Real Analysis (I)

Shibo Liu (October 7, 2024)

CONTENTS

(1) Both volumes of *Analysis* (T. Tao) can be downloaded from the library. Our textbook is vol.2.
(2) Please read Appendix A of vol.1.

1. Metric spaces and continuous maps

1.1. Metric spaces. Recall that for a sequence $\{a_n\} \subset \mathbb{R}$,

$$a_n \to a \qquad \text{or} \qquad \lim_{n\to\infty} a_n = a$$

means

- $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$, $|a_n - a| < \varepsilon$ for all $n \geq N$.

This ε-N definition can be used to rigorously prove all properties (most of them are intuitive). For example, we prove

- $a_n \to a > 0$, then $a_n > 0$ for $n \gg 1$.

Proof. Let $\varepsilon = \frac{a}{2}$, $\exists N \in \mathbb{N}$ such that if $n \geq N$, then $|a_n - a| < \varepsilon$. Thus

$$a_n > a - \varepsilon = \frac{a}{2} > 0.$$

Limit is a fundamental tool of analysis. To define limit, we need a metric. In $a_n \to a$, $|a_n - a|$ is the distance from a_n to a. On $\mathbb{R}$, we may define a metric (or distance function)

$$d(x, y) = |x - y|.$$

To be a distance function, d needs to satisfy natural conditions.

Definition 1.1. Let $X \neq \emptyset$, $d : X \times X \to [0, \infty)$ is a metric (or distance function) if

(1) $d(x, y) \geq 0$, $d(x, y) = 0$ iff $x = y$.
(2) $d(x, y) = d(y, x)$.
(3) $d(x, z) \leq d(x, y) + d(y, z)$.

We call (X, d) a metric space, also denoted by X for simplicity.

ex1

Example 1.2. The discrete space (X, d), where $d : X \times X \to [0, \infty)$,

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

Example 1.3. If $Y \subset X$, let $d_Y = d|_{Y\times Y}$, that is we set

$$d_Y(x, y) = d(x, y) \qquad \text{for } x, y \in Y.$$

Then (Y, d_Y) is a metric space, called a subspace of (X, d).

Example 1.4. On $\mathbb{R}^n$, we can equip the metrics d_2, d_1 as follow: for

$$x = \left(x^1, \ldots, x^n\right) \qquad \text{and} \qquad y = \left(y^1, \ldots, y^n\right),$$

set

$$d_2(x, y) = \sqrt{\sum_{i=1}^{n} (x^i - y^i)^2}, \qquad d_1(x, y) = \sum_{i=1}^{n} \left|x^i - y^i\right|.$$

We can also define a more general metric (for $p \geq 1$)

$$d_p(x, y) = \left(\sum_{i=1}^{n} \left|x^i - y^i\right|^p\right)^{1/p}.$$

If $p \in \{1, 2\}$, d_p reduces to d_1 and d_2. It can be shown that

$$\lim_{p\to\infty} d_p(x, y) = \max_{i\in\overline{n}} \left|x^i - y^i\right|.$$

Thus, we define $d_\infty(x, y) = \max_{i\in\overline{n}} \left|x^i - y^i\right|$.

Remark 1.5. We can equip many metrics on a given set X.

Example 1.6. Let $X = S^2$,

$$d(p, q) = \inf \{L(\gamma) \mid \gamma \subset X \text{ is a curve from } p \text{ to } q\}.$$

Of course we can take $d(p, q) = |p - q|$, the length of $[p, q]$, but $[p, q] \not\subset X$. That is, this is not intrinsic (you need the ambient space $\mathbb{R}^3$), hence not work for abstract surfaces (manifolds).

Example 1.7. Normed vector space $(X, \|\cdot\|)$, let

$$d(x, y) = \|x - y\|.$$

Definition 1.8. Let $\{x_n\}_{n=1}^\infty \subset X$, we say $x_n \to a$ if $d(x_n, a) \to 0$.

