geq 1}$.Our goal is to show this assumption leads to a contradiction.

$a_1 \quad a_2 \qquad a_3 \qquad\qquad b_3 \quad b_2 \quad b_1$

证明. Argue by contradiction. Suppose $\mathbb{R}=\{r_1,r_2,r_3,\dots\}=\{r_k\}_{k\geq 1}$. Define a sequence of (strictly) nested closed intervals as follows. Let

$$\begin{aligned}
a_1 &= r_1, \quad b_1 = r_{k_1}, && \text{where } k_1 = \min\{k\geq 1 : r_k > a_1\},\\
a_2 &= r_{m_2}, \quad \text{where } m_2 && = \min\{k\geq k_1 : r_k \in (a_1,b_1)\},\\
b_2 &= r_{k_2}, \quad \text{where } k_2 && = \min\{k\geq m_2 : r_k \in (a_2,b_1)\},
\end{aligned}$$

and in general, inductively for $n\geq 2$ define

$$\begin{aligned}
a_n &= r_{m_n}, \quad m_n = \min\{k\geq k_{n-1} : r_k \in (a_{n-1},b_{n-1})\},\\
b_n &= r_{k_n}, \quad k_n = \min\{k\geq m_n : r_k \in (a_n,b_{n-1})\}.
\end{aligned}$$

Then $\{[a_n,b_n]\}_{n\geq 1}$ is a nested family of closed intervals with $[a_{n+1},b_{n+1}]\subset(a_n,b_n)$ for all n.

By the nested interval theorem, there exists $r\in\bigcap_{n=1}^{\infty}[a_n,b_n]$. We claim $r\neq r_k$ for all k, a contradiction. Indeed, $r>a_1=r_1$, hence $r\neq r_1$. Then, $r\notin\{r_1,r_2,\dots,r_{k_1-1}\}$ (since

k_1 is the smallest index with $r_k > a_1$). Also, $r \notin \{r_{k_1}, r_{k_1+1}, \ldots, r_{m_2-1}\}$ (since m_2 is the smallest index with $r_k \in (a_1, b_1)$). Arguing inductively, we conclude r is not among any r_k. □

In short: Assume $\mathbb{R}$ is countable, build nested intervals from the list, find a real number in all intervals that' s missing from the list—hence $\mathbb{R}$ is uncountable.

We will discuss a brand new topic: Fundamentals of Topology

0.9 Metric Space

Instead of treating a set X merely as a collection of elements, we often equip X with additional structure.

Definition

A function $d : X \times X \to \mathbb{R}$ is a ***distance*** (or ***metric***) on X if

1. (*positive-definiteness*) $d(p, q) > 0$ for $p \neq q$ and $d(p, p) = 0$;
2. (*symmetry*) $d(p, q) = d(q, p)$;
3. (*triangle inequality*) $d(p, q) \leq d(p, r) + d(r, q)$ for all $p, q, r \in X$.

Example

1. If $X = \mathbb{R}$, then $d(x, y) = |x - y|$.
2. If $X = \mathbb{R}^k$, then $d(x, y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$ (Euclidean metric).
3. If $X = \mathbb{C}$, then $d(z_1, z_2) = |z_1 - z_2|$ where for $z = a + bi$, $|z| = \sqrt{a^2 + b^2}$.

Then, we will discuss 3 foundamental items in Topology: Interior, adherent, exterior points:

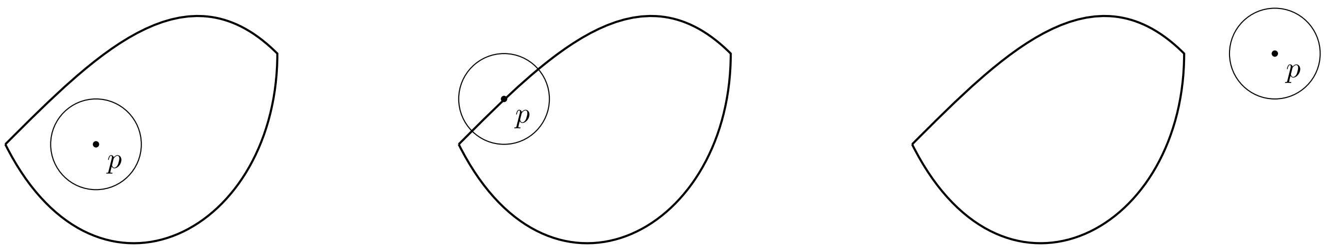
Let (X, d) be a metric space and $A \subset X$ nonempty.

Definition

1. A point p is called an **interior point** of A if there exists $r > 0$ such that **all** points q satisfying $d(p,q) < r$ belong to the set A.

2. A point p is an **adherent point** of A if $\forall \varepsilon > 0$, $\exists q \in A$ such that $d(p,q) < \varepsilon$.

3. A point p is called an **exterior point** of A if there exists $r > 0$ such that **all** points q satisfying $d(p,q) < r$ do not belong to the set A.



interior　　adherent　　exterior

【Remark】:

An **interior point** is a ***special case*** of an **adherent point**.

Notations.

$\text{int}(A) = \mathring{A} = \{p \in X : p \text{ is an interior point of } A\}$;

Closure: $\overline{A} = \{p \in X : p \text{ is an adherent point of } A\}$;

$\text{ext}(A) = \{p \in X : p \text{ is an exterior point of } A\}$.

And we can gain this property below:

Property

$$\mathring{A} \subset A \subset \overline{A}, \quad \text{ext}(A) = \text{int}(A^c).$$

For $p \in X$ and $r > 0$, we can define the **(open) neighborhood**.(we will prove it is open later.)

Definition

(Open) Neighborhood: $N_r(p) = \{q \in X : d(p,q) < r\}$.

For $p \in X$ and $A \subset X$,

$$p \in \mathring{A} \iff \exists r > 0,\ N_r(p) \subset A,$$
$$p \in \overline{A} \iff \forall r > 0,\ N_r(p) \cap A \neq \varnothing,$$
$$p \in \operatorname{ext}(A) \iff \exists r > 0,\ N_r(p) \subset A^c.(or(N_r(p) \cap A = \varnothing)$$

Definition

Let (X, d) be a metric space and nonempty $A \subset X$.

1. A is open if $\mathring{A} = A$.
2. A is closed if $\overline{A} = A$.

Mathematical Description:

A is **open** if $\mathring{A} = A$ means: Every point in A is an interior point, which is equivalent to: For *every* $x \in A$, there exists $\rho > 0$ such that the open ball $U_\rho(x) = \{y \in X \mid d(y, x) < \rho\}$ is entirely contained in A (the "**open ball**" definition).

$\overline{A} = A$ means: Every adherent point of A lies in A which means: a set $A \subset X$ is **closed** if for every $x \in X, \forall \rho > 0,\ U_\rho(x) \cap A \neq \varnothing$, then $x \in A$.

【Remark】:

A **subset of an open set** is **not** necessarily **open** in general.

Example

Let $U = (0, 2)$, which is open.

Take the subset $V = (0, 1] \subseteq U.V = (0, 1]$ is not open in the standard topology

Theorem

For $A \subset X$, A is closed $\iff$ A^c is open.

Hint: "If there exists $r > 0$ with $N_r(p) \subset A^c$",is equal to the proposition: "Existing an $\varepsilon > 0,\ N_\varepsilon(p) \cap A = \varnothing$".

证明. ($\Longrightarrow$: A closed $\Longrightarrow$ A^c open)

Let A be closed ($\overline{A} = A$). Take any $p \in A^c$.If every $r > 0$ had $N_r(p) \cap A \neq \emptyset$, then $p \in \overline{A} = A$ (contradiction, since $p \in A^c$).

Thus, there exists $r > 0$ with $N_r(p) \cap A = \emptyset$, so $N_r(p) \subseteq A^c$.Since $p \in A^c$ is arbitrary, A^c is open.

($\Longleftarrow$: A^c open $\Longrightarrow$ A closed)Let A^c be open.

We show $\overline{A} = A$:Take any $p \in \overline{A}$ (so every neighborhood of p intersects A).If $p \notin A$, then $p \in A^c$. Since A^c is open, $\exists r > 0 : N_r(p) \subseteq A^c$, so $N_r(p) \cap A = \emptyset$—contradicting $p \in \overline{A}$.

Thus $p \in A$, so $\overline{A} \subseteq A$. Since $A \subseteq \overline{A}$ always, $\overline{A} = A$, so A is closed. □

And we can also consider some useful properties of open and closed sets:

Property

1. For any family of open sets $\{G_\alpha\}_{\alpha \in J}$, $\bigcup_{\alpha \in J} G_\alpha$ is open.
2. For finitely many open sets G_i, $i = 1, 2, \cdots, N$, $\bigcap_{i=1}^{N} G_i$ is open.(Hint: You can prove it when n=2 and then use mathematical induction.)

证明. 1. Union of open sets is open:Take $x \in \bigcup_\alpha G_\alpha$: $x \in G_\alpha$ (some α), so G_α (open) gives a neighborhood of x in $G_\alpha \subseteq \bigcup_\alpha G_\alpha$. Thus $\bigcup_\alpha G_\alpha$ is open.

2.Base Case ($n = 2$)Let G_1, G_2 be open ($\mathring{G_1} = G_1, \mathring{G_2} = G_2$). Take $x \in G_1 \cap G_2$:$x \in G_1 \implies \exists r_1 > 0, N_{r_1}(x) \subseteq G_1$.$x \in G_2 \implies \exists r_2 > 0, N_{r_2}(x) \subseteq G_2$.Let $r = \min(r_1, r_2)$. Then $N_r(x) \subseteq G_1 \cap G_2$, so $x \in (G_1 \mathring{\cap} G_2)$. Since $\mathring{S} \subseteq S$ for any S, $(G_1 \mathring{\cap} G_2) = G_1 \cap G_2$—so $G_1 \cap G_2$ is open.

Inductive StepAssume: Intersection of k open sets is open (i.e., $\bigcap_{i=1}^{k} G_i$ is open, so $\mathring{\left(\bigcap_{i=1}^{k} G_i\right)} = \bigcap_{i=1}^{k} G_i$).For $k+1$ open sets:$\bigcap_{i=1}^{k+1} G_i = \left(\bigcap_{i=1}^{k} G_i\right) \cap G_{k+1}$.By induction, $\bigcap_{i=1}^{k} G_i$ is open; by the base case, the intersection of two opens is open. Thus $\bigcap_{i=1}^{k+1} G_i$ is open.By induction, finite intersections of open sets are open.

□

【Remark】:

Let $X = \mathbb{R}$, $d(x, y) = |x - y|$, consider $I_n = (-1/n, 1/n)$. I_n is open for any n, but $\bigcap_{n=1}^{\infty} I_n$ is not open. This shows that (b) is not true for ***infinite*** intersection.

0.10 Limit Isolated Boundary Points

Definition

Let (X, d) be a metric space, $p \in X$ and $A \subset X$.

1. p is a **limit point** of A if $\forall \varepsilon > 0$, $\exists q \in A \setminus \{p\}$ with $d(p, q) < \varepsilon$.
2. p is an **isolated point** of A if $p \in A$ and p is not a limit point of A (equivalently, $\exists r > 0$ s.t. $N_r(p) \cap (A \setminus \{p\}) = \varnothing$).
3. p is a **boundary point** of A if $p \in \overline{A}$ and $p \notin \mathring{A}$.

We can understand Boundary points in this way:

A boundary point is a point that neither completely belongs to A nor completely belongs to the exterior of A. **In any neighborhood** of such a point, there are both points of A and points not of A.

Example

1. **Limit point:** Let $X = \mathbb{R}$ (with the usual metric $d(x, y) = |x - y|$) and $A = (0, 1)$. The point $p = 0$ is a limit point of A: for any $\varepsilon > 0$, there exists $q = \varepsilon/2 \in A \setminus \{0\}$ with $d(0, q) < \varepsilon$.
2. **Isolated point:** Let $X = \mathbb{R}$ and $A = \{1\} \cup (2, 3)$. The point $p = 1$ is an isolated point of A: take $r = 0.5$, then $N_{0.5}(1) = (0.5, 1.5)$, and $N_{0.5}(1) \cap (A \setminus \{1\}) = \varnothing$, so $p = 1$ is not a limit point of A.
3. **Boundary point:** Let $X = \mathbb{R}$ and $A = [0, 1)$. The point $p = 0$ is a boundary point of A: $p \in \overline{A} = [0, 1]$, but $p \notin \mathring{A} = (0, 1)$.

And we Denote :

Definition

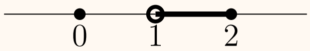
Boundary: $\partial A = \{\text{boundary points of } A\}$
Derived set: $A' = \{$ limit points of A $\}$.

And you can understand these items in one example:

Example

Let $X = \mathbb{R}$, $A = \{0\} \cup (1, 2]$. Then $0 \in A$ but $0 \notin A'$ (isolated). $1 \notin A$ but 1 is a limit point. $2 \in A$ and 2 is a limit point. Hence

$$\mathring{A} = (1, 2), \quad A' = [1, 2], \quad \overline{A} = \{0\} \cup [1, 2], \quad \partial A = \{0, 1, 2\}.$$



For $A \subset X$,we have:

A closure is A itself unions all points that are "infinitely close to A" (limit points),in short: $\overline{A} = A \cup A'$.(So the closure not only contains the limit points of A, but also all the points of isolated points.

So we yield: $A' \subseteq \overline{A}$

0.11 Boundedness and Convergence

Definition

Bounded set:A is said to be **bounded** if $\exists M > 0$ and $p \in X$ such that $d(p, q) \leq M$ for every $q \in A$. Equivalently, $A \subset U_M(p)$

So if the set A can **be covered by a ball** with **finite radius**,then we say A is **bounded**.

Claim

Claim1: The union of bounded sets is bounded.

Claim2: Trivially, any subset of a bounded set is bounded.

Hint: for claim1, only prove the two-set case is ok.

证明. Easy to prove

□

Then, let's talk about the ***dense set***:

Definition

Let (X, d) be a metric space and $A \subset X$. We say that A is a **dense subset** of X if **the closure of A equals X,** i.e., $\overline{A} = X$.

【Remark】:

Denseness depends on both the **ambient space** X and the **metric** d defined on it.

Example

The set of rational numbers $\mathbb{Q}$ is dense in $\mathbb{R}$ with the usual metric.

Easy to gain:

Example

The set of irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$ with the usual metric.

Sequential convergence in X

Recall for a real sequence: $u_n \to u^*$ iff $\forall \varepsilon > 0$, $\exists N$ s.t. $\forall n \geq N$, $|u_n - u^*| < \varepsilon$. The concept generalizes to metric spaces by replacing absolute value with distance.

Definition

Convergence A sequence $(p_n) \subset X$ is said to converge to $p \in X$ *iff* $d(p_n, p) \to 0$; *i.e.* $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n \geq N$, $d(p_n, p) < \varepsilon$.

And we need to prove an important property:

Theorem

Limit characterization of closure: $p \in \overline{A} \iff$ there exists a sequence $(p_n) \subset A$ such that $p_n \to p$ in X.

证明. ($\Longrightarrow$) By definition of closure/adherent point: $\forall \varepsilon > 0, U_\varepsilon(p) \cap A \neq \emptyset$. For each $n \in \mathbb{N}$, pick $u_n \in U_{1/n}(p) \cap A$ and set $p_n = u_n$. Then $d(p_n, p) \leq 1/n \to 0$.

($\Longleftarrow$) Suppose $p_n \to p$ with $p_n \in A$. For any $\varepsilon > 0$, choose n large so that $d(p_n, p) < \varepsilon$. Then $p_n \in U_\varepsilon(p) \cap A \neq \emptyset$, hence $p \in \overline{A}$. □

Corollary & Secondary Conclusion

A is dense in $X \iff$ for every $x \in X$ there exists a sequence $(p_n) \subset A$ with $p_n \to x$.

And ultimately in this section, we will consider 2 types of "Convergence":

Euclidean space $\mathbb{R}^k$. Write $x = (x_1, \ldots, x_k)$ and $d(x, y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$. For a sequence $x_n = (x_{n,1}, \dots, x_{n,k})$,

$$x_n \to x \iff (\forall i) \lim_{n\to\infty} x_{n,i} = x_i.$$

The convergence can also be characterized by the metric

$$d_\infty(x,y) := \max_{1\le i\le k} |x_i - y_i|,$$

which satisfies $d_\infty(x,y) \le d(x,y) \le \sqrt{k} d_\infty(x,y)$ for all $x, y \in \mathbb{R}^k$.

The inequalities $d_\infty \le d \le \sqrt{k}\, d_\infty$ imply:

If $x_n \to x$ in the Euclidean metric ($d(x_n,x) \to 0$), then $d_\infty(x_n,x) \le d(x_n,x) \to 0$, so $x_n \to x$ in d_∞.

If $x_n \to x$ in d_∞ ($d_\infty(x_n,x) \to 0$), then $d(x_n,x) \le \sqrt{k}\, d_\infty(x_n,x) \to 0$, so $x_n \to x$ in the Euclidean metric.

Thus, a sequence converges in d if and only if it converges in d_∞ (to the same limit). So the two metrics "describe convergence in the same way" for $\mathbb{R}^k$.

0.12 Compactness

Definition

A collection of open sets $\{O_\alpha\}_{\alpha\in I}$ is an **open cover** of a set K if: $K \subset \bigcup_{\alpha\in I} O_\alpha$.

In other words, every point $p \in K$ lies in at least one open set O_α from the collection.

From the concept of Open Covering, we can have:

Definition

Compact set: Let (X,d) be a metric space and $A \subset X$. A is compact **if every open cover** of A has a **finite** subcover;

i.e., Whenever $\{E_\alpha\}_{\alpha\in J}$ is a family of open sets with $A \subset \bigcup_{\alpha\in J} E_\alpha$, there exist $\alpha_1, \dots, \alpha_N$ such that $A \subset \bigcup_{i=1}^N E_{\alpha_i}$.

Fact and proposition

Why compactness matters: it provides a way to *find limits* in a finite way.

Compactness turns many problems about **infinite coverings into problems about finite coverings**.

【Remark】:

Compactness **doesn't just mean exsiting a finite open cover** for set A.

Example

Example: Finite Set in $\mathbb{R}$. Let $X = \mathbb{R}$ (with the usual distance $|x-y|$) and $A = \{1, 2, 3\}$.

Why compact?

If the open cover is very "large" (for example, a single open set $(-1, 10)$ can cover A), then its finite subcover is itself, which certainly holds.

If the open cover is very "fine" (for example, covering A with three open sets $U_1 = (0.5, 1.5)$ (covering 1), $U_2 = (1.5, 2.5)$ (covering 2), $U_3 = (2.5, 3.5)$ (covering 3)), since A has only 3 points, no matter how the open cover is constructed, it is sufficient to select "the open set corresponding to each point" from the cover (at most 3), which can cover the entire A.

Therefore, any open cover can find a **finite** subcover, so A is compact.

Example

Counter-Example: Open Interval $(0, 1)$ in $\mathbb{R}$. Let $X = \mathbb{R}$ (with the usual metric $d(x, y) = |x - y|$) and $A = (0, 1)$ (the open interval from 0 to 1).

To show $(0, 1)$ is not compact, we can construct an open cover of • $(0, 1)$ • **with no finite subcover**. While some open covers of $(0, 1)$ (like $(-1, 0.5) \cup (0.4, 2)$) do have finite subcovers, not all open covers do.

Consider open sets $U_n = \left(\frac{1}{n}, 1 - \frac{1}{n}\right)$ for all integers $n \geq 3$. The family $\{U_n \mid n \geq 3\}$ covers $(0, 1)$ because for any $x \in (0, 1)$, there exists some $n \geq 3$ such that $\frac{1}{n} < x < 1 - \frac{1}{n}$.

However, no finite subcollection of $\{U_n \mid n \geq 3\}$ can cover $(0, 1)$. Suppose we pick finitely many U_n, say $U_{n_1}, \ldots, U_{n_k}$, and let $N = \max\{n_1, \ldots, n_k\}$. Their union is $\left(\frac{1}{N}, 1 - \frac{1}{N}\right)$, which misses points like $\frac{1}{2N}$ (since $0 < \frac{1}{2N} < \frac{1}{N}$, so $\frac{1}{2N} \notin \left(\frac{1}{N}, 1 - \frac{1}{N}\right)$).

Thus, this open cover has no finite subcover, so $(0, 1)$ is not compact.

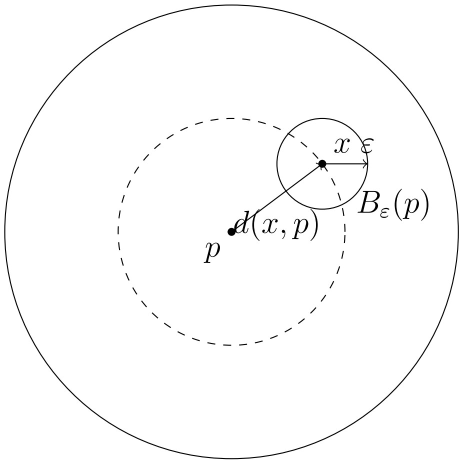
Before this section's main theorem, we first prove a useful lemma.

Lemma

Open balls are open: Let (X, d) be a metric space. For any $p \in X$ and $r > 0$, the open ball

$$U_r(p) = \{x \in X : d(x, p) < r\}$$

is an ***open*** set in (X, d).



证明. Fix $x \in U_r(p)$, so $d(x,p) < r$. Set $\rho := \frac{r-d(x,p)}{2} > 0$. We claim $U_\rho(x) \subset U_r(p)$.

Indeed, if $y \in U_\rho(x)$, then by the triangle inequality

$$d(y,p) \leq d(y,x) + d(x,p) < \rho + d(x,p) = \frac{r-d(x,p)}{2} + d(x,p) < r.$$

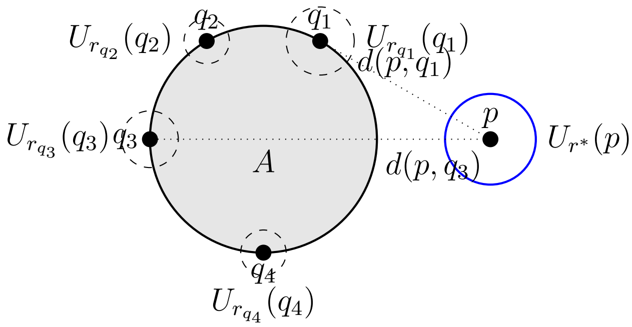
Thus $y \in U_r(p)$. Since such a $\rho > 0$ exists for every $x \in U_r(p)$, the ball $U_r(p)$ is open.

□ □

Theorem

A **compact** set is **bounded** and **closed**.

Hint: For closed part, it is sufficient to show that its complement A^c is an open set.



And we noticed that the only variable is the distance between q(q belongs to A) and p(p belongs to A^c) ; and what we need to deal with is to :

1.find a suitable radius r s.t. fulfills the condition;

2.Constructing a suitable open Ball-cover to help us find a good r.

证明. We prove both properties (boundedness, closedness) of A:

Step 1: A is bounded. To show A is bounded, fix any $\varepsilon > 0$. Note that for each $p \in A$, the open ball $U_\varepsilon(p) = \{x \in X \mid d(x,p) < \varepsilon\}$ contains p, so:$A \subseteq \bigcup_{p\in A} U_\varepsilon(p)$.

This is an open cover of A. By compactness of A, there exists a finite subcover: there exist $p_1, p_2, \ldots, p_N \in A$ such that$A \subseteq \bigcup_{i=1}^N U_\varepsilon(p_i)$.Each $U_\varepsilon(p_i)$ is bounded (it is a ball of radius ε), and a finite union of bounded sets is bounded. Thus A is bounded.

Step 2: A is closed. To show A is closed, we prove its complement $A^c = X \setminus A$ is open (i.e., for every $p \in A^c$, there exists an open ball around p contained in A^c).Let $p \in A^c$.

For each $q \in A$, define $r_q = \frac{1}{2}d(p,q)$ (so $d(p,q) = 2r_q$, meaning $p \notin U_{r_q}(q)$). The collection$\{U_{r_q}(q) \mid q \in A\}$is an open cover of A (since every $q \in A$ is in $U_{r_q}(q)$). By compactness of A, there exists a finite subcover: there exist $q_1, q_2, \ldots, q_N \in A$ such that$A \subseteq \bigcup_{i=1}^{N} U_{r_{q_i}}(q_i)$.Let $r^* = \min\{r_{q_1}, r_{q_2}, \ldots, r_{q_N}\}$, and consider the open ball $U_{r^*}(p) = \{x \in X \mid d(x,p) < r^*\}$.

For any $x \in U_{r^*}(p)$, and for any $i = 1, \ldots, N$, the triangle inequality gives:$d(x,q_i) \geq d(p,q_i) - d(x,p) > 2r_{q_i} - r^* \geq 2r_{q_i} - r_{q_i} = r_{q_i}$.Thus $x \notin U_{r_{q_i}}(q_i)$ for all i. Since $A \subseteq \bigcup_{i=1}^{N} U_{r_{q_i}}(q_i)$, this implies $x \notin A$ (so $x \in A^c$). Therefore:$U_{r^*}(p) \subseteq A^c$.Since $p \in A^c$ was arbitrary, A^c is open —so A is closed.

Combining Step 1 and Step 2, A (compact) is bounded and closed. □

And for the subset of a compact set, we have:

Theorem

If $A \subset X$ is *compact* and $K \subset A$ is *closed*, then K is compact.

Hint : we notice that K^c is open and $K \cap K^c = \varnothing$. And what we need to do is How to use the condition of A's compactness?

证明. Let $K \subset \bigcup_{\alpha \in J} E_\alpha$ be an open cover.

Since K is closed, K^c is open and $A \subset \left(\bigcup_{\alpha \in J} E_\alpha\right) \cup K^c$ is an open cover of A. By compactness of A, choose $\alpha_1, \ldots, \alpha_N$ so that $A \subset \left(\bigcup_{i=1}^{N} E_{\alpha_i}\right) \cup K^c$.

Cause K is a subset of A,intersect with K to obtain a finite subcover of K: that is $K \subset \left(\bigcup_{i=1}^{N} E_{\alpha_i}\right) \cup K^c$.

Since the intersection of K and K^C is empty, so $K \subset \left(\bigcup_{i=1}^{N} E_{\alpha_i}\right)$. □

And we will introduce our main theorem:

Theorem

Heine–Borel in $\mathbb{R}$: For $X = \mathbb{R}$ with $d(x,y) = |x-y|$, a set $A \subset \mathbb{R}$ is compact $\iff$ it is bounded and closed.

证明. proof:

Let $\{G_\alpha\}$ be an open cover of $[a,b]$. Suppose (for contradiction) $[a,b]$ has no finite subcover. Define a nested sequence of closed intervals:

Set $I_0 = [a,b]$.For each n, split I_n into two equal closed subintervals;

let I_{n+1} be the subinterval that cannot be covered by finitely many G_α (such a subinterval exists—else both halves have finite subcovers, so I_n does too, a contradiction).

By construction:$I_0 \supset I_1 \supset I_2 \supset \cdots$,Length $|I_n| = \frac{b-a}{2^n} \to 0$ as $n \to \infty$.By the Nested Interval Theorem, there is a unique $x \in \bigcap_{n=0}^{\infty} I_n$. Since $\{G_\alpha\}$ covers $[a, b]$, $x \in G_{\alpha_0}$ for some α_0; since G_{α_0} is open, there exists $r_0 > 0$ with $(x - r_0, x + r_0) \subset G_{\alpha_0}$.

