数学分析实验班笔记 2022秋

陈风凌

2023 年 2 月 2 日

目录

1 Continuous functions

1.1 Definitions

Definition 1.1 (continuous at a point)**.** Given two metric spaces X, Y and a map $f : X \to Y$, for $x_0 \in X$, if $\forall \varepsilon > 0, \exists \delta > 0$, such that

$$\forall d_X(x, x_0) < \delta, d_Y(f(x), f(x_0)) < \varepsilon$$

we say f is **continuous** at x_0.

Proposition 1.2 (sequential definition of continuity)

The above definition is equivalent as the following definition:

Given two metric spaces X, Y and a map $f : X \to Y$, if for every sequence $\{x_n\}$ that converges to x_0, $\{f(x_n)\}$ converges to $f(x_0)$, then f is continuous at x_0.

Definition 1.3 (continuous maps)**.** If a map $f : X \to Y$ is continuous at every point in X, then f is a continuous map.

Denote by $C([a, b])$ the set of continuous functions on $[a, b]$.

Proposition 1.4

If f, g are continuous functions, we have:

- $af + bg$ is continuous, $\forall a, b \in \mathbb{R}$.
- $f \cdot g$ is continuous.
- if $g(x_0) \neq 0$, $\frac{f(x)}{g(x)}$ is continuous at x_0.
- $g \circ f$ is continuous.(by sequential definition)

Example 1.5

Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2}, & x^2 + y^2 \neq 0 \\ 0, & x = y = 0 \end{cases}$$

Note that for every fixed $x_0 \in \mathbb{R}$, $f(x_0, y)$ is continuous, but $f(x, y)$ is **not** continuous at point $(0, 0)$.

Example 1.6

Consider Dirichlet's function $f : \mathbb{R} \to \mathbb{R}$ such that

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

f is not continuous at any point.

$$g(x) = \begin{cases} \frac{1}{p}, & x = \frac{q}{p} \in \mathbb{Q}, (p, q) = 1 \\ 0, & x \notin \mathbb{Q} \end{cases}$$

g is continuous at every irrational point.

1.2 Properties of continuous functions

Theorem 1.7

An increasing function $f : (a, b) \to \mathbb{R}$ has at most countable many non-continuous points.

Proof. f is non-continuous at $x_0 \iff \lim_{x\to x_0^-} f(x) < \lim_{x\to x_0^+} f(x)$. We map x_0 to the interval $\left(\lim_{x\to x_0^-} f(x), \lim_{x\to x_0^+} f(x)\right)$. Note that these intervals are disjoint from each other. □

Theorem 1.8 (Intermediate value theorem)

f is continuous on $[a, b] \implies \forall c \in [f(a), f(b)]$ (or $[f(b), f(a)]$), $\exists x_0 \in [a, b]$ such that $f(x_0) = c$.

Proof. Use bisection and nested intervals. □

Theorem 1.9

f is continuous on $[a, b] \implies f$ achieves its maximum value on $[a, b]$.

Sketch of the proof. There are two things to prove:

- The image of f is bounded.
- Take the supremum of the image set, prove it can be achieved by f.

both of them can be proved by sequential compactness of $[a, b]$. □

Remark 1.10 — The above two theorem grants that any continuous function's range on a closed interval is a closed interval as well.

Theorem 1.11

The inverse function f^{-1} of a continuous and strictly incresing function $f : [a, b] \to \mathbb{R}$ is also continuous and strictly incresing.

Proof. Clearly f is bijection, so f^{-1} is bijection $\implies$ f^{-1} is strictly inceasing.

For continuity, fix $y_0 = f(x_0)$, we need to show that $\sup\limits_{y<y_0} f^{-1}(y) = \inf\limits_{y>y_0} f^{-1}(y) = x_0$. However this is trivial (assume for contradiction that $\sup\limits_{y<y_0} f^{-1}(y) < x_0$, take $x_1 \in \left(\sup\limits_{y<y_0} f^{-1}(y), x_0\right)$). □

Remark 1.12 — With this theorem we can define functions like $\ln x, \sqrt{x}$ etc.

Example 1.13 ($o(x)$)

- $\ln(1+x) = x + o(x)$ because

$$\lim_{x\to 0} \frac{\ln(1+x)}{e^{\ln(1+x)} - 1} = \lim_{x\to 0} \frac{x}{e^x - 1} = 1.$$

- $(1+x)^\mu = 1 + \mu x + o(x), \mu \in \mathbb{R}, \mu \neq 0$, because

$$\lim_{x\to 0} \frac{e^{\mu \ln(1+x)} - 1}{\mu \ln(1+x)} \cdot \frac{\ln(1+x)}{x} = 1.$$

2 Basic topology

Here we only focus on the topology on Eucild space $\mathbb{R}^n$.

2.1 Definitions

Definition 2.1 (Topological spaces)**.** Given a pair (X, A), where X is a set and A is a collection of subset of X (these subsets are called the open sets). It's a **topological space** if:

- $\varnothing, X \in A$;
- The union of any number of open sets is still an open set.
- The intersection of finite number of open sets is still an open set.

Definition 2.2 (Open sets in $\mathbb{R}^n$ (or general metric spaces))**.** A subset $A \in \mathbb{R}^n$ is an open set if $\forall x_0 \in A$, $\exists \varepsilon > 0$, *s.t.* $B_\varepsilon(x_0) := \{x \mid |x - x_0| < \varepsilon\} \subset A$ (that is to say, there exists a neighborhood of x_0 contained in A).

We can check that the above definition of open sets in $\mathbb{R}^n$ forms a topology.

Proposition 2.3

Any open set in $\mathbb{R}$ can be written as a union of countable many disjoint of open intervals.

Proof. Let I be our open set. $\forall x \in I$, define

$$a_x := \inf\{z \mid (z, x) \subset I\}, \quad b_x := \{z \mid (x, z) \subset I\}$$

(note that I is open, so these sets are nonempty). Then we have $(a_x, b_x) \subset I$, $a_x, b_x \notin I$.

Thus $I = \bigcup\limits_{x \in I} (a_x, b_x)$.

Claim — $\forall x, y \in I$, either $(a_x, b_x) = (a_y, b_y)$ or $(a_x, b_x) \cap (a_y, b_y) = \varnothing$.

WLOG $a_x \leq a_y$, if $a_y \geq b_x$, the claim is already true. Otherwise $a_y < b_x$, we deduce that $a_y = a_x$(otherwise $a_y \in I$). Similarly $b_x = b_y$, this completes the proof. □

Hence I can be written as a union of disjoint open intervals, the number is clearly at most countable many. □

2.2 More definitions

Definition 2.4 (accumulation points)**.** For $X \subset \mathbb{R}^n$, $x_0 \in \mathbb{R}^n$ is an **accumulation point** of X if:

$$\exists \{x_k\}_{k \geq 1} \subset X, x_k \neq x_0, \ s.t. \ x_k \to x_0.$$

The **derived set** X' of X is the set of all the accumulation points of X, $\overline{X} := X \cup X'$ is the **closure** of X.

For $x \in X$, if there's a neighborhood of x which does not contain any other point in X, we say x is an **isolated point**. It's easy to see that a point in X is either an isolated point or an accumulation point.

Definition 2.5 (Closed sets)**.** A **closed set** is the completion of an open set.

Proposition 2.6

X is closed iff for every convergent sequence in X converges to a point in X.

Proof. Assume for contradiction that X^c is not open. Then there exists $y_0 \in X^c$, $\forall n > 1, B_{\frac{1}{n}}(y_0) \cap X$ is nonempty.

Take $y_n \in B_{\frac{1}{n}}(y_0) \cap X$, $y_n \to y_0 \implies y_0 \in X$, contradiction!

The other direction is trivial. □

Proposition 2.7

The closure of a set X is the smallest closed set containing X.

Proof. Need to prove that $\forall x_0 \in \overline{X}, \exists \varepsilon > 0,\ s.t.\ B_\varepsilon(x_0) \cap \overline{X} = \varnothing$.

Otherwise we have $x_n \in X \cup X'$ such that $x_n \to x_0$. if $x_n \in X'$, take $y_n \in X, |y_n - x_n| < \frac{1}{n}$, $\implies y_n \to x_0 \implies x_0 \in X'$.

Therefore $\overline{X}$ is closed. It's clear that any closed set containing X must contains $\overline{X}$. □

The **interior points** $\mathring{X}$ is the largest open set contained in X.

The **boundary** $\partial X := \overline{X} \backslash \mathring{X}$.

Definition 2.8 (Continuous map for topological spaces)**.** X, Y are topological spaces, we say a function $f : X \to Y$ is continuous if the pre-image of any open set in Y is open in X, i.e. $B \subset Y$ is open $\implies f^{-1}(B) \subset X$ is open.

Remark 2.9 — One can check that the definitions of continuous map in metric spaces and topological spaces are equivalent.

Proof. There are two directions to prove:

- metric $\implies$ topological

  Take an open set $U \in Y$, $\forall x \in f^{-1}(U), f(x) \in U$,

  $$\implies \exists \varepsilon > 0\ s.t.\ B_\varepsilon(f(x)) \subset U$$

  $$\implies \exists \delta > 0\ s.t.\ f(B_\delta(x)) \subset B_\varepsilon(f(x))$$

  $$\implies B_\delta(x) \subset f^{-1}(U)$$

- topological $\implies$ metric

  If $x_n \to x_0$ in X, $\forall \varepsilon > 0, f^{-1}(B_\varepsilon(f(x_0)))$ is open in X. Since $x_0 \in f^{-1}(B_\varepsilon(f(x_0)))$, there exists $\delta > 0, B_\delta(x_0) \subset f^{-1}(B_\varepsilon(f(x_0)))$.

  $$\exists N > 1, x_n \in B_\delta(x_0) \subset f^{-1}(B_\varepsilon(f(x_0))), \forall n > N.$$

  This implies $|f(x_n) - f(x_0)| < \varepsilon$, hence $f(x_n) \to f(x_0)$.

□

If $A \subset X$ and $\overline{A} = X$, then we say A is **dense** in X.

Theorem 2.10

A continuous map $f : X \to Y$ of metric spaces is uniquely determined by $f|_A$, where A is dense in X.

Corollary 2.11

$card(C([0,1])) = card(\mathbb{R})$, because $\mathbb{Q}$ is dense in $\mathbb{R}$.

3 Compactness

3.1 Definition of compactness

Example 3.1

Recall: Bolzano-Weierstrass theorem:

In $\mathbb{R}^d$, every bounded infinity seqeunce has a convergent subsequence.

We can see that in $C[0,1]$ this is not true. If we take $\|f-g\| = \sup\limits_{0\le x\le 1}|f-g|$, and f_n has non-zero value near $\frac{1}{n}$, $f_n(\frac{1}{n}) = 1$. We have $\{f_n\}$ is bounded but $\|f_m - f_n\| = 1$ for all $m \neq n$. Which means it doesn't have a convergent subsequence.

Definition 3.2 (Open covering)**.** For $A \subset X$, a collection of open sets $\{U_\alpha\}$ is an **open covering** if $A \subset \bigcup_\alpha U_\alpha$.

Definition 3.3 (Compactness)**.** A is **compact** if every open covering of A has a finite sub-covering.

Theorem 3.4

If $f : X \to Y$ is continuous, then for any compact subset $K \subset X$, $f(K)$ is compact in Y.

Proof. Easily deduced from definitions of continuity and compactness. □

Theorem 3.5 (Cantor nested closed sets)

If F_n is bounded and closed in $\mathbb{R}^n$, and $F_n \subset F_{n-1}$, then $\bigcap\limits_{n=1}^{\infty} F_n \neq \varnothing$.

Proof. Take $x_n \in F_{n-1} \backslash F_n$(Assume first that F_n's are different from each other).

$\{x_n\}$ is bounded $\implies$ it has a accumulation point $y \implies y \in \bigcap\limits_{n=1}^{\infty} F_n$. □

Proposition 3.6

Compact in $\mathbb{R}^n \iff$ bounded and closed.

Proof. $\implies$:

$\{B_n(0)\}$ is an open covering $\implies$ bounded.

Assume for contradiction that there exists a convergent sequence whose limit $x_0 \notin K$.

Note that $\{x \mid 2^{n-1} < |x - x_0| < 2^{n+1}, n \in \mathbb{Z}\}$ is an open covering, it has a finite sub-covering $\implies \exists N, B_{2^{-N}}(x_0) \cap K = \varnothing$.

This is a contradiction because x_0 is the convergency point of a sequence in K.

$\impliedby$:

Take an open covering $\{U_\alpha\}$, since K is bounded, we can put K into a hypercube and split it into several smaller cubes.

If K can't be covered by finitely many U_α's, we obtain a closed set $K_1 \subset K$ and K_1 can't be covered by finitely many open sets in $\{U_\alpha\}$.

Repeat this process and by Cantor nested closed sets, there exists $x_0 \in K_n$. But as n grows larger, the open set containing x_0 must contain K_n as well, contradction! □

Definition 3.7 (Sequencial compactness)**.** A subset $K \subset X$ (X is a metric space) is **sequencial compact** if for all sequence in K, it has a subsequence converging to a point inside K.

Theorem 3.8

Sequencial compactness is equivalent to compactness (in a metric sapce).

3.2 Uniform continuity

Definition 3.9 (Uniform continuity)**.** f is a continuous function on a metric space X. It is **uniformly continuous** if $\forall \varepsilon > 0, \exists \delta > 0$ such that

$$\forall |x - y| < \delta, |f(x) - f(y)| < \varepsilon.$$

Theorem 3.10

If K is compact in a metric space X, f is continuous on K, then f is uniformly continuous on K.

Proof. $\forall x \in K$, exists $\delta(x)$, $\{B_{\delta(x)}(x)\}$ is an open covering of K. Take a finite sub-covering, and let $\delta = \min\{\delta(x_1), \dots, \delta(x_n)\}$ □

Definition 3.11. A sequence of functions $f_n(x)$ is uniformly convergent to f if $\forall \varepsilon > 0, \exists N$ such that

$$\forall n > N, \forall x, |f_n(x) - f(x)| < \varepsilon.$$

Theorem 3.12

f_n is continuous functions on X, $\{f_n\}$ uniformly converges to $f \implies f$ continuous.

Proof.

$$|f(x) - f(x_0)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)|$$

Let n be sufficiently large and $|x - x_0|$ sufficiently small. □

Remark 3.13 — The above theorem suggests that the "uniform" condition is kind of like commutativity of taking limits.

Definition 3.14 (Normal vector spaces)**.** A normal vector space $(X, \|\cdot\|)$ where X is a vector space, $\|\cdot\|$ is called norm satisfies:

- $\|x\| \ge 0$, $\|x\| = 0 \iff x = 0$;
- $\|\lambda x\| = |\lambda|\|x\|$;
- $\|x + y\| \le \|x\| + \|y\|$.

The norm gives the distant function in X, so a normal vector space automatically has a metric structure.

A **Banach space** is a complete normal vector space.