Remark 1.9. The labels can start at any m, $\{x_n\}_{n=m+1}^\infty = \{x_{m+k}\}_{k=1}^\infty$.

Remark 1.10. A sequence is a function $x : \mathbb{N} \to X$. It is *different* to the set $\{x_n \mid n \in \mathbb{N}\}$. For example, there are infinitely many terms in the constant sequence $\{x_n\}$ with $x_n = a$, but as a set it is a singleton $\{a\}$.

If d_i are two metrics on X, it may happen

$$d_1(x_n, a) \to 0, \qquad d_2(x_n, a) \not\to 0.$$

That is, $\{x_n\}$ may converge to a with respect to d_1 (we write $x_n \xrightarrow{d_1} a$) but not d_2. ex2

Example 1.11. $X = C[0, 1]$,

$$d_1(f, g) = \max_{[0,1]} |f - g|, \qquad d_2(f, g) = \int_0^1 |f - g|.$$

Then for $\{f_n\} \subset X$ and $f \in X$,

$$f_n \xrightarrow{d_1} f \quad \Rightarrow \quad f_n \xrightarrow{d_2} f.$$

The converse is not true. Example: Define $f_n : [0, 1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1 - nx & x \in [0, \frac{1}{n}), \\ 0 & x \in (\frac{1}{n}, 1]. \end{cases}$$

Then $f_n \xrightarrow{d_2} 0$ but $f_n \overset{d_1}{\not\longrightarrow} 0$.

We prove that in the above example d_1 verifies the triangle inequality. Take f, g and h from X. For $x \in [0, 1]$ we have

$$\begin{aligned} |f(x) - h(x)| &\le |f(x) - g(x)| + |g(x) - h(x)| \\ &\le \max_{[0,1]} |f - g| + \max_{[0,1]} |g - h| \\ &= d_1(f, g) + d_1(g, h). \end{aligned}$$

Since x is arbitrary, this implies

$$d_1(f, h) = \max_{[0,1]} |f - h| \le d_1(f, g) + d_1(g, h).$$

As an exercise, show that d_2 is a metric as well.

Example 1.12. Let $\{x_i\} \subset \mathbb{R}^n$, then $x_i \to a$ w.r.t. d_p (or d_1, d_2, d_∞) iff $x_i^k \to a^k$ for $k \in \overline{n}$. Thus $x_i \xrightarrow{d_2} a$ is *equivalent* to $x_i \to a$ w.r.t. d_1, d_∞ or d_p.

Proof. This follows from

$$|x_i^k - a^k|^p \le d_p^p(x_i, a) = \sum_{j=1}^n |x_i^j - a^j|^p \qquad \text{for all } k \in \overline{n}$$

and

$$|x_i^k - a^k| \le d_\infty(x_i, a) \le d_1(x_i, a), \qquad \text{for all } k \in \overline{n}.$$

Example 1.13. Let $f : [0,1] \to [0,1]$ be defined by

$$f(x) = x \quad \text{for } x \in (0,1), \qquad f(0) = 1, \qquad f(1) = 0.$$

For $X = [0,1]$, set

$$\rho(x, y) = d_2(f(x), f(y)) = |f(x) - f(y)|.$$

For $x_n = 1/n$,

$$x_n \xrightarrow{d_2} 0, \qquad \text{but} \qquad x_n \xrightarrow{\rho} 1.$$

Therefore, with respect to different metrics, sequences may converge to different points.

Let (X, d) be a metric space, $r > 0$. We call

$$B_r(a) = \{x \in X \mid d(x, a) < r\}$$

the ball centered at a with radius r, or simply r-neighborhood of a, r-ball at a. We also write $B_r^{(X,d)}(a)$ or $B_r^d(a)$, $B_r^X(a)$ if necessary. We also call

$$B_r[a] = \{x \in X \mid d(x, a) \le r\}$$

the closed ball centered at a with radius r.