For large n, $|I_n| < r_0$. Since $x \in I_n$, this implies $I_n \subset (x - r_0, x + r_0) \subset G_{\alpha_0}$—contradicting I_n "having no finite subcover" (it is covered by one set G_{α_0}).Thus $[a, b]$ has a finite subcover (so $[a, b]$ is compact).

For any bounded closed $A \subset \mathbb{R}$: A is bounded (so $A \subset [\inf A, \sup A]$) and closed (so A is a closed subset of the compact interval $[\inf A, \sup A]$). Closed subsets of compact sets are compact, so A is compact. □

And we can extend the theorem to:

Corollary & Secondary Conclusion

For $X = \mathbb{R}^k$, and $d(x, y)$ being the Euclidean distance, a set $A \subset X$ is compact $\iff$ A is bounded and closed.

【Remark】:

We must notice that:Heine-Borel is a special property of $\mathbb{R}^k$ (and similar spaces).
In metric spaces, compact sets are always closed and bounded, but the **converse fails**: a closed and bounded set in a metric space is not necessarily compact.
In general metric spaces (or even **subspaces** of $\mathbb{R}$), "closed + bounded" does not imply compactness.

Compactness depends on **the global structure of the ambient space**, while "closed/bounded in a subspace" are local properties.

we can give a counter-example:

Example

The open interval $(0, 1)$ as a subspace of $\mathbb{R}$
Consider $X_0 = (0, 1)$ with the standard metric $d(x, y) = |x - y|$.
And we consider the subset: $K = (0, 1) \subseteq X_0$

The set $K = (0, 1) \subseteq X_0$ is closed in X_0: Its complement in X_0 is $\varnothing$, which is open in any topology.

And the set $K = (0, 1) \subseteq X_0$ is Bounded: All points lie within distance 1 of each other.

But $K = (0,1)$ is not compact: the open cover $\left\{\left(\frac{1}{n}, 1\right) \mid n \geq 2\right\}$ has no finite subcover

And we will discuss about some facts about the compact sets:

Property

For **finitely** many compact sets K_i, $i = 1, 2, \cdots, N$, $\bigcup_{i=1}^N K_i$ is compact.

证明. Let $K = \bigcup_{i=1}^N K_i$, where each K_i is compact. Let $\mathcal{U} = \{U_\alpha\}_{\alpha \in J}$ be an **open cover** of K.

1. For each K_i: Since $K_i \subseteq K \subseteq \bigcup \mathcal{U}$, $\mathcal{U}$ is an open cover of K_i. By compactness of K_i, there exists a **finite subcover** $\mathcal{U}_i \subseteq \mathcal{U}$ such that $K_i \subseteq \bigcup \mathcal{U}_i$.

2. For the union K: Define $\mathcal{U}' = \bigcup_{i=1}^N \mathcal{U}_i$. Since $\mathcal{U}'$ is a **finite union of finite sets**, $\mathcal{U}'$ is finite. Moreover:

$$K = \bigcup_{i=1}^N K_i \subseteq \bigcup_{i=1}^N \left(\bigcup \mathcal{U}_i\right) = \bigcup \mathcal{U}'.$$

Thus $\mathcal{U}'$ is a finite subcover of $\mathcal{U}$ for K. By definition, $K = \bigcup_{i=1}^N K_i$ is compact.

□

On $X = \mathbb{R}$, $d(x,y) = |x - y|$, consider $I_n = [0, 1 - 1/n]$. we can show that I_n is compact for any n but $\bigcup_{n=1}^\infty I_n$ is not compact. This shows the conclusion above is not true for **infinite** union.

Property

For any family of compact set $\{K_\alpha\}_{\alpha \in J}$, $\bigcap_{\alpha \in J} K_\alpha$ is compact.

(Hint: show that the intersection is a closed subset of some compact set.)

证明. proof: □

And finally we will check two questions in subspace topology:

Exercise

(a) If $K \subset X_0$, show that K is compact in (X_0, d) if and only if it is compact in (X, d).

(b) Suppose X_0 is itself closed in X. Show that for any compact set $K \subset X$, $K \cap X_0$ is compact in X_0.

证明. Part(a):

To show $K \subset X_0$ is compact in (X_0, d) iff compact in (X, d):

($\Rightarrow$) Let K be compact in X_0. For an open cover $\{U_\alpha\}$ of K in X, $\{U_\alpha \cap X_0\}$ covers K in X_0. By compactness in X_0, extract a finite subcover $\{U_{\alpha_1} \cap X_0, \ldots, U_{\alpha_n} \cap X_0\}$; then $\{U_{\alpha_1}, \ldots, U_{\alpha_n}\}$ covers K in X, so K is compact in X.

($\Leftarrow$) Let K be compact in X. For an open cover $\{V_\alpha\}$ of K in X_0 ($V_\alpha = U_\alpha \cap X_0$, U_α open in X), $\{U_\alpha\}$ covers K in X. By compactness in X, extract a finite subcover $\{U_{\alpha_1}, \ldots, U_{\alpha_n}\}$; then $\{V_{\alpha_1}, \ldots, V_{\alpha_n}\}$ covers K in X_0, so K is compact in X_0.

Part(b):

In metric space X, compact sets are closed. Thus:

- K (compact in X) is closed in X.
- X_0 is closed in X, so $K \cap X_0$ (intersection of closed sets) is closed in X.

A closed subset of a compact set (K) is compact in X. By part (a), $K \cap X_0$ (compact in X) is compact in X_0. □

0.13 Sequentially Compactness

Definition

Sequentially Compact: Let (X, d) be a metric space. A set $K \subset X$ is said to be *sequentially compact* if every sequence $\{u_n\}_{n\geq 1} \subset K$ admits a subsequence $\{u_{n_k}\}_{k\geq 1}$ converging to some $u \in K$.

Then we pick an **arbitrary** sequence $\{u_n\}_{n\geq 1} \subset K$ and denote the (countable) set as:

$$A := \{u_n : n \in \mathbb{N}\} \subset K.$$

Easy to notice that:

【Remark】:

The set A is a **countable** set (since it's indexed by natural numbers $\mathbb{N}$).
And set A is also a **subset** of K: $A \subseteq K$.

Our target is to show : Compact $\iff$ sequentially compact, but we first depart this main theorem into two parts(two lemmas below):

Lemma

(Compact $\implies$ sequentially compact) Let $K \subset X$ be compact. Then K is sequentially compact.
(Sequentially compact $\implies$ compact): Let $K \subset X$ be sequentially compact. Then K is compact.

We first start with the Compact $\implies$ sequentially compact side, and we will start with two claims:

Claim

If a set A is contained in a union $\bigcup S_i$, then $A \cap \bigcup S_i = A$.

easy to prove it.

Claim

Set A has a limit point $u \in K$.

Otherwise, for every $p \in K$ there exists $r_p > 0$ such that

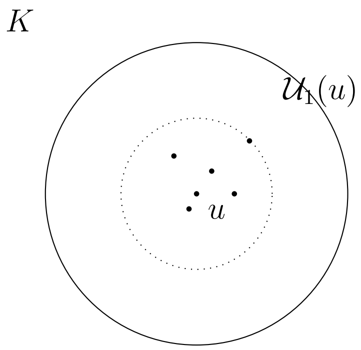
$$U_{r_p}(p) \cap A = \begin{cases} \{p\}, & p \in A, \\ \varnothing, & p \notin A. \end{cases}$$

Using the claims above, we can yield this important lemma:

Lemma

(Compact $\implies$ sequentially compact) Let $K \subset X$ be compact. Then K is sequentially compact.

To prove this, we construct a convergent subsequence for an arbitrary sequence $\{u_n\}_{n\geq 1} \subset K$ using induction:



- *Base Case:* Since u is a limit point of set A, the open ball $U_1(u)$ (centered at u with radius 1) must contain points from A. So we can pick an index n_1 such that the term u_{n_1} satisfies $d(u_{n_1}, u) < 1$.

- *Inductive Step:* Suppose we've already chosen an index n_k. Again, because u is a limit point, the open ball $U_{1/(k+1)}(u)$ (centered at u with radius $\frac{1}{k+1}$) contains infinitely many points of A. So we can pick an index $n_{k+1} > n_k$ such that $d(u_{n_{k+1}}, u) < \frac{1}{k+1}$.

By repeating this process, we obtain a subsequence $\{u_{n_k}\}$. As $k \to \infty$, $\frac{1}{k+1} \to 0$, so $d(u_{n_k}, u) \to 0$, meaning $u_{n_k} \to u$. This shows K is sequentially compact.

IDEA: It's a familiar picture on: $(\mathbb{R}, |x-y|)$.By Bolzano–Weierstrass, any sequence in K has a convergent subsequence. By closedness, the limit stays in K.

And now we wanna prove the other side: **Sequentially compact $\implies$ compact:** Let $K \subset X$ be sequentially compact， then K should be compact.

To prove this we use two claims derived from sequential compactness.

Claim

Claim 1: Let $K \subset X$ be sequentially compact. For every $\varepsilon_0 > 0$, there exist **finitely** many points $x_1, \ldots, x_N \in K$ such that

$$K \subset \bigcup_{i=1}^{N} U_{\varepsilon_0}(x_i).$$

证明. Suppose not. Then for some fixed $\varepsilon_0 > 0$ no finite union of ε_0-balls covers K. Choose $x_1 \in K$. Since $U_{\varepsilon_0}(x_1)$ does not cover K, choose $x_2 \in K \setminus U_{\varepsilon_0}(x_1)$. Inductively, having chosen $x_1, \ldots, x_{n-1}$, pick

$$x_n \in K \setminus \bigcup_{i=1}^{n-1} U_{\varepsilon_0}(x_i).$$

Then the sequence $\{x_n\}$ satisfies $d(x_m, x_n) \geq \varepsilon_0$ for all $m \neq n$. By sequential compactness, there is a convergent subsequence $x_{n_k} \to x \in K$. For large $k > j$,

$$d(x_{n_k}, x_{n_j}) \leq d(x_{n_k}, x) + d(x, x_{n_j}) < \frac{\varepsilon_0}{2} + \frac{\varepsilon_0}{2} = \varepsilon_0,$$

a contradiction. □

Claim

Claim 2: Let $K \subset X$ be sequentially compact and let $\{G_\alpha\}_{\alpha\in J}$ be an open cover of K ($K \subset \bigcup_{\alpha\in J} G_\alpha$).
Then there exists a **uniform** radius $\delta_0 > 0$ such that for every $x \in K$ there is an index $\alpha(x)$ with

$$U_{\delta_0}(x) \subset G_{\alpha(x)}.$$

Prove by contradiction: in other words, just prove: $\forall\delta > 0 \left(\exists x \in K\, \forall\alpha \in J\, \left(U_\delta(x) \not\subseteq G_\alpha\right)\right)$ is totally wrong.

Lemma

[Lebesgue Number Lemma] Let $\{G_\alpha\}_{\alpha\in J}$ be an open cover of K. Then there exists $\delta_0 > 0$ such that for every $x \in K$ there is an index $\alpha(x)$ with

$$U_{\delta_0}(x) \subset G_{\alpha(x)}.$$

证明. Suppose otherwise. Then for each $n \in \mathbb{N}$ there exists $x_n \in K$ such that

$$U_{1/n}(x_n) \not\subset G_\alpha \quad \text{for all } \alpha \in J.$$

By sequential compactness, pass to a subsequence $x_{n_k} \to x \in K \subset \bigcup_{\alpha\in J} G_\alpha$. Choose α_0 with $x \in G_{\alpha_0}$. By openness of G_{α_0}, there exists $r > 0$ with $U_r(x) \subset G_{\alpha_0}$.

Recall that $x_{n_k} \to x$. By definition of convergence in a metric space: for any $\epsilon > 0$, there exists an integer $N \in \mathbb{N}$ such that for all $k > N$,$d(x_{n_k}, x) < \epsilon$.Choose $\epsilon = r/2$. Then there exists some $K_1 \in \mathbb{N}$ such that for all $k > K_1$,$d(x_{n_k}, x) < \frac{r}{2}$. and $1/n_k < r/2$, hence

$$U_{1/n_k}(x_{n_k}) \subset U_r(x) \subset G_{\alpha_0},$$

contradiction. □

Example

Let $K = [0, 1]$. Take a proper open cover that includes $\frac{1}{4}$ and $\frac{3}{4}$, say:$\left\{\left(-\frac{1}{8}, \frac{3}{8}\right), \left(\frac{1}{8}, \frac{5}{8}\right), \left(\frac{3}{8}, \frac{9}{8}\right)\right\}$
Now, apply claim with $\delta_0 = \frac{1}{8}$.

For $x = \frac{1}{4}$: $U_{\frac{1}{8}}\left(\frac{1}{4}\right) = \left(\frac{1}{4} - \frac{1}{8}, \frac{1}{4} + \frac{1}{8}\right) = \left(\frac{1}{8}, \frac{3}{8}\right) \subset \left(-\frac{1}{8}, \frac{3}{8}\right)$ (first set in the cover);
For$x = \frac{3}{4}$:$U_{\frac{1}{8}}\left(\frac{3}{4}\right) = \left(\frac{3}{4} - \frac{1}{8}, \frac{3}{4} + \frac{1}{8}\right) = \left(\frac{5}{8}, \frac{7}{8}\right) \subset \left(\frac{1}{8}, \frac{5}{8}\right)$(second set in the cover).
So using these two claims, we can prove the lemma below:

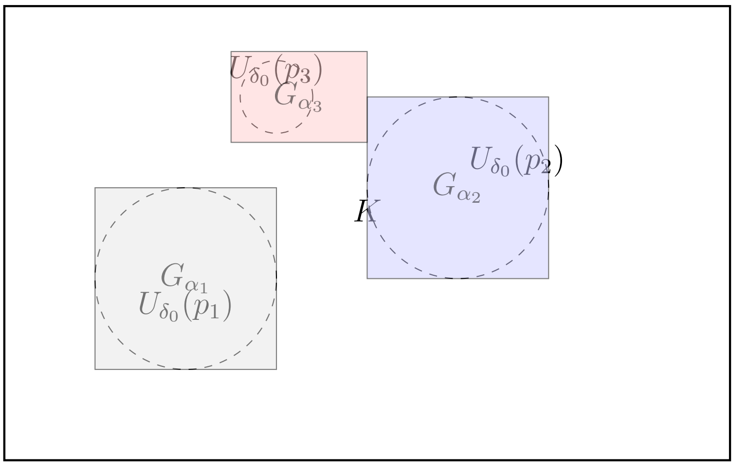
Lemma

(Sequentially compact $\implies$ compact): Let $K \subset X$ be sequentially compact. Then K is compact.

证明. proof:

$\square$

Finite covering of K by $\bigcup_{i=1}^{N} U_{\delta_0}(p_i)$ (Claim 1)



Open cover $\{G_\alpha\}_{\alpha \in J}$: $U_{\delta_0}(p_i) \subset G_{\alpha_i}$ by Claim 2

证明. Let $\{G_\alpha\}_{\alpha \in J}$ be any open cover of K.

By Lemma above (with $\varepsilon_0 = \delta_0$), pick points $p_1, \ldots, p_N \in K$ so that $K \subset \bigcup_{i=1}^{N} U_{\delta_0}(p_i)$.

Then choose $\delta_0 > 0$ with the stated property.For each i, choose α_i such that $U_{\delta_0}(p_i) \subset G_{\alpha_i}$. Then $\{G_{\alpha_i}\}_{i=1}^{N}$ is a finite subcover of K. $\square$

And we now conclude the core results in metric space:

Compact $\iff$ sequentially compact.

0.14 Connectedness & Subspace Topology

Definition

Connected metric space: A metric space (X, d) is **connected** if there **do not exist** two **non-empty** and **disjoint open** sets $E, F \subset X$ such that
$X = E \cup F$.

Otherwise X is **disconnected**.

Intuitive imagination: Think of a metric space as a "geometric figure". If it is a single whole without breakpoints, it is connected; if it can be "split into two separate open blocks", it is disconnected.

Example

- **Connected example:** The real line $\mathbb{R}$ (with the usual metric $|\cdot|$).

  Suppose we want to find two non-empty disjoint open sets E, F covering $\mathbb{R}$. For instance, let $E = (-\infty, a)$ and $F = (a, +\infty)$, but they **do not contain** a, so they cannot cover $\mathbb{R}$.

  If we forcefully put a into E, then $E = (-\infty, a]$, but $(-\infty, a]$ is not an open set (because the neighborhood of a is not entirely in E). Therefore, $\mathbb{R}$ is connected.

  In fact, you cannot find such E and F.

- **Disconnected example:** The space $X = (-\infty, 0) \cup (1, +\infty)$ (with the usual metric).

  Here, $E = (-\infty, 0)$ and $F = (1, +\infty)$ are both open sets in X (because in the "subspace topology" of X, we will discuss later). Moreover, $E \cap F = \emptyset$ and $E \cup F = X$, so X is disconnected.

If (X, d) is disconnected, then one may even find a **"disconnected" ball** $U_r(p)$ (its intersection with the two components splits), so a whole space being disconnected is unusual but it may happen.

Intuitive understanding: The entire space can be split into two open blocks, and naturally, the local area will also be "split" by these two open blocks.

For example, in the above example $X = (-\infty, 0) \cup (1, +\infty)$, take the ball $U_2(0) = (-2, 2)$. Its intersection with X is $U_X = (-2, 0) \cup (1, 2)$, which is obviously two "split" open sets.

And we consider another example:

Example

To determine whether the set $[-1,2] \cup [5,10]$ is connected.

Denote the set as $X = [-1,2] \cup [5,10]$, which is a subspace of the real number set $\mathbb{R}$ (inheriting the usual metric of $\mathbb{R}$).

And of course,this set is disconnected. However if you are in the perspective of the WHOLE space $\mathbb{R}$, it's literally hard to depart this set as two open sets.(e.g. E=[-1,2] and F=[5,10],but E and F are closed in the WHOLE space $\mathbb{R}$)

So, we need to introduce the concept of **Relatively open of subspace topology**.That is in the perspective of X itself, E and F are both "**open**".

And we call it relatively open.

Example

$X = \mathbb{R},\ A = [0,1),\ U = [0, \frac{1}{2})$:

In $\mathbb{R}$, $[0, \frac{1}{2})$ is not open Because $0 \in [0, 1/2)$, and any neighborhood of 0 in $\mathbb{R}$ (such as $(-\epsilon, \epsilon)$) contains negative numbers, which are not in $[0, 1/2)$. Therefore, 0 is not an interior point of $[0, 1/2)$, so $[0, 1/2)$ is not an open set in $\mathbb{R}$.

In A, $[0, \frac{1}{2})$ is open: $U^A_{\frac{1}{2}}(0) = \{q \in [0,1) \mid |0-q| < \frac{1}{2}\} = [0, \frac{1}{2})$, so U is open in A.

But it's too early to give the precise definition of relatively open right now, so we can first introduce an important theorem:

Theorem

If space A is a subspace of X. For any $U \subset A$ we have: U is **open *in*** A ***iff*** there exists open set V in X such that $U = A \cap V$.

【Remark】:

Here, the whole space is still X instead of V, and V is just a open set.

证明. We prove both directions of the equivalence:

Direction 1: ($\Rightarrow$) If U is open in A, then $U = A \cap V$ for some open $V \subseteq X$.

By definition of the subspace topology: if U is open in A, then for every $p \in U$, there exists $r_p > 0$ such that $B^A_{r_p}(p) \subseteq U$. Define

$$V = \bigcup_{p \in U} B_{r_p}(p).$$

V is open in X (unions of open balls are open in metric spaces). We verify $U = A \cap V$:

Step 1: $U \subseteq A \cap V$: For any $p \in U$, $p \in B_{r_p}(p) \subseteq V$ and $p \in A$, so $p \in A \cap V$.

Step 2: $A \cap V \subseteq U$: For any $q \in A \cap V$, there exists $p \in U$ with $q \in B_{r_p}(p)$. Since $q \in A$, $q \in A \cap B_{r_p}(p) = B^A_{r_p}(p) \subseteq U$.

Thus $U = A \cap V$.

Direction 2: (⇐) If $U = A \cap V$ for some open $V \subseteq X$, then U is open in A.

By openness of V in X: for any $p \in V$, there exists $r > 0$ such that $B_r(p) \subseteq V$. Take any $p \in U = A \cap V$: since $p \in V$, there exists $r > 0$ with $B_r(p) \subseteq V$. The subspace open ball

$$B^A_r(p) = A \cap B_r(p) \subseteq A \cap V = U.$$

By definition of the subspace topology, U is open in A.

Both directions hold, so the equivalence is proven. □

This proposition characterizes the **"relativity" of the subspace topology**:

An open set in the subspace A is essentially the **intersection of an open set in the original space X *and* A**;

conversely, the intersection of an **open set in the original space** *and* **A** must be an open set **within A**.

Example

In $(\mathbb{R}, |\cdot|)$, take $A = [0, 1]$. The set $[0, 1/2)$ is open in A, but not open in $X = \mathbb{R}$. Open in A: In the subspace topology, an "open set" requires that there exists an open set V in X such that $U = V \cap A$. For $[0, 1/2)$, take the open set $V = (-1, 1/2)$ in X, then $V \cap A = [0, 1/2)$, so $[0, 1/2)$ is an open set in A.

So we can gain the an essential way in the field of subspace topology to **Relatively Open Subset**.And it's time to get the definition of Relatively open set:

Definition

Relatively open subset: A subset $U \subset A$ is said to be **relatively open** if there exists an open set $V \subset X$ such that $U = A \cap V$.

And if A itself is open in X, we can gain more properties:

Corollary & Secondary Conclusion

If $A \subset X$ is open, then $E \subset A$ is relatively open in A if and only if E is open in X.

证明. E is relatively open in $A \iff E = A \cap V$ for some open $V \subset X$.

($\Rightarrow$) If $E = A \cap V$: A and V are open in X, so their intersection (a finite intersection of open sets) is open in X. Thus E is open in X.

($\Leftarrow$) If E is open in X: Since $E \subset A$, $E = A \cap E$ which implies E is relatively open in A.

$\square$

Above our discussion we can have the def. of the connected subset

Definition

Connected subsets: $A \subset X$ is said to be connected if (A, d) is connected as a metric space.

【Remark】:

The open sets we will discuss will be all relative open.

So that means a subset $A \subset X$ is connected *iff* it cannot be decomposed as the disjoint union of two non-empty sets that are relatively open.

we will show an example:

Example

- To determine whether the set $[-1, 2] \cup [5, 10]$ is connected.
- Analyzing the structure of $[-1, 2] \cup [5, 10]$
  - Denote the set as $X = [-1, 2] \cup [5, 10]$, which is a subspace of the real number set $\mathbb{R}$ (inheriting the usual metric of $\mathbb{R}$).
  - According to the definition of subspace topology: An open set in the subspace X is the intersection of an open set in $\mathbb{R}$ with X.
- Constructing a decomposition into disjoint open sets
  - Take:
    * $U = [-1, 2]$,
    * $V = [5, 10]$.
  - Verify that U and V are open sets in X:
    * For U: Take the open set $(-2, 3)$ in $\mathbb{R}$, then $(-2, 3) \cap X = [-1, 2] = U$, so U is an open set in X.
    * For V: Take the open set $(4, 11)$ in $\mathbb{R}$, then $(4, 11) \cap X = [5, 10] = V$, so V is an open set in X.
- Verifying the conditions for the decomposition
  - Both U and V are non-empty;
  - $U \cap V = \emptyset$ (because $[-1, 2]$ and $[5, 10]$ have no intersection);
  - $U \cup V = X$ (which is exactly the original set).

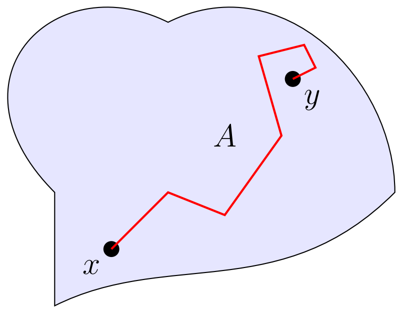
By a segment we mean $[x, y] = \{tx + (1-t)y : 0 \leq t \leq 1\}$.

A polygonal chain from x to y is a finite sequence of segments $[x_1, y_1], \ldots, [x_m, y_m]$ with

$$x_1 = x, \quad y_i = x_{i+1} \ (1 \leq i < m), \quad y_m = y.$$

Theorem

If $A \subset \mathbb{R}^k$ is **open and connected**, then for any $x, y \in A$ there exists a polygonal chain joining x and y whose **every segment lies in A.**



证明. Fix $x \in A$ and define

$$E := \{y \in A : x \text{ and } y \text{ can be joined by a polygonal chain lying in } A\}.$$

We claim both E and $F := A \setminus E$ are open in A.

- **E is open in A:** If $y \in E$, pick $r > 0$ such that $U_r(y) \subset A$ (since A is open, such an r exists). For any $y' \in U_r(y)$, the line segment $[y, y'] \subset U_r(y) \subset A$. Concatenate a polygonal chain from x to y (which exists by $y \in E$) with the segment $[y, y']$: this gives a polygonal chain from x to y' lying in A. Thus $y' \in E$, so $U_r(y) \subset E$. Hence E is open in A.

- **F is open in A:** If $y \in F$, choose $r > 0$ with $U_r(y) \subset A$. Suppose for contradiction there exists $y' \in U_r(y) \cap E$: since $y' \in E$, there is a polygonal chain from x to y' in A; and the segment $[y', y] \subset U_r(y) \subset A$. Concatenating these gives a polygonal chain from x to y in A, which contradicts $y \in F$. Thus $U_r(y) \cap E = \emptyset$, so $U_r(y) \subset F$. Hence F is open in A.

Since A is connected, the only subsets of A that are both open and closed (in A) are $\emptyset$ and A. Note $x \in E$, so $E \neq \emptyset$. Thus $F = A \setminus E = \emptyset$, which implies $E = A$. □

0.15 Continuity

Key Points in this section: Continuous functions (mappings) **preserve topological structure** (open/closed /compact/connected).

After equipping special structures on abstract sets (e.g. metric), it is natural to study mappings which are consistent with the structure.

And the concept of limit is the fundamental of the concept of continuity.

We start with $f : I \to \mathbb{R}$, where $I = [a, b]$, $[a, b)$, $(a, b]$ or (a, b). For now assume $-\infty < a < b < +\infty$; the endpoint cases $a = -\infty$ or $b = +\infty$ can be incorporated similarly. In all cases $I \subset \mathbb{R}$.

Definition

$f : I \to \mathbb{R}$ *has limit* ℓ at $p \in \overline{I}$ if

$$\forall \varepsilon > 0\, \exists \delta > 0 \text{ such that if } x \in I \text{ and } |x - p| < \delta, \quad \text{then } |f(x) - \ell| < \varepsilon.$$

We write $\lim_{x \to p} f(x) = \ell$.

【Remark】:

(1) No matter how close $f(x)$ is required to be to ℓ, as long as x and p are sufficiently close this can be achieved.
(2) Since p need not belong to the domain of f, it may happen that $\ell = \pm\infty$ (e.g. $f(x) = 1/x$ near $p = 0$).
(3) For $b = +\infty$ we define $\lim_{x \to \infty} f(x) = \ell$ (e.g. $\lim_{x \to \infty} 1/x = 0$).