Corollary 3.15

Define $\|f(x)\|_\infty = \sup_{x\in K} |f(x)|$, for a function sequence $\{f_n(x)\}$ satisfying

$$\sum_{n=1}^{\infty} \|f_n(x)\|_\infty \text{ converges,}$$

then

$$\sum_{n=1}^{\infty} f_n(x) \text{ uniformly converges.}$$

Example 3.16 (Weierstrass' function)

$$W_{a,b}(x) = \sum_{n=1}^{\infty} a^n \cos(b^n \pi x), 0 < a < 1$$

is a continuous function.

Example 3.17

There exists a continous map from $[0,1]$ to $[0,1]^2$.

We can partition the unit square to smaller squares with side length 3^{1-n}, let the image of $f_n(x)$ be diagnals of these squares.

We can see that $|f_{n-1}(x) - f_n(x)| \le 3^{1-n}\sqrt{2}$, so $\{f_n\}$ uniformly converges to a function f. This is the desired continuous function.

4 Derivatives and differentiation

4.1 Definitions & basic properties

Definition 4.1 (derivatives)**.** f is **differentiable** at a point $x_0 \in (a,b)$, if

$$f'(x_0) := \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \text{ exists,}$$

and $f'(x_0)$ is called the **derivative** of f at x_0.

We say f is differentiable on (a,b) if f is differentiable at every point in (a,b).

Remark 4.2 — If f is differentiable at x_0,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)$$

which means we can approximatly view f as a linear function near x_0.

If f has the k-th derivative on (a,b) and $f^{(k)}$ is continuous, we say $f \in C^k(a,b)$. f is called **smooth** if $f \in C^\infty(a,b)$.

Proposition 4.3

Basic rules of taking derivatives:

- $(af + bg)' = af' + bg'$;
- $(fg)' = f'g + fg'$;
- $(f \circ g)' = (f' \circ g) \cdot g'$.

They can be proved using the above remark or by definition.

Example 4.4

Derivatives of elementary functions:

- $f(x) = x^n, f'(x) = nx^{n-1}$;
- $f(x) = e^x, f'(x) = e^x$,

$$\lim_{x \to x_0} \frac{e^x - e^{x_0}}{x - x_0} = \lim_{x \to x_0} \frac{e^{x_0}(e^{x-x_0} - 1)}{x - x_0} = e^{x_0}$$

- $f(x) = \ln x, f'(x) = \frac{1}{x}$,

$$\lim_{x \to x_0} \frac{\ln x - \ln x_0}{x - x_0} = \lim_{x \to x_0} \frac{\ln(\frac{x}{x_0} - 1)}{\frac{x}{x_0} - 1} \cdot \frac{1}{x_0} = \frac{1}{x_0}$$

- $f(x) = x^\alpha, f'(x) = \alpha x^{\alpha-1}$, by the exponential function and the chain rule.
- $(\sin x)' = \cos x, (\cos x)' = -\sin x$.

Theorem 4.5 (Leibniz's rule)

$$(f \cdot g)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(k)} g^{(n-k)}$$

Proposition 4.6

$f(x)$ is differentiable at x_0, $f'(x_0) > 0 \iff \exists \delta > 0$, such that $f(x_0 - h) < f(x_0) < f(x_0 + h'), \forall 0 < h, h' < \delta$.

Corollary 4.7

A differentiable function f on (a, b) is strictly increasing if $f'(x) > 0 \forall x \in (a, b)$.

local maximum/minimum

Theorem 4.8

If f has a local extrema x_0, and f is differentiable at x_0, then $f'(x_0) = 0$.

4.2 Mean value theorems

Theorem 4.9 (Rolle's theorem)

Suppose f is continuous on $[a, b]$, differentiable on (a, b), and $f(a) = f(b)$. Then there's a point $\xi \in (a, b)$ such that $f'(\xi) = 0$.

Proof. Assume that f is not constant. Take the global maximum and minimun of f. Since $f(a) = f(b)$, either of them must lie inside (a, b), this is the desired ξ. □

The direct corollary of Rolle's theorem is the following famous theorem:

Theorem 4.10 (Lagrange's mean value theorem)

Suppose f is continuous on $[a, b]$, differentiable on (a, b). Then there's a point $\xi \in (a, b)$ such that $f'(\xi) = \frac{f(b)-f(a)}{b-a}$.

Corollary 4.11

A continuous function $f(x)$ on $[a, b]$ is differentiable on (a, b), and $f'(x) \geq 0$, then we have f is increasing.

If $f'(x) > 0$, f is striclty increaing.

Theorem 4.12

A function $f \in C^1(a, b)$, $f'(x_0) \neq 0$, then there exists $\delta > 0$ such that

$$f : (x_0 - \delta, x_0 + \delta) \to (c, d) \text{ is a } C^1 \text{ homeomorphism}$$

i.e. f is a bijection and f^{-1} is a C^1 function.

Proof. WLOG $f'(x_0) > 0$, then f is increasing and continuous on $(x_0 - \delta, x_0 + \delta)$. Thus f has a continuous inverse f^{-1}.

We only need to prove $(f^{-1})'$ is a C^1 function.

$$\lim_{x \to x_0} \frac{f^{-1}(x) - f^{-1}(x_0)}{x - x_0} = \lim_{u \to u_0} \frac{u - u_0}{f(u) - f(u_0)} = \frac{1}{f'(u)}$$

Here we used f^{-1} is continuous. □

Remark 4.13 — The above theorem can be generalized to C^k using induction.

Theorem 4.14 (Darboux)

f is differentiable on $[a,b]$, $\forall u \in [\min(f'(a), f'(b)), \max(f'(a), f'(b))], \exists c \in [a,b]$ such that $f'(c) = u$.

Proof. WLOG $f'(a) < 0 < f'(b), u = 0$ (replace $f(x)$ with $f(x) - ux$).

Take the minimum of $f(x)$, say c. We have $c \neq a, b$ and $f'(c) = 0$. □

Theorem 4.15 (Cauchy)

f, g are continuous functions on $[a,b]$, differentiable on (a,b), and $g'(x) \neq 0, \forall x \in (a,b)$.

Then

$$\exists x_0 \in (a,b), \text{ such that } \frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(x_0)}{g'(x_0)}.$$

Proof. Darboux $\implies g' > 0$ (or $g' < 0$), thus g is stricly increasing (or decreasing).

C^1 homeomorphism $\implies \exists g^{-1}$ differentiable on $[g(a), g(b)]$.

By lagrange's mean value theorem we're done. □

Remark 4.16 — Alternative proof:

Use Rolle's theorem on $f(x) - \frac{f(b)-f(a)}{g(b)-g(a)}(g(x) - g(a))$ (here we need $g(b) \neq g(a)$).

In fact we can use Rolle's on $F(x) := (g(b)-g(a))f(x) - (f(b)-f(a))(g(x)-g(a))$, and get $(f(b)-f(a))g'(x_0) = (g(b)-g(a))f'(x_0)$, where we don't even use the condition that $g'(x) \neq 0$.

4.3 On different definitions of trig functions

Theorem 4.17

Suppose $F : [a,b] \to \mathbb{R}^n$, differentiable on (a,b) (every entry is differentiable). If $F'(x) = 0 \forall x$, we have that F is constant.

Proof. $F = (f_1, \dots, f_n)$, $f_k'(x) = 0$, from Lagrange we deduce each f_k is constant, so F is constant as well. □

Consider the exponent map $e : M_n(\mathbb{R}) \to M_n(\mathbb{R})$:

$$A \mapsto \sum_{k=0}^{\infty} \frac{A^k}{k!}, \quad \|A\| = \sup |a_{ij}|$$

Fix a matrix A, then $e^{xA} : \mathbb{R} \to M_n(\mathbb{R})$, since xA, yA are commutative, $e^{xA} \cdot e^{yA} = e^{(x+y)A}$.

$$\begin{aligned}\lim_{h\to 0}\frac{e^{(x+h)A}-e^{xA}}{h} &= \lim_{h\to 0}\frac{e^{xA}(e^{hA}-I_n)}{h}\\ &= e^{xA}\lim_{h\to 0}\sum_{k=1}^{\infty}\frac{h^{k-1}A^k}{k!}\\ &= e^{xA}\cdot A\end{aligned}$$

Consider $F:\mathbb{R}\to\mathbb{R}^n$ satisfying $F'(x)=AF(x)$, where $A\in M_n(\mathbb{R})$. We want to find all such functions F.

Since

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(e^{-xA}F(x)\right)=e^{-xA}(-AF(x)+F'(x))=0,$$

Hence $e^{-xA}F(x)$ is a constant.

Now we come back to look at our trignometry fuctions defined by $\sin x=\dfrac{e^{ix}-e^{-ix}}{2i}$, $\cos x=\dfrac{e^{ix}+e^{-ix}}{2}$. Let

$$F(x)=\begin{pmatrix}\sin x\\ \cos x\end{pmatrix}$$

$$\implies F'(x)=\begin{pmatrix}\cos x\\ -\sin x\end{pmatrix}=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix}\sin x\\ \cos x\end{pmatrix}=JF(x)$$

$$\implies F(x)=e^{xJ}F(0)=e^{xJ}\begin{pmatrix}0\\ 1\end{pmatrix}\text{ is the unique function satisfying }F'(x)=JF(x).$$

By geometry definition of $\sin x,\cos x$, we can also deduce the above defferentiaition relations (high school maths) and $F(0)$, so the series definition coinsides with the geometry definition.

4.4 Differetiating infinite series

Theorem 4.18

f_k are differentiable functions on $[a,b]$, $\sum_{k=1}^{\infty} f_k'(x)$ uniformly converges, and $\exists x_0$ such that $\sum_{k=1}^{\infty} f_k(x_0)$ converges.

Then $\sum_{k=1}^{\infty}$ uniformly converges on $[a,b]$, and

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{k=1}^{\infty} f_k(x)\right)=\sum_{k=1}^{\infty}\frac{\mathrm{d}}{\mathrm{d}x}f_k(x).$$

Proof.

$$\left|\sum_{k=n}^{m}(f_k(x)-f_k(x_0))\right|=\left|(x-x_0)\sum_{k=n}^{m}f_k'(\xi)\right|<\varepsilon|x-x_0|$$

$\implies \sum f_k(x)$ uniformly converges.

$\forall \varepsilon > 0, \exists N$ such that

$$\left|\sum_{k=n}^{m} f_k'(x)\right| < \frac{\varepsilon}{3}, \quad \forall n, m > N$$

By Lagrange's theorem,

$$\left|\sum_{k=n}^{m} \frac{f_k(x+h) - f_k(x)}{h}\right| = \left|\sum_{k=n}^{m} f_k'(x+\xi)\right| < \frac{\varepsilon}{3}$$

Fix h, let $m \to \infty$, we get

$$\left|\sum_{k=n}^{\infty} \frac{f_k(x+h) - f_k(x)}{h}\right| < \frac{\varepsilon}{3}, \quad \forall n > N$$

Thus

$$\begin{aligned}\lim_{h\to 0}\left(\frac{\sum_{k=1}^{\infty} f_k(x) - \sum_{k=1}^{\infty} f_k(x+h)}{h}\right) &= \lim_{h\to 0}\left(\sum_{k=1}^{n} \frac{f_k(x+h) - f_k(x)}{h} + \sum_{k>n} \frac{f_k(x+h) - f_k(x)}{h}\right) \\ &= \sum_{k=1}^{n} f_k'(x) + \lim_{h\to 0}\sum_{k>n} \frac{f_k(x+h) - f_k(x)}{h}\end{aligned}$$

Since the second part is uniformly small, thus by taking $n \to \infty$ we get $\sum f_k(x)$ is differentiable and its derivative is precisely $\sum f_k'(x)$. □

Now we'll give an example where f_k's are differentiable but $\sum f_k$ is not.

Let

$$W_{a,b}(x) = \sum_{k=1}^{\infty} a^n \cos(b^n \pi x),\ 0 < a < 1,\ ab > 1 + \frac{3}{2}\pi,\ 2 \nmid b$$

Claim 4.19 — Then $W_{a,b}(x)$ is continuous on $\mathbb{R}$, but is not differetiable at any point.

Clearly $W_{a,b}$ is continuous as the sum uniformly converges.

Fix $y_) \in \mathbb{R}$, we'll prove that $W_{a,b}$ is not differentiable at y_0. To prove this, we take a sequence $\{y_n\} \to y_0$ as follow:

Let $\delta_n = b^n y_0 - z_n \in [\frac{1}{2}, \frac{3}{2})$, where $z_n \in \mathbb{Z}$, let $y_n = \frac{z_n}{b}$ so that $y_n \to y_0$.

$$\frac{W(y_n) - W(y_0)}{y_n - y_0} = \sum_{k=1}^{n-1} \frac{a^k(\cos(b^k\pi y_n) - \cos(b^k \pi y_0))}{y_n - y_0} + \sum_{k\geq n} \frac{a^k(\cos(b^k\pi y_n) - \cos(b^k \pi y_0))}{y_n - y_0}$$

While we have (by Lagrange's theorem)

$$\sum_{k=1}^{n-1} \frac{a^k(\cos(b^k\pi y_n) - \cos(b^k \pi y_0))}{y_n - y_0} = \sum_{k=1}^{n-1} (ab)^k \pi \sin\theta_k$$

and

$$\begin{aligned}\sum_{k\geq n} \frac{a^k(\cos(b^k\pi y_n) - \cos(b^k \pi y_0))}{y_n - y_0} &= \sum_{k\geq n} \frac{a^k(\cos(b^{k-n} z_n \pi) - \cos(b^{k-n}\pi z_n + b^{k-n}\pi\delta_n))}{-\frac{\delta_n}{b^n}} \\ &= \sum_{k\geq n} \frac{a^k b^n((-1)^{z_n} - \cos(b^{k-n}\delta_n \pi))}{-\delta_n} \\ &= \frac{a^n b^n (-1)^{z_n+1}}{\delta_n} \sum_{k\geq n} a^{k-n}(1 - (-1)^{z_n}\cos(b^{k-n}\delta_n\pi))\end{aligned}$$

So taking absolute value of the above we deduce

$$\begin{aligned}
\left|\frac{W(y_n)-W(y_0)}{y_n-y_0}\right| &\geq \left|\sum_{k\geq n}\frac{a^k(\cos(b^k\pi y_n)-\cos(b^k\pi y_0))}{y_n-y_0}\right| - \left|\sum_{k=1}^{n-1}\frac{a^k(\cos(b^k\pi y_n)-\cos(b^k\pi y_0))}{y_n-y_0}\right| \\
&\geq \left|\frac{a^nb^n}{\delta_n}\sum_{k\geq n}a^{k-n}(1-(-1)^{z_n}\cos(b^{k-n}\delta_n\pi))\right| - \sum_{k=1}^{n-1}(ab)^k\pi \\
&\geq \frac{a^nb^n}{\delta_n}(1-\cos(\delta_n\pi)) - \frac{a^nb^n\pi}{ab-1} \\
&\geq (ab)^n\left(\frac{2}{3}-\frac{\pi}{ab-1}\right)
\end{aligned}$$

where the last inequality is taking the $k=n$ term in the sum. (This is valid because the terms in the summation are all non-negative)

Since $ab-1\geq\frac{3}{2}\pi$, as $n\to\infty$, $(ab)^n$ can be arbitarily large, $W(x)$ is not differentiable at y_0. This implies the desired result.