When $X = \mathbb{R}^n$ and $a = 0$, we write B_r for $B_r(0)$. To indicate the dimension we also write B_r^n.

Example 1.14. In $\mathbb{R}^2$, $B_1^{d_2}(0)$, $B_1^{d_1}(0)$, $B_1^{d_\infty}(0)$.

Example 1.15. If $Y \subset X$, $Y \neq \emptyset$, then Y is a subspace of X. Let $a \in Y$, r-ball in Y at a is

$$\begin{aligned} B_r^Y(a) &= \{x \in Y \mid d_Y(x, a) < r\} \\ &= \{x \in Y \mid d(x, a) < r\}. \end{aligned}$$

We have $B_r^Y(a) = B_r(a) \cap Y$.

For $E \subset X$, we say that E is bounded if $E \subset B_r(a)$ for some $a \in X$ and $r > 0$. This is equivalent to

$$\operatorname{diam} E := \sup_{x,y \in E} d(x, y) < \infty.$$

Proposition 1.16. *If $x_n \to a$, then $\{x_n\}$ is bounded. If moreover $x_n \to b$, then $a = b$.*

Definition 1.17. Let (X, d) be a metric space, $E \subset X$.

(1) a is an interior point of E if $B_r(a) \subset E$ for some $r > 0$. We denote by E° (the interier of E) the set of all interior points.

(2) a is an exterior point of E if $a \in (E^c)^\circ$. That is, there is $r > 0$ such that $B_r(a) \cap E = \emptyset$. We denote by E^e (the exterior of E) the set of all exterior points.

(3) a is a boundary point of E if $a \in X \backslash (E^\circ \cup E^e)$. Namely, for $\forall r > 0$

$$E \cap B_r(a) \neq \emptyset, \qquad E^c \cap B_r(a) \neq \emptyset.$$

The set of all bdry pts is denoted by ∂E (the boundary of E).

(4) a is an adherent point of E, if $E \cap B_r(a) \neq \emptyset$ for $\forall r > 0$. The set of such a is denoted $\overline{E}$ (the closure of E).

(5) a is an accumulation point of E, if $(E \backslash \{a\}) \cap B_r(a) \neq \emptyset$ for $\forall r > 0$. That is $a \in \overline{E \backslash \{a\}}$. The set of such a is denoted by E' (the derivative of E).

It is clear that $E' \subset \overline{E}$,

$$\partial E = \partial E^c = X \backslash (E^\circ \cup E^e), \tag{1.1}$$ x2

and $(E \backslash \{a\}) \cap B_r(a)$ is infinite if $a \in E'$.

lt

Example 1.18. Find E° and E^e for $E = [0, 1) \times (0, 1)$.

Proof. It is easy to see that $E^\circ = (0, 1) \times (0, 1)$. We also have

$$E^e = \{x^1 > 1\} \cup \{x^1 < 0\} \cup \{x^2 > 1\} \cup \{x^2 < 0\}.$$

To see "$\supset$", let $a \in$ RHS. We may assume $a \in \{x^1 > 1\}$, that is $a^1 > 1$. If $x \in B_{a^1-1}(a)$, then $x^1 > 1$ hence $x \notin E$, we conclude $B_{a^1-1}(a) \cap E = \emptyset$, so $a \in E^e$.

To see "$\subset$" we argue by contradiction. Suppose $a \notin$ RHS, then by de Morgan's law

$$\begin{aligned} a &\in \left(\{x^1 > 1\} \cup \{x^1 < 0\} \cup \{x^2 > 1\} \cup \{x^2 < 0\}\right)^c \\ &= \{x^1 > 1\}^c \cap \{x^1 < 0\}^c \cap \{x^2 > 1\}^c \cap \{x^2 < 0\} \\ &= \{x^1 \le 1\} \cap \{x^1 \ge 0\} \cap \{x^2 \le 1\} \cap \{x^2 \ge 0\} \\ &= ([0, 1] \times \mathbb{R}) \cap (\mathbb{R} \times [0, 1]) = [0, 1] \times [0, 1]. \end{aligned}$$

It is now clear that $a \notin E^e$.