Definition

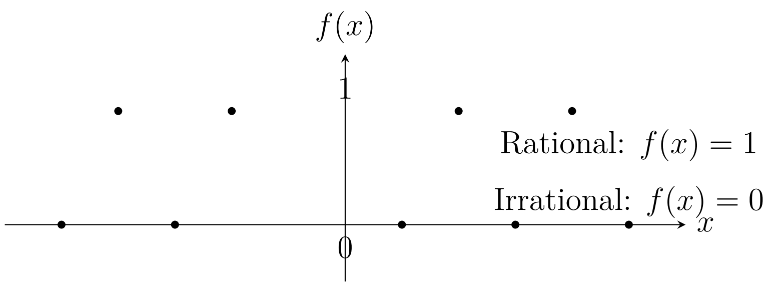
For $p \in \overline{I}$, f is *continuous at* p if $\lim_{x \to p} f(x) = f(p)$.
It is *continuous on* I if it is *continuous at every* $p \in I$.

Single-variable facts extend to general metric spaces. This includes $f : \mathbb{R}^k \to \mathbb{R}^m$ (vector-valued) and $f : \mathbb{C} \to \mathbb{C}$ (complex variable). For continuity alone we only need a metric.

Dirichlet function. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & x \notin \mathbb{Q}. \end{cases}$$

Then f is discontinuous at every $p \in \mathbb{R}$, and each discontinuity is of the second kind.



Now we will introduce the concept of one-side limit:
Let $I \subset \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$. Fix $p \in I$.

(Left/right limits at p). We say f has *left limit* ℓ at p (written $f(p-) = \ell$) if

$$\forall \varepsilon > 0 \ \exists \delta > 0 \text{ s.t. } 0 < p - x < \delta, \ x \in I \implies |f(x) - \ell| < \varepsilon.$$

We say f has *right limit* r at p (written $f(p+) = r$) if

$$\forall \varepsilon > 0 \ \exists \delta > 0 \text{ s.t. } 0 < x - p < \delta, \ x \in I \implies |f(x) - r| < \varepsilon.$$

Corollary & Secondary Conclusion

f is continuous at p if and only if both one–sided limits exist and

$$f(p-) = f(p+) = f(p).$$

explaination. ($\Rightarrow$) If f is continuous at p, then the ε–δ definition with $x \to p$ through $x < p$ (resp. $x > p$) gives $f(p-) = f(p)$ and $f(p+) = f(p)$.

($\Leftarrow$) If the one–sided limits exist and equal $f(p)$, then for any $\varepsilon > 0$ there are $\delta_-, \delta_+ > 0$ such that $|f(x) - f(p)| < \varepsilon$ whenever $p - \delta_- < x < p$ and whenever $p < x < p + \delta_+$. With $\delta = \min\{\delta_-, \delta_+\}$ we have $|f(x) - f(p)| < \varepsilon$ for all $|x - p| < \delta$, so f is continuous at p. $\square$

Definition

Discontinuities of first/second kind. Let $p \in I$ and suppose f is *not* continuous at p.

- If both $f(p-)$ and $f(p+)$ exist (finite), we say f has a *discontinuity of the first kind* (also called a *jump* or *simple discontinuity*) at p.
- Otherwise we say f has a *discontinuity of the second kind* at p.

0.16 Continuity in Metric Space

Then we will discuss the continuity in metric space:
Let (X, d_X) and (Y, d_Y) be metric spaces and $f : E \subset X \to Y$.

Definition

(1) **Limit:** For $p \in \overline{E}$, f has *limit ℓ at p* if

$$\forall \varepsilon > 0\, \exists \delta > 0 \text{ such that } x \in E,\ d_X(x,p) < \delta \Rightarrow d_Y\big(f(x), \ell\big) < \varepsilon.$$

(2) **Continuity:** For $p \in E$, f is *continuous at p* if $\lim_{x\to p} f(x) = f(p)$.
And f is *continuous on E* if it is continuous at every $p \in E$.

Theorem

Let $f : E \to Y$ and $p \in \overline{E}$.
(Sequential characterization of *limits*).
$\lim_{x\to p} f(x) = \ell$ *iff* for **every sequence** $(p_n) \subset E$ with $p_n \to p$ in X, we have $f(p_n) \to \ell$ in Y.
(Sequential characterization of *continuity*).
The mapping f is *continuous at p iff* for every sequence $(p_n) \subset E$ with $p_n \to p$, we have $f(p_n) \to f(p)$.

Example

(Vector-valued functions). Let $Y = \mathbb{R}^k$ with Euclidean distance $\|\cdot\|$. Writing $f(x) = (f_1(x), \ldots, f_k(x))$, the following are equivalent:
(1) $\lim_{x\to p} f(x) = \ell = (\ell_1, \ldots, \ell_k)$;
(2) for each $i = 1, \ldots, k$, $\lim_{x\to p} f_i(x) = \ell_i$.
$f = (f_1, \ldots, f_k)$ is continuous iff each component f_i is continuous.

Consider a continuous function $f : \mathbb{R}^k \to \mathbb{R}$. Fix a $x = (x_1, x_2, \cdots, x_k) \in \mathbb{R}^k$. For $i = 1, 2, \cdots, k$, define a function $f_i : \mathbb{R} \to \mathbb{R}$ by

$$f_i(t) = f(x_1, x_2, \cdots, x_{i-1}, t, x_{i+1}, \cdots, x_k).$$

Show that, if f is continuous at x, then f_i is continuous at x_i for each $i = 1, 2 \cdots, k$.

Let $f, g : E \to \mathbb{R}$ with $\lim_{x\to p} f(x) = A$ and $\lim_{x\to p} g(x) = B$. Then

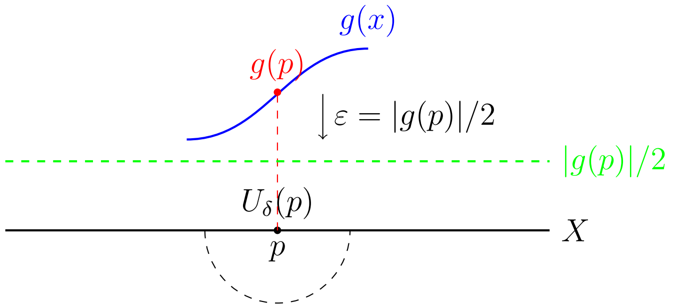
$$\lim_{x\to p}(f+g)(x) = A + B, \quad \lim_{x\to p}(fg)(x) = AB, \quad B \neq 0 \Rightarrow \lim_{x\to p} \frac{f}{g}(x) = \frac{A}{B}.$$

Claim If f, g are continuous at p, then $f+g$ and fg are continuous at p. If, in addition, $g \neq 0$ in a neighborhood of p, then f/g is continuous at p.

Corollary & Secondary Conclusion

If g is continuous at p and $g(p) \neq 0$, then $g \neq 0$ in some neighborhood of p.

Put $\varepsilon = |g(p)|/2 > 0$



证明. By continuity of g at p, for $\varepsilon = \frac{|g(p)|}{2} > 0$ (valid since $g(p) \neq 0$), there exists $\delta > 0$ such that:

$$x \in U_\delta(p) \implies |g(x) - g(p)| < \frac{|g(p)|}{2}.$$

By the reverse triangle inequality $\big||g(x)| - |g(p)|\big| \leq |g(x) - g(p)|$, we get:

$$|g(x)| \geq |g(p)| - |g(x) - g(p)| > |g(p)| - \frac{|g(p)|}{2} = \frac{|g(p)|}{2} > 0.$$

Thus $g(x) \neq 0$ for all $x \in U_\delta(p)$. □

0.17 Topological characterization of continuity

Theorem

$f : X \to Y$ is **continuous** *iff* for every open $G \subset Y$, the **preimage** $f^{-1}(G) \subset X$ is open.

证明. Let $p \in f^{-1}(V)$. Since V is open, choose $\varepsilon > 0$ with $U_\varepsilon(f(p)) \subset V$. By continuity, there exists $\delta > 0$ such that $d_X(x,p) < \delta$ implies $d_Y(f(x), f(p)) < \varepsilon$. Hence $U_\delta(p) \subset f^{-1}(U_\varepsilon(f(p))) \subset f^{-1}(V)$, so $f^{-1}(V)$ is open.

Fix $p \in X$ and $\varepsilon > 0$. The set $V = U_\varepsilon(f(p))$ is open in Y, so $f^{-1}(V)$ is open in X and $p \in f^{-1}(V)$. There exists $\delta > 0$ with $U_\delta(p) \subset f^{-1}(V)$. Thus $d_X(x,p) < \delta$ implies $f(x) \in V$, i.e. $d_Y(f(x), f(p)) < \varepsilon$. Hence f is continuous at p. □ □

Use $f^{-1}(G^c) = (f^{-1}(G))^c$ and the theorem above,we gain:

Corollary & Secondary Conclusion

$f : X \to Y$ is continuous *iff* for every closed $G \subset Y$, the preimage $f^{-1}(G) \subset X$ is closed.

【Remark】:

To clarify: A continuous function **does not** directly "keep" (preserve) open/closed sets in the forward direction(i.e., f(open set in X) is not necessarily open in Y, and f(closed set in X) is not necessarily closed in Y).

Instead, continuous functions preserve open/closed sets in the ***reverse direction*** (via preimages):

f is continuous $\iff$ preimages of open sets in Y are open in X.

f is continuous $\iff$ preimages of closed sets in Y are closed in X.

In short: Continuity guarantees that ***pulling back*** open/closed sets from the codomain Y to the domain X preserves openness/closedness —not that pushing forward sets from X to Y does.

Before we go to the relationship between compactness and continuous function, we will first mention three properties regarding the preimage:

For any function (mapping) $f : X \to Y$, and $U \subset X$, $V \subset Y$, define

$$f^{-1}(V) = \{x \in X : f(x) \in V\},$$

$$f(U) = \{y \in Y : y = f(x) \text{ for some } x \in U\}.$$

we have:

Property

(a) If $V \subset V' \subset Y$, then $f^{-1}(V) \subset f^{-1}(V')$; if $U \subset U' \subset X$, then $f(U) \subset f(U')$.

(b) For any $V \subset Y$, $f(f^{-1}(V)) \subset V$.

(c) If $V_\alpha \subset Y$ for each $\alpha \in J$, then

$$f^{-1}\left(\bigcup_{\alpha\in J} V_\alpha\right) = \bigcup_{\alpha\in J} f^{-1}(V_\alpha).$$

证明. (a)(b): easy to understand and prove;

(c) $f^{-1}\left(\bigcup_{\alpha\in J} V_\alpha\right) = \bigcup_{\alpha\in J} f^{-1}(V_\alpha)$ for $V_\alpha \subseteq Y$We prove set equality (two inclusions):

$f^{-1}\left(\bigcup_{\alpha\in J} V_\alpha\right) \subseteq \bigcup_{\alpha\in J} f^{-1}(V_\alpha)$Let $x \in f^{-1}\left(\bigcup_{\alpha\in J} V_\alpha\right)$. By preimage definition: $f(x) \in \bigcup_{\alpha\in J} V_\alpha$, so $f(x) \in V_\alpha$ for some $\alpha \in J$. Thus $x \in f^{-1}(V_\alpha)$ (preimage definition), so $x \in \bigcup_{\alpha\in J} f^{-1}(V_\alpha)$.$\bigcup_{\alpha\in J} f^{-1}(V_\alpha) \subseteq f^{-1}\left(\bigcup_{\alpha\in J} V_\alpha\right)$Let $x \in \bigcup_{\alpha\in J} f^{-1}(V_\alpha)$. Then $x \in f^{-1}(V_\alpha)$ for some $\alpha \in J$, so $f(x) \in V_\alpha$ (preimage definition). Since $V_\alpha \subseteq \bigcup_{\alpha\in J} V_\alpha$, we get $f(x) \in \bigcup_{\alpha\in J} V_\alpha$. By preimage definition, $x \in f^{-1}\left(\bigcup_{\alpha\in J} V_\alpha\right)$.

□

【Remark】:

For property (b), we should notice that $f(f^{-1}(V)) \subset V$ instead of $f(f^{-1}(V)) = V$.

Example

Consider the function $f : \mathbb{R} \to \mathbb{R}$, defined as $f(x) = x^2$ (not surjective), and take the set $V = [-1, +\infty)$.
Step 1: Find $f^{-1}(V)$ $f^{-1}(V) = \{x \in \mathbb{R} \mid x^2 \in [-1, +\infty)\}$. Since for all real numbers x, $x^2 \geq 0$, $x^2 \in [-1, +\infty)$ holds for all $x \in \mathbb{R}$. Therefore: $f^{-1}(V) = \mathbb{R}$.
Step 2: Find $f(f^{-1}(V))$ Substitute the elements in $\mathbb{R}$ into $f(x) = x^2$, and the resulting set is $[0, +\infty)$ (due to the non-negativity of square numbers).

WE also need to notice a common mistake:

【Remark】:

If $f : D \to I$ is continuous, then its inverse mapping $f^{-1} : I \to D$ may not be continuous.

Compactness and Continuous function

Theorem

If $K \subseteq X$ is compact and $f : K \to Y$ is continuous, then $f(K)$ is compact in Y.

证明. (open-cover proof) Let $\{V_\alpha\}_{\alpha\in J}$ be an open cover of $f(K)$ in Y; that is, $f(K) \subset \bigcup_{\alpha\in J} V_\alpha$ and each V_α is open in Y. Since f is continuous, $f^{-1}(V_\alpha)$ is open in K for each α, and

$$K \subseteq f^{-1}\left(\bigcup_{\alpha\in J} V_\alpha\right) = \bigcup_{\alpha\in J} f^{-1}(V_\alpha).$$

Thus $\{f^{-1}(V_\alpha)\}_{\alpha\in A}$ is an open cover of K. By compactness of K, there exist $\alpha_1, \ldots, \alpha_m \in A$ such that

$$K \subseteq \bigcup_{i=1}^{m} f^{-1}(V_{\alpha_i}).$$

Applying f gives

$$f(K) \subseteq \bigcup_{i=1}^{m} f\left(f^{-1}(V_{\alpha_i})\right) \subseteq \bigcup_{i=1}^{m} V_{\alpha_i},$$

so $\{V_{\alpha_i}\}_{i=1}^m$ is a finite subcover of $f(K)$. Hence $f(K)$ is compact. □

Definition

Bounded function: f is *bounded* on E if there exists $C > 0$ such that $|f(x)| \le C$ for all $x \in E$.

Equivalently, f is bounded on E iff the image set $f(E) \subset \mathbb{R}$ is bounded.

(1) (Bounded $\not\Rightarrow$ continuous) With $E = X = \mathbb{R}$,

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

is bounded but discontinuous.

(2) (Continuous $\not\Rightarrow$ bounded) On $E = (0, \infty)$, $f(x) = 1/x$ is continuous but unbounded.

Property

If E is **compact** and $f : E \to \mathbb{R}$ is **continuous**, then f is **bounded** on E.

So we have:

Theorem

Extreme value theorem: If E is **compact** and $f : E \to \mathbb{R}$ is **continuous**, then there exist $p, q \in E$ such that $f(p) = max$ and $f(q) = min$.

Example

Take $E = (0, 1)$ (**not compact**) and $f(x) = \frac{1}{x}$. f is continuous on $(0, 1)$, but f is unbounded above (approaches $+\infty$ as $x \to 0^+$) and has no maximum value.

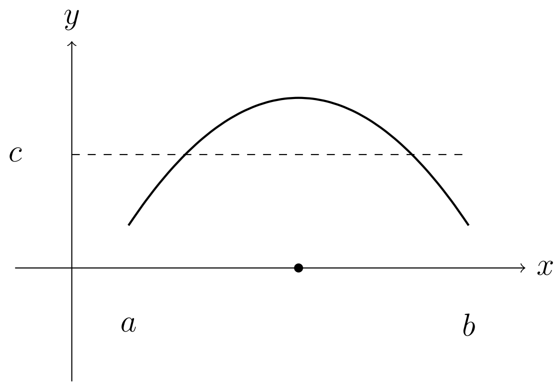
Lemma

If $E \subset X$ is **connected** and $f : E \to Y$ is **continuous**, then $f(E)$ is **connected**.

证明. Regard f as a map $(E, d_X) \to (f(E), d_Y)$. If $f(E) = A \cup B$ with nonempty, disjoint A, B open in $f(E)$, then $E = f^{-1}(A) \cup f^{-1}(B)$ is a separation of E by nonempty open sets (preimages of open sets are open), contradicting connectedness. □

Theorem

(Intermediate Value Theorem). Let $E \subset X$ be **connected** and $f : E \to \mathbb{R}$ **continuous**. Then $f(E)$ is connected in $\mathbb{R}$, hence an interval. Consequently, if $a, b \in E$ and c lies between $f(a)$ and $f(b)$, there **exists** $p \in E$ with $f(p) = c$.



The only connected subsets of $\mathbb{R}$ are intervals. A common application: if $f : [a, b] \to \mathbb{R}$ is continuous and $f(a)f(b) < 0$, there exists $\xi \in (a, b)$ with $f(\xi) = 0$.

0.18 Uniform continuity

Definition

Uniform continuity. $f : E \to Y$ is *uniformly continuous* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in E$, $d_X(x, y) < \delta \implies d_Y\big(f(x), f(y)\big) < \varepsilon$.

Ordinary continuity is local: the permissible δ may depend on the point;

Uniform continuity requires a single δ that works for all points simultaneously. This is especially important near the *boundary* of the domain (e.g. $f(x) = 1/x$ on $(0, \infty)$).

Example

$f(x) = 1/x$ on $(0, \infty)$ is not uniformly continuous. Take $\varepsilon_0 = 1$. For any $\delta \in (0, 1)$ choose $x = \delta/2$, $y = \delta/3$. Then $|x - y| = \delta/6 < \delta$ while $|f(x) - f(y)| = \left|\frac{2}{\delta} - \frac{3}{\delta}\right| = \frac{1}{\delta} > 1$.

Lemma

(Uniform continuity preserves Cauchy sequences). If $f : E \to Y$ is uniformly continuous and $(x_n) \subset E$ is Cauchy in (X, d_X), then $\big(f(x_n)\big)$ is Cauchy in (Y, d_Y).

This fact lets us **extend uniformly continuous functions to the boundary of intervals** by defining the value at an endpoint as the limit along any sequence approaching that endpoint.

Corollary & Secondary Conclusion

(Uniformly continuous extension on an open interval). Let $I = (a, b) \subset \mathbb{R}$ (allow $a = -\infty$ or $b = +\infty$). If $f : I \to \mathbb{R}$ is **uniformly continuous**, there exists a **continuous** $\overline{f} : \overline{I} \to \mathbb{R}$ with $\overline{f}(x) = f(x)$ for $x \in I$.

In short,Uniform continuity of $f : (a, b) \to \mathbb{R}$ guarantees that limits of f exist at the endpoints of I

证明. Assume $a, b \in \mathbb{R}$; the infinite-endpoint cases are similar. Let $x_n = b - \frac{1}{n} \in I$. Then (x_n) is Cauchy in $\mathbb{R}$, so $(f(x_n))$ is Cauchy and convergent; define $\overline{f}(b) := \lim_n f(x_n)$. Define $\overline{f}(a)$ analogously and set $\overline{f} = f$ on I.

To prove continuity at b, let $\varepsilon > 0$. Uniform continuity gives $\delta_0 > 0$ such that $|u - v| < \delta_0 \implies |f(u) - f(v)| < \varepsilon/2$. If $(p_n) \subset I$ with $p_n \to b$, then for large n we have $|p_n - x_n| < \delta_0$, hence $|f(p_n) - f(x_n)| < \varepsilon/2$.

Since $f(x_n) \to \overline{f}(b)$, we get $|f(p_n) - \overline{f}(b)| < \varepsilon$ for large n. Thus $\overline{f}$ is continuous at b. □

This construction above works in any *complete* metric target (because Cauchy sequences converge). This is a central tool in analysis (e.g. in complex analysis).

If a continuous function is already defined on a compact set, then it must be uniformly continuous (Heine–Cantor below). Together with the previous remark, this provides a route to extend continuous functions from open sets to compact sets.

Theorem

Heine–Cantor: If $E \subset X$ is *compact* and $f : E \to Y$ is *continuous*, then f is ***uniformly continuous*** on E.

证明. We construct the $\delta > 0$ (for uniform continuity) using compactness of E:

Step 1: Fix $\varepsilon > 0$ We seek a $\delta > 0$ (independent of points) such that:

$$\forall x, y \in E : \quad d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon.$$

Step 2: Use continuity to build open balls For each $p \in E$, continuity of f at p gives $\delta_p > 0$ with:

$$\forall x \in E : \quad d_X(x, p) < \delta_p \implies d_Y(f(x), f(p)) < \frac{\varepsilon}{2}.$$

Define the open ball (radius $\delta_p/2$):

$$U_p = \left\{ x \in X \mid d_X(x,p) < \frac{\delta_p}{2} \right\}.$$

The collection $\{U_p \mid p \in E\}$ is an *open cover* of E (every $p \in E$ lies in U_p).

Step 3: Apply compactness for a finite subcover Since E is compact, there exist $p_1, p_2, \ldots, p_N \in E$ such that:

$$E \subseteq \bigcup_{i=1}^{N} U_{\delta_{p_i}/2}(p_i).$$

Step 4: Define the uniform δ Let:

$$\delta = \min\left\{ \frac{\delta_{p_1}}{2}, \frac{\delta_{p_2}}{2}, \ldots, \frac{\delta_{p_N}}{2} \right\}.$$

Since δ is the minimum of finitely many positive numbers, $\delta > 0$.

Step 5: Verify uniform continuity Take any $x, y \in E$ with $d_X(x,y) < \delta$: - Since $E \subseteq \bigcup_{i=1}^{N} U_{\delta_{p_i}/2}(p_i)$, $x \in U_{\delta_{p_i}/2}(p_i)$ for some i, so $d_X(x,p_i) < \frac{\delta_{p_i}}{2}$. - By the triangle inequality:

$$d_X(y,p_i) \le d_X(y,x) + d_X(x,p_i) < \delta + \frac{\delta_{p_i}}{2} \le \frac{\delta_{p_i}}{2} + \frac{\delta_{p_i}}{2} = \delta_{p_i}.$$

- By continuity (Step 2):

$$d_Y(f(x), f(p_i)) < \frac{\varepsilon}{2} \quad \text{and} \quad d_Y(f(y), f(p_i)) < \frac{\varepsilon}{2}.$$

- By the triangle inequality for d_Y:

$$d_Y(f(x), f(y)) \le d_Y(f(x), f(p_i)) + d_Y(f(p_i), f(y)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus f is uniformly continuous on E. □

Starting from this lecture we investigate the derivative of functions (on intervals of R).

Differentiation is a structure specific to such spaces and is not defined in a general metric space without additional structure.

0.19 Basic Definition

Let $I \subset \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ a real–valued function. Fix $x \in I$ and define, for $t \in I$, $t \neq x$,

$$\varphi(t) := \frac{f(t) - f(x)}{t - x}.$$

Definition

Derivative at a point: We say that f is *differentiable* at x if the limit $\lim_{t \to x} \varphi(t)$ exists. In this case we write

$$f'(x) := \lim_{t \to x} \frac{f(t) - f(x)}{t - x}.$$

If f is differentiable at every $x \in I$, we say f is *differentiable on* I, and the map $f' : I \to \mathbb{R}$ is called the *derivative* of f.

Fact and proposition

[Differentiable $\implies$ continuous] If f is differentiable at $x \in I$, then f is continuous at x.

Hint: Triangle inequality

证明. Let $\varepsilon > 0$. By differentiability there exists $\delta_0 > 0$ such that for $0 < |t - x| < \delta_0$ we have

$$\left| \frac{f(t) - f(x)}{t - x} - f'(x) \right| < \varepsilon.$$

Hence $|f(t) - f(x)| \leq |t - x|(|f'(x)| + \varepsilon)$, which tends to 0 as $t \to x$; therefore $f(t) \to f(x)$. □

However the converse is not necessarily correct.

Example

The function $f(t) = |t|$ is continuous on $\mathbb{R}$ but is not differentiable at 0.

Linearization characterization

Theorem

Linear approximation: For a function $f : I \to \mathbb{R}$ and a point $x \in I$ the following are equivalent:

1. f is differentiable at x and $f'(x) = A$;
2. $\lim_{t \to x} \frac{f(t) - f(x) - A(t - x)}{t - x} = 0$.

So we gain a useful corollary:

Corollary & Secondary Conclusion

Derivative as a local linearization: Near x we have $f(t) \approx f(x) + f'(x)(t-x)$; i.e. f is **locally** well approximated by its tangent line.

Derivative's Algebraic rules

In what follows assume $f, g : I \to \mathbb{R}$ are differentiable at $x \in I$.

Theorem

Operational properties

1. $(f+g)'(x) = f'(x) + g'(x)$.
2. $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$.
3. If $g(x) \neq 0$, then $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x)-f(x)g'(x)}{g(x)^2}$.

Just similar to sequence's proof

0.20 Chain rule

Theorem

Chain rule: Let $f : I \to \mathbb{R}$ and $g : J \to \mathbb{R}$ with $f(I) \subset J$. If f is **differentiable** at $x \in I$ and g is **differentiable** at $y = f(x)$, then $h = g \circ f$ is **differentiable** at x and

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

证明. Detailed ε-δ proof. Write $y = f(x)$. For $t \neq x$ we use the algebraic identity

$$\frac{g(f(t)) - g(f(x))}{t-x} - g'(y)f'(x) = \left(\frac{g(f(t)) - g(y)}{f(t)-y} - g'(y)\right) \cdot \frac{f(t)-f(x)}{t-x} \tag{1}$$
$$+g'(y)\left(\frac{f(t)-f(x)}{t-x} - f'(x)\right).$$

(If $f(t) = f(x)$, interpret the first factor as 0; then (1) still holds.)

Step 1: control the difference quotient of g. Since g is differentiable at y, for any $\varepsilon > 0$ there exists $\eta > 0$ such that

$$0 < |u-y| < \eta \implies \left|\frac{g(u)-g(y)}{u-y} - g'(y)\right| < \frac{\varepsilon}{2(1+|f'(x)|)}. \tag{2}$$

Step 2: control the difference quotient of f and get a uniform bound. Since f is differentiable at x, there exists $\delta_1 > 0$ such that

$$0 < |t - x| < \delta_1 \implies \left| \frac{f(t) - f(x)}{t - x} - f'(x) \right| < \min\left\{1, \frac{\varepsilon}{2(1 + |g'(y)|)}\right\}. \tag{3}$$

Consequently,

$$0 < |t - x| < \delta_1 \implies \left| \frac{f(t) - f(x)}{t - x} \right| \leq |f'(x)| + 1. \tag{4}$$

Step 3: link $t \to x$ to $u \to y$. Because differentiability implies continuity, there exists $\delta_2 > 0$ such that

$$|t - x| < \delta_2 \implies |f(t) - f(x)| < \eta. \tag{5}$$

Step 4: combine. Let $\delta = \min\{\delta_1, \delta_2\}$. If $0 < |t - x| < \delta$, then by (2), (3), (4), (5) and the identity (1),

$$\begin{aligned}
\left| \frac{g(f(t)) - g(f(x))}{t - x} - g'(y)f'(x) \right| &\leq \left| \frac{g(f(t)) - g(y)}{f(t) - y} - g'(y) \right| \left| \frac{f(t) - f(x)}{t - x} \right| \\
&\quad + |g'(y)| \left| \frac{f(t) - f(x)}{t - x} - f'(x) \right| \\
&< \frac{\varepsilon}{2(1 + |f'(x)|)}(|f'(x)| + 1) + |g'(y)| \cdot \frac{\varepsilon}{2(1 + |g'(y)|)} \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Hence

$$\lim_{t \to x} \frac{g(f(t)) - g(f(x))}{t - x} = g'(y)f'(x) = g'(f(x))f'(x),$$

□

0.21 Local extrema

Let $p \in I$.