5 Applications of differentiation

5.1 L'Hopital's rule

This well-known theorem has a reputation of "洛神" among students in high school and non-mathematics major. But it's not as useful as Taylor series to those who study mathematics.

Theorem 5.1 (L'Hôpital's rule, $\frac{0}{0}$)

f,g are differentiable functions on (a,b) satisfying

- $\lim\limits_{x\to a^+}f(x)=\lim\limits_{x\to a^+}g(x)=0$;
- $\forall x\in(a,b),\ g'(x)\neq 0$;
- $\lim\limits_{x\to a^+}\dfrac{f(x)}{g(x)}$ exists (possibly infinity).

Then we have

$$\lim_{x\to a^+}\frac{f(x)}{g(x)}=\lim_{x\to a^+}\frac{f'(x)}{g'(x)}$$

Proof. Let $F(a)=G(a)=0$, $F(x)=f(x),G(x)=g(x),x>a$.

F,G are continuous on $[a,\frac{a+b}{2}]$ and differetiable on $(a,\frac{a+b}{2})$, by Cauchy's theorem:

$$\lim_{x\to a^+}\frac{F(x)-F(a)}{G(x)-G(a)}=\lim_{x\to a^+}\frac{F'(\xi)}{G'(\xi)}=\lim_{x\to a^+}\frac{f'(\xi)}{g'(\xi)}=\lim_{x\to a^+}\frac{f'(x)}{g'(x)}$$

□

Theorem 5.2 (L'Hôpital's rule, $\frac{\infty}{\infty}$)

f, g are differentiable functions on (a, b) satisfying

- $\lim_{x\to a^+} |f(x)| = \lim_{x\to a^+} |g(x)| = +\infty$;
- $\forall x \in (a, b),\ g'(x) \neq 0$;
- $\lim_{x\to a^+} \dfrac{f(x)}{g(x)}$ exists (possibly infinity).

Then we have

$$\lim_{x\to a^+} \frac{f(x)}{g(x)} = \lim_{x\to a^+} \frac{f'(x)}{g'(x)}$$

Proof. WLOG $g'(x) < 0$ (Darboux grants that g' does not change sign), let $y = g(x)$, since we have g stricly decreasing, g is a bijection:$(a, b) \to (g(b), +\infty)$.

Note that:

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

$$\frac{\mathrm{d}}{\mathrm{d}y} f(g^{-1}(y)) = f'(x) \cdot \frac{1}{g'(x)}$$

This shows that we may assume $g(x) = x$ and $x \in (c, +\infty)$ to simplify the situation. Suppose $\lim_{x\to a^+} f'(x) = A$ (possibly infinity),

$$\frac{f(x)}{x} = \frac{f(x_0) + f'(\xi)(x - x_0)}{x}, \xi \in (x_0, x)$$

Take x_0 large, so that $f'(\xi)$ is close to A; then let x suffiently large,

$$\frac{f(x)}{x} = f'(\xi)(1 - \frac{x_0}{x}) + \frac{f(x_0)}{x} \to A$$

□

Remark 5.3 — To prove the case where $a = \infty$, use the map $x \mapsto \frac{1}{x}$.

Remark 5.4 — The condition that $g'(x) \neq 0$ in L'hopital's rule is necessary, since we can take $f(x) = x + \sin x \cos x, g(x) = e^{\sin x}(x + \sin x \cos x)$ and let $x \to +\infty$.

We have

$$\lim_{x\to+\infty} \frac{f'(x)}{g'(x)} = 0 \text{ but } \lim_{x\to+\infty} \frac{f(x)}{g(x)} \text{ doesn't exist.}$$

5.2 Taylor series

This is the idea of approximating general functions using polynomials.

Theorem 5.5 (Taylor series)

There are 3 types of remainder:

1. Peano: Let $f(x)$ be a function that has n-th derivative at point a,
$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + o((x-a)^n)$$
2. Lagrange: Let $f(x) \in C^n[a,b]$ be a function that has $(n+1)$-th derivative on interval (a,b), then $\forall x \in [a,b]$, $\exists \xi \in [a,b]$,
$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$
3. Cauchy: Same conditions as above, then $\forall x \in [a,b]$, $\exists \xi \in [a,b]$,
$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n(x-a)$$

Proof.

- Peano: f has $(n-1)$-th derivative on a neighborhood of a, use L'Hopital's rule ($n-1$ times) on
$$\lim_{x\to a} \frac{f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k}{(x-a)^n}$$
to get it is 0.

- Lagrange & Cauchy: Fix x, define $F(t)$ on $[a,x]$:
$$F(t) := \sum_{k=0}^{n} \frac{f^{(k)}(t)}{k!}(x-t)^k, F'(t) = \frac{f^{(n+1)}(t)}{n!}(x-t)^n.$$
Use Cauchy's mean value theorem, for any $G(t)$
$$\frac{F(x)-F(a)}{G(x)-G(a)} = \frac{F'(\xi)}{G'(\xi)}$$
The remainder
$$f(x) - F(a) = \frac{G(x)-G(a)}{G'(\xi)} \cdot \frac{f^{(n+1)}(\xi)}{n!}(x-\xi)^n$$
Take $G(t) = x - t$ to get Cauchy remainder, take $G(t) = (x-t)^{n+1}$ to get Lagrange remainder.

$\square$

Remark 5.6 — The polynomial in Taylor series is the optimal approximation to f.

Example 5.7

Taylor series of e^x at 0:

$$e^x = \sum_{k=0}^{n} \frac{x^k}{k!} + o(x^n)$$

5.3 Convexity

Definition 5.8 (Convex functions)**.** We say a function f is **convex** on (a, b), if $\forall 0 \le t \le 1, \forall x, y \in (a, b)$,

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$$

In other words, the function's image is below the segment joining any two points on it.

The definition of convex function is equivalent to $\forall x < z < y$,

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

Definition 5.9 (Convex sets)**.** For a vector space V, we say $X \subset V$ is **convex** if $\forall x, y \in X, t \in [0, 1]$, $tx + (1-t)y \in X$.

Theorem 5.10 (Jensen's inequality)

If f is convex on (a, b), $t_1, \ldots, t_n \ge 0, \sum t_i = 1$, then

$$f\left(\sum_{i=1}^{n} t_i x_i\right) \le \sum_{i=1}^{n} t_i f(x_i).$$

Theorem 5.11

If f is convex on (a, b), $a < x < y < b$, we have

$$D_-(f)(x) \le D_+(f)(x) \le \frac{f(x) - f(y)}{x - y} \le D_-(f)(y) \le D_+(f)(y)$$

where D_-, D_+ are left and right derivatives.

Proof. It's obvious by the alternative definition of convex functions.

i.e. take $h \to 0$, $a < x - h < x < x + h < y - h < y < y + h < b$. Monotone and bounded implies the derivatives exist. □

Theorem 5.12

A function f is continuous on (a, b) and $D_+(f)$ is increasing $\iff$ f is convex.

Proof. " $\Longleftarrow$ " is already proved.

Now assume f is continous and $D_+(f)$ is increasing.

We're going to prove that for $x < z < y$,

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

Without loss of generality, suppose that $D_+(f)(z) = 0 = f(z)$ (otherwise we subtract a linear function from $f(x)$).

Thus $D_+(f)(x) \leq 0, \forall x \leq z$ and $D_+(f)(y) \geq 0, \forall y \geq z$.

If $\exists y_0 > z, f(y_0) < 0$, let $y_* := \inf\{y \mid f(y) < 0, z \leq y \leq y_0\}$, from continuity we have $f(y_*) = 0$, and $\exists \{y_n\} \to y_*, f(y_n) < 0$.

$$\Longrightarrow D_+(f)(y_*) = \lim_{n\to\infty} \frac{f(y_n)-f(y_*)}{y_n - y_*} \leq 0$$

Let $g_\varepsilon(x) = f(x) + \varepsilon(x-z), D_+(g_\varepsilon) > 0$, apply the above to g_ε, we have $g_\varepsilon(y) > g_\varepsilon(z)$.

Therefore $f(y) + \varepsilon(y-z) > 0, \forall \varepsilon > 0 \Longrightarrow f(y) \geq 0$.

$\forall x_1 < x_2 < z$, f has a minimum point x' in $[x_1, x_2]$.

If $x' \neq x_2$, we deduce $D_+(f)(x') \geq 0$.

Similarly define $g_{-\varepsilon}$, then $g_{-\varepsilon}(x) > g_{-\varepsilon}(z), \forall \varepsilon > 0 \Longrightarrow f(x) \geq 0$.

Hence

$$\frac{f(x)-f(z)}{x-z} \leq 0 \leq \frac{f(y)-f(z)}{y-z}.$$

□

Example 5.13 (Minkowski's inequality)

Define a norm in $\mathbb{R}^n$:

$$x = (x_1, \ldots, x_n), \quad \|x\|_{l_p} := \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}.$$

We want to prove $\|x\| + \|y\| \geq \|x+y\|$.

By Jensen's inequality, x^p is convex when $p \geq 1$, so

$$\begin{aligned}\sum_{i=1}^n (x_i+y_i)^p &\leq \sum_{i=1}^n \left(t\left(\frac{x_i}{t}\right)^p + (1-t)\left(\frac{y_i}{1-t}\right)^p\right) \\ &= \sum_{i=1}^n x_i^p t^{1-p} + \sum_{i=1}^n y_i^p (1-t)^{1-p}.\end{aligned}$$

By choosing t such that the above two terms are equal, we get the result.

6 Integration

6.1 Riemann-Stieltjes integral

Definition 6.1. A partition of a closed interval $[a,b]$ $\sigma = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$. Given an increasing function $\mu(x)$ and a bounded function $f(x)$.

Define

$$\Delta_i = \mu(x_i) - \mu(x_{i-1}) \geq 0, \quad M_i = \sup_{x \in [x_{i-1}, x_i]} f(x), \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x).$$

The step length $|\sigma| = \max_{1 \leq i \leq n} |x_i - x_{i-1}|$

Define the **upper and lower Darboux sums**:

$$\overline{S}_\mu(f;\sigma) = \sum_{i=1}^n M_i \Delta_i = \sum_{i=1}^n M_i(\mu(x_i) - \mu(x_{i-1}))$$

$$\underline{S}_\mu(f;\sigma) = \sum_{i=1}^n m_i \Delta_i = \sum_{i=1}^n m_i(\mu(x_i) - \mu(x_{i-1}))$$

Define the **upper and lower integrals** of f:

$$\overline{\int_a^b} f \mathrm{d}\mu = \inf_\sigma \overline{S}_\mu(f;\sigma), \quad \underline{\int_a^b} f \mathrm{d}\mu = \sup_\sigma \underline{S}_\mu(f;\sigma)$$

Definition 6.2. Given an increasing function $\mu(x)$ on $[a,b]$, we say a bounded function f is **Riemann-Stieltjes integrable**, if

$$\overline{\int_a^b} f \mathrm{d}\mu = \underline{\int_a^b} f \, \mathrm{d}\mu$$

and denote the value by $\int_a^b f \mathrm{d}\mu$.

All integrable functions on $[a,b]$ form a vector space, denoted by $\mathcal{R}([a,b])$.

Given partitions σ, σ_1, if all the partition points of σ_1 are in σ, i.e. $\sigma_1 \subset \sigma$, we say σ is a **refinement** of σ_1.

Denote by $\sigma_1 \cup \sigma_2$ the partition that contains the points in the union of σ_1 and σ_2.

Theorem 6.3

If $\sigma_1 \subset \sigma_2$, then

$$\overline{S}_\mu(f;\sigma_1) \geq \overline{S}_\mu(f;\sigma_2), \quad \underline{S}_\mu(f;\sigma_1) \leq \underline{S}_\mu(f;\sigma_2)$$

Proof. WLOG $\sigma_2 = \sigma_1 \cup \{y_0\}$, $y_0 \in (x_0, x_1)$, where $\sigma_1 = \{x_0 < x_1 < \cdots < x_n\}$.

$$\begin{aligned}\overline{S}_\mu(f;\sigma_1) - \overline{S}_\mu(f;\sigma_2) = &\sup_{x \in [x_0,x_1]} f(x)(\mu(x_1) - \mu(x_0)) \\ &- \sup_{x \in [x_0,y_0]} f(x)(\mu(y_0) - \mu(x_0)) - \sup_{x \in [y_0,x_1]} f(x)(\mu(x_1) - \mu(y_0)) \\ &\geq 0.\end{aligned}$$

□

Corollary 6.4

The upper integral is greater than or equal to the lower integral. i.e.

$$\overline{\int_a^b} f \, \mathrm{d}\mu \geq \underline{\int_a^b} f \, \mathrm{d}\mu$$

Proof. For any partition σ_1, σ_2,

$$\overline{S}_\mu(f;\sigma_1) \geq \overline{S}_\mu(f;\sigma_1 \cup \sigma_2) \geq \underline{S}_\mu(f;\sigma_1 \cup \sigma_2) \geq \underline{S}_\mu(f;\sigma_2)$$

□

Corollary 6.5

A function f is integrable $\iff$ $\forall \varepsilon > 0$, $\exists \sigma$, such that

$$\overline{S}_\mu(f;\sigma) - \underline{S}_\mu(f;\sigma) < \varepsilon.$$

Remark 6.6 — This is equivalent to

$$\sum_{i=1}^n (M_i - m_i)(\mu(x_i) - \mu(x_{i-1})) < \varepsilon.$$

If f is integrable, we have $\forall \varepsilon > 0$, $\exists$ partition σ, $\forall \xi_i \in [x_{i-1}, x_i]$,

$$\left| \sum_{i=1}^n f(\xi_i)(\mu(x_i) - \mu(x_{i-1})) - \int_a^b f \, \mathrm{d}\mu \right| < \varepsilon$$

The first term is called the **Riemann sum**.

6.2 Criteria for integrable functions

Theorem 6.7

All continuous functions on $[a,b]$ are integrable.

Proof. Let $\omega_i(f) := \sup_{x,y \in [x_{i-1}, x_i]} |f(x) - f(y)|$.

f is uniformly continuous, so as $|\sigma| \to 0$, $\omega_i(f)$ can be arbitarily small, from this we deduce f is intergrable. □

Theorem 6.8

If f is monotone and bounded on $[a,b]$, μ is continuous (actually we only need that the discontinuous points of μ and f do not coincide), then f is integrable.

Proof. μ is uniformly continuous, Δ_i can be arbitarily small. □

Theorem 6.9

If Φ is continuous, f is integrable, then $\Phi(f(x))$ is integrable.