Using the above results and (1.1),

$$\begin{aligned} \partial E &= \mathbb{R}^2 \backslash (E^\circ \cup E^e) \\ &= (\{0, 1\} \times [0, 1]) \cup ([0, 1] \times \{0, 1\}). \end{aligned}$$

p5

Proposition 1.19. *Let $E \subset X$, $a \in X$.*

(1) *$a \in \overline{E}$ iff there is $\{x_n\} \subset E$ such that $x_n \to a$.*

(2) *$a \in E'$ iff there is $\{x_n\} \subset E \backslash a$ such that $x_n \to a$ (exercise).*

(3) *$\overline{E} = E^\circ \sqcup \partial E$.*

(4) *$(E^c)^\circ = (\overline{E})^c$.*

1r

Remark 1.20. Because $E^\circ \subset E \subset \overline{E}$, we also have $\overline{E} = E \cup \partial E$.

Proof. (1) ($\Rightarrow$) For $n \in \mathbb{N}$, $E \cap B_{1/n}(a) \neq \emptyset$. Take x_n from this set we get $\{x_n\} \subset E$ s.t. $x_n \to a$.

($\Leftarrow$) For $r > 0$, since $x_n \to a$, $\exists m \in \mathbb{N}$ and such that $d(x_m, a) < r$, or $x_m \in E \cap B_r(a)$. Thus $E \cap B_r(a) \neq \emptyset$. We conclude $a \in \overline{E}$.

(3) It is clear that $\overline{E} \supset E^\circ \cup \partial E$. To see $\overline{E} \subset E^\circ \cup \partial E$, let $a \in \overline{E}$. If $a \notin \partial E$, then

$$E^c \cap B_r(a) = \emptyset$$

for some $r > 0$ (because $E \cap B_r(a) \neq \emptyset$). Hence $B_r(a) \subset E$, $a \in E^\circ$.

(4) If $a \in (E^c)^\circ$, then $\exists r > 0$ s.t. $B_r(a) \subset E^c$. Thus $B_r(a) \cap E = \emptyset$, so $a \in (\overline{E})^c$. If $a \in (\overline{E})^c$, $\exists r > 0$ s.t. $B_r(a) \cap E = \emptyset$. Thus $B_r(a) \subset E^c$, so $a \in (E^c)^\circ$.

Using this proposition, for the E given in Example 1.18, we have

$$E' = \overline{E} = [0,1] \times [0,1] \,.$$

Because $E' \subset \overline{E}$, it suffices to show

$$[0,1] \times [0,1] \subset E' \qquad \text{and} \qquad \overline{E} \subset [0,1] \times [0,1] \,.$$

We prove the first. For $a \in [0,1] \times [0,1]$, set

$$x_n = \left(\frac{n a^1}{n+1}, \frac{n^2 a^2 + 1}{n^2 + 2n} \right).$$

Then $\{x_n\} \subset E \setminus \{a\}$, $x_n \to a$. Hence $a \in E'$.

Definition 1.21. Let (X, d) be a metric space, $E \subset X$. We say that E is closed if $\partial E \subset E$, E is open if $\partial E \cap E = \emptyset$. r1

Remark 1.22. E can be neither open nor closed (for example, the E given in Example 1.18); or both open and closed. From the definitions and $\partial E = \partial E^c$ it is clear that

- E is open iff E^c is closed.

Example 1.23. X and $\emptyset$ are open and closed, $B_r(a)$ is open, $\{a\}$ is closed.