Definition

We say that f has a *local maximum* at p if there **exists** $\delta > 0$ such that $f(q) \leq f(p)$ for all $q \in I$ with $|q - p| < \delta$.
Similarly for a *local minimum*.

Lemma

Fermat's lemma : If x is an interior point of I and f is ***differentiable*** at x, then any local maximum or minimum at x forces $f'(x) = 0$.

证明. Assume x is an interior local maximum (the minimum case is symmetric). Since $x \in \mathrm{int}(I)$, there exists $\delta > 0$ with $(x-\delta, x+\delta) \subseteq I$, and $f(t) \le f(x)$ for all $t \in (x-\delta, x+\delta)$.

- For $t \in (x, x+\delta)$: $\frac{f(t)-f(x)}{t-x} \le 0$. Taking $t \to x^+$:

$$f'(x) = \lim_{t \to x^+} \frac{f(t)-f(x)}{t-x} \le 0.$$

- For $t \in (x-\delta, x)$: $\frac{f(t)-f(x)}{t-x} \ge 0$. Taking $t \to x^-$:

$$f'(x) = \lim_{t \to x^-} \frac{f(t)-f(x)}{t-x} \ge 0.$$

Thus $f'(x) = 0$. □

0.22 Mean value theorems

Lemma

Rolle' s Lemma: Let $f : [a, b] \to \mathbb{R}$ be a function satisfying three conditions:
f is continuous on the closed interval $[a, b]$;
f is differentiable on the open interval (a, b);
$f(a) = f(b)$.
Then, there exists **at least one** point $\xi \in (a, b)$ such that $f'(\xi) = 0$.

Intuitive Interpretation Geometrically, Rolle' s Lemma says: If a continuous, differentiable curve starts and ends at the same height, there must be at least one point on the curve where its tangent line is horizontal (i.e., the derivative is zero).

证明. proof:

□

Theorem

Mean value theorem: If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then there exists $\xi \in (a, b)$ with

$$f(b) - f(a) = f'(\xi)\,(b - a).$$

It can be proved by Rolle's lemma also, it can directly be yeiled by Cauchy's MVT(below).

证明. proof:

□

Corollary & Secondary Conclusion

(Nonvanishing derivative implies injectivity). *If* $f : (a, b) \to \mathbb{R}$ *is differentiable and* $f'(x) \neq 0$ *for all* $x \in (a, b)$*, then* f *is injective, i.e.,* $f(x) \neq f(y)$ *if* $x \neq y$*.*

Theorem

Cauchy mean value theorem: Let $f, g : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $\xi \in (a, b)$ such that

$$(f(b) - f(a))g'(\xi) = (g(b) - g(a))f'(\xi).$$

证明. Consider $h(t) = \big(f(b) - f(a)\big)g(t) - \big(g(b) - g(a)\big)f(t)$, continuous on $[a, b]$. Let $p, q \in [a, b]$ be points where h attains its max/min. If one of them lies in (a, b), then $h'(p) = 0$ or $h'(q) = 0$, giving the desired equality. If both extrema occur at the endpoints, then h is constant on $[a, b]$ because $h(a) = h(b)$, again giving $h'(t) = 0$ for all $t \in (a, b)$. □

0.23 Vector–valued functions

Let $\mathbf{f} = (f_1, \ldots, f_k) : I \to \mathbb{R}^k$.

Definition

We say $\mathbf{f}$ is differentiable at $x \in I$ if there exists $\mathbf{A} \in \mathbb{R}^k$ such that

$$\lim_{t \to x} \frac{\mathbf{f}(t) - \mathbf{f}(x)}{t - x} = \mathbf{A}.$$

We then write $\mathbf{A} = \mathbf{f}'(x)$.

Claim

Componentwise characterization: $\mathbf{f}$ is differentiable at x with derivative $\mathbf{f}'(x) = (f_1'(x), \ldots, f_k'(x))$ if and only if each component f_i is differentiable at x.

Fact and proposition

Vector mean value estimate: Suppose $\mathbf{f}$ is continuous on $[a, b]$ and differentiable on (a, b). Then there exists $\xi \in (a, b)$ such that

$$\|\mathbf{f}(b) - \mathbf{f}(a)\| \le (b - a)\|\mathbf{f}'(\xi)\|.$$

证明. **Step 1: Define the auxiliary scalar function.** We introduce a scalar-valued function to link $\mathbf{f}$ (a vector function) to the scalar Mean Value Theorem (MVT). Let:

$$\phi(t) = \big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}(t),$$

where $\cdot$ denotes the **dot product** (so $\phi(t) \in \mathbb{R}$ for all $t \in [a, b]$).

Step 2: Verify $\phi(t)$ satisfies the scalar MVT. Recall the *scalar Mean Value Theorem*: If $g : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then there exists $\xi \in (a, b)$ such that

$$g(b) - g(a) = g'(\xi)(b - a).$$

For $\phi(t)$: - Since $\mathbf{f}$ is **continuous** on $[a, b]$, the dot product $\phi(t) = \big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}(t)$ is continuous on $[a, b]$ (dot products of continuous vector functions are continuous). - Since $\mathbf{f}$ is **differentiable** on (a, b), the derivative of $\phi(t)$ (via the dot product product rule) is:

$$\phi'(t) = \frac{d}{dt}\left[\big(\mathbf{f}(b) - \mathbf{f}(a)\big)\right] \cdot \mathbf{f}(t) + \big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}'(t).$$

The term $\frac{d}{dt}[\mathbf{f}(b) - \mathbf{f}(a)] = \mathbf{0}$ (as $\mathbf{f}(b) - \mathbf{f}(a)$ is constant), so:

$$\phi'(t) = \big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}'(t).$$

Thus $\phi(t)$ is differentiable on (a, b).

Step 3: Apply the scalar MVT to $\phi(t)$. By the scalar MVT, there exists $\xi \in (a, b)$ such that:

$$\phi(b) - \phi(a) = \phi'(\xi)(b - a).$$

Step 4: Compute $\phi(b) - \phi(a)$. Evaluate ϕ at $t = b$ and $t = a$: - $\phi(b) = \big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}(b)$, - $\phi(a) = \big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}(a)$.

Subtracting these gives:

$$\phi(b) - \phi(a) = \big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \big(\mathbf{f}(b) - \mathbf{f}(a)\big).$$

By definition of the **Euclidean norm** ($\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$ for any vector $\mathbf{v}$), this simplifies to:

$$\phi(b) - \phi(a) = \|\mathbf{f}(b) - \mathbf{f}(a)\|^2.$$

Step 5: Substitute $\phi'(\xi)$ into the MVT equation. From Step 2, $\phi'(\xi) = \big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}'(\xi)$. Substitute this and the result of Step 4 into the scalar MVT equation (Step 3):

$$\|\mathbf{f}(b) - \mathbf{f}(a)\|^2 = \left[\big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}'(\xi)\right](b-a).$$

Step 6: Apply the Cauchy-Schwarz Inequality. The *Cauchy-Schwarz Inequality* states that for any vectors $\mathbf{u}, \mathbf{v}$,

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

Let $\mathbf{u} = \mathbf{f}(b) - \mathbf{f}(a)$ and $\mathbf{v} = \mathbf{f}'(\xi)$. Then:

$$\left|\big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}'(\xi)\right| \leq \|\mathbf{f}(b) - \mathbf{f}(a)\| \cdot \|\mathbf{f}'(\xi)\|.$$

Taking absolute values of both sides of the equation from Step 5 (noting $b - a > 0$, so it does not affect the absolute value):

$$\|\mathbf{f}(b) - \mathbf{f}(a)\|^2 = \left|\big(\mathbf{f}(b) - \mathbf{f}(a)\big) \cdot \mathbf{f}'(\xi)\right|(b-a).$$

Step 7: Simplify to obtain the inequality. Substitute the Cauchy-Schwarz bound into the equation above:

$$\|\mathbf{f}(b) - \mathbf{f}(a)\|^2 \leq \|\mathbf{f}(b) - \mathbf{f}(a)\| \cdot \|\mathbf{f}'(\xi)\| \cdot (b-a).$$

- If $\|\mathbf{f}(b) - \mathbf{f}(a)\| \neq 0$, divide both sides by $\|\mathbf{f}(b) - \mathbf{f}(a)\|$:

$$\|\mathbf{f}(b) - \mathbf{f}(a)\| \leq (b-a)\|\mathbf{f}'(\xi)\|.$$

- If $\|\mathbf{f}(b) - \mathbf{f}(a)\| = 0$, then $\mathbf{f}(b) = \mathbf{f}(a)$, and the inequality $\mathbf{0} \leq (b-a)\|\mathbf{f}'(\xi)\|$ is *trivially true* (the right-hand side is non-negative).

Conclusion. In either case, there exists $\xi \in (a, b)$ such that

$$\|\mathbf{f}(b) - \mathbf{f}(a)\| \leq (b-a)\|\mathbf{f}'(\xi)\|.$$

□

0.24 L'Hospital's rule

Theorem

(L'Hospital's rule):
Then the limit $\lim_{x\to c}\frac{f(x)}{g(x)}$ exists and equals L.*Let* $f, g : (a, b) \to \mathbb{R}$ *be differentiable and let* $a < c < b$. *Assume that* $g'(x) \neq 0$ *for all* $x \in (a, b)$ *and that the limit*

$$\lim_{x\to c}\frac{f'(x)}{g'(x)} = L \in \mathbb{R} \cup \{\pm\infty\}$$

exists. Suppose in addition that either

1. $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$, *or*

2. $\lim_{x\to c} |g(x)| = \lim_{x\to c} |f(x)| = +\infty$.

Then the limit $\lim_{x\to c}\frac{f(x)}{g(x)}$ *exists and equals* L.

Hint: The core idea is to **decompose complex fractional differences into a controllable sum of several terms** through splitting and transformation, and then use the **triangle inequality** to amplify them, thereby obtaining an estimation formula that is convenient for analysis.

证明. By *Cauchy's mean value theorem*, for any $x \neq y$ in a punctured neighborhood of c, there exists t (between x and y) with

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(t)}{g'(t)}.$$

Case 1: $L \in \mathbb{R}$ (finite limit). Fix $\varepsilon > 0$. By the hypothesis on $\frac{f'}{g'}$, choose $\delta > 0$ so that whenever $x, y \in (c - \delta, c + \delta) \setminus \{c\}$, the t (from Cauchy's MVT) satisfies

$$\left|\frac{f'(t)}{g'(t)} - L\right| < \varepsilon.$$

If (i) (the $0/0$ case: $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$) holds, pick y near c so $|f(y)| < \varepsilon$ and $|g(y)| < \varepsilon$. If (ii) (the ∞/∞ case: $\lim_{x\to c} |g(x)| = \infty$) holds, pick x near c so $|g(x)|$ is arbitrarily large. In both cases, $g(x) \neq 0$, $g(y) \neq 0$, and:

$$\left|\frac{f(x)}{g(x)} - L\right| = \left|\frac{f(x) - f(y)}{g(x)} + \frac{f(y)}{g(x)} - L\right| \le \left|\frac{g(x) - g(y)}{g(x)}\right|\left|\frac{f(x) - f(y)}{g(x) - g(y)} - L\right| + \frac{|f(y)|}{|g(x)|} + \frac{|g(y)|}{|g(x)|}|L|.$$

The first factor $\left(\left|\frac{g(x)-g(y)}{g(x)}\right|\right)$ is bounded by $1 + \left|\frac{g(y)}{g(x)}\right|$; we can make this ≤ 2 by choosing x/y appropriately. The middle factor is $< \varepsilon$ (by Cauchy's MVT). By choosing x (depending on ε), the second/third terms are $< \varepsilon$. Thus:

$$\left|\frac{f(x)}{g(x)} - L\right| \le \left(1 + \left|\frac{g(y)}{g(x)}\right|\right)\varepsilon + \varepsilon \le 4\varepsilon.$$

Case 2: $L = +\infty$ (the $-\infty$ case is analogous). Given $A > 0$, choose $\delta > 0$ so that $\frac{f'(t)}{g'(t)} > A$ (for the t from Cauchy's MVT) whenever x, y are in a δ-neighborhood of c. Then:

$$\frac{f(x) - f(y)}{g(x) - g(y)} > A.$$

If (i) holds, send $y \to c$ to get $\frac{f(x)}{g(x)} > A$ for x near c. If (ii) holds, choose y with $|g(y)|$ large (and same sign as $g(x)$); the conclusion follows. Thus $\lim_{x\to c} \frac{f(x)}{g(x)} = +\infty$. □

【Remark】:

The rule is local and maybe iterated when the hypotheses remain valid, i.e., for higher order derivatives.

0.25 Darboux Property

Theorem

(Darboux property for derivatives): *Let $f : (a, b) \to \mathbb{R}$ be differentiable. If $c < d$ and λ lies strictly between $f'(c)$ and $f'(d)$, then there exists $s \in (c, d)$ with $f'(s) = \lambda$. Equivalently, f' has the intermediate value property on (a, b).*

证明. Assume $f'(c) < \lambda < f'(d)$ (the other order is symmetric). Set $g(t) = f(t) - \lambda t$. Then

$$g'(c) = f'(c) - \lambda < 0 \quad \text{and} \quad g'(d) = f'(d) - \lambda > 0.$$

By definition of g', there exist $\delta_c, \delta_d > 0$ such that

$$\frac{g(t) - g(c)}{t - c} < 0, \quad \forall t \in (c - \delta_c, c + \delta_c),$$

$$\frac{g(t) - g(d)}{t - d} > 0, \quad \forall t \in (d - \delta_c, d + \delta_c).$$

In particular, g decreases to the right of c and increases to the left of d, so g attains a global minimum at some $s \in (c, d)$.

By Fermat' s lemma for extrema: we exsit a s,s.t. $g'(s) = 0$, which means $f'(s) = \lambda$. □

Corollary & Secondary Conclusion

A derivative ***cannot have a discontinuity of the first kind*** (no jump discontinuity), because any two one-sided limits would force f' to *take all intermediate values.*

证明. Suppose otherwise, then we discuss the following two cases:
Case 1: $f'(x+) \neq f'(x-)$. Assume WLOG $f'(x+) > f'(x-)$. Take $\varepsilon = \frac{f'(x+)-f'(x-)}{4}$. By definition of limit, we can pick $\delta > 0$ such that for any $x < y < x + \delta$ we have

$$f'(y) > \frac{3f'(x+) + f'(x-)}{4} = L_+,$$

and for any $x - \delta < y < x$,

$$f'(y) < \frac{3f'(x-) + f'(x+)}{4} = L_-.$$

Then, taking $y_1 = x - \delta/2$, $y_2 = x + \delta/2$, we have $f'(y_1) < L_- < L_+ < f'(y_2)$. But if we choose $\lambda = (L_-, L_+) \setminus \{f'(x)\}$, there exists no $y \in (y_1, y_2) \subset (x - \delta, x + \delta)$ such that $f'(y) = \lambda$.
Case 2: $f'(x+) = f'(x-) \neq f'(x)$. Suppose $L = f'(x+) = f'(x-) < f'(x)$. Then take $\varepsilon = \frac{f'(x)-L}{2}$, and find a $\delta > 0$ such that $\forall y \in (x - \delta, x + \delta),\ y \neq x$,

$$f'(y) < \frac{f'(x) + L}{2}.$$

So $f'(x - \delta/2) < f'(x)$, but for any $y \in (x - \delta/2, x)$,

$$f'(y) \notin \left(\frac{f'(x) + L}{2}, f'(x)\right).$$

□

0.26 Taylor's theorem

Definition

If f is n times differentiable on (a, b), write $f^{(0)} = f$ and say that f is *smooth of order n.*
For $\alpha \in (a, b)$, define the $(n-1)$-st Taylor polynomial of f at α by$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(t - \alpha)^k$.

And we will show an important Theorem:

Theorem

(Taylor's theorem). *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and n times differentiable on (a, b). Fix $\alpha, \beta \in [a, b]$ with $\alpha < \beta$ and let P be the Taylor polynomial of order $n-1$ at α. Then there exists $\xi \in (\alpha, \beta)$ such that* $f(\beta) = P(\beta) + \frac{f^{(n)}(\xi)}{n!}(\beta - \alpha)^n$.

证明. Define $M = \frac{f(\beta)-P(\beta)}{(\beta-\alpha)^n}$, $g(t) = f(t) - P(t) - M(t-\alpha)^n$, $t \in [a, b]$. Then $g(\alpha) = g(\beta) = 0$. By the Mean Value Theorem (MVT), there exists $\xi_1 \in (\alpha, \beta)$ with $g'(\xi_1) = 0$. Applying MVT n times (to $g, g', \ldots, g^{(n-1)}$), we obtain a point $\xi \in (\alpha, \beta)$ with $g^{(n)}(\xi) = 0$. As $P^{(n)} \equiv 0$, this reads

$$f^{(n)}(\xi) - n!\, M = 0, \quad \text{i.e.} \quad M = \frac{f^{(n)}(\xi)}{n!},$$

and the formula follows. □

【Remark】:

This Taylor's Theorem is in Lagrange form of the remainder

0.27 Basic Concepts

We now prepare the language used for **Riemann and Riemann–Stieltjes integrals.**

Definition

(Partitions and mesh). A **partition** of $[a, b]$ is a **finite** set $P = \{(t_{i-1}, t_i) : i = 1, 2, \cdots, m, a = t_0 < t_1 < \cdots < t_m = b\}$.
Its **mesh** is $\|P\| = \max_{1\le i\le m}(t_i - t_{i-1})$. We also write $\Delta t_i = t_i - t_{i-1}$.

Example

Consider the interval $[0, 4]$ as an example.
A partition $P = \{0, 2, 4\}$ is essentially a set of **partition points** $0, 2, 4$, corresponding to subintervals $[0, 2]$ and $[2, 4]$.

Definition

(Upper and lower sums for Riemann–Stieltjes). Let $\alpha : [a,b] \to \mathbb{R}$ be **increasing** and $f : [a,b] \to \mathbb{R}$ be **bounded**. For a **partition** P define

$$m_i = \inf_{t\in[t_{i-1},t_i]} f(t), \quad M_i = \sup_{t\in[t_{i-1},t_i]} f(t), \quad \Delta\alpha_i = \alpha(t_i) - \alpha(t_{i-1}).$$

The lower sum and upper sum are

$$L(P,f,\alpha) = \sum_{i=1}^{m} m_i \Delta\alpha_i, \quad U(P,f,\alpha) = \sum_{i=1}^{m} M_i \Delta\alpha_i.$$

Fact and proposition

The Role of α: The $\alpha : [a,b] \to \mathbb{R}$ is an **increasing function** (key property: monotone non-decreasing). It replaces the ”independent variable x” in the Riemann integral and acts as the ”reference” for the integral.

The familiar Riemann integral (for area calculation) is a **special case** of the Riemann-Stieltjes integral: when $\alpha(x) = x$, $\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1}) = x_i - x_{i-1} = \Delta x_i$. (We will discuss it in detail in next part)

Definition

A **partition** P' is said to be a *refinement* of a partition P if **all partition points in P** are contained in P' (and we are allowed to add more points).
For two partitions P_1 and P_2, we say P is the *common partition* if it contains all partition points of both P_1 and P_2. (We may denote $P = P_1 \cup P_2$.)

Example

- $P = \{0, 2, 4\}$ (partition points: $0, 2, 4$; corresponding subintervals: $[0, 2], [2, 4]$)
- $P' = \{0, 1, 2, 3, 4\}$ (partition points: $0, 1, 2, 3, 4$; corresponding subintervals: $[0, 1], [1, 2], [2, 3], [3, 4]$)

Clearly, all partition points of P (i.e., $0, 2, 4$) are contained in P', so $P \subseteq P'$.

Example

Let the interval of interest be $[a, b] = [0, 5]$. We define two arbitrary partitions of $[0, 5]$:

- Let $P_1 = \{0, 2, 5\}$ (partition points: $0, 2, 5$). The subintervals corresponding to P_1 are: $[0, 2]$, $[2, 5]$.
- Let $P_2 = \{0, 1, 3, 5\}$ (partition points: $0, 1, 3, 5$). The subintervals corresponding to P_2 are: $[0, 1]$, $[1, 3]$, $[3, 5]$.

The *common partition* $P = P_1 \cup P_2$ is formed by taking *all distinct partition points* from P_1 and P_2, sorted in increasing order.
For our example:

$$P_1 \cup P_2 = \{0, 2, 5\} \cup \{0, 1, 3, 5\} = \{0, 1, 2, 3, 5\}$$

Key Verification
- $P = \{0, 1, 2, 3, 5\}$ contains *all partition points* of P_1 (i.e., $0, 2, 5$) and P_2 (i.e., $0, 1, 3, 5$), satisfying the definition of a common partition.

【Remark】:

Notice that P is also a **refinement** of both P_1 and P_2 (it adds new points to each original partition).

And we have:

Property

Additivity of α-Increments: Let $P = \{y_0, y_1, \ldots, y_n\}$ be a partition of an interval $[a, b]$, and let $[y_{j-1}, y_j]$ be a subinterval of P. Let P' be a refinement of P (i.e., $P \subseteq P'$), so P' inserts additional partition points between y_{j-1} and y_j, denoted as:

$$y_{j-1} = x_k < x_{k+1} < \cdots < x_l = y_j$$

for some integers $k < l$.
The sum of α-increments over the subintervals of P' contained in $[y_{j-1}, y_j]$ is additive:

$$\sum_{i=k}^{l} (\alpha(x_i) - \alpha(x_{i-1})) = \alpha(y_j) - \alpha(y_{j-1})$$

The sum on the left-hand side is a **telescoping sum**—all intermediate terms cancel out.

证明. By definition of the partition points:

- $x_k = y_{j-1}$ (the left endpoint of the subinterval),
- $x_l = y_j$ (the right endpoint of the subinterval).

Expanding the sum explicitly confirms the cancellation:

$$\begin{aligned}\sum_{i=k}^{l} (\alpha(x_i) - \alpha(x_{i-1})) &= (\alpha(x_k) - \alpha(x_{k-1})) + (\alpha(x_{k+1}) - \alpha(x_k)) + (\alpha(x_{k+2}) - \alpha(x_{k+1})) + \cdots + (\alpha(x_l) - \alpha(\\ &= -\alpha(x_{k-1}) + \alpha(x_k) - \alpha(x_k) + \alpha(x_{k+1}) - \alpha(x_{k+1}) + \cdots + \alpha(x_l) \\ &= \alpha(x_l) - \alpha(x_k)\end{aligned}$$

Substituting $x_k = y_{j-1}$ and $x_l = y_j$ gives:

$$\sum_{i=k}^{l} (\alpha(x_i) - \alpha(x_{i-1})) = \alpha(y_j) - \alpha(y_{j-1})$$

This completes the proof. □

Property

If P' is a refinement of P, then$L(P, f, \alpha) \leq L(P', f, \alpha)$ and $U(P', f, \alpha) \leq U(P, f, \alpha)$.

证明. We prove the statement for U; the proof for L is similar. Let $P' = \{x_0 < \cdots < x_n\}$ and $P = \{y_0 < \cdots < y_m\}$ with $P \subseteq P'$. Then

$$U(P', f, \alpha) = \sum_{i=1}^{n} (\alpha(x_i) - \alpha(x_{i-1})) \sup_{x \in [x_{i-1}, x_i]} f(x).$$

Group together the subintervals of P' that lie inside a fixed (y_{j-1}, y_j):

$$U(P', f, \alpha) = \sum_{j=1}^{m} \sum_{[x_{i-1}, x_i] \subseteq [y_{j-1}, y_j]} (\alpha(x_i) - \alpha(x_{i-1})) \sup_{x \in [x_{i-1}, x_i]} f(x).$$

Since $\sup_{[x_{i-1}, x_i]} f \leq \sup_{[y_{j-1}, y_j]} f$ and the α-increments over the pieces add up to $\alpha(y_j) - \alpha(y_{j-1})$, we obtain

$$U(P', f, \alpha) \leq \sum_{j=1}^{m} (\alpha(y_j) - \alpha(y_{j-1})) \sup_{x \in [y_{j-1}, y_j]} f(x) = U(P, f, \alpha).$$

□

Definition

(Upper/lower integral and integrability). The **upper integral** is $\overline{\int_a^b} f d\alpha = \inf_P U(P, f, \alpha)$ and the **lower integral** is $\underline{\int_a^b} f d\alpha = \sup_P L(P, f, \alpha)$, where the *infimum* and *supremum* run over all partitions P of $[a, b]$.
If these two numbers ***coincide***, we say f is ***Riemann–Stieltjes integrable*** with respect to α and write $\int_a^b f d\alpha$ for the common value.

We do not assume f is *continuous*; later we will give convenient sufficient conditions (continuity suffices but is not necessary). Also α need not be continuous; the central structural assumption for now is that α is increasing.

【Remark】:

The Riemann integral is a ***special case*** of the Riemann-Stieltjes integral, when the integrator $\alpha(t) = t$, $\Delta\alpha_i = \Delta t_i$, and in this case, $\mathcal{R}_\alpha[a, b]$ degenerates to $\mathcal{R}[a, b]$. Conversely, the Riemann-Stieltjes integral is a generalized extension of the Riemann integral: it allows the **use of "non-length" weights** (characterized by the increment of $\alpha(t)$) to accumulate the values of the function f, thus covering a wider range of application scenarios.

And we will show a classical example:

Example

Define the Dirichlet Function:
The Dirichlet function $f : \mathbb{R} \to \{0, 1\}$ is defined as: $f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \text{ (rationals)}, \\ 0 & \text{if } x \notin \mathbb{Q} \text{ (irrationals)}. \end{cases}$ We aim to show $f \notin \mathcal{R}_\alpha[a, b]$ for any $a < b$ and any increasing function $\alpha : [a, b] \to \mathbb{R}$.

证明. A key property of $\mathbb{Q}$ and $\mathbb{Q}^c$ is that they are dense in $\mathbb{R}$.

This means:In any non-empty interval $[t_{i-1}, t_i]$, there exists a rational number $q_i \in \mathbb{Q} \cap [t_{i-1}, t_i]$.In any non-empty interval $[t_{i-1}, t_i]$, there exists an irrational number $r_i \in \mathbb{Q}^c \cap [t_{i-1}, t_i]$.

Fix an Arbitrary Partition P of $[a, b]$: Let $P = \{t_0, t_1, \ldots, t_m\}$ be a partition of $[a, b]$ with $a = t_0 < t_1 < \cdots < t_m = b$. For each subinterval $[t_{i-1}, t_i] \in P$ compute the Upper

Sum $U(P, f, \alpha)$

The upper sum for f with respect to α is:$U(P, f, \alpha) = \sum_{i=1}^{m} M_i^f \Delta\alpha_i$,where $M_i^f = \sup_{x\in[t_{i-1},t_i]} f(x)$ and $\Delta\alpha_i = \alpha(t_i) - \alpha(t_{i-1})$.