Proof. Φ uniformly continuous $\implies \forall\varepsilon > 0, \exists\delta > 0$

$$|\Phi(f(x)) - \Phi(f(y))| < \varepsilon, \forall |f(x) - f(y)| < \delta.$$

f integrable $\implies \forall\varepsilon_0 > 0$, $\exists\sigma$,

$$\sum_{i=1}^{n} \omega_i(f)\Delta_i < \varepsilon_0.$$

$$\begin{aligned}
\sum_{i=1}^{n} \omega_i(\Phi \circ f)\Delta_i &= \sum_{i=1}^{n} \sup |\Phi(f(x)) - \Phi(f(y))| \cdot \Delta_i \\
&= \sum_{\omega_i(f)<\delta} \varepsilon\Delta_i + \sum_{\omega_i(f)\geq\delta} 2M\Delta_i \qquad\qquad (M = \sup|\Phi|) \\
&\leq \varepsilon|\mu(b) - \mu(a)| + 2M \sum_{\omega_i(f)\geq\delta} \Delta_i \\
&\leq \varepsilon|\mu(b) - \mu(a)| + 2M\delta^{-1} \sum_{i=1}^{n} \omega_i(f)\Delta_i \\
&\leq \varepsilon|\mu(b) - \mu(a)| + 2M\delta^{-1}\varepsilon_0
\end{aligned}$$

We're done by letting $\varepsilon_0 < \delta\varepsilon$. □

Proposition 6.10

Properties of integrable functions:

- $\mathcal{R}(\mu)$ is a vector space;
- $\forall f \leq g$, $\int_a^b f \,\mathrm{d}\mu \leq \int_a^b g \,\mathrm{d}\mu$;
- f is integrable on $[a, b]$, $\forall c \in (a, b)$, f is integrable on $[a, c]$ and $[c, b]$, $\int_a^b f \,\mathrm{d}\mu = \int_a^c f \,\mathrm{d}\mu + \int_c^b f \,\mathrm{d}\mu$;
- f integrable $\implies |f|$ integrable, $\int_a^b f \,\mathrm{d}\mu \leq \int_a^b |f| \,\mathrm{d}\mu$;
- $\alpha > 0, \beta > 0$, μ, ν increasing, then $\int_a^b f \,\mathrm{d}(\alpha\mu + \beta\nu) = \alpha\int_a^b f \,\mathrm{d}\mu + \beta\int_a^b f \,\mathrm{d}\nu$;
- f, g integrable $\implies fg$ integrable.

Example 6.11

Given $0 \le \alpha < \beta$, define jump function:

$$j(x) = \begin{cases} 0, & x < 0 \\ \alpha, & x = 0 \\ \beta, & x > 0 \end{cases}$$

For a function f which is continuous at s, we have

$$\int_a^b f(x)\,\mathrm{d}j(x-s) = f(s)\beta.$$

proof: $s \in \sigma, \overline{S}_\mu(f;\sigma) = M_{i-1}(\mu(s)-\mu(x_{i-1})) + M_i(\mu(sx_i)-\mu(s)) = M_{i-1}\alpha + M_i(\beta-\alpha)$.

Example 6.12

Given $\gamma_n \le \beta_n \le \alpha_n$, $\sum(\alpha_n - \gamma_n)$ converges, $x_n \in [a,b]$ are distinct.

Define:

$$J(x) = \sum_{n=1}^{\infty} j_n(x), \text{ where } j_n(x) = \begin{cases} \gamma_n, & x < x_n \\ \beta_n, & x = x_n \\ \alpha_n, & x > x_n \end{cases}$$

Then for f continuous at all the x_n's, we have

$$\int_a^b f\,\mathrm{d}J = \sum_{n=1}^{\infty} f(x_n)(\alpha_n - \gamma_n) = \sum_{n=1}^{\infty} \int_a^b f\,\mathrm{d}j_n$$

Proof. $\forall \varepsilon > 0, \exists N$ such that $\sum\limits_{n>N}(\alpha_n - \gamma_n) < \varepsilon$.

$$\begin{aligned}
\implies \left| \int_a^b f\,\mathrm{d}J - \sum_{k=1}^{n} f(x_k)(\alpha_k - \gamma_k) \right| &= \left| \int_a^b f\,\mathrm{d}\left(J - \sum_{k=1}^{n} j_k\right) \right| \\
&\le \sup|f| \cdot \left| \left(J - \sum_{k=1}^{n} j_k\right)(b) - \left(J - \sum_{k=1}^{n} j_k\right)(a) \right| \\
&\le \sup|f| \sum_{k>N} (\alpha_k - \gamma_k) < \varepsilon
\end{aligned}$$

We still need to prove f is integrable.

Let $\mu_N = \sum\limits_{k=1}^{N} j_k(x - x_k)$, $\forall \varepsilon > 0$, $\exists \sigma$ such that

$$\begin{aligned}
\sum_{i=1}^{n} \omega_i(f)(J(\xi_i) - J(\xi_{i-1})) &= \sum_{i=1}^{n} \omega_i(f)|\mu_N(\xi_i) - \mu_N(\xi_{i-1})| \\
&\quad + \sum_{i=1}^{n} \omega_i(f)|(J-\mu_N)(\xi_i) - (J-\mu_N)(\xi_{i-1})| \\
&\le \varepsilon + 2\sup|f| \cdot |(J-\mu_N)(b) - (J-\mu_N)(a)| < (1 + 2\sup|f|)\varepsilon.
\end{aligned}$$

□

6.3 Fundamental theorem of calculus

Theorem 6.13 (Fundamental theorem of calculus - 1)

f is Riemann integrable on $[a, b]$, define $F(x) = \int_a^x f(t)\,\mathrm{d}t$. Then F is continuous on $[a, b]$. If f is continuous at x_0, then F is differentiable at x_0 and $F'(x_0) = f(x_0)$.

Proof. $F(x_0 + h) - F(x_0) = \int_{x_0}^{x_0+h} f(t)\,\mathrm{d}t \le \int_{x_0}^{x_0+h} |f(t)|\,\mathrm{d}t \le \sup |f| \cdot h$.

When f is continuous at x_0,

$$\lim_{h \to 0} \frac{F(x_0 + h) - F(x_0)}{h} = \lim_{h \to 0} \frac{1}{h} \int_{x_0}^{x_0+h} f(t)\,\mathrm{d}t = f(x_0)$$

□

Theorem 6.14 (Fundamental theorem of calculus - 2)

Given differentiable function $F(x)$ on $[a, b]$, and F' is Riemann integrable on $[a, b]$, then

$$\int_a^b F'(x)\,\mathrm{d}x = F(b) - F(a)$$

Proof.

$$F(b) - F(a) = \sum_{i=1}^{n} (F(x_i) - F(x_{i-1})) = \sum_{i=1}^{n} F'(\xi_i)(x_i - x_{i-1}) \to \int_a^b F'(x)\,\mathrm{d}x$$

□

This theorem is also called Newton-Lebniz formula. From this and the rules of taking derivatives, we get:

Corollary 6.15 (Intergration by parts)

F, G is differentiable on $[a, b]$, and the derivatives are Riemann integrable.

$$\int_a^b FG'\,\mathrm{d}x = (FG)\big|_a^b - \int_a^b F'G\,\mathrm{d}x$$

Theorem 6.16 (Taylor series with integral remainder)

$f \in C^{m+1}[a, b]$, then

$$f(x) = \sum_{k=0}^{m} \frac{f^{(k)}(a)}{k!}(x - a)^k + \int_a^x \frac{(x - t)^m}{m!} f^{(m+1)}(t)\,\mathrm{d}t.$$

Because of the fundamental theorem of calculus, when we compute integrals, we often want to find the function F. This is the reverse of taking derivatives, and it's called indefinite integrals.

Corollary 6.17 (Intergration by substitution)

$$\int f(y)\,\mathrm{d}y = F(y) + C \implies \int f(y(x))y'(x)\,\mathrm{d}x = F(y(x)) + C$$

$$u'(t) \neq 0, \int f(u(t))u'(t)\,\mathrm{d}t = F(t) + C \implies \int f(y)\,\mathrm{d}y = F(u^{-1}(y)) + C$$

Example 6.18 (Wallis' integral formula)

Let

$$I_n = \int_0^{\frac{\pi}{2}} \sin^n\theta\,\mathrm{d}\theta,$$

we have

$$\lim_{n\to\infty} I_n\sqrt{n} = \sqrt{\frac{\pi}{2}}.$$

Proof.

$$\begin{aligned}
I_{n+2} &= \int_0^{\frac{\pi}{2}} \sin^n\theta(1-\cos^2\theta)\,\mathrm{d}\theta \\
&= I_n - \int_0^{\frac{\pi}{2}} \cos^2\theta\sin^n\theta\,\mathrm{d}\theta \\
&= I_n - \int_0^{\frac{\pi}{2}} \frac{\cos\theta}{n+1}\,\mathrm{d}\sin^{n+1}\theta \\
&= I_n - \left.\frac{\cos\theta\sin^{n+1}\theta}{n+1}\right|_0^{\frac{\pi}{2}} + \frac{1}{n+1}\int_0^{\frac{\pi}{2}} \sin^{n+1}\theta(-\sin\theta)\,\mathrm{d}\theta \\
&= I_n - \frac{1}{n+1}I_{n+2}
\end{aligned}$$

Thus $I_{n+2} = \frac{n+1}{n+2}I_n$, $I_0 = \frac{\pi}{2}, I_1 = 1$.

$$I_{2n} = \frac{(2n)!}{((2n)!!)^2}\cdot\frac{\pi}{2} = \frac{(2n)!}{2^{2n}(n!)^2}\cdot\frac{\pi}{2}$$

$$I_{2n+1} = \frac{2^{2n}(n!)^2}{(2n+1)!}$$

$I_{2n}\cdot I_{2n+1} = \frac{\pi}{2(2n+1)}$, let $x_n = I_n\sqrt{n\cdot\frac{2}{\pi}}$, note that

$$\frac{x_{n+2}}{x_n} = \frac{n+1}{\sqrt{n(n+2)}} > 1, \quad x_{2n}x_{2n+1} = \sqrt{\frac{2n}{2n+1}}$$

so $\{x_{2n}\}, \{x_{2n+1}\}$ are increasing,

$\implies \lim_{n\to\infty} x_{2n} = \lim_{n\to\infty} x_{2n+1} = 1.$ □

Example 6.19 (Stirling's formula)

$$\lim_{n\to\infty}\frac{n!e^n}{n^{n+\frac{1}{2}}} = \sqrt{2\pi}$$

Proof. Let $a_n = \dfrac{e^n n!}{n^{n+\frac{1}{2}}}$.

$$\ln \frac{a_n}{a_{n+1}} = (n+\frac{1}{2})\ln(1+\frac{1}{n}) - 1 > 0$$

Where the inequality follows from Taylor's series: $\ln(1+x) \geq x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4}$.

Therefore $a_n > a_{n+1} > 0$, $\lim\limits_{n\to\infty} a_n$ exists, let it be α.

From Wallis' formula,

$$\begin{aligned}\sqrt{2(2n+1)}I_{2n+1} &= \sqrt{2(2n+1)}\frac{2^{2n}(n!)^2}{(2n+1)!} \\ &= \frac{a_n^2}{a_{2n}\sqrt{2}} \cdot \frac{\sqrt{2n}}{\sqrt{2n+1}} \longrightarrow \sqrt{\pi}\end{aligned}$$

By taking $n \to \infty$, we get $\alpha = \sqrt{2\pi}$, as desired. □

6.4 Mean value theorem for integrals

Theorem 6.20 (First mean value theorem)

Suppose μ is increasing on $[a,b]$ and f is continuous, g is non-negative and integrable on $[a,b]$. Then $\exists \xi \in [a,b]$ such that

$$\int_a^b fg \,\mathrm{d}\mu = f(\xi)\int_a^b g \,\mathrm{d}\mu.$$

Proof.

$$\min\left(f(x)\int_a^b g\,\mathrm{d}\mu\right) \leq \int_a^b fg\,\mathrm{d}\mu \leq \max\left(f(x)\int_a^b g\,\mathrm{d}\mu\right)$$

From intermediate value theorem on f we get the result. □

Theorem 6.21 (Second mean value theorem)

If g is Riemann integrable on $[a,b]$, f is non-negative and decreasing on $[a,b]$. Then $\exists c \in [a,b]$ such that

$$\int_a^b fg\,\mathrm{d}x = f(a)\int_a^c g\,\mathrm{d}x.$$

Proof. We prove that case where g is continuous first.

Let $G(x) = \int_a^x g(t)\,\mathrm{d}t$. By intergration by parts (here we make use of Stieltjes integral),

$$-\int_a^b (-f)\,\mathrm{d}G = (f\cdot G)\big|_a^b + \int_a^b G(x)\,\mathrm{d}(-f).$$

And note that

$$f(a)\min G \leq (f\cdot G)\big|_a^b + \int_a^b G(x)\,\mathrm{d}(-f) \leq f(a)\max G$$

By intermediate theorem on G and we're done.

Now we come back to the general case. Since g is integrable, $\forall \varepsilon > 0$, there exists a continuous (actually we can let it be smooth) function g_ε s.t. $\int_a^b |g - g_\varepsilon| \, \mathrm{d}x < \varepsilon$.

(The process is quite boring, the sketch is that we find a elchelon function first, and use a continuous(smooth) function to approch it)

Hence $\exists c_\varepsilon \in [a, b]$ such that

$$\int_a^b f g_\varepsilon \, \mathrm{d}x = f(a) \int_a^{c_\varepsilon} g_\varepsilon \, \mathrm{d}x.$$

$$\int_a^b fg \, \mathrm{d}x = \int_a^b f(g - g_\varepsilon) \, \mathrm{d}x - f(a) \int_a^{c_\varepsilon} (g - g_\varepsilon) \, \mathrm{d}x + f(a) \int_a^{c_\varepsilon} g \, \mathrm{d}x$$

$$\implies \left| \int_a^b fg \, \mathrm{d}x - f(a) \int_a^{c_\varepsilon} g \, \mathrm{d}x \right| \le 2f(a)(b-a)\varepsilon$$

From here we claim that there exists $c \in [a, b]$ s.t. $\int_a^b fg \, \mathrm{d}x = f(a) \int_a^c g \, \mathrm{d}x$.

Because the range of $G(t) := f(a) \int_a^t g \, \mathrm{d}x$ is a closed interval(intermediate value theorem), and it can be arbitarily close to $\int_a^b fg \, \mathrm{d}x$, so it must contains $\int_a^b fg \, \mathrm{d}x$. □

Remark 6.22 — This proof shows the idea of "approximating general functions using good functions" again. This is an important method in analysis.

Corollary 6.23

If f is monotone and g integrable on $[a, b]$, then $\exists c \in [a, b]$,

$$\int_a^b f(x)g(x) \, \mathrm{d}x = f(a) \int_a^c g(x) \, \mathrm{d}x + f(b) \int_c^b g(x) \, \mathrm{d}x.$$

6.5 Improper integrals

When we talk about Riemann integral $\int_a^b f(x) \, \mathrm{d}x$, we require that $[a, b]$ is a closed interval and f is bounded. There are 2 cases which make the integral "improper", namely when one of a, b is infinity or f is unbounded.

We can combine these 2 cases into one: on interval $[a, b)$ (where b can be infinity), if $\forall c \in [a, b)$, f is Riemann-integrable on $[a, c]$, and $\lim\limits_{c \to b^-} \int_a^c f(x) \, \mathrm{d}x$ exists, we define the improper integral $\int_a^b f(x) \, \mathrm{d}x$ to be the limit.