Proposition 1.24. *Properties of open sets.*

(1) *E is open iff $E = E^\circ$.*
(2) *$E_1 \cap E_2$ is open if E_1 and E_2 are.*
(3) *Let $\{E_\lambda\}_{\lambda \in \Lambda}$ be a collection of open sets, then $\bigcup_\lambda E_\lambda$ is open.*

Proof. (1) ($\Rightarrow$) Let $a \in E$, then $a \notin \partial E$, $\exists r > 0$ such that $E^c \cap B_r(a) = \emptyset$ (because $E \cap B_r(a) \neq \emptyset$), that is $B_r(a) \subset E$, $a \in E^\circ$.

($\Leftarrow$) Let $a \in \partial E$, then $a \notin E^\circ$ (why?). Thus $\partial E \cap E = \partial E \cap E^\circ = \emptyset$.

(2) Let $a \in E_1 \cap E_2$, then $a \in E_i = E_i^\circ$. There are $r_i > 0$ s.t. $B_{r_i}(a) \subset E_i$. Let $r = \min\{r_1, r_2\}$. Then

$$B_r(a) \subset B_{r_1}(a) \cap B_{r_2}(a) \subset E_1 \cap E_2,$$

this means $a \in (E_1 \cap E_2)^\circ$. Consequently $E_1 \cap E_2 = (E_1 \cap E_2)^\circ$ and $E_1 \cap E_2$ is open[(1)].

(3) For $a \in \bigcup_\lambda E_\lambda$, we have $a \in E_{\lambda'}$ for some λ'. Since $E_{\lambda'}$ is open, $B_r(a) \subset E_{\lambda'}$ for some $r > 0$. Hence

$$B_r(a) \subset \bigcup_\lambda E_\lambda$$

and we deduce $a \in \left(\bigcup_\lambda E_\lambda\right)^\circ$.

Using the relation between open and closed sets (Remark 1.22), as corollary we have

Proposition 1.25. *Properties of open sets.*

(1) *F is closed iff $F = \overline{F}$, iff $\{x_n\} \subset F$ and $x_n \to a$ imply $a \in F$.*
(2) *$F_1 \cup F_2$ is closed if F_1 and F_2 are.*
(3) *Let $\{F_\lambda\}_{\lambda \in \Lambda}$ be a collection of closed sets, then $\bigcap_{\lambda \in \Lambda} F_\lambda$ is closed.*

[(1)]It is *not convenient* to prove via definition because it is hard to describe $\partial(E_1 \cap E_2)$.

Proof. (1) By Remark 1.20 the first part is clear. Alternatively, using Proposition 1.19 (4), we have: F closed $\Leftrightarrow$ F^c open $\Leftrightarrow$ $F^c = (F^c)^\circ$ $\Leftrightarrow$

$$\overline{F} = [(F^c)^\circ]^c = [F^c]^c = F.$$

Now we prove the second part.

($\Rightarrow$) Assume $F = \overline{F}$. By Proposition 1.19, $\{x_n\} \subset F$ and $x_n \to a$ implies $a \in \overline{F}$, thus $a \in F$.

($\Leftarrow$) Let $a \in \overline{F}$. By Proposition 1.19, there is $\{x_n\} \subset F$ s.t. $x_n \to a$. By assumption $a \in F$. Hence $\overline{F} = F$.

(3) Since all F_λ^c are open, $\bigcup_{\lambda\in\Lambda} F_\lambda^c$ is open. Being complement of open set,

$$\bigcap_{\lambda\in\Lambda} F_\lambda = \left(\bigcup_{\lambda\in\Lambda} F_\lambda^c\right)^c \qquad \text{is closed.}$$

Or, assume $\{x_n\} \subset \bigcap_{\lambda\in\Lambda} F_\lambda$, $x_n \to a$. Then for all λ we have $\{x_n\} \subset F_\lambda$. We conclude $a \in F_\lambda$ because F_λ is closed. Thus $a \in \bigcap_{\lambda\in\Lambda} F_\lambda$ and by (1), $\bigcap_{\lambda\in\Lambda} F_\lambda$ is closed.

p6

Proposition 1.26. *Let (X, d) be a metric space, $E \subset X$. Then*

$$(1) \quad E^\circ = \bigcup_{U\subset E,\, U \text{ open}} U, \qquad\qquad (2) \quad \overline{E} = \bigcap_{C\supset E,\, C \text{ closed}} C.$$

Remark 1.27. From this we see that E° is the largest *open* set contained in E, $\overline{E}$ is the smallest *closed* set containning E.