By density, $[t_{i-1}, t_i]$ contains a rational q_i, so $f(q_i) = 1$. Thus, $\sup_{[t_{i-1},t_i]} f(x) = 1$ (since f only takes values 0 or 1).

Therefore, $M_i^f = 1$ for all i, and:$U(P, f, \alpha) = \sum_{i=1}^{m} 1{\cdot}\Delta\alpha_i = \sum_{i=1}^{m} (\alpha(t_i) - \alpha(t_{i-1})) = \alpha(b) - \alpha(a)$.

The lower sum for f with respect to α is:$L(P, f, \alpha) = \sum_{i=1}^{m} m_i^f \Delta\alpha_i$,where $m_i^f = \inf_{x\in[t_{i-1},t_i]} f(x)$.

By density, $[t_{i-1}, t_i]$ contains an irrational r_i, so $f(r_i) = 0$. Thus, $\inf_{[t_{i-1},t_i]} f(x) = 0$ (since $f \geq 0$).

Therefore, $m_i^f = 0$ for all i, and:$L(P, f, \alpha) = \sum_{i=1}^{m} 0 \cdot \Delta\alpha_i = 0$.

Define Upper and Lower IntegralsThe upper integral is $\overline{\int_a^b} f\, d\alpha = \inf_P U(P, f, \alpha)$.

Since every upper sum is $\alpha(b) - \alpha(a)$, we have $\overline{\int_a^b} f\, d\alpha = \alpha(b) - \alpha(a)$.

The lower integral is $\underline{\int_a^b} f\, d\alpha = \sup_P L(P, f, \alpha)$.

Since every lower sum is 0, we have $\underline{\int_a^b} f\, d\alpha = 0$.

So:f is not Riemann-Stieltjes integrable on $[a, b]$ for any $a < b$ and any increasing α. □

Lemma

The largest possible lower sum (over all partitions) is never larger than the smallest possible upper sum (over all partitions).
In short:
$\underline{\int_a^b} f\, d\alpha \leq \overline{\int_a^b} f\, d\alpha$.

The core logic here is to formalize that for every partition P_1, its lower sum $L(P_1, f, \alpha)$ is bounded above by the infimum of all upper sums, whatever How P_2 chosen (the upper integral).

证明. For any partitions P_1, P_2, let $P = P_1 \cup P_2$. Then

$$L(P_1, f, \alpha) \leq L(P, f, \alpha) \leq U(P, f, \alpha) \leq U(P_2, f, \alpha).$$

Taking the supremum over all partitions P_1 on the left and the infimum over all partitions P_2 on the right gives the claim. □

This is a foundational result for integrability for f to be Riemann-Stieltjes integrable, we need these two integrals to be equal.

Theorem

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

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

证明. If (*) holds, then

$$L(P,f,\alpha) \le \underline{\int_a^b} f\,d\alpha \le \overline{\int_a^b} f\,d\alpha \le U(P,f,\alpha),$$

so $\overline{\int_a^b} f\,d\alpha - \underline{\int_a^b} f\,d\alpha \le U(P,f,\alpha) - L(P,f,\alpha) < \varepsilon$. Since ε is arbitrary, the lower and upper integrals are equal.

Conversely, if $\underline{\int_a^b} f\,d\alpha = \overline{\int_a^b} f\,d\alpha$, pick partitions P_1, P_2 so that

$$U(P_1,f,\alpha) < \overline{\int_a^b} f\,d\alpha + \frac{\varepsilon}{2}, \quad L(P_2,f,\alpha) > \underline{\int_a^b} f\,d\alpha - \frac{\varepsilon}{2}.$$

Let $P = P_1 \cup P_2$. Then

$$U(P,f,\alpha) \le U(P_1,f,\alpha), \quad L(P,f,\alpha) \ge L(P_2,f,\alpha),$$

and hence $U(P,f,\alpha) - L(P,f,\alpha) < \varepsilon$. □

And we also have:

Theorem

(Riemann sums approximate the integral). *If $f \in \mathcal{R}_\alpha[a,b]$, then for every $\varepsilon > 0$ there exists a partition $P = \{a = t_0 < t_1 < \cdots < t_n = b\}$ such that, for any choice of $\xi_i \in [t_{i-1}, t_i]$,*

$$\left| \sum_{i=1}^{n} f(\xi_i)\Delta\alpha_i - \int_a^b f\,d\alpha \right| < \varepsilon, \quad \text{where } \Delta\alpha_i = \alpha(t_i) - \alpha(t_{i-1})$$

.

证明. By integrability, choose a partition P with $U(P,f,\alpha) - L(P,f,\alpha) < \varepsilon$, and write $M_i = \sup_{x\in[t_{i-1},t_i]} f(x)$, $m_i = \inf_{x\in[t_{i-1},t_i]} f(x)$. Then

$$L(P,f,\alpha) = \sum_i m_i \Delta\alpha_i \le \sum_i f(\xi_i)\Delta\alpha_i \le \sum_i M_i\Delta\alpha_i = U(P,f,\alpha),$$

whence the conclusion. □

And we will show an important theorem:

Theorem

If f is continuous on $[a, b]$, then $f \in \mathcal{R}_\alpha[a, b]$.

Hint: f is uniformly continuous

证明. Fix $\varepsilon > 0$. Since f is uniformly continuous on $[a, b]$, choose $\delta_0 > 0$ such that $|f(s) - f(t)| < \varepsilon$ whenever $|s - t| < \delta_0$. Let P be a uniform partition with mesh $\|P\| < \delta_0$. Write $M_i = \sup_{[t_{i-1}, t_i]} f$ and $m_i = \inf_{[t_{i-1}, t_i]} f$. Then $M_i - m_i < \varepsilon$ for each i, so

$$U(P, f, \alpha) - L(P, f, \alpha) = \sum_i (M_i - m_i)\Delta\alpha_i \leq \varepsilon \sum_i \Delta\alpha_i = \varepsilon(\alpha(b) - \alpha(a)).$$

Thus $f \in \mathcal{R}_\alpha[a, b]$. □

Corollary & Secondary Conclusion

Suppose $f \in \mathcal{R}_\alpha[a, b]$, $m \leq f \leq M$, *and* ϕ *is continuous on* $[m, M]$. *Then* $h := \phi \circ f \in \mathcal{R}_\alpha[a, b]$.

证明. We proceed by verifying the integrability criterion (upper/lower sums can be made arbitrarily small).

Step 1: Uniform Continuity of ϕ Fix $\varepsilon > 0$. Since ϕ is continuous on the compact interval $[m, M]$, it is *uniformly continuous* on $[m, M]$. Thus, there exists $\delta \in (0, \varepsilon)$ such that:

$$|x - y| < \delta \implies |\phi(x) - \phi(y)| < \varepsilon \quad \forall x, y \in [m, M].$$

Step 2: Choose a Partition for f Since $f \in \mathcal{R}_\alpha[a, b]$, there exists a partition $P = \{t_0, t_1, \ldots, t_n\}$ of $[a, b]$ with:

$$U(P, f, \alpha) - L(P, f, \alpha) < \delta^2,$$

where $M_i^f = \sup_{[t_{i-1}, t_i]} f$, $m_i^f = \inf_{[t_{i-1}, t_i]} f$, and $\Delta\alpha_i = \alpha(t_i) - \alpha(t_{i-1})$.

Step 3: Bound α-Increments on "Large Gap" Subintervals Split the subintervals of P into: - $S_1 = \{i : M_i^f - m_i^f < \delta\}$ (small variation in f), - $S_2 = \{i : M_i^f - m_i^f \geq \delta\}$ (large variation in f).

For S_2:

$$\delta \sum_{i \in S_2} \Delta\alpha_i \leq \sum_{i \in S_2} (M_i^f - m_i^f)\Delta\alpha_i \leq U(P, f, \alpha) - L(P, f, \alpha) < \delta^2.$$

Dividing by δ gives:

$$\sum_{i\in S_2} \Delta\alpha_i < \delta < \varepsilon.$$

Step 4: Bound Upper/Lower Sum Gap for $h = \phi \circ f$ Let $M_i^h = \sup_{[t_{i-1},t_i]}(\phi \circ f)$ and $m_i^h = \inf_{[t_{i-1},t_i]}(\phi \circ f)$. We split the sum for $U(P,h,\alpha) - L(P,h,\alpha)$:

$$U(P,h,\alpha) - L(P,h,\alpha) = \sum_{i\in S_1}(M_i^h - m_i^h)\Delta\alpha_i + \sum_{i\in S_2}(M_i^h - m_i^h)\Delta\alpha_i.$$

- For S_1: By uniform continuity of ϕ, $M_i^h - m_i^h < \varepsilon$, so:

$$\sum_{i\in S_1}(M_i^h - m_i^h)\Delta\alpha_i < \varepsilon \sum_{i\in S_1} \Delta\alpha_i \le \varepsilon(\alpha(b) - \alpha(a)).$$

- For S_2: Let $M = \sup_{x\in[a,b]} |\phi(x)|$ (finite, as ϕ is continuous on compact $[m, M]$). Then $M_i^h - m_i^h \le 2M$, so:

$$\sum_{i\in S_2}(M_i^h - m_i^h)\Delta\alpha_i \le 2M \sum_{i\in S_2} \Delta\alpha_i < 2M\varepsilon.$$

Step 5: Conclude Integrability Combining these bounds:

$$U(P,h,\alpha) - L(P,h,\alpha) < \varepsilon(\alpha(b) - \alpha(a)) + 2M\varepsilon = \varepsilon\left(\alpha(b) - \alpha(a) + 2M\right).$$

Since $\varepsilon > 0$ is arbitrary, $h = \phi \circ f \in \mathcal{R}_\alpha[a,b]$. □

The theorem shows that *composing* a Riemann-Stieltjes integrable function f with a continuous function ϕ preserves integrability.

Properties of the integral

Let $f, f_1, f_2 \in \mathcal{R}_\alpha[a,b]$ and $c \in \mathbb{R}$. Then:

(a) $f_1 + f_2 \in \mathcal{R}_\alpha[a,b]$ and $cf \in \mathcal{R}_\alpha[a,b]$, with $\int_a^b (f_1+f_2)\,d\alpha = \int_a^b f_1\,d\alpha + \int_a^b f_2\,d\alpha$, $\int_a^b cf\,d\alpha = c\int_a^b f\,d\alpha$.

(b) If $f_1 \le f_2$ on $[a,b]$, then $\int_a^b f_1\,d\alpha \le \int_a^b f_2\,d\alpha$.

(c) If $a < c < b$ and $f \in \mathcal{R}_\alpha[a,b]$, then $f \in \mathcal{R}_\alpha[a,c] \cap \mathcal{R}_\alpha[c,b]$ and $\int_a^b f\,d\alpha = \int_a^c f\,d\alpha + \int_c^b f\,d\alpha$.

(d) If $f \in \mathcal{R}_\alpha[a,b]$ and $|f| \le M$ on $[a,b]$, then $\left|\int_a^b f\,d\alpha\right| \le M\left(\alpha(b) - \alpha(a)\right)$.

And we also have these properties:

Property

If $f, g \in \mathcal{R}_\alpha[a,b]$, then

(a) $fg \in \mathcal{R}_\alpha[a,b]$;

(b) $\left|\int_a^b f\,d\alpha\right| \le \int_a^b |f|\,d\alpha$.

证明. (a) Since the function $x \mapsto x^2$ is continuous, Theorem 11 (composition of continuous and integrable functions) implies $f^2, g^2 \in \mathcal{R}_\alpha[a,b]$. Using the algebraic identity:

$$fg = \frac{1}{4}\left((f+g)^2 - (f-g)^2\right),$$

and the linearity of the Riemann-Stieltjes integral (integrable functions are closed under linear combinations), we conclude $fg \in \mathcal{R}_\alpha[a,b]$.

(b) Since the function $x \mapsto |x|$ is continuous, Theorem 11 gives $|f| \in \mathcal{R}_\alpha[a,b]$. Define $c = 1$ if $\int_a^b f\,d\alpha \ge 0$, and $c = -1$ otherwise. By definition of absolute value:

$$cf \le |f|.$$

By the monotonicity property of the Riemann-Stieltjes integral (if $h \le k$ and both are integrable, then $\int_a^b h\,d\alpha \le \int_a^b k\,d\alpha$):

$$\left|\int_a^b f\,d\alpha\right| = c\int_a^b f\,d\alpha = \int_a^b cf\,d\alpha \le \int_a^b |f|\,d\alpha.$$

□

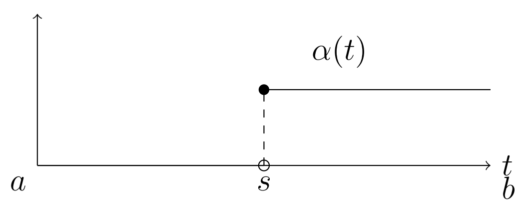
0.28 Integrals against step functions

Let $a < b$ and fix $s \in (a,b)$. Define the step function

$$\alpha(t) = \begin{cases} 0, & t < s, \\ 1, & t \ge s. \end{cases}$$

Claim

Claim If f is continuous at s, then $f \in \mathcal{R}_\alpha[a,b]$ and $\displaystyle\int_a^b f\,d\alpha = f(s)$.



证明. Take a partition $P = \{t_0 = a < t_1 < s < t_2 < t_3 = b\}$ (any partition that isolates s by one subinterval works). Since α is constant on each subinterval except $[t_1, t_2]$, the upper and lower α-sums reduce to

$$U(P, f, \alpha) = (\alpha(t_2) - \alpha(t_1)) \sup_{x\in[t_1,t_2]} f(x) = \sup_{x\in[t_1,t_2]} f(x),$$

and

$$L(P, f, \alpha) = (\alpha(t_2) - \alpha(t_1)) \inf_{x\in[t_1,t_2]} f(x) = \inf_{x\in[t_1,t_2]} f(x).$$

By continuity of f at s, for any $\varepsilon > 0$ we can choose t_1, t_2 with $t_1 < s < t_2$ so close to s that $\sup_{[t_1,t_2]} f - \inf_{[t_1,t_2]} f < \varepsilon$. Hence $U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$, proving integrability. Letting $t_1, t_2 \to s$, both U and L tend to $f(s)$, so the integral equals $f(s)$. □

In the above claim, we can understand it in this way: When $\alpha$is in the continue interval belongs to the $t < s$ and $t > s$, the $d\alpha$ should be 0(Cause $\alpha$is always a constant).Only in the "Jump Point" can they "abruptly" change their values from 0 to 1 and we can get the definite integral value.

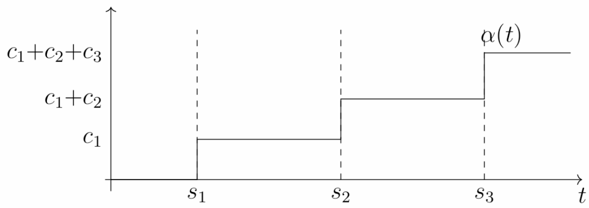
$\alpha(t)$: It is a function composed of a series of "jump points" s_n, and the **jump amplitude** at each jump point s_n is c_n. For example, if $\alpha(t)$ jumps from 0 to c_1 at s_1, from c_1 to $c_1 + c_2$ at s_2, and so on, then c_n is the **increment** at the n-th jump point s_n.

And we can yeild an important theorem:

Theorem

If f is continuous on $[a, b]$, then f is Riemann-Stieltjes integrable with respect to α, and the integral is given by:

$$\int_a^b f\, d\alpha = \sum_{n=1}^{\infty} c_n f(s_n)$$



That is: The Riemann-Stieltjes (R-S) integral of a continuous function with respect to the step function α is equal to the sum over all jump points, where each term is the function value $f(s_n)$ at jump point s_n multiplied by the jump amplitude c_n at that point.

【Remark】:

Simplified Key Takeaway:

You can have **infinite jump points**—as long as each jump is so small that their total size is finite (like infinite tiny steps adding up to 1 meter).

You **cannot have infinite total jump**—even with infinite jump points, if their total size is infinite, the integral "blows up" and is meaningless.

0.29 Differentiable integrator

Assume α is increasing on $[a,b]$, differentiable on (a,b) and $\alpha' \in \mathcal{R}[a,b]$ (usual Riemann sense).

Theorem

Let f be bounded on $[a,b]$. Then $f \in \mathcal{R}_\alpha[a,b]$ if and only if $f\alpha' \in \mathcal{R}[a,b]$, and in that case

$$\int_a^b f\,d\alpha = \int_a^b f(t)\alpha'(t)\,dt.$$

The significance of this theorem lies in establishing a **connection** between the Riemann-Stieltjes integral and the ordinary Riemann integral

Notice that the **mean value theorem** plays an important role in the proof.

证明. Let $M = \sup_{[a,b]} |f|$. For a partition $P = \{t_i\}$ write $\Delta t_i = t_i - t_{i-1}$ and $\Delta\alpha_i = \alpha(t_i) - \alpha(t_{i-1})$. By the mean value theorem, for each i there exists $\xi_i \in (t_{i-1}, t_i)$ such that $\Delta\alpha_i = \alpha'(\xi_i)\Delta t_i$. Hence,

$$U(P,f,\alpha) = \sum_i \left(\sup_{x\in[t_{i-1},t_i]} f(x) \right) \Delta\alpha_i = \sum_i \left(\sup_{x\in[t_{i-1},t_i]} f(x) \right) \alpha'(\xi_i)\Delta t_i,$$

$$L(P,f,\alpha) = \sum_i \left(\inf_{x\in[t_{i-1},t_i]} f(x) \right) \alpha'(\xi_i)\Delta t_i.$$

Because α' is Riemann integrable, there exists a partition P such that $U(P,\alpha',\mathrm{id}) - L(P,\alpha',\mathrm{id}) < \varepsilon$. Choosing any $\eta_i \in [t_{i-1}, t_i]$ we get

$$\begin{aligned}
\left|\sum f(\eta_i)\Delta\alpha_i - \sum f(\eta_i)\alpha'(\eta_i)\Delta t_i\right| = \left|\sum f(\eta_i)\left(\alpha'(\xi_i) - \alpha'(\eta_i)\right)\Delta t_i\right| &\le \sum |f(\eta_i)|\,|\alpha'(\xi_i) - \alpha'(\eta_i)|\,\Delta t_i \\
&\le M \sum |\alpha'(\xi_i) - \alpha'(\eta_i)|\,\Delta t_i \\
&\le M\left(U(P,\alpha',\mathrm{id}) - L(P,\alpha',\mathrm{id})\right) \\
&\le M\varepsilon.
\end{aligned}$$

So,

$$\sum f(\eta_i)\Delta\alpha_i \leq U(P, f\alpha', \mathrm{id}) + M\varepsilon.$$

Taking sup over $\eta_i \in [t_{i-1}, t_i]$ gives

$$U(P, f, \alpha) \leq U(P, f\alpha', \mathrm{id}) + M\varepsilon.$$

But note that for any partition P', because $P \cup P'$ is a refinement of P, we still have

$$U(P \cup P', \alpha', \mathrm{id}) - L(P \cup P', \alpha', \mathrm{id}) < \varepsilon.$$

Therefore,

$$U(P', f\alpha') \geq U(P \cup P', f\alpha') \geq U(P \cup P', f, \alpha) - M\varepsilon \geq \overline{\int_a^b} f d\alpha - M\varepsilon.$$

By arbitrariness of P' and ε, we have

$$\overline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f\alpha' dt.$$

Repeating the same argument to lower sum, we conclude

$$\overline{\int_a^b} f d\alpha = \overline{\int_a^b} f\alpha' dt.$$

We can use the same method to prove the identity of lower integrals. Thus the integrability are equivalent and the value of integrals are the same. □

0.30 Integration and differentiation(usual Riemann case)

In this section take $\alpha(t) = t$ and write $\int f\, d\alpha = \int f(t)\, dt$.

We can apply MVT to prove one of the most important theorems in Calculus:

Theorem

Newton-Leibniz Formula(Fundamental Theorem of Calculus). If $f \in \mathcal{R}[a, b]$ and there **exists** a *differentiable* F on $[a, b]$ with $F' = f$, then

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

证明. Let $\varepsilon > 0$ and choose a partition $P = \{t_i\}$ so that the Riemann upper and lower sums of f satisfy $U(P, f) - L(P, f) < \varepsilon$. By the mean value theorem, for each subinterval there is $s_i \in (t_{i-1}, t_i)$ with $F(t_i) - F(t_{i-1}) = F'(s_i)(t_i - t_{i-1}) = f(s_i)(t_i - t_{i-1})$. Therefore

$$\sum_i \inf_{[t_{i-1},t_i]} \times f\,(t_i - t_{i-1}) \le F(b) - F(a) \le \sum_i \sup_{[t_{i-1},t_i]} \times f\,(t_i - t_{i-1}).$$

Letting the mesh of P go to 0 (hence $U - L \to 0$) yields $F(b) - F(a) = \int_a^b f$. $\square$

Fact and proposition

Let $f \in \mathcal{R}[a, b]$ and define $F(x) = \int_a^x f(t)\,dt$ for $x \in [a, b]$,then F is continuous on $[a, b]$.

If f is continuous at $x_0 \in (a, b)$, then F is differentiable at x_0 and $F'(x_0) = f(x_0)$.

证明. Let $M = \sup_{[a,b]} |f|$. For $x, y \in [a, b]$ with $x < y$,

$$|F(y) - F(x)| = \left| \int_x^y f(t)\,dt \right| \le M|y - x|,$$

so F is (uniformly) continuous. If f is continuous at x_0, given $\varepsilon > 0$ choose $\delta > 0$ so that $|f(t) - f(x_0)| < \varepsilon$ whenever $|t - x_0| < \delta$. For $0 < |h| < \delta$,

$$\left| \frac{F(x_0 + h) - F(x_0)}{h} - f(x_0) \right| = \left| \frac{1}{h} \int_{x_0}^{x_0+h} (f(t) - f(x_0))\,dt \right| \le \sup_{|t-x_0|<\delta} |f(t) - f(x_0)| < \varepsilon,$$

which proves $F'(x_0) = f(x_0)$. $\square$

Method

(Integration by parts). Suppose F and G are differentiable on (a, b) and let $F' = f$, $G' = g$, with $f, g \in \mathcal{R}[a, b]$. Then

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

And we can extend this theorem above to:

Method

(Integration by parts). Suppose α is continuous on $[a, b]$ and that α, β are increasing on $[a, b]$. Then

$$\int_a^b \beta(t)\,d\alpha(t) = \beta(b)\alpha(b) - \beta(a)\alpha(a) - \int_a^b \alpha(t)\,d\beta(t).$$

We know $\beta \in \mathcal{R}_\alpha[a,b]$ (we do not need β to be continuous). On the other hand, α is continuous, hence $\alpha \in \mathcal{R}_\beta[a,b]$.

Fact and proposition

(Change of variable formula). Let $\varphi : [A,B] \to [a,b]$ be strictly increasing and continuous, mapping $[A,B]$ onto $[a,b]$. Suppose α is increasing on $[a,b]$ and $f \in \mathcal{R}_\alpha[a,b]$. Define

$$\beta(y) = \alpha(\varphi(y)), \quad g(y) = f(\varphi(y)).$$

Then $g \in \mathcal{R}_\beta[A,B]$ and

$$\int_A^B g\,d\beta = \int_a^b f\,d\alpha.$$

This builds a bridge between integrals on different spaces. In particular, when φ is differentiable, the formula reduces to the usual change of variables

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

This theorem is a generalized change of variable formula for Riemann-Stieltjes integrals. And the "first substitution method" (also called "integration by substitution" or "u-substitution") is a special case of this theorem

0.30.1 The space of integrable functions.

Definition

For $f \in \mathcal{R}_\alpha[a,b]$, define

$$\|f\|_2 = \left(\int_a^b |f(t)|^2\,d\alpha(t)\right)^{1/2}, \quad d(f,g) = \|f-g\|_2.$$

Theorem

(Approximation by continuous functions) :If $f \in \mathcal{R}_\alpha[a,b]$, then for any $\varepsilon > 0$ there ***exists*** a continuous function g on $[a,b]$ such that $d(f,g) < \varepsilon$.

Hint: Construct a continuous function through "linear interpolation" as an approximation object for f.

证明. Pick a partition $P = \{t_0, \ldots, t_n\}$ such that

$$U(P, f, \alpha) - L(P, f, \alpha) = \sum_{i=1}^{n} \left(M_i^f - m_i^f \right) \Delta\alpha_i < \varepsilon,$$

where $M_i^f = \sup_{[t_{i-1}, t_i]} f$ and $m_i^f = \inf_{[t_{i-1}, t_i]} f$. Define g by linear interpolation on each $[t_{i-1}, t_i]$:

$$g(t) = \frac{t - t_{i-1}}{\Delta t_i} f(t_i) + \frac{t_i - t}{\Delta t_i} f(t_{i-1}), \quad t \in [t_{i-1}, t_i], \ \Delta t_i = t_i - t_{i-1}.$$

Then g is continuous on $[a, b]$. Using that for $t \in [t_{i-1}, t_i]$ we have $|f(t) - f(t_{i-1})| \leq M_i^f - m_i^f$ and $|f(t) - f(t_i)| \leq M_i^f - m_i^f$, one obtains

$$\int_a^b |g - f|^2 d\alpha \leq \sum_{i=1}^{n} \left(M_i^f - m_i^f \right)^2 \Delta\alpha_i \leq 2M \sum_{i=1}^{n} \left(M_i^f - m_i^f \right) \Delta\alpha_i < 2M\varepsilon,$$

where $M = \sup_{t \in [a,b]} |f(t)|$. Hence $\|g - f\|_2 < \sqrt{2M\varepsilon}$. Given $\varepsilon > 0$, scale the preceding choice to ensure $\|g - f\|_2 < \varepsilon$. $\square$

0.31 Convergence in sequence or series of functions

Definition

Let E be a nonempty set and $(f_n)_{n \in \mathbb{N}}$ a **sequence of functions** $f_n : E \to \mathbb{R}$.

(1) We say that (f_n) **converges *pointwise*** to a function $f : E \to \mathbb{R}$ if for every $x \in E$,

$$f_n(x) \to f(x) \quad \text{(as a sequence in } \mathbb{R}).$$

(2) The series $\sum_{n=1}^{\infty} f_n$ is said to **converge *pointwise*** to a function $S : E \to \mathbb{R}$ if for every $x \in E$ the series $\sum_{n=1}^{\infty} f_n(x)$ converges (in $\mathbb{R}$) and

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

Example

Let $E = [0, \infty)$ and $f_n(x) = e^{-nx}$. Then:

(i) $f_n \to f$ pointwise with

$$f(x) = \begin{cases} 0, & x > 0, \\ 1, & x = 0. \end{cases}$$

(ii) For $x > 0$ we have the geometric series

$$\sum_{n=1}^{\infty} f_n(x) = \sum_{n=1}^{\infty} e^{-nx} = \frac{e^{-x}}{1 - e^{-x}},$$

while the series diverges at $x = 0$. Hence it does *not* converge pointwise on $[0, \infty)$.
(iii) However, it converges pointwise to the function $S(x) = \frac{e^{-x}}{1-e^{-x}}$ on $(0, \infty)$.

Property

(Pointwise convergence, in ε–N form). (f_n) *converges to* f *pointwise on* E *if and only if* for every $x \in E$ and every $\varepsilon > 0$ there exists $N = N_{\varepsilon,x} \in \mathbb{N}$ such that for all $n \geq N$,

$$|f_n(x) - f(x)| < \varepsilon.$$

【Remark】:

The index N above generally depends on both ε and the *point* x.
If x changes, the same N may fail (recall a similar situation for continuity).