For an open interval (a, b), $\int_a^b f(x) \, \mathrm{d}x = \lim\limits_{\varepsilon_2, \varepsilon_1 \to 0^+} \int_{a+\varepsilon_2}^{b-\varepsilon_1} f(x) \, \mathrm{d}x$, here $\varepsilon_1, \varepsilon_2$ are independent variables.

Example 6.24

$$\int_1^{+\infty} \frac{\mathrm{d}x}{x^\alpha} = \lim_{N \to +\infty} \int_1^N \frac{\mathrm{d}x}{x^\alpha} = \lim_{N \to +\infty} \frac{N^{1-\alpha} - 1}{1 - \alpha}$$

When $\alpha > 1$ it is convergent, and divergent otherwise.

If an improper integral $\int_a^b |f| \, \mathrm{d}x$ is convergent, we say f is **absolutely convergent**. This implies that $\int_a^b f \, \mathrm{d}x$ converges.(by Cauchy convergence criterion)

If $\int_a^b f(x) \, \mathrm{d}x$ is not absolutely convergent but it converges, we say it's **conditionally convergent**.

Example 6.25

$$\begin{aligned}
\int_1^{+\infty} \frac{\sin x}{x} \, \mathrm{d}x &= \lim_{N \to +\infty} \left(-\int_1^N \frac{\mathrm{d}\cos x}{x} \right) \\
&= \lim_{N \to +\infty} \left(-\left. \frac{\cos x}{x} \right|_1^N - \int_1^N \frac{\cos x}{x^2} \, \mathrm{d}x \right) \\
&= \cos 1 - \int_1^{+\infty} \frac{\cos x}{x^2} \, \mathrm{d}x
\end{aligned}$$

So it converges, but obviously it's not absolutely convergent.

Example 6.26 (The Γ function)

$$\Gamma(s) := \int_0^{+\infty} e^{-x} x^{s-1} \, \mathrm{d}x$$

$$\int_0^1 e^{-x} x^{s-1} \, \mathrm{d}x \le \int_0^1 x^{s-1} \, \mathrm{d}x = \left. \frac{x^s}{s} \right|_0^1 = \frac{1}{s}$$

$$\int_1^{+\infty} e^{-x} x^{s-1} \le C(s) \int_1^{+\infty} \frac{\mathrm{d}x}{x^2} < +\infty$$

where $C(s)$ is a constant, so it is well-defined (i.e. the integral converges).

Some facts about Γ function:

$$\Gamma(s+1) = \int_0^{+\infty} e^{-x} x^s \, \mathrm{d}x = -\int_0^{+\infty} x^s \, \mathrm{d}e^{-x} = -\left. x^s e^{-x} \right|_0^{+\infty} + \int_0^{+\infty} e^{-x} s x^{s-1} \, \mathrm{d}x = s\Gamma(s).$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma(p)\Gamma(1-p) = \frac{\pi}{\sin(p\pi)}, 0 < p < 1.$$

6.6 Integration and differentiation

Theorem 6.27 (Integrating infinte series)

Let f_n be integrable functions on $[a, b]$, $\sum\limits_{n=1}^{\infty} f_n$ uniformly converges and the sum is integrable, then

$$\int_a^b \left(\sum_{n=1}^{\infty} f_n(x) \right) \mathrm{d}x = \sum_{n=1}^{\infty} \int_a^b f_n(x) \, \mathrm{d}x.$$

Proof. $\forall \varepsilon > 0, \exists N$ such that $|\sum_{k=n}^{\infty} f_k(x)| < \varepsilon$, $\forall x \in [a, b], n > N$.

Thus
$$\left|\int_a^b \sum_{n=1}^{\infty} f_n(x)\,\mathrm{d}x - \sum_{n=1}^{N}\int_a^b f_n(x)\,\mathrm{d}x\right| < \varepsilon(b-a)$$

□

Example 6.28

$$\int_0^x \frac{1}{1-t}\,\mathrm{d}t = \int_0^x (1+t+t^2+\cdots)\,\mathrm{d}t,\ |x|<1$$
$$\implies -\ln(1-x) = x + \frac{x^2}{2} + \cdots,\ |x|<1$$

Similarly,
$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - \cdots,\ |x|<1$$
$$\implies \arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots,\ |x|<1$$

Taking the limit $x \to 1$ we get:
$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \cdots$$

Proposition 6.29 (Differentiate an intergral with a parameter)

Suppose a function $f(x,t) : [a,b]\times[c,d] \to \mathbb{R}$ is continuous, and $\frac{\partial f}{\partial t}(x,t)$ is continuous. Let
$$F(t) = \int_a^b f(x,t)\,\mathrm{d}x$$

Then
$$F'(t) = \int_a^b \frac{\partial f}{\partial t}(x,t)\,\mathrm{d}x,\quad t\in[c,d]$$

Proof. By Lagrange's mean value theorem,
$$\frac{F(t+h)-F(h)}{h} - \int_a^b \frac{\partial f}{\partial t}(x,t)\,\mathrm{d}x = \int_a^b \left(\frac{\partial f}{\partial t}(x,t+\theta h) - \frac{\partial f}{\partial t}(x,t)\right)\mathrm{d}x$$
where $\theta = \theta(x,h) \in [0,1]$.

From the continuity of $\frac{\partial f}{\partial t}(x,t)$ and the compactness of $[a,b]\times[c,d]$, we know $\frac{\partial f}{\partial t}(x,t)$ is uniformly continuous.

Thus $\forall\varepsilon,\ \exists\delta,\ \forall|h|<\delta,\ |\frac{\partial f}{\partial t}(x,t+\theta h) - \frac{\partial f}{\partial t}(x,t)| < \varepsilon$.
$$\left|\frac{F(t+h)-F(h)}{h} - \int_a^b \frac{\partial f}{\partial t}(x,t)\,\mathrm{d}x\right| < \int_a^b \varepsilon\,\mathrm{d}x = (b-a)\varepsilon$$

□

Proposition 6.30 (Differentiate an improper intergral with a parameter)

Suppose a function $f(x,t)$ and $\frac{\partial f}{\partial t}(x,t)$ is continuous on $[a,b) \times [c,d]$. The integral of $\frac{\partial f}{\partial t}(x,t)$ on $[a,b)$ is uniformly convergent with respect to t.

If there exists $t_0 \in [c,d]$ such that $f(x,t_0)$ is integrable on $[a,b)$ (the improper integral converges), then $\int_a^b f(x,t)\,\mathrm{d}x$ is uniformly convergent with respect to t.

Let

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

we have

$$F'(t) = \int_a^b \frac{\partial f}{\partial t}(x,t)\,\mathrm{d}x, \quad t \in [c,d]$$

Proof. Take a sequence $\{b_n\} \to b$, let

$$F_n(t) = \int_{b_{n-1}}^{b_n} f(x,t)\,\mathrm{d}x.$$

From the above proposition, we know

$$F_n'(t) = \int_{b_{n-1}}^{b_n} \frac{\partial f}{\partial t}(x,t)\,\mathrm{d}x.$$

By the conditions, $\sum F_n'(t)$ uniformly converges, and $\exists t_0$ s.t. $\sum F_n(t_0)$ converges. We deduce the conclusion from Theorem 4.18:

$$F'(t) = \left(\sum_{i=1}^{\infty} F_n(t)\right)' = \sum_{n=1}^{\infty} F_n'(t) = \int_a^b \frac{\partial f}{\partial t}(x,t)\,\mathrm{d}x$$

□

Example 6.31

We have

$$\int_0^{+\infty} \frac{\sin x}{x}\,\mathrm{d}x = \frac{\pi}{2}$$

by computing the derivative of

$$F(t) = \int_0^{+\infty} e^{-tx}\frac{\sin x}{x}\,\mathrm{d}x$$

and proving F is continuous at 0.

7 Applications of integration

7.1 The uniqueness of the solution of ODE(Optional)

We need some preparations before we actually handle it:

Theorem 7.1 (Contraction mapping principle)

Let (X, d) be a complete metric space. If a continuous map $T : X \to X$ and a constant $0 < \gamma < 1$ satisfying

$$d(T(x), T(x')) \leq \gamma d(x, x').$$

Then T has a unique fixed point x_*.

Proof. Let x_0 be an arbitarty point in X, and $x_{n+1} = f(x_n)$, then $\{x_n\}$ is Cauchy therefore convergent.

Let y be the convergency point, we have $f(y) = y$ since f is contiuous. The uniqueness of y is trivial. □

Now comes the main theorem:

Theorem 7.2 (Cauchy-Lipschitz)

Let $\Omega \subset \mathbb{R}^n$ be an open set, f be a continuous function on $\Omega \times (a, b)$ satisfying that there exists a constant $C > 0$, s.t. $\forall t \in (a, b)$,

$$|f(x, t) - f(y, t)| \leq C|x - y|, \quad \forall x, y \in \Omega, t \in (a, b)$$

Then for any $(x_0, t_0) \in \Omega \times (a, b)$, there exists $\delta = \delta(x_0, t_0) > 0$, such that there's a unique map $x(t) : (t_0 - \delta, t_0 + \delta) \to \Omega$ satisfying:

$$\begin{cases} x'(t) = f(x(t), t) \\ x(t_0) = x_0 \end{cases}$$

Proof. First we solve the equation

$$x'(t) = f(x_0, t), \quad x(t_0) = x_0$$

By Newton-Leibniz formula we get a solution

$$x_1(t) = x_0 + \int_{t_0}^{t} f(x_0, \tau)\,\mathrm{d}\tau, \quad t \in (a, b)$$

If we substitute x_1 into the equation we'll get another solution x_2, and a series of function $x_n(t)$, we wish to prove it converges to the desired function. This process is called *Picard iteration.*

Let

$$T(x(t)) = x_0 + \int_{t_0}^{t} f(x(\tau), \tau)\,\mathrm{d}\tau$$

be a map.

We want to use Contraction mapping principle on this map to get the "fixed point".

To do this, we need to find a complete metric space X and prove T is indeed a contraction map.

Construction of space X:

$$X = \{x(t) \mid x : [t_0 - \delta, t_0 + \delta] \to \Omega \text{ continuous}, \ x(t_0) = x_0\}$$

where the distance is given by L^∞ norm:

$$\|x(t)\| := \sup_{t \in [t_0 - \delta, t_0 + \delta]} |x(t)|.$$

Remark 7.3 — We can check this is indeed a metric space. But X isn't complete, because the image set Ω is open (there can be a series of function $\{x_n(t)\}$ with $x_n(t_1) \to y \in \partial\Omega$ for some $t_1 \in [t_0 - \delta, t_0 + \delta]$).

To solve this problem, we need to restrict the image of x to a closed subset of Ω.

Since Ω is open, $\exists \delta_0 > 0$ such that $B_{\delta_0}(x_0) \subset \Omega$.

Let $X' = \{x \in X \mid \|x(t) - x_0\| \le \delta_1 < \delta_0\}$, X' is a closed subset of X and thus complete.

Now we're going to prove $T : X' \to X'$ is a contraction map.

First note that $T(x)$ is continuous (f, x both are continuous), $T(x(t_0)) = t_0$.

Check $T(x) \in X'$: Since f is continuous on the closed set $\overline{B_{\delta_1}(x_0)} \times [t_0 - \delta, t_0 + \delta]$, $|f|$ has a bound M.

$$\|T(x(t)) - x_0\| \le |t - t_0| M \le \delta M$$

So we need $\delta < \frac{\delta_1}{M}$.

From the Lipschitz condition of f,

$$\begin{aligned}
|T(x(t)) - T(y(t))| &= \left| \int_{t_0}^{t} (f(x(\tau), \tau) - f(y(\tau), \tau)) \, \mathrm{d}\tau \right| \\
&\le \left| \int_{t_0}^{t} C |x(\tau) - y(\tau)| \, \mathrm{d}\tau \right| \\
&\le C\delta \|x - y\|, \quad \forall t \in [t_0 - \delta, t_0 + \delta]
\end{aligned}$$

Thus if we take $\delta < \min\{\frac{1}{2C}, \frac{\delta_1}{M}\}$, T is a contraction map. Hence the conclusion follows from Theorem 7.1. □

But usually we'll see the differential equations containing terms like $f''(x)$, the solution is: to view the equations as equations of a vector-valued function $F(x) = (f(x), f'(x))$.

With this tool we'll prove the famous Kepler's laws of planetary motion.

- Each planet's orbit about the Sun is an ellipse. The Sun is at one focus.
- The imaginary line joining a planet and the Sun sweeps equal areas of space during equal time intervals as the planet orbits.

- The squares of the orbital periods of the planets are directly proportional to the cubes of the semi-major axes of their orbits.

These laws are derived from a lot of observations by Kepler. Now we're trying to imply these laws from Newton's laws by solving differetial equations.

The location of a planet is a function with respect to the time t: $\gamma : \mathbb{R} \to \mathbb{R}^3$. We write $\gamma(t) = (x(t), y(t), z(t))$.

The velocity and acceleration are denoted by $\dot{\gamma}$ and $\ddot{\gamma}$ (where the dots represents the derivative).

By Newton's laws we have

$$m\ddot{\gamma}(t) = -\frac{GMm}{r^2} \cdot \frac{\gamma(t)}{r}.$$

WLOG $z(0) = z'(0) = 0$ (by choosing a suitable coordinate system).

$$\frac{\mathrm{d}}{\mathrm{d}t}(\gamma \times \dot{\gamma}) = \dot{\gamma} \times \dot{\gamma} + \gamma \times \ddot{\gamma} = 0$$

This tells us $\gamma \times \dot{\gamma} = \gamma(0) \times \dot{\gamma}(0), \forall t$. We'll prove the third coordinate(z) is always 0 and reduce it to a two-dimensional problem.

$$\gamma \times \dot{\gamma} = (yz' - zy', zx' - xz', xy' - yx') = \gamma(0) \times \dot{\gamma}(0) = c(0, 0, 1)$$

We may assume $xy' - yx' \neq 0$ (otherwise the locus lies within a straight line), hence we have $z = z' = 0$ by solving linear equations.

Let $\gamma(t) = (x(t), y(t)) = r(\cos\theta, \sin\theta)$. Compute:

$$\dot{\gamma}(t) = (r'\cos\theta - r\sin\theta \cdot \theta', r'\sin\theta + r\cos\theta \cdot \theta')$$

$$\begin{aligned}\ddot{\gamma}(t) = (&r''\cos\theta - 2r'\sin\theta \cdot \theta' - r(\sin\theta \cdot \theta'' + \cos\theta \cdot (\theta')^2),\\ &r''\sin\theta + 2r'\cos\theta \cdot \theta' + r(\cos\theta \cdot \theta'' - \sin\theta \cdot (\theta')^2))\end{aligned}$$

The equation gives

$$r''\cos\theta - 2r'\sin\theta \cdot \theta' - r(\sin\theta \cdot \theta'' + \cos\theta \cdot (\theta')^2) = -\frac{GM}{r^2}\cos\theta \tag{1}$$

$$r''\sin\theta + 2r'\cos\theta \cdot \theta' + r(\cos\theta \cdot \theta'' - \sin\theta \cdot (\theta')^2) = -\frac{GM}{r^2}\sin\theta \tag{2}$$

Observe that $(1) \times \sin\theta - (2) \times \cos\theta = 0$ and $(1) \times \cos\theta + (2) \times \sin\theta = -\frac{GM}{r^2}$:

$$2r'\theta' + r\theta'' = 0 \tag{3}$$

$$r'' - r(\theta')^2 = -\frac{GM}{r^2} \tag{4}$$

Now (3) gives $0 = 2rr'\theta' + r^2\theta'' = (r^2\theta')' \implies r^2\theta' = c$ which is Kepler's 2nd law.