Proof. (1) If $a \in E^\circ$, then $B_r(a) \subset E$ for some $r > 0$. Since $B_r(a)$ is open we conclude

$$E^\circ \subset \bigcup_{U\subset E,\, U \text{ open}} U.$$

Now let $a \in \bigcup_{U\subset E,\, U \text{ open}} U$. Then $a \in U$ for some open $U \subset E$, there is $r > 0$ such that $B_r(a) \subset U \subset E$. Hence $a \in E^\circ$.

(2) Using Proposition 1.19 (4) and de Morgan's law

$$\left(\bigcap_{C\supset E,\, C \text{ closed}} C\right)^c = \bigcup_{C\supset E,\, C \text{ closed}} C^c = \bigcup_{U\subset E^c,\, U \text{ open}} U$$
$$= (E^c)^\circ = \left(\overline{E}\right)^c.$$

Alternative Proof. Let $a \in \overline{E}$. Given closed $C \supset E$, for all $r > 0$ we have

$$C \cap B_r(a) \supset E \cap B_r(a) \neq \emptyset.$$

Thus $a \in \overline{C} = C$. This yields

$$\overline{E} \subset \bigcap_{C\supset E,\, C \text{ closed}} C.$$

On the other hand, if $a \notin \overline{E}$, $\exists r > 0$ such that $B_r(a) \cap E = \emptyset$. Thus $C := [B_r(a)]^c$ is a closed set containning E. Noting that $a \notin C$, we see that

$$a \notin \bigcap_{C\supset E,\, C \text{ closed}} C.$$

Hence

$$\overline{E} \supset \bigcap_{C \supset E,\, C \text{ closed}} C. \tag{1.2}$$

e3

Remark 1.28. It seems difficult to prove (1.2) by showing that every point on the right hand side is in $\overline{E}$.

yi

Example 1.29. $E \subset X$ is open iff E is union of some balls.

Proof. Since $E = E^\circ$, for $a \in E$, $\exists r_a > 0$ such that

$$\{a\} \subset B_{r_a}(a) \subset E.$$

We conclude

$$E = \bigcup_{a \in E} \{a\} \subset \bigcup_{a \in E} B_{r_a}(a) \subset E.$$

Thus $E = \bigcup_{a \in E} B_{r_a}(a)$ is union of balls $B_{r_a}(a)$.

Let Y be a subspace of X and $E \subset Y$. Then there are two meanings for the openness of E: open in (the subspace) Y or open in (the ambient space) X. In the former case we may simply say that E is Y-open.

p1

Proposition 1.30. *Let Y be a subspace of X. $E \subset Y$ is Y-open iff $E = Y \cap U$ for some X-open set U.*

Proof. ($\Rightarrow$) If E is Y-open, it must be union of some Y-balls $B_\lambda^Y = B_\lambda \cap Y$, where B_λ are some X-balls, see Example 1.29. We deduce

$$E = \bigcup_\lambda B_\lambda^Y = \bigcup_\lambda (Y \cap B_\lambda) = Y \cap \left(\bigcup_\lambda B_\lambda \right) = Y \cap U,$$

where $U = \bigcup_\lambda B_\lambda$ is X-open.

($\Leftarrow$) If $E = Y \cap U$ for X-open U, since U are union of X-balls B_λ we see that

$$E = Y \cap \left(\bigcup_\lambda B_\lambda \right) = \bigcup_\lambda (B_\lambda \cap Y) = \bigcup_\lambda B_\lambda^Y$$

is union of some Y-open balls B_λ^Y, thus Y-open.