Definition

Let E be a **nonempty set**, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions $f_n : E \to \mathbb{R}$.
We say $f_n \to f$ *uniformly* on E if

$$\forall \varepsilon > 0\, \exists N \in \mathbb{N}\, \text{s.t.}\, \forall x \in E,\, \forall n \geq N, \quad |f_n(x) - f(x)| < \varepsilon.$$

【Remark】:

The index N above **only depends** on both ε and the function f_n, **not** the variable x.

We similarly define the *uniform convergence* of series of functions.

Definition

For each $n \in \mathbb{N}$, define the *partial sum* of the series $\sum_{n=1}^{\infty} f_n$ as

$$S_n(x) = \sum_{k=1}^{n} f_k(x) \quad (x \in E).$$

We say the series $\sum_{n=1}^{\infty} f_n$ converges *uniformly* to a function $S : E \to \mathbb{R}$ on E if the sequence of partial sums $(S_n)_{n\in\mathbb{N}}$ converges uniformly to S on E, i.e.,

$$\forall \varepsilon > 0 \, \exists N \in \mathbb{N} \, \text{s.t.} \, \forall x \in E, \, \forall n \geq N, \quad |S_n(x) - S(x)| < \varepsilon.$$

Equivalently, since $S(x) = \sum_{k=1}^{\infty} f_k(x)$ (pointwise limit of partial sums), the above expression can be written as

$$\forall \varepsilon > 0 \, \exists N \in \mathbb{N} \, \text{s.t.} \, \forall x \in E, \, \forall n \geq N, \quad \left| \sum_{k=1}^{n} f_k(x) - \sum_{k=1}^{\infty} f_k(x) \right| < \varepsilon.$$

The key feature remains: for fixed ε, **the same** N **works for** ***all*** $x \in E$**.** This differs from pointwise convergence of the series, where N may depend on both ε and x.

1. If we vary ε, N may change, this is true for any kind of convergence.
2. If we *fix* ε, the same index N must work for *all* $x \in E$. This is the key difference from pointwise convergence.

Example

For $f_n(x) = e^{-nx}$ on $[0, \infty)$, the convergence $f_n \to f$ described above is *not* uniform. Indeed, taking $\varepsilon = \frac{1}{2}$, for any N put $x = \frac{\log 2}{2N}$; then

$$|f_N(x) - f(x)| = e^{-Nx} > \frac{1}{2}.$$

Method

(Detecting failure of uniform convergence). $f_n \nrightarrow f$ *uniformly* on E if and only if there exists a sequence $(x_n) \subset E$ such that

$$\limsup_{n\to\infty} |f_n(x_n) - f(x_n)| > 0.$$

A metric viewpoint

Definition

For any function $f : E \to \mathbb{R}$, define the (supremum) norm

$$\|f\|_\infty = \sup_{x \in E} |f(x)| \in [0, \infty],$$

And for f, g define:

$$d_\infty(f, g) = \|f - g\|_\infty.$$

It measures the **largest difference** between $f_n(x)$ and $f(x)$ across all points $x \in E$.

【Remark】:

d_∞ satisfies metric axiom whenever it is finite.

证明. Just apply 3 axioms of metric to verify it. □

Fact and proposition

(Norm/metric characterization of uniform convergence). For any functions f_n, f we have

$$f_n \to f \quad \text{uniformly on } E \quad \iff \quad d_\infty(f_n, f) = \sup_{x \in E} |f_n(x) - f(x)| \xrightarrow{n \to \infty} 0.$$

证明. **Necessity** ($\Longrightarrow$): If $f_n \to f$ uniformly on E, then for any $\varepsilon > 0$, there exists N such that for all $n > N$ and $x \in E$, $|f_n(x) - f(x)| < \varepsilon$. Taking supremum over $x \in E$, we get $d_\infty(f_n, f) = \sup_{x \in E} |f_n(x) - f(x)| \le \varepsilon$, so $d_\infty(f_n, f) \to 0$.

Sufficiency ($\Longleftarrow$): If $d_\infty(f_n, f) \to 0$, then for any $\varepsilon > 0$, there exists N such that for all $n > N$, $d_\infty(f_n, f) < \varepsilon$. Thus for all $x \in E$, $|f_n(x) - f(x)| \le d_\infty(f_n, f) < \varepsilon$, so $f_n \to f$ uniformly on E. □

And naturally we can get:To sum up, "$\{f_n\}$ converges in $(C(X), d_\infty)$ with respect to d_∞" *and* "$\{f_n\}$ converges uniformly to f" are **completely equivalent statements**

they just describe the same phenomenon of "a sequence of functions converging synchronously over the entire set" from two different perspectives: "convergence in a metric space" and "uniform convergence in analysis".

A convergence is ***metrizable*** if we can define a distance (metric) such that convergence in the metric is equivalent to the original convergence. Here, d_∞ is that metric for uniform convergence: **uniform convergence is just "convergence in the d_∞ metric."**

So the Fact shows that uniform convergence is a type of convergence that can be metrized.

Almost all convergences we have studied are metrizable, **except pointwise convergence.**

The essence lies in **the contradiction between the "pointwise locality" of its convergence and the "global consistency" of metrics**, and the corresponding topology does not satisfy the necessary conditions for a metrizable topology.

Recall Some Definitions

Cauchy Sequence in a Metric Space: A sequence (f_n) in a metric space (X, d) is Cauchy if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(f_n, f_m) < \varepsilon$ for all $m, n \geq N$.

Completeness of a Metric Space: A metric space (X, d) is complete if every Cauchy sequence in X converges to a point within X.

Theorem

Let $\{f_n\} \subset C([a, b])$ (continuous functions on the ***compact*** interval $[a, b]$). Then:

$\{f_n\}$ converges *uniformly* on $[a, b]$ $\iff$ $\{f_n\}$ is a Cauchy sequence in $(C([a, b]), d_\infty)$.

证明. We prove both directions of the equivalence:

Direction 1: Uniformly convergent $\implies$ Cauchy sequence Suppose $\{f_n\}$ converges uniformly to $f \in C([a, b])$. We need to show $\{f_n\}$ is Cauchy in $(C([a, b]), d_\infty)$.

Fix $\varepsilon > 0$. By uniform convergence:

$$\exists N \in \mathbb{N} \text{ such that } \forall n \geq N, \quad d_\infty(f_n, f) < \frac{\varepsilon}{2}.$$

For any $n, m \geq N$, use the *triangle inequality for the uniform metric*:

$$d_\infty(f_n, f_m) \leq d_\infty(f_n, f) + d_\infty(f, f_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

By definition, this means $\{f_n\}$ is a Cauchy sequence in $(C([a, b]), d_\infty)$.

Direction 2: Cauchy sequence $\implies$ Uniformly convergent Suppose $\{f_n\}$ is a Cauchy sequence in $(C([a, b]), d_\infty)$. We need to show $\{f_n\}$ converges uniformly to some $f \in C([a, b])$.

We split this into 3 substeps:

Substep 2.1: to prove $\{f_n\}$ has a pointwise limit f: For any fixed $x \in [a, b]$, the sequence $\{f_n(x)\} \subset \mathbb{R}$ is a Cauchy sequence in $\mathbb{R}$:

$$|f_n(x) - f_m(x)| \leq \sup_{t \in [a,b]} |f_n(t) - f_m(t)| = d_\infty(f_n, f_m) < \varepsilon \quad (\forall n, m \geq N).$$

Since $\mathbb{R}$ is complete, $\{f_n(x)\}$ converges to a real number—call it $f(x)$. This defines a pointwise limit function $f : [a, b] \to \mathbb{R}$.

Substep 2.2: $\{f_n\}$ converges uniformly to f: We now show the convergence is uniform (not just pointwise). Fix $\varepsilon > 0$. Since $\{f_n\}$ is Cauchy:

$$\exists N \in \mathbb{N} \text{ such that } \forall n, m \geq N, \quad \sup_{t \in [a,b]} |f_n(t) - f_m(t)| < \varepsilon.$$

Fix $n \geq N$ and let $m \to \infty$. For any $t \in [a, b]$, $f_m(t) \to f(t)$, so:

$$|f_n(t) - f(t)| = \lim_{m \to \infty} |f_n(t) - f_m(t)| \leq \varepsilon.$$

This inequality holds for *all* $t \in [a, b]$, so:

$$\sup_{t \in [a,b]} |f_n(t) - f(t)| \leq \varepsilon \quad (\forall n \geq N).$$

By definition, this means $\{f_n\} \to f$ uniformly on $[a, b]$.

Substep 2.3: $f \in C([a, b])$ (so the limit is in the space) A key theorem: *The uniform limit of continuous functions is continuous.* Since each $f_n \in C([a, b])$ and $\{f_n\} \to f$ uniformly, f is continuous on $[a, b]$ (i.e., $f \in C([a, b])$).

Combining Substeps 2.1–2.3: $\{f_n\}$ converges uniformly to $f \in C([a, b])$. □

Key takeaway: The equivalence holds *precisely because* $(C([a, b]), d_\infty)$ is complete (thanks to compactness of $[a, b]$). For non-compact E, we only have:

$$\text{Uniformly convergent} \implies \text{Cauchy}, \quad \text{but not vice versa.}$$

To get the **uniform continuity** of f, we need an additional assumption—most commonly, that **E is a compact metric space** (since any continuous function on a compact set is uniformly continuous).

And we notice an important fact:

Fact and proposition

Let $C(E)$ denote the space of continuous functions on E.
If E is compact, then $(C(E), d_\infty)$ is a complete metric space.

证明. When E is compact:

Every Cauchy sequence $(f_n) \subset C(E)$ converges uniformly to a limit f.

f is continuous (so $f \in C(E)$, the limit stays inside our space).

f is bounded (thanks to E being compact).

This means $(C(E), d_\infty)$ satisfies the definition of a complete metric space! □

Indeed, the uniform limit of continuous functions is continuous (on the whole E), and consequently bounded.

0.32 The properties of Uni-convergence

The advantage of uniform convergence is that it allows us to freely exchange the limit with other operations (such as the Riemann–Stieltjes integral). In this lecture we look at differentiation and integration.

Interchanging limits

Property

(Uniform convergence and limit). Let (X, d) be a metric space and $E \subset X$. For simplicity you may keep $(\mathbb{R}, |\cdot|)$ in mind.
Let (f_n) be functions $E \to \mathbb{R}$, and let x be a limit point of E. Assume for each n the limit

$$A_n = \lim_{t \to x} f_n(t)$$

exists. If $f_n \to f$ uniformly on E, then (A_n) converges, and moreover

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

证明. Fix $\varepsilon > 0$ and take N such that $d_\infty(f_n, f) < \varepsilon$ for all $n \geq N$. For $m, n \geq N$, by the definition of A_m and A_n we can choose t close to x so that $|f_m(t) - A_m| < \varepsilon$ and $|f_n(t) - A_n| < \varepsilon$. Then

$$|A_n - A_m| \leq |A_n - f_n(t)| + |f_n(t) - f_m(t)| + |f_m(t) - A_m| < \varepsilon + d_\infty(f_n, f_m) + \varepsilon \leq 3\varepsilon.$$

Hence (A_n) is Cauchy in $\mathbb{R}$, thus convergent; write $A = \lim_{n\to\infty} A_n$. Using again $d_\infty(f_n, f) < \varepsilon$ for $n \geq N$,

$$|f(t) - A| \leq |f(t) - f_n(t)| + |f_n(t) - A_n| + |A_n - A| < \varepsilon + |f_n(t) - A_n| + |A_n - A|.$$

Let $t \to x$ (so that $|f_n(t) - A_n| \to 0$) and then $n \to \infty$ to deduce $\limsup_{t\to x} |f(t) - A| \leq \varepsilon$. Since $\varepsilon > 0$ is arbitrary, $\lim_{t\to x} f(t) = A$. □

This theorem addresses the problem of the *interchange of limits*: If the sequence of functions (f_n) satisfies certain conditions, then the two operations—“first take the limit with respect to t (yielding the sequence A_n), then take the limit with respect to n” and “first take the limit with respect to n (yielding the function $f(t)$), then take the limit with respect to t”—are interchangeable, i.e.,

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

Conditions:

1. (X, d) is a metric space, $E \subset X$, and x is a *limit point* of E (i.e., there exists a sequence in E that converges to x);

2. For each $n \in \mathbb{N}$, the limit $A_n = \lim_{t \to x} f_n(t)$ exists (i.e., each f_n has a limit at x);

3. $f_n \to f$ is *uniformly convergent* on E.

Conclusion: The sequence (A_n) converges, and $\lim_{n\to\infty} A_n = \lim_{t\to x} f(t)$.

Uniform convergence and continuity

Property

Suppose each f_n is ***continuous*** at $x_0 \in E$, and $f_n \to f$ **uniformly** on E. Then f is also **continuous** at x_0.

证明. Given $\varepsilon > 0$, choose n such that $d_\infty(f_n, f) < \varepsilon$. By continuity of f_n at x_0 there exists $\delta > 0$ with $|f_n(t) - f_n(x_0)| < \varepsilon$ whenever $d(t, x_0) < \delta$. Then for such t,

$$|f(t) - f(x_0)| \leq |f(t) - f_n(t)| + |f_n(t) - f_n(x_0)| + |f_n(x_0) - f(x_0)| < 3\varepsilon.$$

$\square$

Property

Let α be increasing on $[a, b]$. Suppose $f_n \in \mathcal{R}_\alpha[a, b]$ for all $n \in \mathbb{N}$ and $f_n \to f$ uniformly on $[a, b]$. Then $f \in \mathcal{R}_\alpha[a, b]$ and

$$\int_a^b f \, d\alpha = \lim_{n\to\infty} \int_a^b f_n \, d\alpha.$$

证明. Let $\varepsilon_n = \|f_n - f\|_\infty$. Then $\varepsilon_n \to 0$. For any partition $P = \{a = t_0 < t_1 < \cdots < t_m = b\}$, let $m_i' = \inf_{t\in[t_{i-1},t_i]} f_n(t)$ and $m_i = \inf_{t\in[t_{i-1},t_i]} f(t)$. From $|f_n - f| \leq \varepsilon_n$ we have $m_i' \geq m_i - \varepsilon_n$ and $m_i' \leq m_i + \varepsilon_n$. Hence

$$\begin{aligned} L(P, f_n, \alpha) &= \sum_{i=1}^m m_i'(\alpha(t_i) - \alpha(t_{i-1})) \geq \sum_{i=1}^m (m_i - \varepsilon_n)(\alpha(t_i) - \alpha(t_{i-1})) \\ &= L(P, f, \alpha) - \varepsilon_n \left(\alpha(b) - \alpha(a)\right). \end{aligned}$$

Similarly $U(P, f_n, \alpha) \leq U(P, f, \alpha) + \varepsilon_n \left(\alpha(b) - \alpha(a)\right)$. Taking suprema (resp. infima) over P gives

$$\int_a^b f \, d\alpha - \varepsilon_n \left(\alpha(b) - \alpha(a)\right) \leq \int_a^b f_n \, d\alpha \leq \int_a^b f \, d\alpha + \varepsilon_n \left(\alpha(b) - \alpha(a)\right).$$

Letting $n \to \infty$ yields the claim.

□

Method

(Integration by terms). If $f_n \in \mathcal{R}_\alpha[a,b]$ and the series $\sum_{n=1}^\infty f_n$ converges uniformly on $[a,b]$, then its sum f belongs to $\mathcal{R}_\alpha[a,b]$ and

$$\int_a^b \left(\sum_{n=1}^\infty f_n\right) d\alpha = \sum_{n=1}^\infty \int_a^b f_n \, d\alpha.$$

Property

Suppose f_n are differentiable on $[a,b]$. Assume there exists $x_0 \in [a,b]$ such that $\{f_n(x_0)\}$ converges, and that $f_n' \to g$ uniformly on $[a,b]$ for some function g. Then f_n converges uniformly on $[a,b]$ to a function f, and f is differentiable on (a,b) with $f' = g$.

Shortly, we first clarify the theorem' s components (hypotheses → conclusions) and intuition:

Hypotheses:
1. f_n are *differentiable* on $[a,b]$ (so each f_n is continuous on $[a,b]$, since differentiability implies continuity).
2. There exists some $x_0 \in [a,b]$ such that the sequence $\{f_n(x_0)\}$ (values of f_n at x_0) converges to a real number.
3. The derivatives $\{f_n'\}$ converge *uniformly* on $[a,b]$ to a function g.

Conclusions:
1. The sequence $\{f_n\}$ converges *uniformly* on $[a,b]$ to a function f (so f is continuous on $[a,b]$, by uniform convergence preserving continuity).
2. The limit function f is *differentiable* on (a,b), and its derivative equals g (i.e., $f' = g$).

Key Intuition: Uniform convergence of f_n' controls the "rate of change" of f_n across $[a,b]$. Combining this with convergence at a single point x_0 ensures f_n stay close *everywhere* (uniform convergence). We then show the derivative of the uniform limit is the uniform limit of the derivatives (a non-trivial result—this fails if f_n' only converges pointwise!).

证明. We split the proof into two main steps: (1) Prove f_n converges uniformly to f; (2) Prove $f' = g$.

Step 1: Show f_n converges uniformly on $[a,b]$

To prove uniform convergence, we show $\{f_n\}$ is a *Cauchy sequence* in the complete metric space $(C([a,b]), d_\infty)$ (where $d_\infty(f,h) = \sup_{x\in[a,b]} |f(x) - h(x)|$).

Fix $\varepsilon > 0$. By hypotheses:

- $\{f_n(x_0)\}$ is convergent (hence Cauchy): There exists $N_1 \in \mathbb{N}$ such that for all $n, m \geq N_1$, $|f_n(x_0) - f_m(x_0)| < \frac{\varepsilon}{2}$.

- $f_n' \to g$ uniformly (hence $\{f_n'\}$ is Cauchy): There exists $N_2 \in \mathbb{N}$ such that for all $n, m \geq N_2$, $\|f_n' - f_m'\|_\infty < \frac{\varepsilon}{2(b-a)}$.

Let $N = \max\{N_1, N_2\}$. For any $x \in [a,b]$ and $n, m \geq N$, apply the **Mean Value Theorem** to $f_n - f_m$ (which is differentiable on $[a,b]$):

$$f_n(x) - f_m(x) - (f_n(x_0) - f_m(x_0)) = (f_n'(\xi) - f_m'(\xi))\,(x - x_0)$$

for some ξ between x and x_0. Taking absolute values:

$$|f_n(x) - f_m(x)| \leq |f_n(x_0) - f_m(x_0)| + |f_n'(\xi) - f_m'(\xi)| \cdot |x - x_0|.$$

By our choices of N_1 and N_2:

$$|f_n(x) - f_m(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2(b-a)} \cdot (b-a) = \varepsilon.$$

Since this holds for *all* $x \in [a,b]$, we have $\|f_n - f_m\|_\infty < \varepsilon$ for $n, m \geq N$. Thus $\{f_n\}$ is Cauchy in $(C([a,b]), d_\infty)$. By completeness of $C([a,b])$ (compact domain implies completeness), $\{f_n\}$ converges uniformly to some $f \in C([a,b])$.

Step 2: Show f is differentiable on (a,b) with $f' = g$

Fix $x \in (a,b)$; we aim to show $f'(x) = g(x)$. For $t \in [a,b]$, $t \neq x$, define:

$$F_n(t) = \frac{f_n(t) - f_n(x)}{t - x} \quad (\text{for } t \neq x), \quad F_n(x) = f_n'(x).$$

We first show $\{F_n\}$ converges uniformly on $[a,b]$ to:

$$F(t) = \frac{f(t) - f(x)}{t - x} \quad (\text{for } t \neq x), \quad F(x) = g(x).$$

For $t \neq x$: By uniform convergence of $f_n \to f$,

$$|F_n(t) - F(t)| = \frac{|f_n(t) - f(t) - (f_n(x) - f(x))|}{|t - x|}.$$

Apply MVT to $f_n - f$ (continuous on $[a,b]$, differentiable on (a,b)):

$$f_n(t) - f(t) - (f_n(x) - f(x)) = (f_n'(\eta) - g(\eta))(t - x)$$

for some η between t and x. Thus:

$$|F_n(t) - F(t)| = |f_n'(\eta) - g(\eta)| \leq \|f_n' - g\|_\infty.$$

For $t = x$: $|F_n(x) - F(x)| = |f_n'(x) - g(x)| \leq \|f_n' - g\|_\infty$.

Since $f_n' \to g$ uniformly, $\|f_n' - g\|_\infty \to 0$ as $n \to \infty$. Thus $F_n \to F$ **uniformly** on $[a, b]$.

Now, take the limit as $t \to x$: - For each n, $F_n(t) \to F_n(x) = f_n'(x)$ as $t \to x$ (by definition of $f_n'(x)$). - By uniform convergence preserving limits (a theorem: if $F_n \to F$ uniformly and each $F_n(t) \to c_n$ as $t \to x$, then $F(t) \to \lim c_n$ as $t \to x$):

$$\lim_{t \to x} F(t) = \lim_{n \to \infty} F_n(x) = \lim_{n \to \infty} f_n'(x) = g(x).$$

But by definition of $F(t)$:

$$\lim_{t \to x} F(t) = \lim_{t \to x} \frac{f(t) - f(x)}{t - x} = f'(x).$$

Thus $f'(x) = g(x)$ for all $x \in (a, b)$. □

Method

(Differentiation term by term). Suppose:

(i) each f_n is differentiable on $[a, b]$;

(ii) there exists $x_0 \in [a, b]$ such that $\sum_{n=1}^{\infty} f_n(x_0)$ converges;

(iii) $\sum_{n=1}^{\infty} f_n'$ converges uniformly on $[a, b]$.

Then the series $\sum_{n=1}^{\infty} f_n$ converges uniformly on $[a, b]$ to a differentiable function f, and $(\sum_{n=1}^{\infty} f_n)' = \sum_{n=1}^{\infty} f_n'$.

Power series

Let $\sum_{n=0}^{\infty} c_n x^n$ be a power series, and let $R \in [0, \infty]$ denote its radius of convergence (e.g. by the Cauchy–Hadamard formula $R^{-1} = \limsup_{n \to \infty} \sqrt[n]{|c_n|}$).

First we need to recall:

The **Weierstrass M-test** is a sufficient condition for the uniform and absolute convergence of a series of functions.

Method

Weierstrass M-test: Let $\sum_{n=1}^{\infty} f_n(x)$ be a series of functions defined on a set E. There exists a sequence of non-negative constants $\{M_n\}$ such that $|f_n(x)| \leq M_n$ for all $x \in E$ and $n \in \mathbb{N}$.
If the positive-term series $\sum_{n=1}^{\infty} M_n$ converges, then $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly and absolutely on E.

【Remark】:

It is a sufficient but not necessary condition.

Property

For a power series with radius of convergence R the following hold.

(1) For every $0 < r < R$, the series $\sum_{n=0}^{\infty} c_n x^n$ converges uniformly on $[-r, r]$.

(2) The radius of convergence of $\sum_{n=0}^{\infty} n c_n x^{n-1}$ equals R.

(3) The function $f(x) = \sum_{n=0}^{\infty} c_n x^n$ is differentiable for $|x| < R$ and $f'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1}$.

(4) f is C^∞ on $(-R, R)$.

(5) If $-R < a < b < R$, then

$$\int_a^b f(t)\,dt = \sum_{n=0}^{\infty} c_n \frac{b^{n+1} - a^{n+1}}{n+1}.$$

0.33 Arzelà–Ascoli theorem

Uniform convergence is good, but how to obtain such convergence?

Recall: For a compact set, it is sequentially compact: any sequence admits a convergent subsequence (with limit in the set).

Uniform convergence is under the metric space $(C(E), d_\infty)$.

We know in $\mathbb{R}^k$ (with the Euclidean metric) that bounded and closed sets are compact. Is this still true in $(C(E), d_\infty)$?

Boundedness in $(C(E), d_\infty)$

Definition

A sequence $(f_n) \subset C(E)$ is said to be ***uniformly bounded*** *if* there exists a constant C such that
$$\|f_n\|_\infty \le C, \quad \forall n \ge 1.$$
Equivalently, (f_n) is ***uniformly bounded*** *if*
$$\sup_{n\ge 1}\sup_{x\in E}|f_n(x)| < \infty.$$

Question. Does every uniformly bounded sequence have a convergent subsequence? Equivalently, is every bounded closed set compact in the metric space $(C(E), d_\infty)$?

Example

Let $f_n(x) = x^n$, $E = [0,1] \subset \mathbb{R}$.
Then $\|f_n\|_\infty \le 1$, $\forall n \ge 1$.

But for any subsequence (f_{n_k}) we have
$$f_{n_k}(x) \to f(x) := \begin{cases} 0, & 0 \le x < 1, \\ 1, & x = 1. \end{cases}$$
Hence $f \notin C(E)$. Therefore any subsequence of (f_n) **does not** converge under d_∞.

Example

Let
$$f_n(x) = \frac{x^2}{x^2 + (1-nx)^2}, \qquad x \in [0,1].$$
Note that $\|f_n\|_\infty \le 1$. And $f_n(x) \to 0$ pointwise on $[0,1]$.

However, if we take $x_n := 1/n$, then $f_n(x_n) = 1$, so any subsequence cannot converge uniformly.

Criterion of precompactness in $(C(E), d_\infty)$

Definition

A subset $F \subset C(E)$ is said to be ***precompact*** if any sequence $(f_n) \subset F$ has a subsequence (f_{n_k}) which converges *uniformly* (although we do not know whether the limit is still in F).

This definition extends to any metric space (Y, d_Y). **Moreover, one can show that *any set* $F \subset Y$ is *precompact iff* $\overline{F}$ is *compact*.**

Equi-continuity

Definition

(Equi-continuous). Let (X, d) be a metric space, and $F \subset C(X)$. F is said to be **(uniformly) equi-continuous** if for every $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $x, y \in X$ with $d(x, y) < \delta$, we have

$$|f(x) - f(y)| \leq \varepsilon, \quad \forall f \in F.$$

【Remark】:

The key point is that the choice of δ is ***independent of*** f. Equi-continuity forces each $f \in F$ to be uniformly continuous, but it is stronger than that.

Equivalently, F is equi-continuous *iff* for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\sup_{f \in F} \sup_{\substack{x,y \in X \\ d(x,y) < \delta}} |f(x) - f(y)| < \varepsilon.$$

Property

Suppose (f_n) satisfies:

1. For all $n \geq 1$, f_n is differentiable on (a, b) and continuous on $[a, b]$;
2. (f_n') is uniformly bounded, i.e. $|f_n'(x)| \leq C$ for all $x \in (a, b)$ and all $n \geq 1$.

Then $(f_n)_n$ is equi-continuous.

Proof. For any $n \geq 1$ and any $x, y \in [a, b]$, by the mean value theorem there exists $\xi \in (x, y)$ such that

$$f_n(x) - f_n(y) = f_n'(\xi)(x - y).$$

Thus, for any $\varepsilon > 0$, taking $\delta = \varepsilon / C$, whenever $|x - y| < \delta$ we have

$$|f_n(x) - f_n(y)| = |f_n'(\xi)||x - y| \leq C\delta < \varepsilon, \quad \forall n \geq 1.$$

Hence (f_n) is equi-continuous. □

Claim

If (f_n) **converges uniformly** and X is assumed compact, then f_n is **equi-continuous**.