(The tiny sector which was swept through with radius r has area $\pi r^2(\theta(t + \Delta t) - \theta(t))$, so $\pi r^2\theta'$ is the derivative of the area swept through by the segment)

From (4):

$$r'' - \frac{c^2}{r^3} = -\frac{GM}{r^2}$$

$$2r'r'' - 2r'c^2r^{-3} = -\frac{GM}{r^2}2r'$$
$$((r')^2)' + c^2(r^{-2})' = 2GM(r^{-1})'$$

Thus $\exists C_1, C_2$ such that

$$(r')^2 + c^2r^{-2} = 2GMr^{-1} + C_2$$
$$(r')^2 = C_1^2 - \left(\frac{c}{r} - \frac{GM}{c}\right)^2$$

Set

$$\begin{cases} \frac{c}{r} - \frac{GM}{c} = C_1\cos\beta(t) \\ r' = C_1\sin\beta(t) \end{cases}$$

Taking the derivative of the first equation,

$$-\frac{cr'}{r^2} = -r'\beta'(t) \implies \beta'(t) = \frac{c}{r^2}$$

This is saying $\beta'(t) = \theta'(t)$ (recall that $r^2\theta' = c$), and we can WLOG assume $\theta(0) = \beta(0)$ (just rotate the axes to change θ by a constant).

Therefore we can get

$$r = \frac{l}{1 + e\cos\theta}, \quad e = \frac{C_1 c}{GM}, l = \frac{c^2}{GM}$$

Which means the locus is an ellipse (techiquely it can also be a parabola or hyperbola, and you can see it makes sense too) with a focus on the origin (this is Kepler's 1st law).

And the area of the ellipse $T = \frac{\pi ab}{c}$ satisfies $T^2 = \frac{\pi^2 a^3}{GM}$, which proves Kepler's 3rd law.

7.2 The length of curves and Brachistochrone problem

Definition 7.4. A **curve** is a map $\gamma : [0,1] \to \mathbb{R}^d$. We say a curve is **rectifiable** if

$$\sup_{0=t_0<t_1<\cdots<t_n=1} \sum_{i=1}^n |\gamma(t_i) - \gamma(t_{i-1})| < +\infty$$

The supremum is called the **length** of the curve $L(\gamma)$.

If $\gamma \in C^1$,

$$|\gamma(t_i) - \gamma(t_{i-1})| = \sqrt{\sum_{j=1}^d (x_j(t_i) - x_j(t_{i-1}))^2} = (t_i - t_{i-1})\sqrt{\sum_{j=1}^d (x_j'(\xi_{ji}))^2}$$

From the uniform continuity of x_j', we can prove $\sum |\gamma(t_i) - \gamma(t_{i-1})| \to \int_0^1 |\gamma'(t)|\,\mathrm{d}t$.

Theorem 7.5

If $\gamma \in C^1$, then

$$L(\gamma) = \int_0^1 |\gamma'(t)|\,\mathrm{d}t.$$

Proof. The skectch are decribed above, the details are omitted (you might need mean value theorem of integrals). □

This induces the arc length parameter:

$\varphi : [0,1] \to [0, L(\gamma)]$ with $\varphi(t) = \int_0^t |\gamma'(x)| \, dx$, obviously φ is increasing and C^1, so by implicit function theorem φ is a C^1 homeomorphism.

If we consider the curve with a new parameter $\gamma \circ \varphi^{-1} : [0, L(\gamma)] \to \mathbb{R}^n$ (this is called the arc length parameter).

$$\frac{d\gamma(\varphi^{-1}(y))}{dy} = \gamma'(\varphi^{-1}(y))(\varphi^{-1})'(y) = \frac{\gamma'(\varphi^{-1}(y))}{|\gamma'(\varphi^{-1}(y))|}$$

So in this parameter we always have $|\gamma'| = 1$.

Remark 7.6 — The continuity cannot grant theat the curve is rectifiable, and rectifiable curve need not be continouous.

Definition 7.7. A real function F on $[a,b]$ is called a **bounded variation function**, if its total variation

$$T_F(a,b) = \sup_\sigma \sum_{j=1}^{n} |F(x_j) - F(_{j-1})|$$

is finite. All such functions form a space $BV([a,b])$.

Theorem 7.8

A curve $\gamma = (x_1, \ldots, x_n)$ is rectifiable if each of x_i is a bounded variation function.

Example 7.9

Some examples of BV functions:

- If F is increasing and bounded, $T_F([a,b]) = F(b) - F(b)$;
- If F is differentiable on $[a,b]$ and F' is bounded, then $T_F([a,b]) \le (b-a) \sup F'$;
- If F' is Riemann integrable, $T_F([a,b]) = \int_a^b |F'| \, dx$.

We can divide the total variation into positive part and negative part:

$$P_F(a,b) = \sup_\sigma \sum_{j=1}^{n} (F(x_j) - F(x_{j-1}))_+$$

$$N_F(a,b) = \sup_\sigma \sum_{j=1}^{n} (F(x_j) - F(x_{j-1}))_-$$

where $a_+ = \max a, 0, a_- = -\min a, 0$.

Lemma 7.10

If $F \in BV([a,b])$, $a \leq x \leq b$, we have

$$F(x) - F(a) = P_F(a,x) - N_F(a,x), \quad T_F(a,b) = P_F(a,b) + N_F(a,b).$$

Proof. WLOG assume $x = b$.

By definition,

$$F(x_j) - F(x_{j-1}) = (F(x_j) - F(x_{j-1}))_+ - (F(x_j) - F(x_{j-1}))_-,$$

$$|F(x_j) - F(x_{j-1})| = (F(x_j) - F(x_{j-1}))_+ + (F(x_j) - F(x_{j-1}))_-.$$

As we've proven in Darboux sums, refining the partition will make the sum larger. By taking refinement of the partition we're more or less done. □

Theorem 7.11 (Jordan's decomposition of BV functions)

A function F is bounded variation function iff it can be written as a difference between two bounded increasing functions.

Proof. If $F = F_1 - F_2$ obviously $F \in BV([a,b])$.

If $F \in BV([a,b])$, let $F_1 = F(a) + P_F(a,x), F_2 = N_F(a,x)$. □

Remark 7.12 — $BV([a,b])$ is an algebra and a Banach space with norm $\|F\| = \sup F + T_F(a,b)$.

Let's come to the problem of brachistochrone.

8 Lebesgue measure theory

8.1 Cantor set and σ-algebra

Definition 8.1 (Cantor set)**.** Divide the interval $[0,1]$ equally into three intervals and remove the one in the middle, we get two non-intersect closed set

$$\left[0, \frac{1}{3}\right], \left[\frac{2}{3}, 1\right].$$

Apply the same process to these intervals, we get four sets:

$$\left[0, \frac{1}{9}\right], \left[\frac{2}{9}, \frac{1}{3}\right], \left[\frac{2}{3}, \frac{7}{9}\right], \left[\frac{8}{9}, 1\right].$$

Repeating this process, we get 2^n intervals with length 3^{-n} at the n-th step. Denote the union of these sets by F_n, it's clear F_n is a series of nested closed sets, the limit is known as the **Cantor set**

$$\mathcal{C} = \bigcap_{n=1}^{\infty} F_n = \lim_{n \to \infty} F_n.$$

The Cantor set has the following properties:

- $\mathcal{C}$ is a nonempty closed set. (From the nested closed sets theorem)
- $\mathcal{C}$ doesn't have inner points, i.e. $\mathring{\mathcal{C}} = \varnothing$.

  Note that F_n consists of intervals of length 3^{-n}, so any given interval (a,b) cannot be contained in $\mathcal{C}$.
- $\mathcal{C}$ is a complete set, i.e. $\mathcal{C}' = \mathcal{C}$.

  Because $x \in \mathcal{C} \iff x \in [a_n, b_n]$, we know $a_n, b_n \in \mathcal{C}$, so $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = x$, $\mathcal{C}$ doesn't have isolated points.
- $\mathcal{C}$ is completely disconnected, i.e. $\forall x, y \in \mathcal{C}$, $\exists z \notin \mathcal{C}$, $x < z < y$.

  Because when n gets sufficiently large, x, y are in distinct intervals of length 3^{-n}.
- $\mathcal{C} = \{\sum_{n=1}^{\infty} \frac{a_n}{3^n} \mid a_n = 0, 2\}$.
- $card(\mathcal{C}) = c = card(\mathbb{R})$, this follows from it's a complete set. (Also follows from the above expression)
- The intervals removed have a total length of

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

  This means Cantor set has "length" zero.

Consider a map $F : \mathcal{C} \to [0,1]$ which shows $card(\mathcal{C}) = c$:

$$x = \sum_{n=1}^{\infty} \frac{a_n}{3^n} \mapsto F(x) = \sum_{n=1}^{\infty} \frac{\frac{1}{2}a_n}{2^n}, a_n \in \{0, 2\}.$$

We can see that F is surjective and increasing.
Moreover, note that for an interval (e, f) removed at the n-th step, we have

$$e = \sum_{k=1}^{n-1} \frac{a_k}{3^k} + \sum_{k=n+1}^{\infty} \frac{2}{3^k}, \quad f = \sum_{k=1}^{n-1} \frac{a_k}{3^k} + \frac{2}{3^n}.$$

So by definition, $F(e) = F(f)$, we can extend F to the interval (e, f) with $F(x) = F(e), \forall x \in (e, f)$.

Therefore we can extend F to $[0,1]$, this function is called the **Cantor-Lebesgue function**. Clearly F is continuous, and has derivative 0 on $\mathcal{C}^c$. (because it's constant on each interval of $[0,1]\backslash\mathcal{C}$)

At the point $x \in \mathcal{C}$, we have

$$\lim_{n\to\infty} \frac{\sum_{k=n}^{\infty} \frac{\frac{1}{2}a_k}{2^k}}{\sum_{k=n}^{\infty} \frac{a_k}{3^k}} = +\infty.$$

This shows that although F has derivative zero almost everywhere, but $F(x)$ isn't constant and doesn't satisfy the Fundamental theorem of Calculus.

It's also known as "demon's staircase", a continuous function consisting of infinitely many staircase.

To measure the size of more complex sets, we allow the union of countable many closed sets to be a F_σ set, and the intersection of countable many open sets to be a G_δ set. (F comes from a French word, G comes from a German word)

We can also define $F_{\sigma\delta}$ sets and $G_{\delta\sigma}$ sets.

Example 8.2

The continuous point of a function f on $\mathbb{R}^n$ is a G_δ set.

In fact, it can be represented as

$$\bigcap_{k=1}^{\infty}\left\{x : \lim_{\delta\to 0}\sup_{y,z\in B_x(\delta)}|f(y)-f(z)|<\frac{1}{k}\right\}.$$

and each set above is open.

Recall that a set E is dense if $\overline{E}=\mathbb{R}^n$, and **nowhere dense** if $\mathring{\overline{E}}=\varnothing$. The countable union of nowhere dense sets are called **meager set** or **first category set**, and other sets are **second category sets**.

Note that subsets of first category cannot be the entire set (Baire's category theorem), this can be used to prove some existence problem.

Example 8.3

Aussme that $f_k(x)\in C(\mathbb{R}^n)$, $\lim_{k\to\infty}f_k(x)=f(x)$. Then the set of discontinuous points of f is of first category, which shows f must have continuous points.

Proof. f discontinuous at point $x \iff$ exists an open interval $I\subset\mathbb{R}$ s.t. $x\in f^{-1}(I)\backslash(\mathring{f^{-1}}(I))$.

Now we prove $f^{-1}(I)$ is a F_σ set: Let $\{I_n\}$ be the open intervals with rational center and radius, we also require $\overline{I_n}\subset I$. Then we have

$$f^{-1}(I)=f^{-1}\left(\bigcup_n I_n\right)=f^{-1}\left(\bigcup_n \overline{I_n}\right)=\bigcup_{n,m\geq 1}\bigcap_{k\geq m}f_k^{-1}\left(\overline{I_n}\right).$$

The reasons are as follows:

For $x\in f^{-1}(I)$, exists I_n s.t. $f(x)\in I_n$. Because f_k converges to f, $f_m(x)\in\overline{I_n}$ for sufficiently large m.

Conversely, if exists n, for all sufficiently large m, $f_m(x)\in\overline{I_n}$ always holds, then from convergence $x\in f(\overline{I_n})$.

$f^{-1}(I)$ is a F_σ set $\implies f^{-1}(I)\backslash(\mathring{f^{-1}}(I))$ is of first category (it is a subset of a countable union of boundaries of closed sets), from Baire's theorem we deduce the discontinuous points of f are of first category. □

We define Baire-0 function to be continuous functions, and Baire-(n+1) function to be the limit of Baire-n functions.

Definition 8.4. We say a family Γ of subsets of X is a σ-algebra, if

- $\varnothing \in \Gamma$;
- $A \in \Gamma \implies A^c \in \Gamma$;
- $A_n \in \Gamma \implies \bigcup_{n=1}^{\infty} A_n \in \Gamma$.

It's clear σ-algebra is closed under countable intersections. In the sense of generated groups/ideals/spaces, we define the smallest σ-algebra containing Σ to be the generated σ-algebra $\Gamma(\Sigma)$.

Let $\mathcal{B}$ denote the σ-algebra generated by all the open sets (called Borel σ-algebra), the sets in $\mathcal{B}$ are called Borel sets.

Definition 8.5. An increasing sequence of subsets in X is

$$A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$$

define $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

Similarly a decreasing sequence is

$$B_1 \supset B_2 \supset \cdots \supset B_n \supset \cdots$$

define $\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n$.

In the same spirit we can define the superior limit and inferior limit for general sequences of sets, namely

$$\limsup_{k\to\infty} E_k := \bigcap_{k=1}^{\infty} \bigcup_{l=k}^{\infty} E_k, \quad \liminf_{k\to\infty} E_k := \bigcup_{k=1}^{\infty} \bigcap_{l=k}^{\infty} E_k.$$

8.2 Outer measure

We want to define a "measure" on a family of sets $\mathscr{A}$ to describe how large each set is, and we want it to satisfy **countable additivity**. This requires $\mathscr{A}$ is closed under countable unions. This leads us to σ-algebras.

Definition 8.6 (Measure spaces)**.** If $\mathscr{A}$ is a σ-algebra on X, we say $(X, \mathscr{A})$ is a measurable space. If there's a non-negative function μ on $\mathscr{A}$ satisfies:

- $\mu(\varnothing) = 0$;
- For any pairwise disjoint sets A_i, we have

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Then we say the triple $(X, \mathscr{A}, \mu)$ is a **measure space**.