Remark 1.31. Similarly, we may define Y-closed sets, and show that E is Y-closed iff $E = Y \cap C$ for some X-closed set C (exercise).

1.2. Cauchy sequences, completeness, continuous maps. A sequence $\{x_n\}$ in X is simply a map $x : \mathbb{N} \to X$, we then denote $x_n = x(n)$. If $n : \mathbb{N} \to \mathbb{N}$ is strictly increasing, the composition

$$y = x \circ n : \mathbb{N} \to X, \qquad i \mapsto x(n(i))$$

is a sequence $\{y_i\}$ in X (here $y_i = x(n(i)) = x_{n_i}$), called a subsequence of $\{x_n\}$ and denoted by $\{x_{n_i}\}_{i=1}^\infty$.

It is then easy to see that if $x_n \to a$, then $x_{n_i} \to a$ (because $\{d(x_{n_i}, a)\}$ is a subsequence of $\{d(x_n, a)\}$).

Definition 1.32. Let (X, d) be a metric space. A sequence $\{x_n\} \subset X$ is a Cauchy sequence, if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ s.t. $d(x_m, x_n) < \varepsilon$ for all $m, n \geq N$. We say that (X, d) is complete, if every Cauchy sequence in X converges.

Proposition 1.33. *Let $\{x_n\}$ be a sequence in X.*

(1) *If $x_n \to a$, then $\{x_n\}$ is Cauchy.*
(2) *If $\{x_n\}$ is Cauchy, $x_{n_i} \to a$, then $x_n \to a$.*

Example 1.34. $(\mathbb{R}^n, d_2)$ is complete, $(\mathbb{Q}^n, d_2)$ is not. The space (X, d) in Example 1.2 is also complete.

Example 1.35. In Example 1.11, (X, d_1) is complete but (X, d_2) is not.

Example 1.36. For $a \in \mathbb{R}^N$, $r > 0$, let X be the set of all continuous $x : (-h, h) \to \overline{B}_r(a)$ equipped with the metric

$$d(x, y) = \sup_{t \in (-h,h)} |x(t) - y(t)|, \qquad x, y \in X.$$

Then X is complete.

Proof. It is clear that d is a metric on X (similar to the paragraph after Example 1.11). To see that X is complete, let $\{x_k\}$ be a Cauchy sequence in X, namely $d(x_i, x_j) \to 0$ as $i, j \to \infty$. Given any $t \in (-h, h)$, from

$$\left|x_i(t) - x_j(t)\right| \le d(x_i, x_j),$$

we see that $\{x_k(t)\}$ is a Cauchy sequence in $\mathbb{R}^N$. Therefore, we may define a map

$$x : (-h, h) \to \mathbb{R}^N \qquad \text{via} \qquad x(t) = \lim_{k\to\infty} x_k(t),$$

because for every $t \in (-h, h)$ the limit exists. We claim that:

(1) $x(t) \in \overline{B}_r(a)$ for all $t \in (-h, h)$.
Because $x_k \in X$, $x_k(t) \in \overline{B}_r(a)$. That is

$$|x_k(t) - a| \le r \qquad \text{for all } t \in (-h, h)\,.$$

Let $k \to \infty$ we deduce $|x(t) - a| \le r$, that is $x(t) \in \overline{B}_r(a)$.

(2) x is continuous.
Let $t_0 \in (-h, h)$ and $\{t_i\} \subset (-h, h)$ with $t_i \to t_0$. Given $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that if $\ell \ge k \ge N$ then

$$|x_k(t) - x_\ell(t)| \le d(x_k, x_\ell) < \varepsilon \qquad \text{for all } t \in (-h, h)\,.$$