证明. Suppose $f_n \to f$ uniformly, with each $f_n \in C(X)$. Then $f \in C(X)$. In particular, it is uniformly continuous. Thus, for any $\varepsilon > 0$, there exists $\delta_1 > 0$ such that whenever $d(x, y) < \delta_1$ we have

$$|f(x) - f(y)| < \varepsilon.$$

By uniform convergence, there exists N such that $\|f_n - f\|_\infty < \varepsilon$ for all $n \geq N$. In this case, for all $n \geq N$ and all $x, y \in X$ with $d(x, y) < \delta_1$, we have

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f(x)| + |f(x) - f(y)| + |f(y) - f_n(y)| < 3\varepsilon.$$

For $n = 1, 2, \ldots, N-1$, since each f_n is continuous on the compact set X, it is uniformly continuous. Thus we can choose $\delta_2 > 0$ such that $d(x, y) < \delta_2$ implies $|f_n(x) - f_n(y)| < \varepsilon$ for all $1 \leq n \leq N-1$.

Taking $\delta := \min\{\delta_1, \delta_2\}$, we see that whenever $d(x, y) < \delta$,

$$|f_n(x) - f_n(y)| < 3\varepsilon, \forall n \geq 1.$$

Hence (f_n) is equi-continuous. □

Arzelà–Ascoli theorem

Theorem

Let (X, d) be a metric space, and $(f_n) \subset C(X)$. Suppose that (1) X is compact; (2) (f_n) is uniformly bounded; (3) (f_n) is equi-continuous.
Then $(f_n) \subset C(X)$ is precompact, i.e., there exists a subsequence f_{n_k} such that $f_{n_k} \to f$ for some $f \in C(X)$.

证明. We only prove the case $X = [a, b]$; however, we only use the property that X has a countable dense set. In fact, this is satisfied by every compact metric space.

Let $\mathbb{Q} \cap [a, b]$ have the following properties: (1) $\mathbb{Q} \cap [a, b] = \{x_i\}_{i \geq 1}$, i.e., it is countable; (2) For all $x \in [a, b]$, there exists a sequence $(x_{i_k})_k \subset \mathbb{Q} \cap [a, b]$ such that $x_{i_k} \to x$ (i.e., it is dense).

Now fix each $i \geq 1$. The sequence $\{f_n(x_i)\}_{n \geq 1}$ is bounded in $\mathbb{R}$ (because (f_n) is uniformly bounded). By Bolzano–Weierstrass, there exists a convergent subsequence, which we denote by $\{f_{n_k^{(1)}}(x_1)\}_{k \geq 1}$ for $i = 1$.

Note that each x_i may correspond to a different subsequence, but we may construct a common one as follows. Inductively we assume that we have chosen subsequences so that: (i) for each i, $\{f_{n_k^{(i)}}(x_i)\}_{k\geq 1}$ converges; (ii) for each i, $\{f_{n_k^{(i)}}\}_{k\geq 1}$ is a subsequence of $\{f_{n_k^{(i-1)}}\}_{k\geq 1}$.

Then we obtain a diagonal subsequence $f_{m_k} := f_{n_k^{(k)}}$ such that for every $i \geq 1$, $\{f_{m_k}(x_i)\}_{k\geq 1}$ is convergent.

We now aim to extend the convergence on $\{x_i\}_{i\geq 1}$ to the whole interval $[a,b]$, where compactness plays a role.

Fix $\varepsilon > 0$, and pick $\delta_0 > 0$ such that for all $y \in [a,b]$, whenever $|x-y| < \delta_0$ we have

$$|f_n(x) - f_n(y)| < \varepsilon, \forall n \geq 1,$$

by equi-continuity.

Hence we expect that the good convergence property can be extended near $\{x_i\}_{i\geq 1}$ up to a δ_0-radius. Note that

$$[a,b] \subset \bigcup_{i=1}^{\infty} (x_i - \delta_0, x_i + \delta_0)$$

is an open cover; by compactness, we may assume that

$$[a,b] \subset \bigcup_{i=1}^{N} (x_i - \delta_0, x_i + \delta_0)$$

for some finite N.

Now, for any $\varepsilon > 0$, we aim to find $K \in \mathbb{N}$ such that whenever $k, l \geq K$, we have

$$\|f_{m_k} - f_{m_l}\|_\infty < \varepsilon,$$

which proves that (f_{m_k}) is a Cauchy sequence in $(C([a,b]), d_\infty)$ and thus converges uniformly.

Because each $\{f_{m_k}(x_i)\}_{k\geq 1}$ is Cauchy in $\mathbb{R}$, there exists K_i such that for all $k, l \geq K_i$,

$$|f_{m_k}(x_i) - f_{m_l}(x_i)| < \varepsilon, 1 \leq i \leq N.$$

Let $K := \max\{K_1, \dots, K_N\}$. Then for any $k, l \geq K$ and any $x \in [a,b]$, there exists some $i \in \{1, \dots, N\}$ with $|x - x_i| < \delta_0$. Hence

$$\begin{aligned} |f_{m_k}(x) - f_{m_l}(x)| &\leq |f_{m_k}(x) - f_{m_k}(x_i)| + |f_{m_k}(x_i) - f_{m_l}(x_i)| + |f_{m_l}(x_i) - f_{m_l}(x)| \\ &< \varepsilon + \varepsilon + \varepsilon = 3\varepsilon. \end{aligned}$$

Thus $\|f_{m_k} - f_{m_l}\|_\infty < 3\varepsilon$ for all $k, l \geq K$. This proves the uniform convergence, and hence the theorem. □

Weierstrass approximation theorem

Theorem

If f is continuous on $[a,b]$, then there exists a sequence of polynomials $\{P_n\}$ such that $P_n \to f$ on $[a,b]$ uniformly.

证明. First, we may without loss of generality assume that $[a, b] = [0, 1]$ and that $f(a) = f(b) = 0$. Then, by setting $f = 0$ outside $[0, 1]$, we obtain a uniformly continuous function on the whole $[-1, 1]$.

(Think: give the full details of this paragraph.)

Let

$$Q_n(x) = c_n \left(1 - x^2\right)^n, c_n = \frac{1}{\int_{-1}^{1} \left(1 - x^2\right)^n dx},$$

so that

$$\int_{-1}^{1} Q_n(x) dx = 1, Q_n(x) \geq 0.$$

Now let

$$P_n(x) := \int_{-1}^{1} f(x + t) Q_n(t) dt.$$

(Equivalently, $P_n(x) = \int_{-1}^{1} f(t) Q_n(t - x) dt$.)

Then P_n is a polynomial for all $n \geq 1$.

Moreover,

$$\begin{aligned} |P_n(x) - f(x)| &\leq \int_{-1}^{1} |f(x + t) - f(x)| Q_n(t) dt \\ &\leq \int_{|t|<\delta} |f(x + t) - f(x)| Q_n(t) dt + \int_{\delta \leq |t| \leq 1} |f(x + t) - f(x)| Q_n(t) dt. \end{aligned}$$

The first term is controlled by ε if $|f(x + t) - f(x)| < \varepsilon$. By uniform continuity, this is feasible if δ is small enough, because $|t| < \delta \Rightarrow |x + t - x| < \delta$.

The second term is controlled by

$$2\|f\|_\infty c_n \left(1 - \delta^2\right)^n.$$

Note that, by an estimate (related to the Bernoulli/Midterm problem),

$$\left(1 - x^2\right)^n \geq 1 - nx^2 \text{ for } |x| \leq 1.$$

Hence

$$c_n = \frac{1}{\int_{-1}^{1} \left(1 - x^2\right)^n dx} \leq \frac{1}{2 \int_{0}^{1} \left(1 - x^2\right)^n dx} \leq \frac{3}{4\sqrt{n}},$$

so for this choice of δ (which only depends on ε) we can pick N large enough such that

$$c_n \left(1 - \delta^2\right)^n < \varepsilon, \forall n \geq N.$$

Consequently,

$$\|P_n - f\|_\infty < 2\varepsilon, \forall n \geq N.$$

□

【Remark】:

Why choosing Q_n? Note that

$$\int_{-1}^{1} Q_n(x)dx = 1,$$

we keep the total mass unchanged (since $\int_{-1}^{1} Q_n(x)dx = 1$), but $Q_n(x) \to 0$ for $x \neq 0$ and $Q_n(x) \to \infty$ when $x = 0$. So, when n gets larger and larger, the integral

$$\int_{-1}^{1} f(x+t)Q_n(t)dt$$

concentrates on $f(x)$ when $n \to \infty$. This is related to the Dirac "delta" function in physics.

0.34 Multivariable Functions.

Let $f : E \subset \mathbb{R}^n \to \mathbb{R}$.

Another viewpoint. In one dimension, differentiability at x means we can approximate f near x by an affine function with an $o(|t-x|)$ error. This motivates the multivariable definition below.

In multiple dimensions, $f(x+h)$ and $f(x)$ are real numbers but h is a vector. We therefore expect the first-order term to be the inner product with some vector in $\mathbb{R}^n$.

Definition

(Fréchet differentiability at a point). Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$. For $x_0 \in D$, we say that f is differentiable at x_0 if there exists $A \in \mathbb{R}^n$ such that

$$\lim_{h \to 0} \frac{|f(x_0+h) - f(x_0) - A \cdot h|}{|h|} = 0.$$

We then write $A = Df(x_0)$ and call it the (total) derivative at x_0.

Why the definition makes sense:Because D is open, when $|h|$ is small we have $x_0 + h \in D$, so the left-hand side is well-defined. The limit above expresses that f is well-approximated by the affine map $L(y) = A \cdot (y - x_0) + f(x_0)$ near x_0:

$$\lim_{y \to x_0} \frac{|f(y) - L(y)|}{|y - x_0|} = 0.$$

For $n = 1$ the graphs of affine maps are straight lines; for $n = 2$ they are planes; for $n \geq 3$ they are hyperplanes.

Definition

Partial derivatives: For e_i the i-th standard basis vector, we say f has the i-th *partial derivative* at x if the limit

$$\partial_i f(x) := \lim_{t\to 0} \frac{f(x+te_i)-f(x)}{t}$$

exists.

We denote $\nabla f(x) = (\partial_1 f(x), \ldots, \partial_n f(x))^\top$ and call it the **gradient** of f at x.

Corollary & Secondary Conclusion

Differentiability $\implies$ existence of partials and identification: If f is differentiable at x_0, then each $\partial_i f(x_0)$ exists and

$$Df(x_0) = \nabla f(x_0), \quad f(x_0+h) - f(x_0) = \nabla f(x_0)\cdot h + o(|h|).$$

证明. Take $h = te_i$. Then

$$\frac{f(x_0+te_i)-f(x_0)}{t} = \frac{f'(x_0)\cdot(te_i)}{t} + \frac{o(|te_i|)}{t} \to f'(x_0)\cdot e_i.$$

Thus $\partial_i f(x_0) = f'(x_0)\cdot e_i$, and collecting all coordinates gives $f'(x_0) = \nabla f(x_0)$. □ □

Corollary & Secondary Conclusion

Differentiability implies continuity:If f is differentiable at x_0 then f is continuous at x_0.

证明. From the definition, $f(x_0+h)-f(x_0) = f'(x_0)\cdot h + o(|h|)$. Hence $|f(x_0+h)-f(x_0)| \leq |f'(x_0)|\,|h| + o(|h|) \to 0$. □

【Remark】:

(Existence of partials does not imply differentiability).Consider the scalar function

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2}, & (x,y)\neq(0,0), \\ 0, & (x,y)=(0,0). \end{cases}$$

Then $\partial_x f(0,0) = \partial_y f(0,0) = 0$, but f is not continuous at $(0,0)$ (e.g. along $y=x$ we get $f(x,x)=\frac{1}{2}$), hence not differentiable at $(0,0)$.

A sufficient condition:Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$. If all partial derivatives $\partial_i f$ exist in a neighborhood of $x_0 \in D$ and are continuous at x_0 (i.e. $f \in C^1$ near x_0), then f is differentiable at x_0.

Vector-valued functions

If $F = (f^1, \ldots, f^m) : \mathbb{R}^n \to \mathbb{R}^m$ is vector-valued, we look for a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$F(x+h) - F(x) = Ah + o(|h|).$$

Everything is now in vector norm, and the linear map A is represented by matrix multiplication.

Definition

(Jacobian): If F is differentiable at x, then the derivative (Jacobian) is

$$DF(x) = \begin{bmatrix} \partial_1 f^1 & \partial_2 f^1 & \cdots & \partial_n f^1 \\ \partial_1 f^2 & \partial_2 f^2 & \cdots & \partial_n f^2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f^m & \partial_2 f^m & \cdots & \partial_n f^m \end{bmatrix}_x .$$

Consider $f : D \to \mathbb{R}^m$ and $g : E \to \mathbb{R}^k$, where $D \subset \mathbb{R}^n$ and $E \subset \mathbb{R}^m$ are open with $f(D) \subset E$.

Theorem

(Chain rule). Fix $x_0 \in D$. If f is differentiable at x_0 and g is differentiable at $f(x_0)$, then $h := g \circ f$ is differentiable at x_0 and

$$D(g \circ f)(x_0) = Dg(f(x_0))\, Df(x_0) \quad \text{(matrix multiplication)}.$$

证明.

$$\begin{aligned} |h(x) - h(x_0) - Dg(f(x_0))Df(x_0)(x-x_0)| \le{}& |g(f(x)) - g(f(x_0)) - Dg(f(x_0))\,(f(x) - f(x_0))| \\ &+ |Dg(f(x_0))| \cdot |f(x) - f(x_0) - Df(x_0)(x-x_0)|\,. \end{aligned}$$

For every $\varepsilon > 0$, by differentiability of f and we choose $\delta_0 > 0$ s.t. when $|y - y_0| < \delta$, we have $|g(y) - g(y_0) - Dg(y_0)(y - y_0)| \le \varepsilon |y - y_0|$; and for $|x - x_0| < \delta_0$, we have $|f(x) - f(x_0) - Df(x_0)(x - x_0)| < \varepsilon |x - x_0|$.

Further choose $\delta' > 0$ s.t. when $|x - x_0| < \delta'$ we have $|f(x) - f(x_0)| < \delta_0$. Thus, if $|x - x_0| < \min\{\delta_0, \delta'\}$,

$$\begin{aligned} |h(x) - h(x_0) - Dg(f(x_0))Df(x_0)(x-x_0)| &< \varepsilon\,(|f(x) - f(x_0)| + |Dg(f(x_0))|\,|x - x_0|) \\ &\le \varepsilon\,(|f(x) - f(x_0) - Df(x_0)(x-x_0)| + |Df(x_0)|\,|x - x_0|) + \\ &\le \varepsilon\,(\varepsilon + |Df(x_0)| + |Dg(f(x_0))|)\,|x - x_0|. \end{aligned}$$

□

Theorem

MVT - Mean Value Type Inequality: Suppose $D \subset \mathbb{R}^n$ is convex, i.e. for any $p, q \in D$ the segment $[p, q] \subset D$. If $f : D \to \mathbb{R}$ is differentiable on D and $\sup_{x \in D} \|\nabla f(x)\| \leq M$, then

$$|f(p) - f(q)| \leq M\,|p - q| \quad \forall\, p, q \in D.$$

证明. Let $g(t) = f(p + t(q - p))$ on $[0, 1]$. Then $g'(t) = \nabla f(p + t(q - p)) \cdot (q - p)$. By the one-variable mean value theorem,

$$|f(p) - f(q)| = |g(0) - g(1)| \leq \sup_{t \in [0,1]} |g'(t)| \leq \sup_{x \in D} \|\nabla f(x)\|\,|p - q| \leq M\,|p - q|.$$

□

【Remark】:

For vector-valued maps $F : D \to \mathbb{R}^m$ with $\sup_{x \in D} \|DF(x)\| \leq M$, the same argument yields $\|F(p) - F(q)\| \leq M\,\|p - q\|$.

Fact and proposition

The gradient of a multivariate function is a "vector of directional derivatives". Unlike in the case of univariate functions, it is impossible to find a precise point such that the difference equals the "exact contribution" of the gradient at that point. Instead, we can only estimate the upper bound of the change using the upper bound of the gradient norm M. Therefore, what we obtain is an inequality.

In the previous lecture we introduced differentiability of functions $f : \mathbb{R}^n \to \mathbb{R}$ and $f : \mathbb{R}^n \to \mathbb{R}^m$, partial derivatives, gradients, Jacobian matrices, and the chain rule. In this lecture we move to two fundamental results about functions of several variables: the Inverse Function Theorem - when a C^1 map has a C^1 inverse near a point; the Implicit Function Theorem - when an equation $F(x, y) = 0$ defines y as a function of x

Both theorems express the same informal idea: a C^1 map behaves "almost like" its derivative, i.e. an invertible linear map.

Standing notation. Throughout this lecture, $D \subset \mathbb{R}^n$ will denote an open set. If

$f: D \to \mathbb{R}^m$ is differentiable, we write $Df(x)$ for the Jacobian matrix at x:

$$Df(x) = \begin{pmatrix} \partial_1 f^1 & \partial_2 f^1 & \cdots & \partial_n f^1 \\ \partial_1 f^2 & \partial_2 f^2 & \cdots & \partial_n f^2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f^m & \partial_2 f^m & \cdots & \partial_n f^m \end{pmatrix} (x)$$

A matrix $A \in \mathbb{R}^{m \times n}$ can be viewed as a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$. We will use a matrix norm $\|A\|$, defined by

$$\|A\| := \sup_{|h|=1} |Ah|.$$

Different reasonable choices of such norms are all equivalent in finite dimensions, so for our purposes it is not important which one we choose. For this particular choice, we use the fact that, for each matrix $A \in \mathbb{R}^{m \times n}$ and vector $h \in \mathbb{R}^n$:

$$|Ah| \le \|A\| |h|.$$

C^1 maps and differentiability

Recall from last time: If all partial derivatives $\partial_i f^j$ exist on D and are continuous, then f is differentiable on D (the proof used the Mean Value Theorem coordinate by coordinate).

0.35 Inverse Function Theorem

Definition

Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}^m$. We say f is of class C^1 on D if all partial derivatives $\partial_i f^j(x)$ exist for all $x \in D$, $i = 1, ..., n$, $j = 1, ..., m$; every partial derivative $\partial_i f^j$ is continuous on D.

Contraction Mapping Theorem (fixed point theorem)

The proof of the Inverse Function Theorem we use is based on a general fixed point principle.

Definition

Let (X, d) be a metric space. A map $\varphi : X \to X$ is called a contraction if there exists a constant $c \in [0, 1)$ such that

$$d(\varphi(x), \varphi(y)) \le c d(x, y) \quad \text{for all } x, y \in X.$$

Definition

A metric space (X, d) is called complete if every Cauchy sequence in X converges to a limit in X. The standard Euclidean space $(\mathbb{R}^n, |\cdot|)$ is complete.

Theorem

(Contraction Mapping Theorem). Let (X, d) be a complete metric space and let $\varphi : X \to X$ be a contraction. Then there exists a unique point $x \in X$ such that

$$\varphi(x) = x.$$

Such an x is called the fixed point of φ.

证明. This is a midterm problem. □

【Remark】:

Heuristically: the contraction mapping theorem says that if we have a map that "shrinks distances" on a complete space, then repeated iteration eventually converges to a unique equilibrium point. The Inverse Function Theorem will turn the equation $f(x) = y$ into such a fixed point problem.

Inverse Function Theorem

We now restrict to the case $m = n$. So $f : D \to \mathbb{R}^n$ and $Df(x)$ is an $n \times n$ matrix. We write $\|A\|$ for some operator norm on matrices (induced by the Euclidean norm).

Theorem

(Inverse Function Theorem). Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}^n$ be of class C^1. Fix $a \in D$ and set $b = f(a)$. Assume that the Jacobian matrix $Df(a)$ is invertible. Then there exist open sets U, $V \subset \mathbb{R}^n$ such that: $a \in U \subset D$ and $b \in V$; $f : U \to V$ is bijective; the inverse map $g := f^{-1} : V \to U$ is of class C^1 and for every $y \in V$,

$$Dg(y) = (Df(g(y)))^{-1}.$$

【Remark】:

This theorem formalizes the idea that if $Df(a)$ is invertible, then near a the map f behaves like the linear mapping $Df(a)$, which has an inverse. In particular, it is locally injective, and we can use f as a change of coordinates.

【Remark】:

In dimension one, this reduces to: if $f : (a-\delta, a+\delta) \to \mathbb{R}$ is C^1 and $f'(a) \neq 0$, then f is strictly monotone (hence invertible) in some smaller interval and the inverse is C^1 with $(f^{-1})'(f(a)) = 1/f'(a)$.

Idea of the proof

We only give a detailed proof of the existence and uniqueness of x solving $f(x) = y$ for y near b. Differentiability of the inverse can be checked either directly or using the chain rule.

Define the linear isomorphism

$$A := Df(a).$$

To solve $f(x) = y$, we rewrite the equation as

$$x + \underbrace{A^{-1}(y - f(x))}_{\text{small correction}} = x.$$

In other words, x must be a fixed point of the map

$$\Phi_y(x) := x + A^{-1}(y - f(x)).$$

We will choose a small closed ball B around a such that, for all y in a small ball around b, the map $\Phi_y : B \to B$ is a contraction. Then the result follows from the Contraction Mapping Theorem.

Lemma

(Mean-value theorem for multi-variable, vector valued functions). Suppose D is open and convex, i.e., for each $p, q \in D$, the segment connecting p and q, denoted by $[p, q]$, remains in D. Then, if f is differentiable in D and Df is bounded in D, then for any $p, q \in D$,

$$|f(p) - f(q)| \leq \|Df\|_\infty |p - q|.$$

证明. Let $g(t) = f(tp + (1-t)q)$. Then by chain rule, g is differentiable and

By MVT (an inequality) for vector valued, but single variable function, for some $\xi \in (0, 1)$. □

Lemma

(Control on Df near a). Since f is C^1 at a, there exists $r > 0$ such that $U(a, r) \subset D$ and $\|Df(x) - A\| < \frac{1}{2\|A^{-1}\|}$ for all $x \in U(a, r)$.

证明. By continuity of Df at a, $\|Df(x) - A\| \to 0$ as $x \to a$, so we can choose $r > 0$ such that the inequality holds whenever $|x - a| < r$. $\square$

Lemma

(Lipschitz estimate for Φ_y). Let $B := \overline{U_r(a)}$ as above. Then for any fixed $y \in \mathbb{R}^n$ and any $x_1, x_2 \in B$,

$$|\Phi_y(x_1) - \Phi_y(x_2)| \leq \frac{1}{2}|x_1 - x_2|.$$

In particular, Φ_y is a contraction on B (with constant $c = \frac{1}{2}$, independent of y).

证明. For $x_1, x_2 \in B$,

$$\begin{aligned}\Phi_y(x_1) - \Phi_y(x_2) &= (x_1 - x_2) - A^{-1}(f(x_1) - f(x_2)) \\ &= -A^{-1}(f(x_1) - f(x_2) - A(x_1 - x_2)).\end{aligned}$$

By the Mean Value Theorem (multi-variable, vector-valued) used with $f(x) - Ax$, and therefore

$$\begin{aligned}|\Phi_y(x_1) - \Phi_y(x_2)| &\leq \left\|A^{-1}\right\| \|Df(z) - A\| |x_1 - x_2| \\ &\leq \left\|A^{-1}\right\| \cdot \frac{1}{2\left\|A^{-1}\right\|}|x_1 - x_2| \\ &= \frac{1}{2}|x_1 - x_2|.\end{aligned}$$

$\square$

Lemma

(Mapping B into itself). There exists $\rho > 0$ such that for all y with $|y - b| < \rho$ and all $x \in B$, $\Phi_y(x) \in B$.

证明. For $x \in B$ and y near b,

Choose $\rho > 0$ so small that $\|A^{-1}\| \rho \leq \frac{r}{2}$. Then, since $|x - a| \leq r$ for $x \in B$, so $\Phi_y(x) \in B$. $\square$

证明. (Proof of the Inverse Function Theorem (existence and uniqueness)). Let $B = \overline{U_r(a)}$ as above, and let $V := U_\rho(b)$. For each $y \in V$, the map $\Phi_y : B \to B$ is a contraction, so by the Contraction Mapping Theorem it has a unique fixed point $x \in B$. By definition of Φ_y, this fixed point satisfies $f(x) = y$. Thus every $y \in V$ has a unique preimage $x \in B$, so f is bijective from $U := B^\circ$ to V.

One can then check that the inverse $g = f^{-1} : V \to U$ is differentiable and that the formula

$$Dg(y) = (Df(g(y)))^{-1}$$

follows from the chain rule applied to $g \circ f = \text{id}$. We omit these (standard) details here. $\square$

【Remark】:

Geometrically, the theorem says: if the tangent map $Df(a)$ is invertible, then near a the map f is a smooth change of coordinates. It preserves local dimension, local orientation, etc. In more advanced courses, when learning differential geometry, this will be the starting point for defining smooth charts of manifolds.

Implicit Function Theorem

The Implicit Function Theorem is a consequence of the Inverse Function Theorem applied to a suitable map in a larger space.

We first consider a linear case:

Linear model :The IFT is a nonlinear version of a very simple linear fact.

Consider a linear map $F(x,y) = Ax + By - w$, where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and A, B are matrices of appropriate sizes. Assume B is invertible. Then for every x and every $w \in \mathbb{R}^m$ we can solve $Ax + By = w$ uniquely for y: $y = -B^{-1}Ax + B^{-1}w$.

So, **in the linear world, invertibility of B means we can express y as a linear function of x and w.**

IFT says: if a smooth nonlinear function $F(x,y)$ has derivative in y given by an invertible matrix at some point (a,b), then near (a,b) the same phenomenon holds, but now with a nonlinear function g: $F(x, g(x)) = 0$ for x near a, and $g(a) = b$.

In other words: **locally the nonlinear problem behaves like its linearization.**

Theorem

(Implicit Function Theorem) Let $n, m \in \mathbb{N}$, $U \subset \mathbb{R}^{n+m}$ be open, and $F : U \to \mathbb{R}^m$ be C^1. Write $(x,y) \in \mathbb{R}^{n+m}$ with $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$.

Suppose $(a,b) \in U$ satisfies $F(a,b) = 0$ and the $m \times m$ partial derivative matrix $D_yF(a,b)$ is invertible.

Then there exist open sets $U_0 \ni a$ (in $\mathbb{R}^n$) and $V_0 \ni b$ (in $\mathbb{R}^m$), and a unique C^1 map $g : U_0 \to V_0$ such that:$g(a) = b$;For all $x \in U_0$, $F(x, g(x)) = 0$ (and $y = g(x)$ is the only $y \in V_0$ with $F(x,y) = 0$).

Moreover, the derivative of g is:$Dg(x) = -\left(D_yF(x, g(x))\right)^{-1} D_xF(x, g(x))$, where D_xF is the $m \times n$ partial derivative matrix of F with respect to x .

【Remark】:

$D_yF(a,b)$ is invertible means $F_y(a,b) \neq 0$” for $m = 1$, scalar case.