Lebesguue measure is a measure space on $X = \mathbb{R}^n$ satisfying $\mu(I) = |I|$, where I is a bounded cuboid in $\mathbb{R}^n$. In what follows we'll give the construction of Lebesgue measure.

Let $\mathscr{E}$ be the set of all the finite unions of bounded cuboid, we call it the base set.

There's a natural function m such that $m(I) = |I|$ which is finitely additive. From m we can induce the **outer measure** m^*

$$m^*(E) = \inf\left\{\sum_{k=1}^{\infty} m(I_k), \quad E \subset \bigcup_{k=1}^{\infty} I_k, I_k \in \mathscr{E}\right\}.$$

The outer measure m^* in $\mathbb{R}^n$ has the following properties:

- It's non-negative;
- It's monotone, i.e. $m^*(A) \leq m^*(B)$ for $A \subset B$;
- **countable subadditivity**: $m^*\left(\bigcup_{k=1}^{\infty} E_k\right) \leq \sum_{k=1}^{\infty} m^*(E_k)$;
- If $d(E_1, E_2) > 0$, then $m^*(E_1 \cup E_2) = m^*(E_1) + m^*(E_2)$;
- $m^*(E + \{x_0\}) = m^*(E)$, $m^*(\lambda E) = |\lambda|^n m^*(E)$.

More generally, if a non-negative function satisfies the first three properties, we say it's an **outer measure**.

Proof of the third property. WLOG $\sum m^*(E_k) < \infty$, for any $\varepsilon > 0$, we can find $I_{k,i} \in \mathcal{E}$ such that

$$\sum_{i=1}^{\infty} |I_{k,i}| < m^*(E_k) + \frac{\varepsilon}{2^k}.$$

$\{I_{k,i}\}$ is a cover of $\bigcup E_k$, hence

$$m^*\left(\bigcup_{k=1}^{\infty} E_k\right) \leq \sum_{k,i=1}^{\infty} |I_{k,i}| = \sum_{k=1}^{\infty} M^*(E_k) + \varepsilon.$$

Since ε can be arbitarily small, the conclusion follows. □

Example 8.7

$m^*(\mathcal{C}) = 0$.

Proof. $\mathcal{C} \subset F_n$, hence

$$m^*(\mathcal{C}) \leq m^*(F_n) = 3^{-n}2^n \to 0.$$

□

This induces a "distance" $d^*(A, B) := m^*(A\Delta B)$, where Δ is the symmetric difference.

We can check d^* satisfies symmetry and triangle inequality, but $d^*(A, B) = 0$ cannot imply $A = B$.

8.3 Measurable sets

Because of the sepciality of Eucildean space, measurable sets are defined to be the limit set of open sets at first. But later in abstract measure theory, the measurable sets are defined to satisfy *Caratheodory condition*, that is:

Definition 8.8. We say a set E is **measurable** (Lebesgue measurable) if

$$m^*(T) = m^*(E \cap T) + m^*(T \cap E^c), \quad \forall T \subset \mathbb{R}^n.$$

Let $\mathscr{U}$ be the collection of all the measurable sets in $\mathbb{R}^n$.

Example 8.9

If $m^*(E) = 0$, then E is measurable. We say such E is **null**. Because

$$m^*(E \cap T) + m^*(T \cap E^c) = m^*(T \cap E^c) \leq m^*(T)$$

We say a measure space is **complete** if all the null sets are measurable.

From this fact we know $card(\mathscr{U}) = 2^c$. It follows from all the subset of Cantor set $\mathcal{C}$ is null, hence measurable. $card(\mathcal{C}) = c \implies card(\mathscr{U}) = 2^c$.

Proposition 8.10

$\mathscr{U}$ is a σ-algebra, and $(\mathbb{R}^n, \mathscr{U}, m^*)$ is a measure space.

Proof. It's obvious that $\varnothing \in \mathscr{U}$, and $E \in \mathscr{U} \implies E^c \in \mathscr{U}$.

Now we need to prove the countable additivity.

First we prove it's finitely additive.

$$\begin{aligned} m^*(T) &= m^*(T \cap E_1) + m^*(T \cap E_1^c) \\ &= m^*(T \cap E_1 \cap E_2) + m^*(T \cap E_1 \cap E_2^c) + m^*(T \cap E_1^c \cap E_2) + m^*(T \cap E_1 \cap E_2^c) \\ &\geq m^*(T \cap (E_1 \cup E_2)) + m^*(T \cap (E_1 \cup E_2)^c) \geq m^*(T) \end{aligned}$$

Hence $\mathscr{U}$ is closed under unions and intersections.

In particular, if $E_1 \cap E_2 = \varnothing$,

$$\begin{aligned} m^*(T \cap (E_1 \cup E_2)) &= m^*(T \cap (E_1 \cup E_2) \cap E_1) + m^*(T \cap (E_1 \cup E_2) \cup E_1^c) \\ &= m^*(T \cup E_1) + m^*(T \cup E_2). \end{aligned}$$

Now assume E_k measurable and pairwise disjoint. Let E be their union and F_k be the union of first k sets.

$$\begin{aligned} m^*(T) &= m^*(T \cap F_k) + m^*(T \cap F_k^c) \\ &= \sum_{i=1}^{k} m^*(T \cap E_i) + m^*(T \cap F_k^c) \\ &\geq \sum_{i=1}^{k} m^*(T \cap E_i) + m^*(T \cap E^c) \end{aligned}$$

Taking the limit $k \to \infty$:

$$m^*(T) \geq \sum_{i=1}^{\infty} m^*(T \cap E_i) + m^*(T \cap E^c) \geq m^*(T \cap E) + m^*(T \cap E^c) \geq m^*(T)$$

which forces all the equality hold. (the middle inequality follows from countable subadditivity)

This implies $E \in \mathscr{U}$, and

$$m^*(E) = \sum_{i=1}^{\infty} m^*(E \cap E_i) + m^*(E \cap E^c) = \sum_{i=1}^{\infty} m^*(E_i).$$

At last for general E_i's (not necessarily disjoint), we have

$$\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} \left(E_i \backslash \bigcup_{k=1}^{i-1} E_k \right) \in \mathscr{U}.$$

□

So we can restrict m^* on $\mathscr{U}$ to get the **Lebesgue measure**, denoted by m. Thus $(\mathbb{R}^n, \mathscr{U}, m)$ is the complete measure space by extending $(\mathscr{E}, m^*)$.

As a corollary, Lebesgue measure satisfies *monotone convergence theorem*:

Theorem 8.11

For a monotone sequence $\{E_k\}$, we have

$$m\left(\lim_{k\to\infty} E_k\right) = \lim_{k\to\infty} m(E_k),$$

if $\{E_k\}$ is decreasing we require $m(E_k)$ is finite for some k.

Proof. Just some abstract nonsense:

If $\{E_k\}$ is increasing,

$$\begin{aligned} m\left(\lim_{k\to\infty} E_k\right) &= m\left(\bigcup_{k=1}^{\infty} E_k\right) \\ &= m\left(\bigcup_{k=1}^{\infty} \left(E_k \backslash \bigcup_{i=1}^{k-1} E_i\right)\right) \\ &= \sum_{k=1}^{\infty} m\left(E_k \backslash \bigcup_{i=1}^{k-1} E_i\right) = \lim_{k\to\infty} m(E_k). \end{aligned}$$

If $\{E_k\}$ is decreasing, just take the complement sets. WLOG $m(E_1)$ is finite,

$$\begin{aligned} m\left(\lim_{k\to\infty} E_k\right) &= m\left(\bigcap_{k=1}^{\infty} E_k\right) \\ &= m(E_1) - m\left(\bigcup_{k=1}^{\infty} (E_1 \backslash E_k)\right) \\ &= m(E_1) - \lim_{k\to\infty} m(E_1 \backslash E_k) = \lim_{k\to\infty} m(E_k). \end{aligned}$$

□

For general sequences, we can consider their superior limit and inferior limit instead.

Theorem 8.12 (Fatou's lemma)

For any measurable sequence $\{E_k\}$, we have

$$m\left(\liminf_{k\to\infty} E_k\right) \leq \liminf_{k\to\infty} m(E_k), \quad m\left(\limsup_{k\to\infty} E_k\right) \geq \limsup_{k\to\infty} m(E_k).$$

where in the second inequality we require $m\left(\bigcup_{l=k}^{\infty} E_l\right)$ is finite for some k.

Proof. By definition we have

$$m\left(\liminf_{k\to\infty} E_k\right) = m\left(\bigcup_{k=1}^{\infty}\bigcap_{l=k}^{\infty} E_l\right) = \lim_{k\to\infty} m\left(\bigcap_{l=k}^{\infty} E_l\right) \leq \liminf_{k\to\infty} m(E_k).$$

Simlarly,

$$m\left(\limsup_{k\to\infty} E_k\right) = m\left(\bigcap_{k=1}^{\infty}\bigcup_{l=k}^{\infty} E_l\right) = \lim_{k\to\infty} m\left(\bigcup_{l=k}^{\infty} E_l\right) \geq \limsup_{k\to\infty} m(E_k).$$

□

This lemma implies the following theorem:

Theorem 8.13 (Borel-Cantelli)

Let E_k be measurable sets, satisfying $\sum_{k=1}^{\infty} m(E_k) < \infty$. We have

$$m(\liminf_{k\to\infty} E_k) = m(\limsup_{k\to\infty} E_k) = 0.$$

Proof. By Fatou's lemma this is obvious. □

Remark 8.14 — This theorem is a famous theorem in probability theory. Its inverse needs stronger conditions to hold.

8.4 Properties of measurable sets

Earlier we proved that null sets are measurable, in this section we'll dig deeper into measurable sets, revealing its structure.

Theorem 8.15

Measurable sets have following properties:

- The cuboid I is measurable. In particular, open sets and closed sets are measurable, hence Borel sets are measurable;
- Give $E \in \mathscr{U}$, for any $\varepsilon > 0$ there exists an open set G and a closed set F such that:
$$F \subset E \subset G, \quad m(G \backslash F) < \varepsilon.$$
In particular, we have
$$E = H \backslash Z_1 = K \cup Z_2,$$
where Z_1, Z_2 are null sets, H is G_δ set, K is F_σ set.

Proof. Let I be a bounded cuboid. $\forall T \subset \mathbb{R}^n$, we need to prove

$$m^*(T) = m * (T \cap I) + m^*(T \cap I^c).$$

For all the covers $\{I_k\}$ of T, $\{I_k \cap I\}$ is a cover of $\{T \cap I\}$, $\{I_k \cap I^c\}$ can be divided to a finite union of cuboids, so it's a cover of $T \cap I^c$. Hence

$$m^*(T) + \varepsilon > \sum |I_k| = \sum |I_k \cap I| + |I_k \cap I^c| \leq m^*(T \cap I) + m^*(T \cap I^c).$$

This shows I is measurable.

Now for the second part.

WLOG $m^*(E) < \infty$ (otherwise write E as a countable union of bounded sets), we can take a cover of *open* cuboids of E (otherwise extend each I_k a bit), by definition,

$$m(E) + \varepsilon > \sum_{k=1}^{\infty} |I_k|.$$

So $G = \bigcup I_k \supset E$, G open and

$$m(G \backslash E) < \varepsilon.$$

For closed set we consider E^c and apply the same arguments.

Then the G_δ set $G = \bigcap G_k$ satisfies $m(G \backslash E) \leq m(G_n) - m(E) \leq \frac{1}{2^n}, \forall n \leq 1$. which shows $G \backslash E$ is null. □

Here H is called the **equi-measure hull** of E, and K the **equi-measure kernel** of E. For general a set E, its equi-measure hull also exist (in the outer measure)

This theorem tells us that Borel sets and measurable sets only differs by null sets.

Example 8.16

There exists a second category set which is null.

Proof. Let $\{r_n\}$ be a permutation of rational numbers in $[0, 1]$, and $I_{n,k} = (r_n - 2^{-n-k}, r_n + 2^{-n-k})$.

Consider $E = \bigcap_k \bigcup_n I_{n,k}$. Clearly E is of measure zero.

Note that $(\bigcup_n I_{n,k})^c$ is nowhere dense, so E^c is of first category, (by Baire's theorem) E is of second category. $\square$

Example 8.17

All the reals satisfying the following condition form null a set:

exists $(p_n, q_n), q_n > 1$ pairwise distinct s.t.

$$\left|x - \frac{p_n}{q_n}\right| \leq q_n^{-3}.$$

Theorem 8.18

Let E be a set whose measure is positive. For all $0 < \lambda < 1$, exists cuboid I s.t.

$$\lambda|I| \leq m(I \cap E).$$

Proof. WLOG $m(E) < \infty$, let $\{I_k\}$ be a cover of E.

Assume that $\lambda|I| > m(I \cap E), \forall I$. We have

$$m(E) + \varepsilon > \sum_k |I_k| > \lambda^{-1} \sum_k m(I_k \cap E) > \lambda^{-1} m(E),$$

which gives a contradiction. $\square$

Theorem 8.19 (Steinhaus)

Let E be a set of positive measure in Euclidean space, there exist $\delta > 0$ s.t.

$$B(0, \delta) \subset E - E.$$

Proof. WLOG $0 < m(E) < \infty$, from above we know there exists a cuboid I such that

$$\lambda|I| < m(I \cap E).$$

Hence we may assume $E \subset I$, and

$$m(E) > \lambda|I|.$$

Suppose there exist $x_k \to 0$ such that $x_n \notin E - E$, which implies $(E + x_n) \cap E = \varnothing$.

$$m(E \cup (E + x_k)) = 2m(E)$$

Note that $E \cup (E + x_k) \subset I \cup (I + x_k)$, so

$$m(E \cup (E + x_k)) < (1 + C(I)|x_k|)|I| < (1 + C(I)|x_k|)\lambda^{-1} m(E),$$

where $C(I)$ is a constant depending on I.

Therefore

$$2m(E) < (1 + C(I)|x_k|)\lambda^{-1}m(E)$$

Take $\lambda \to 1$ and $x_k \to 0$ we deduce a contradition. □

Corollary 8.20

The Cauchy's functional equation

$$f(x+y) = f(x) + f(y), \forall x, y \in \mathbb{R}$$

If f is bounded on a set E of positive measure, then $f(x) = cx$.

Proof. First we know $f(x) = cx$ for $x \in \mathbb{Q}$.

From Steinhaus's theorem, $\exists \delta > 0, [-\delta, \delta] \subset E - E$. So f is also bounded on $[-\delta, \delta]$.

For an irrational number α, $\exists q \in \mathbb{Q}$ such that $|\alpha - q| < \delta$,

$$|f(\alpha) - qf(1)| < M$$

$$\implies |f(n\alpha) - q_n f(1)| < M, \forall n \in \mathbb{Z}.$$

Compute

$$\begin{aligned}|f(x) - xf(1)| &= \frac{1}{n}|f(nx) - nxf(1)| \\ &\leq \frac{1}{n}(|f(nx) - q_n f(1)| + |nx - q_n| \cdot |f(1)|) \\ &\leq \frac{1}{n}(M + \delta|f(1)|).\end{aligned}$$

Let $n \to \infty$ and we're done. □

8.5 Non-measurable sets

In this section we'll construct a non-measurable set using Axiom of Choice.