证明. (Idea of proof). Consider the map

$$\Phi(x, y) := (x, F(x, y)).$$

Its derivative at (a, b) is the block matrix

$$D\Phi(a,b) = \begin{pmatrix} I_n & 0 \\ D_xF(a,b) & D_yF(a,b) \end{pmatrix}$$

which is invertible because $D_yF(a, b)$ is invertible (block triangular matrix with invertible diagonal). By the Inverse Function Theorem, there is a local inverse Ψ of Φ near (a, b), and its first component has the form $\Psi_1(u, v)$. By construction, the equation $F(x, y) = 0$ corresponds to $v = 0$, so setting $g(u) := \Psi_2(u, 0)$ produces the desired implicit function. The formula for Dg comes from differentiating $F(x, g(x)) \equiv 0$ and solving for Dg. □

The theorem says: if at a point (a, b) the equation $F(x, y) = 0$ allows us to solve for y in terms of x (because D_yF is invertible), then this remains true in a whole neighborhood, and the solution $y = g(x)$ depends smoothly on x.

In dimension $n = m = 1$, for $F(x, y) = y - f(x)$, the theorem reduces to the usual inverse function theorem.

”At (a, b), $F(x, y) = 0$ allows solving for y in terms of x (because D_yF is invertible)”: The condition $D_yF(a, b)$ (the derivative of F with respect to y, an $m \times m$ matrix) being invertible is the key. In linear algebra, an invertible matrix means the linear map is bijective (one-to-one and onto). For the non-linear function F, this means: near (a, b), the behavior of F (in the y-direction) is ”like” an invertible linear map—so we can ”undo” $F(x, y) = 0$ to write y as a function of x (locally, at (a, b)).

”This remains true in a whole neighborhood”: The IFT does not only guarantee solvability at the single point (a, b). Instead, it says there exist open sets (neighborhoods) U_0 (around in $\mathbb{R}^n$) and V_0 (around b, in $\mathbb{R}^m$) such that:For every $x \in U_0$, there is a unique $y = g(x) \in V_0$ satisfying $F(x, y) = 0$.So solvability is a local property (true near (a, b), not just at (a, b)).

”The solution $y = g(x)$ depends smoothly on x”:”Smoothly” here means g is C^1 (continuo differentiable). This ensures the solution function g behaves nicely: small changes in x lead to small, well-behaved changes in $y = g(x)$ (no abrupt jumps or ”bad” behavior in the function or its derivative).

Example 16 (A simple implicit curve). Let $F(x, y) = x^2 + y^2 - 1$. At the point $(a, b) = (0, 1)$, we have $F(0, 1) = 0$ and

$$D_yF(0, 1) = 2b = 2 \neq 0.$$

Thus near $(0,1)$ we may write the unit circle as the graph of a C^1 function $y = g(x)$ with $g(0) = 1$. The formula above gives

$$g'(x) = -\frac{\partial_x F(x, g(x))}{\partial_y F(x, g(x))} = -\frac{2x}{2g(x)} = -\frac{x}{g(x)},$$

which matches the slope of the circle.

In the previous lecture we stated and proved the Inverse Function Theorem and stated the Implicit Function Theorem as a corollary.

In this lecture we will: • Prove the Implicit Function Theorem (proof in the last lecture notes); • Differentiate under integral sign.

Further understanding of the Implicit Function Theorem

Several simple cases:

Let us specialise **to the simplest nontrivial situation**: $n = m = 1$. So $F : \mathbb{R}^2 \to \mathbb{R}$ and the equation $F(x, y) = 0$ defines a plane curve (at least near some point).

Suppose (a, b) satisfies $F(a, b) = 0$ and $F_y(a, b) \neq 0$. IFT tells us that: ***A*.**near (a, b) the zero set $\{F = 0\}$ is the graph of a function $y = g(x)$ with $g(a) = b$; ***B*.**g is differentiable and $g'(x) = -\frac{F_x(x,g(x))}{F_y(x,g(x))}$.

Near (a, b) we can approximate F by its derivative: $F(a + u, b + v) \approx F_x(a, b)u + F_y(a, b)v$. (Because $F(a, b) = 0$)

This is the equation of a straight line through $(0, 0)$ in the (u, v)-plane. If $F_y(a, b) \neq 0$, we can solve $v = -\frac{F_x(a,b)}{F_y(a,b)}u$, which is a non-vertical line.

If instead $F_y(a, b) = 0$, the tangent line could be vertical, and then we can not solve for y as a function of x (we might still solve for x as a function of y, provided $F_x \neq 0$).

证明. The Implicit Function Theorem (IFT) needs $F_y(a, b) \neq 0$ (invertible) to guarantee uniqueness of y for each x:

If $F_y(a, b) \neq 0$, F is "responsive" to y-changes: small shifts in y push F away from 0, so only one y can keep $F(x, y) = 0$ for a given x.

If $F_y(a, b) = 0$, F is not responsive to y-changes: multiple y-values near b will still keep $F(x, y) = 0$ for the same x. □

And we need several examples:

Example

(Circle). Consider $F(x, y) = x^2 + y^2 - 1$.

On the unit circle, $F(x, y) = 0$. At the point $(0, 1)$, $F_y(0, 1) = 2y\big|_{(0,1)} = 2 \neq 0$.

Thus near $(0, 1)$, the circle can be written as a graph $y = g(x)$. The derivative is $g'(x) = -\frac{F_x(x,g(x))}{F_y(x,g(x))} = -\frac{2x}{2y} = -\frac{x}{y}$. So $g'(0) = 0$.

Geometrically, the circle has a horizontal tangent at $(0, 1)$.

Example

(Failure when the hypothesis breaks). Consider the curve $y^2 = x^3$, i.e., $F(x, y) = y^2 - x^3$. At $(0, 0)$, $F(0, 0) = 0$, but $F_y(0, 0) = 2y\big|_{(0,0)} = 0$.

The IFT does not apply. Indeed, near 0 we have ***two*** branches $y = \pm x^{3/2}$ for $x > 0$, and they do not define a single-valued C^1 function $y = g(x)$ on any interval around 0.

So the failure of $F_y \neq 0$ is really a geometric obstruction.

level sets as smooth surfaces

In general, suppose $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ and $F(a, b) = 0$, and the $m \times m$ matrix $D_yF(a, b)$ is invertible. The zero set $\{F = 0\}$ has (near (a, b)) dimension n and can be parameterised as $(x, g(x))$ for some open set $U_0 \subset \mathbb{R}^n$, where $g : U_0 \to \mathbb{R}^m$ is C^1. This is a local n-dimensional surface in $\mathbb{R}^{n+m}$.

A useful special case: $m = 1$. Then $F : \mathbb{R}^{n+1} \to \mathbb{R}$ and $\{F = 0\}$ is a hypersurface in $\mathbb{R}^{n+1}$. If $\partial_y F(a, b) \neq 0$, IFT says we can write this hypersurface locally as a graph $y = g(x)$.

Let $[a, b]$ and $[c, d]$ be closed intervals in $\mathbb{R}$ and consider a function $\varphi : [a, b] \times [c, d] \to \mathbb{R}$. For each parameter $t \in [c, d]$, we define a function of x: $\varphi_t(x) = \varphi(x, t)$, and the integral $f(t) = \int_a^b \varphi(x, t)dx$.

The basic question is: When can we differentiate $f(t)$ by differentiating under the integral sign? **That is, when do we have $f'(t) = \int_a^b \partial_t\varphi(x, t)dx$?**

Theorem

(Differentiation under the integral sign, local version) Let $[a, b], [c, d] \subset \mathbb{R}$ be closed intervals and let $\varphi : [a, b] \times [c, d] \to \mathbb{R}$. Fix a point $s \in (c, d)$. Assume:
(i) For every $t \in [c, d]$, the function $x \mapsto \varphi(x, t)$ is Riemann integrable on $[a, b]$;
(ii) The partial derivative $\partial_t\varphi(x, s)$ exists for all $x \in [a, b]$, and the function $x \mapsto \partial_t\varphi(x, s)$ is Riemann integrable on $[a, b]$;
(iii) For every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $|t - s| < \delta$ and $x \in [a, b]$, $|\partial_t\varphi(x, t) - \partial_t\varphi(x, s)| < \varepsilon$ (The family $\{\partial_t\varphi(x, \cdot)\}_{x\in[a,b]}$ is equicontinuous).
Define $f(t) = \int_a^b \varphi(x, t)dx$.
Then f is differentiable at $t = s$ and $f'(s) = \int_a^b \partial_t\varphi(x, s)dx$.

证明. Fix $s \in (c, d)$ as in the theorem. We want to compute the limit $\lim_{t\to s} \frac{f(t)-f(s)}{t-s}$.

For $t \neq s$, define $\psi_t(x) = \frac{\varphi(x,t)-\varphi(x,s)}{t-s}$. Then $\frac{f(t)-f(s)}{t-s} = \int_a^b \psi_t(x)dx$.

Now fix $x \in [a, b]$. By the one-variable Mean Value Theorem in the t-variable, there exists a point $u = u(x, t)$ between s and t such that $\varphi(x, t) - \varphi(x, s) = (t - s)\partial_t\varphi(x, u)$. Thus $\psi_t(x) = \partial_t\varphi(x, u)$.

As $t \to s$, we have $u(x, t) \to s$ and, by assumption (iii), $|\psi_t(x) - \partial_t\varphi(x, s)| = |\partial_t\varphi(x, u) - \partial_t\varphi(x, s)| < \varepsilon$ for all $x \in [a, b]$ and all t sufficiently close to s. In other words, ψ_t converges uniformly to $\partial_t\varphi(\cdot, s)$ on $[a, b]$ as $t \to s$.

We now use this uniform convergence to pass the limit inside the integral. Fix $\varepsilon > 0$. By the above, there exists $\delta > 0$ such that whenever $|t - s| < \delta$, $|\psi_t(x) - \partial_t\varphi(x, s)| < \frac{\varepsilon}{b-a}$ for all $x \in [a, b]$. Therefore,

$$\left|\frac{f(t) - f(s)}{t - s} - \int_a^b \partial_t\varphi(x, s)dx\right| = \left|\int_a^b (\psi_t(x) - \partial_t\varphi(x, s))\, dx\right| \leq \int_a^b |\psi_t(x)-\partial_t\varphi(x, s)|dx < \int_a^b \frac{\varepsilon}{b - a}dx =$$

Since $\varepsilon > 0$ is arbitrary, this shows that $\lim_{t\to s} \frac{f(t)-f(s)}{t-s} = \int_a^b \partial_t\varphi(x, s)dx$, which is exactly $f'(s)$. □

2.4 A slightly more global version The theorem above is local at a single point s. In practice, we often want differentiability for all t in an interval. One convenient sufficient condition is:

(Global C^1 version on a rectangle). Let $[a, b], [c, d] \subset \mathbb{R}$ and $\varphi : [a, b] \times [c, d] \to \mathbb{R}$. Assume: φ is continuous on $[a, b] \times [c, d]$; the partial derivative $\partial_t\varphi(x, t)$ exists and is continuous on $[a, b] \times [c, d]$. For $t \in [c, d]$, define $f(t) = \int_a^b \varphi(x, t)dx$. Then f is differentiable on (c, d) and $f'(t) = \int_a^b \partial_t\varphi(x, t)dx$.

【Remark】:

This is just the previous theorem applied at each $s \in (c, d)$, because continuity of $\partial_t \varphi$ on the compact rectangle $[a, b] \times [c, d]$ implies the uniform continuity required in assumption (iii).

0.36 Second Order Derivatives

In the previous lectures we developed the differential calculus of several variables, introduced the Inverse and Implicit Function Theorems, and saw how first derivatives control many local properties of maps.

In this lecture we move one step further and let second derivatives take the lead:

- equality of **mixed second-order partial derivatives** (symmetry of the Hessian);
- **quadratic approximation** and the second derivative test for local extrema;(Taylor)
- constrained extrema and the method of **Lagrange multipliers**.

0.36.1 Intro to SOD

Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$.

Definition

For $i, j \in \{1, \ldots, n\}$ we may consider the iterated partial derivatives

$$D_i f(x) = \frac{\partial f}{\partial x_i}(x), \quad D_j\left(D_i f\right)(x) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(x)\right),$$

when they exist. We write

$$D_{ji} f(x) := D_j\left(D_i f\right)(x), \quad D_{ij} f(x) := D_i\left(D_j f\right)(x).$$

A very natural question is: When do we have $D_{ij} f(x) = D_{ji} f(x)$?

Theorem

(Clairaut’ s theorem):[Equality of mixed partial derivatives] Let $D \subset \mathbb{R}^2$ be open and $f : D \to \mathbb{R}$. Assume that:

- $D_{12} f$ and $D_{21} f$ exist on D;
- $D_{12} f$ and $D_{21} f$ are continuous on D.

Then $D_{12} f(x) = D_{21} f(x)$ for all $x \in D$.

This is sometimes called Clairaut’ s theorem or Schwarz’ s theorem.And it can extend to general n.The same holds in higher dimensions:

Corollary & Secondary Conclusion

If all second-order partial derivatives of f exist and are continuous on $D \subset \mathbb{R}^n$, then $D_{ij}f = D_{ji}f$ for all i, j.

【Remark】:

The **continuity assumption** on the second derivatives is important.
The continuity of mixed partial derivatives ensures that the partial derivatives behave "regularly" near every point in D—they do not oscillate wildly or have abrupt jumps at a specific point.In our setting we will always assume enough smoothness (C^2), so we can safely identify $D_{ij}f$ and $D_{ji}f$.

there are pathological examples where $D_{12}f$ and $D_{21}f$ exist everywhere but disagree at a point if they are not continuous there.

Example

证明. It is enough to prove the identity at a single point. By translating coordinates we may assume that $0 = (0, 0) \in D$ and we prove that $D_{12}f(0, 0) = D_{21}f(0, 0)$.

Let $A = D_{12}f(0, 0)$ and $B = D_{21}f(0, 0)$. Let $\varepsilon > 0$. By continuity of $D_{21}f$ at $(0, 0)$, there exists $\delta > 0$ such that for all $(x, y) \in D$ with $|(x, y)| < \delta$, we have $|D_{21}f(x, y) - B| < \varepsilon$.

For $(h, k) \in \mathbb{R}^2$ define

$$\Delta(h, k) = f(h, k) - f(h, 0) - f(0, k) + f(0, 0).$$

We want to relate $\Delta(h, k)$ to $D_{21}f$. Fix (h, k) and define

$$u(t) := f(t, k) - f(t, 0), \quad t \in \mathbb{R}.$$

Then $\Delta(h, k) = u(h) - u(0)$.

By the one-variable **Mean Value Theorem** there exists x between 0 and h (so $|x| \leq |h|$) such that

$$u(h) - u(0) = hu'(x).$$

And $u'(x) = D_1f(x, k) - D_1f(x, 0)$.

For fixed x, consider the function

$$v(s) := D_1f(x, s), \quad s \in \mathbb{R}.$$

Again by the **Mean Value Theorem** there exists y between 0 and k (so $|y| \leq |k|$) such that

$$v(k) - v(0) = kv'(y) = kD_{21}f(x, y).$$

Thus

$$u'(x) = kD_{21}f(x,y) \implies \Delta(h,k) = hkD_{21}f(x,y),$$

so

$$\left| B - \frac{\Delta(h,k)}{hk} \right| = |B - D_{21}f(x,y)| < \varepsilon. \tag{*}$$

Combining the two steps we obtain for some (x,y) with $|x| \le |h|$ and $|y| \le |k|$:

$$\frac{\Delta(h,k)}{hk} = D_{21}f(x,y).$$

Assume $hk \neq 0$ and $|(h,k)| < \delta/2$. Then $|(x,y)| \le |(h,k)| < \delta$, so by the choice of δ, the inequality (*) holds.

Fix h with $0 < |h| < \delta/2$. For $k \to 0$ with $0 < |k| < \delta/2$, by the definition of the partial derivative $D_2 f$ at $(h,0)$ and $(0,0)$, we have:

$$\lim_{k\to 0} \frac{f(h,k) - f(h,0)}{k} = D_2f(h,0), \quad \lim_{k\to 0} \frac{f(0,k) - f(0,0)}{k} = D_2f(0,0).$$

Hence

$$\lim_{k\to 0} \frac{\Delta(h,k)}{hk} = \frac{D_2f(h,0) - D_2f(0,0)}{h}.$$

Passing to the limit $k \to 0$ in (*) gives

$$\left| B - \frac{D_2f(h,0) - D_2f(0,0)}{h} \right| \le \varepsilon$$

for all sufficiently small h.

But by the definition of $A = D_{12}f(0,0)$, we have

$$\lim_{h\to 0} \frac{D_2f(h,0) - D_2f(0,0)}{h} = A.$$

Taking the limit $h \to 0$ in the previous inequality, we obtain $|B - A| \le \varepsilon$. Since $\varepsilon > 0$ was arbitrary, this implies $A = B$, i.e., $D_{12}f(0,0) = D_{21}f(0,0)$.

This proves the theorem at $(0,0)$; the general case follows by translation. $\square$

reduce the multi-variable problem to single-variable calculus (via the Mean Value Theorem), use a key difference term to connect mixed partial derivatives, and leverage continuity of mixed partials to force their equality at a point

Corollary & Secondary Conclusion

Symmetry of the Hessian: If $f \in C^2(D)$, i.e., all second-order partial derivatives of f exist and are continuous on D, then the matrix

$$D^2 f(a) := \begin{pmatrix} D_{11}f(a) & \cdots & D_{1n}f(a) \\ \vdots & \ddots & \vdots \\ D_{n1}f(a) & \cdots & D_{nn}f(a) \end{pmatrix}$$

is symmetric for every $a \in D$.

From now on we will usually work with C^2 functions, and silently use the fact that the order of differentiation in second-order partials does not matter.

This symmetric matrix $D^2 f(a)$ is called the ***Hessian*** of f at a.

0.36.2 Local extreme

We now turn to local maxima/minima and how derivatives detect them.

Definition

Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$. Let $a \in D$.
- We say that f has a **local minimum** at a if there exists $r > 0$ such that $f(x) \geq f(a)$ for all $x \in U_r(a) \cap D$.
- We say that f has a **local maximum** at a if there exists $r > 0$ such that $f(x) \leq f(a)$ for all $x \in U_r(a) \cap D$.
- A local extremum is either a local minimum or a local maximum.

Definition

Let $f : D \to \mathbb{R}$ be **differentiable** at $a \in D$. We say that a is a ***critical point*** of f if $Df(a) = 0$ (i.e., $\nabla f(a) = 0$).

Theorem

Local extrema are critical points: Let $D \subset \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ be differentiable on D. If $a \in D$ is a local extremum of f, then $Df(a) = 0$, i.e., $\nabla f(a) = 0$.

证明. Easy to prove □

From now on we assume $f \in C^2(D)$.

Hessian and associated quadratic form

Definition

Let $f \in C^2(D)$ and $a \in D$. The ***Hessian matrix*** of f at a is the $n \times n$ matrix

$$D^2 f(a) = (D_{ij} f(a))_{1 \le i,j \le n}.$$

By the previous section, $D^2 f(a)$ is symmetric.

We associate to $D^2 f(a)$ a quadratic form.

Definition

Let M be a symmetric $n \times n$ matrix and $Q(h) = h^\top M h$ the associated quadratic form.
- M (or Q) is **positive definite** if $Q(h) > 0$ for all $h \neq 0$.
- M (or Q) is **negative definite** if $Q(h) < 0$ for all $h \neq 0$.
- M (or Q) is **indefinite** if there exist h, k with $Q(h) > 0$ and $Q(k) < 0$.

Quadratic Taylor expansion

Theorem

Quadratic Taylor approximation: Let $D \subset \mathbb{R}^n$ be open, $f \in C^2(D)$, and $a \in D$. Then there exists $r_0 > 0$ such that $U_{r_0}(a) \subset D$ and for all $h \in \mathbb{R}^n$ with $|h| < r_0$ we have

$$f(a+h) = f(a) + \nabla f(a) \cdot h + \frac{1}{2} h^\top D^2 f(a+ch) h$$

for some $c \in (0,1)$. Equivalently,

$$f(a+h) = f(a) + \nabla f(a) \cdot h + \frac{1}{2} h^\top D^2 f(a) h + R(a,h),$$

where the remainder satisfies $\lim_{h \to 0} \frac{R(a,h)}{|h|^2} = 0$.

Idea of proof. Consider the one-variable function $\varphi(t) = f(a+th)$ for h small enough that $a + th \in D$ for $t \in [0,1]$. Then

$$\varphi'(t) = \nabla f(a+th) \cdot h,$$

$$\varphi''(t) = h^\top D^2 f(a+th) h.$$

Now apply the one-variable Taylor theorem with remainder to φ at $t = 0$: for some $c \in (0,1)$,

$$\varphi(1) = \varphi(0) + \varphi'(0) \cdot 1 + \frac{1}{2} \varphi''(c) \cdot 1^2.$$

This gives

$$f(a+h) = f(a) + \nabla f(a) \cdot h + \frac{1}{2} h^\top D^2 f(a+ch) h.$$

We can write $h^\top D^2 f(a+ch)h = h^\top D^2 f(a)h + h^\top \left(D^2 f(a+ch) - D^2 f(a)\right) h$ and use the continuity of $D^2 f$ to show that the last term is $o(|h|^2)$. □

We can now use the quadratic approximation to classify critical points.

Method

Second derivative test: Let $f \in C^2(D)$ and $a \in D$ be a critical point of f (so $\nabla f(a) = 0$). Let $Q_a(h) = h^\top D^2 f(a) h$ be the quadratic form associated to the Hessian at a.

- If Q_a is **positive definite,** then a is a strict **local minimum** of f.
- If Q_a is **negative definite**, then a is a strict local maximum of f.
- If Q_a is indefinite, then a is a **saddle point** (neither a local minimum nor a local maximum).

证明. Since $\nabla f(a) = 0$, the quadratic approximation reads

$$f(a+h) = f(a) + \frac{1}{2} Q_a(h) + R(a,h),$$

where $\lim_{h\to 0} \frac{R(a,h)}{|h|^2} = 0$.

Assume Q_a is positive definite. Then there exists $\alpha > 0$ such that $Q_a(h) \geq \alpha |h|^2$ for all h (this is a property of positive definite forms). Thus

$$f(a+h) - f(a) \geq \frac{1}{2}\alpha |h|^2 + R(a,h).$$

Given $\varepsilon > 0$, for h small we have $|R(a,h)| \leq \varepsilon |h|^2$. Choosing $\varepsilon < \alpha/4$, we get

$$f(a+h) - f(a) \geq \frac{1}{2}\alpha |h|^2 - \frac{\alpha}{4}|h|^2 = \frac{\alpha}{4}|h|^2 > 0$$

for all sufficiently small $h \neq 0$. Hence $f(a+h) > f(a)$ near a, so a is a strict local minimum.

The negative definite case is analogous.

If Q_a is indefinite, there exist h, k with $Q_a(h) > 0$ and $Q_a(k) < 0$. Taking th and tk with t small and using the same estimate shows that $f(a+th) > f(a)$ for small nonzero t, while $f(a+tk) < f(a)$ for small nonzero t. So a is neither a local minimum nor a local maximum. □

【Remark】:

If Q_a is positive semi-definite or negative semi-definite but not definite, the test is inconclusive in general.

0.36.3 Lagrange Multiplier Theorem

We finally turn to optimization with a constraint.

Constrained extrema

Let $D \subset \mathbb{R}^n$ be open and let $g : D \to \mathbb{R}$ be a C^1 function. We consider the constraint set $M = \{x \in D : g(x) = 0\}$.

We are interested in extrema of $f : D \to \mathbb{R}$ on M, i.e., points $a \in M$ such that $f(a) \geq f(x)$ (or $f(a) \leq f(x)$) for all $x \in M$ near a.

Definition

We say that $a \in M$ is a **local minimum** of f under the constraint $g = 0$ if there exists $r > 0$ such that $f(x) \geq f(a)$ for all $x \in U_r(a) \cap M$.
Local maximum is defined similarly.

Lagrange multiplier theorem

Theorem

Lagrange multipliers, one constraint: Let $D \subset \mathbb{R}^n$ be open and $f, g : D \to \mathbb{R}$ be C^1. Let $a \in D$ satisfy $g(a) = 0$ and $\nabla g(a) \neq 0$. Assume that a is a local extremum of f on the constraint set $M = \{x \in D : g(x) = 0\}$.
Then there exists a scalar $\lambda_0 \in \mathbb{R}$ such that

$$\nabla f(a) + \lambda_0 \nabla g(a) = 0.$$

Equivalently, $\nabla f(a)$ is parallel to $\nabla g(a)$.

Idea of proof in dimension 2. For intuition, assume $n = 2$, $a = (0, 0)$, and D is a neighbourhood of the origin in $\mathbb{R}^2$. Assume $\partial_y g(0, 0) \neq 0$. By the Implicit Function Theorem, near $(0, 0)$ the constraint $g(x, y) = 0$ can be written as $y = \varphi(x)$ for some C^1 function φ with $\varphi(0) = 0$.

Define a one-variable function $h(x) := f(x, \varphi(x))$. The assumption that $(0, 0)$ is a constrained extremum means that $x = 0$ is a local extremum of h, so $h'(0) = 0$.

By the chain rule,

$$h'(x) = \partial_x f(x, \varphi(x)) + \partial_y f(x, \varphi(x)) \cdot \varphi'(x),$$

so $h'(0) = \partial_x f(0, 0) + \partial_y f(0, 0) \cdot \varphi'(0) = 0$.

On the other hand, differentiating $g(x, \varphi(x)) \equiv 0$ with respect to x gives

$$\partial_x g(x, \varphi(x)) + \partial_y g(x, \varphi(x)) \cdot \varphi'(x) = 0,$$

so $\varphi'(0) = -\frac{\partial_x g(0,0)}{\partial_y g(0,0)}$.

Substituting into the expression for $h'(0) = 0$, we find

$$\partial_x f(0,0) + \partial_y f(0,0)\left(-\frac{\partial_x g(0,0)}{\partial_y g(0,0)}\right) = 0,$$

or equivalently

$$\partial_x f(0,0) \cdot \partial_y g(0,0) - \partial_y f(0,0) \cdot \partial_x g(0,0) = 0.$$

This says exactly that the vectors $\nabla f(0,0)$ and $\nabla g(0,0)$ are parallel. Thus there is a λ_0 such that $\nabla f(0,0) + \lambda_0 \nabla g(0,0) = 0$.

In higher dimensions the idea is the same: use the Implicit Function Theorem to parameterise the constraint by $n-1$ free variables, compose f with this parameterisation to get a function of $n-1$ variables, and use the chain rule at a critical point. $\square$

Lagrangian formulation

It is often convenient to package the condition into a single function, the Lagrangian:

$$L(x, \lambda) = f(x) + \lambda g(x),$$

where $x \in D$ and $\lambda \in \mathbb{R}$ is a Lagrange multiplier.

The Lagrange multiplier condition can be written compactly as

$$\nabla_{x,\lambda} L(x, \lambda) = 0,$$

i.e.,

$$\nabla f(x) + \lambda \nabla g(x) = 0 \quad \text{and} \quad g(x) = 0.$$

So constrained critical points of f correspond to unconstrained critical points of L.

【Remark】:

The Lagrange multiplier theorem gives necessary conditions for constrained extrema, not sufficient ones.

To decide whether a constrained critical point is a minimum, maximum, or saddle, one usually combines the Lagrange multiplier condition with a second derivative analysis restricted to the constraint surface. We will not develop the full theory here, but the basic intuition is that one should look at the Hessian of f restricted to directions tangent to the constraint.