Divide $[0,1]$ into equivalent classes:

$$[0,1] = \bigcup A_\alpha$$

where $x \sim y \iff x - y \in \mathbb{Q}$.

From each class take an element x_α(by Axiom of Choice), they form a *Vitali set* $\mathscr{N}$.

Theorem 8.21

$\mathscr{N}$ is non-measurable.

Proof. Assume $\mathscr{N} \in \mathscr{U}$, let $\mathscr{N}_q := (\mathscr{N} + q) \cap [0,1]$, $\forall q \in \mathbb{Q}$. For distinct $p, q \in \mathbb{Q}$,

$$\mathscr{N}_p \cap \mathscr{N}_q = \varnothing.$$

So

$$1 = m([0,1]) = \sum_{q\in\mathbb{Q}} m(\mathscr{N}_q),$$

$\implies \mathscr{N}$ is not null.

Note that

$$m(\mathscr{N}_q) \geq m(\mathscr{N}) - |q|,$$

Consider the rational numbers in $(0, \frac{1}{2}m(\mathscr{N}))$, there are countable many such numbers, each $m(\mathscr{N}_q) > \frac{1}{2}m(\mathscr{N}) > 0$, their sum must be infinity, contradiction! □

8.6 Measurable functions(TODO)

For convenience, we allow a function f on $\mathbb{R}^n$ take the value $\pm\infty$, but we require $0\cdot\infty = 0$.

Definition 8.22 (Measurable maps)**.** We say a map $f : (X, \mathcal{A}) \to (Y, \mathcal{B})$ between measure spaces is **measurable**, if

$$f^{-1}(B) \in \mathcal{A}, \forall B \in \mathcal{B}.$$

This definition is similar to the continuous maps in topological spaces.

We're going to study the measurable functions from Lebesgue measure space to measure spaces on $\mathbb{R}$. Till now we know four kinds of measure on $\mathbb{R}$: the empty set, the power set, Borel σ-algebra and Lebesgue σ-algebra.

Definition 8.23. Let $E \subset \mathbb{R}^n$ be a measurable set. We say a function $f : E \to \mathbb{R}$ is **measurable**, if the set

$$\{x : f(x) > t\}$$

is measurable for all $t \in \mathbb{R}$.

> **Remark 8.24** — The open intervals (t, ∞) can generate the entire Borel σ-algebra, so this definition is actually saying f is a measurable map from E to $(\mathbb{R}, \mathcal{B})$. There are sufficient reasons why we define measurable functions on Borel measurable space instead of Lebesgue measurable space, we shall see them later.

Note that it's equivalent to $\{x : f(x) \leq t\}$ or $\{x : f(x) < t\}$ is measurable. Using the idea of "generators of Borel σ-algebra", we can get more equivalent statements:

Proposition 8.25

The followings are equivalent:

1. f is a measurable function;
2. $\forall t \in \mathbb{R}$, $\{x \mid f(x) > t\}$ is measurable;
3. $\forall t \in \mathbb{R}$, $\{x \mid f(x) \leq t\}$ is measurable;
4. $\forall t \in \mathbb{R}$, $\{x \mid f(x) < t\}$ is measurable;
5. $\{x \mid f(x) = +\infty\}, \{x \mid f(x) = -\infty\}$ are measurable, and the pre-image of open(or closed) sets are measurable;
6. $\{x \mid f(x) = +\infty\}, \{x \mid f(x) = -\infty\}$ are measurable, and $\{x \mid f(x) \in [a, b)\}$ is measurable, $\forall a < b \in \mathbb{R}$.

Proposition 8.26

Criteria for measurable functions:

- Continuous functions are measurable.
- f is finite-valued, measurable and Φ continuous, then $\Phi \circ f$ is measurable. ($f \circ \Phi$ is not necessarily measurable)
- If f_n's are measurable, then $\sup f_n, \inf f_n, \limsup\limits_{n\to\infty} f_n, \liminf\limits_{n\to\infty} f_n$ are all measurable. In particular if f_n converges, then $\lim\limits_{n\to\infty} f_n$ is measurable.
- If f, g are measurable, then f^k is measurable, $f+g$, fg are measurable when they are well-defined (i.e. $\infty \pm \infty, \infty/\infty, 0/0$ don't occur).

Proof. Continuous function: the pre-image of open sets are open, hence measurable.

For open sets O, $(\Phi \circ f)^{-1}(O) = f^{-1} \circ \Phi^{-1}(O)$ is measurable. So $\Phi \circ f$ is measurable.

$$\{\sup_n f_n(x) > t\} = \bigcup_n \{f_n(x) > t\}$$

Thus $\sup_n f_n$ measurable, similarly $\inf_n f_n$ measurable, and $\limsup\limits_{n\to\infty} f_n = \inf_k(\sup_{n>k} f_n)$, $\liminf\limits_{n\to\infty} = \sup_k(\inf_{n>k} f_n)$ are measurable.

For f^k, note that

$$\{f^k > t\} = \begin{cases} \{f > t^{\frac{1}{k}}\}, & 2 \nmid k \\ \{f > t^{\frac{1}{k}}\} \cup \{f < -t^{\frac{1}{k}}\}, & 2 \mid k, t \geq 0 \end{cases}$$

For $f+g$,

$$\{f+g > t\} = \bigcup_{r\in\mathbb{Q}} \{f > r\} \cup \{g > t-r\},$$

and for fg,

$$fg = \frac{(f+g)^2 - (f-g)^2}{4}.$$

□

Since we assume $0 \cdot \infty = 0$, so when two functions only differ on a null set, we can somehow "ignore" this difference.

Remark 8.27 — Basically, in measure theory, we only care about things up to a difference of a null set, as most tools in measure theory will not tell these apart (for example, the Lebesgue integral).

Definition 8.28. We say two measurable functions f, g are equal **almost everywhere**(or equivalent), if f, g only differ on a null set, written as $f(x) = g(x)$ *a.e.*.

Observe that if $f(x) = g(x)$ *a.e.* and f measurable, then g is also measurable.

We hope some simple functions are dense in measurable functions. (Just like step functions or polynomials are dense in continuous functions).

Let χ_E be the indicator function of the set E. If a_n's are constants, R_n's are cuboid, then we say $\sum_{n=1}^{N} a_n \chi_{R_n}$ is a **step function**. If we replace R_n with measurable sets, we say this is a **simple function**. Obviously simple functions are measurable.

Theorem 8.29

For any non-negative function $f(x)$, there exists a series of non-negative, increasing simple functions $\varphi_k(x)$ converges to $f(x)$. (Written as $\varphi_k(x) \nearrow f(x)$)

Proof. Since simple functions are bounded, so we need to approximate $f(x)$ with bounded functions first:

Define

$$F_k(x) = \begin{cases} f(x), & |x| < k, f(x) \le k; \\ k, & |x| < k, f(x) > k; \\ 0, & |x| \ge k. \end{cases}$$

Clearly $F_k(x)$ is an increasing function series converging to f.

Now we approach $F_k(x)$ using simple functions:

Let

$$E_{l,j} = \left\{ x \mid \frac{l}{2^j} < F_k(x) \le \frac{l+1}{2^j}, |x| < k \right\}, \quad \forall 0 \le l < 2^j k.$$

and define

$$F_{k,j} = \sum_{l=0}^{2^j k - 1} \frac{l}{2^j} \chi_{E_{l,j}}.$$

Each $F_{k,k}$ is a simple function, they are increasing and converge to f, as desired. □

Theorem 8.30

Let f be a measurable function, there exists simple functions $\varphi_k(x)$ such that

$$|\varphi_k(x)| \leq |\varphi_{k+1}|, \quad \lim_{k\to\infty} \varphi_k(x) = f(x), \forall x.$$

Proof. Decompose f into $f = f^+ - f^-$, where $f^+(x) = \max\{0, f(x)\}, f^-(x) = -\min\{0, f(x)\}$.

From previous theorem, there exists $\varphi_k^+(x) \nearrow f^+(x)$, $\varphi_k^-(x) \nearrow f^-(x)$.

Hence

$$\varphi_k(x) := \varphi_k^+(x) - \varphi_k^-(x)$$

satisfies the conditions(note that φ_k^+ and φ_k^- cannot both be positive at the same point). □

Theorem 8.31

Let f be a measurable function, there exists step functions $\psi_k(x)$ such that

$$\lim_{k\to\infty} \psi_k(x) = f(x) \ a.e.$$

Proof. From previous theorems we know f can be approached by simple functions, say

$$\varphi_k(x) = \sum_{l=1}^{N_k} a_{k,l}\chi_{E_{k,l}}.$$

For a fixed k, we may assume that $E_{k,l}$'s are pairwise disjoint and $a_{k,l}$'s are disctinct. (This is called the *standard form* of simple functions) Note that from the construction in previous theorems, we can require $E_{k,l}$ are of finite measure.

For each $E_{k,l}$, we can find finitely many disjoint cuboids, denote their union by $I_{k,l,\varepsilon}$, such that

$$m\left(E_{k,l}\Delta I_{k,l,\varepsilon}\right) < 2\varepsilon$$

(where $A\Delta B = A\backslash B \cup B\backslash A$)

Now define step functions

$$\psi_k(x) = \sum_{l=1}^{N_k} a_{k,l}\chi_{I_{k,l,\varepsilon}},$$

we have

$$\{\psi_k(x) \neq \varphi_k(x)\} \subset \bigcup_{l=1}^{N_k} E_{k,l}\Delta I_{k,l,\varepsilon} =: A_k.$$

So once we set $N_k\varepsilon < \frac{1}{2^k}$, we have $m(A_k) < \frac{1}{2^k}$,

$$\sum_{k=1}^{\infty} m(A_k) < \infty,$$

by Borel-Cantelli Theorem (Theorem 8.13), we have

$$m\left(\limsup_{k\to\infty} A_k\right) = 0.$$

This shows that

$$\{\psi_k(x) \nrightarrow f(x)\} \subset \{\text{There exists infintely many } k \text{ such that } \psi_k(x) \neq \varphi_k(x)\}$$

is a null set, which means $\psi_k(x) \to f(x), a.e.$. □

In contrast to continuous functions (which can be approached by step functions at every point), measurable functions can only be approached *almost everywhere*. This is because you can change the value of f on any null sets, while f remains measurable.

Next we'll look at the convergency function of measurable functions. We already know the limit is measurable as well, but we can dig much deeper:

Theorem 8.32 (Egorov)

If $m(E) < \infty$, let f_k be measurable functions on E, converges to f almost everywhere.

For any $\varepsilon > 0$, exists a closed set $F_\varepsilon \subset E$, s.t. $m(E \backslash F_\varepsilon) \leq \varepsilon$ and f_k uniformly converges to f on F_ε.

Proof. WLOG f_k conveges to f after removing a null set.

$$E_k^n := \{x \in E : |f_j(x) - f(x)| < \frac{1}{n}, \forall j > k\}.$$

For a fixed n, E_k^n is increasing and its limit is E. Since $m(E)$ finite, exists k_n s.t. $m(E \backslash E_{k_n}^n) < 2^{-n}$.

$$|f_j(x) - f(x)| < \frac{1}{n}, \forall j > k, x \in E_{k_n}^n.$$

Let N be sufficiently large s.t. $\sum_{n \geq N} 2^{-n} < \varepsilon$, define

$$A_\varepsilon = \bigcap_{n \geq N} E_{k_n}^n.$$

TODO □

Theorem 8.33 (Lusin)

Let f be a finite measurable function on E, $m(E) < \infty$, then for any $\varepsilon > 0$, there exists a closed set F_ε such that

$$F_\varepsilon \subset E, \quad m(E \backslash F_\varepsilon) \leq \varepsilon$$

and f restricts to a continuous function on F_ε.

Proof. TODO

Sketch of the proof: First we can find step functions $f_k \to f, a.e.$. Then we can find sets E_k, f_k continuous on $E \backslash E_k$, and $m(E_k) < 2^{-k}$.

By Egorov's theorem, we can find $A_{\frac{\varepsilon}{3}}$ s.t. f_k uniformly converges on $A_{\frac{\varepsilon}{3}}$, □

The above theorems induce the so-called *Littlewood principles*:

- Measurable sets are almost finite unions of cuboids;
- Measurable functions are almost continuous;
- Convergent measurable functions are almost uniformly convergent.

8.7 Other convergence patterns(TODO)

Now we've seen "converge almost everywhere"($f_k \to f, a.e.$). This is equivalent to

$$\forall \varepsilon > 0, m\left(\limsup_{n\to\infty}\{|f_n(x) - f(x)| \geq \varepsilon\}\right) = 0.$$

Since the points that don't converge are in the supreme limit set of the above sets.

Egorov theorem induces another convergence: "converge almost uniformly", written as $f_k \to f, a.u.$. It is equivalent to

$$\forall \varepsilon > 0, \lim_{n\to\infty} m\left(\bigcup_{k=n}^{\infty}\{|f_k - f| \geq \varepsilon\}\right) = 0.$$

There is still another pattern of convergence: "converge in measure". It's defined as:

$$\forall \varepsilon > 0, \lim_{k\to\infty} m\left(\{|f_k - f| \geq \varepsilon\}\right) = 0.$$

This is written as $f_k \to f$ *in measure*, or $f_k \xrightarrow{m} f$.

The relations between these patterns of convergence:

- $f_k \to f, a.u. \implies f_k \to f, a.e.$ and $f_k \xrightarrow{m} f$;
- If $m(E) < \infty$, by Egorov's theorem, $f_k \to f, a.e. \implies f_k \to f, a.u. \implies f_k \xrightarrow{m} f$.
- If $f_k \xrightarrow{m} f$, we can find a subsequence $f_{k'}$ s.t. $f_{k'} \to f, a.u.$ on E.

Remark 8.34 — For the proof of the last statement, see homework 12.12.

Example 8.35

TODO

9 Lebesgue Integration(TODO)

9.1 Integration of simple functions

Recall we've defined simple functions to be:

$$f = \sum_{k=1}^{N} a_k \chi_{E_k},$$

where a_k are constants, E_k are measurable sets. Moreover we can require that a_k distinct, E_k disjoint.

Define the integration

$$\int \sum_{k=1}^{N} a_k \chi_{E_k} = \sum_{k=1}^{N} a_k m(E_k).$$

We can check this definition is well-defined, i.e. independent on the choices of a_k and E_k.

properties

9.2 Integration of bounded functions on finite measurable sets

Let E be our finite measurable set.

Note that we can approximate f using simple functions φ_k. Define the integration

$$\int f := \lim_{k\to\infty} \int \varphi_k.$$

To check it's well-defined, let $I_k = \int \varphi_k$. $\forall \varepsilon > 0$, $\exists A_\varepsilon \subset E$ such that φ_k uniformly converges on A_ε, and $m(E \backslash A_\varepsilon) < \varepsilon$. Then exists N s.t.

$$|\varphi_n(x) - \varphi_m(x)| < \varepsilon, \quad \forall n, m > N, x \in A_\varepsilon.$$