Letting $\ell \to \infty$ we deduce $|x_k(t) - x(t)| \le \varepsilon$. Thus[2]

$$\sup_{t\in(-h,h)} |x_k(t) - x(t)| \le \varepsilon. \tag{1.3}$$

Therefore

$$\begin{aligned} |x(t_i) - x(t_0)| &\le |x(t_i) - x_k(t_i)| + |x_k(t_i) - x_k(t_0)| + |x_k(t_0) - x(t_0)| \\ &\le 2 \sup_{t\in[-h,h]} |x(t) - x_k(t)| + |x_k(t_i) - x_k(t_0)| \\ &\le 2\varepsilon + |x_k(t_i) - x_k(t_0)|\,. \end{aligned}$$

Since x_k is continuous at t_0, it follows that

$$\varlimsup_{i\to\infty} |x(t_i) - x(t_0)| \le 2\varepsilon.$$

So $x(t_i) \to x(t_0)$ and x is continuous at t_0.

[2] At this point the LHS can not be written as $d(x_k, x)$ because we don't know that $x \in X$.

From these claims, x can be viewed as a continuous map $x : (-h, h) \to \overline{B}_r(a)$. Thus $x \in X$ and the left hand side of (1.3) can be written as $d(x_k, x)$, so that $d(x_k, x) \leq \varepsilon$. Hence $x_k \to x$.

Proposition 1.37. *Let (Y, d_Y) be a subspace of (X, d).* p0

(1) *If (Y, d_Y) is complete, then Y is X-closed.*
(2) *If (X, d) is complete and Y is X-closed, then (Y, d_Y) is complete.*

Proof. (1) Assume $\{x_n\} \subset Y$, $x_n \to a$ in X, we need to prove that $a \in Y$. Since $x_n \to a$ in X, $\{x_n\}$ is a Cauchy sequence in X, therefore it *is also* a Cauchy sequence in Y. Because Y is complete, $x_n \to a'$ in Y for some $a' \in Y$. By the definition of d_Y we get

$$d(x_n, a') = d_Y(x_n, a') \to 0.$$

Hence $x_n \to a'$ in X as well. Thus $a = a'$, $a \in Y$.

(2) Let $\{x_n\}$ be a Cauchy sequence in Y. Then $\{x_n\}$ is also a Cauchy sequence in X, hence $x_n \to a$ for some $a \in X$. Because Y is X-closed and $\{x_n\} \subset Y$, we see that $a \in Y$. From

$$d_Y(x_n, a) = d(x_n, a) \to 0$$

we see that $x_n \to a$ in Y.

Remark 1.38. If non-empty $Y \subset X$ is closed, we call Y a closed subspace of X.

Continuity of maps $f : \mathbb{R}^m \to \mathbb{R}^n$ can be generalized to maps between metric spaces.

Definition 1.39. Let (X, d) and (Y, ρ) be metric spaces, we say that $f : X \to Y$ is continuous at $a \in X$, if $\forall \varepsilon > 0$, $\exists \delta > 0$ such that

$$f(B_\delta^X(a)) \subset B_\varepsilon^Y(f(a)). \tag{1.4}$$ e1

If f is continuous at every $x \in X$, we say that $f : X \to Y$ is continuous.

Remark 1.40. Condition (1.4) means that points near a are mapped to points near $f(a)$, that is

$$d(x, a) < \delta \quad\implies\quad \rho(f(x), f(a)) < \varepsilon.$$

Example 1.41. If $A \subset X$ and $f : X \to Y$ is continuous at $a \in A$, then[3] $f|_A : A \to Y$ is also continuous at a; if f itself is continuous, then $f|_A$ is continuous. As a consequence, for $E \subset X$, the inclusion map $i : E \to X$ defined by $i(x) = x$, is continuous ($i = 1_X|_E$).

Example 1.42. Let $f : X \to Y$ be continuous and $Z \subset Y$. If $f(X) \subset Z$, then we have a continuous map $f^Z : X \to Z$ given by $x \mapsto f(x)$.

Definition 1.43. Let $f : X \to Y$.

(1) If $\forall \varepsilon > 0$, $\exists \delta > 0$ such that for all $x, y \in X$,

$$d(x, y) < \delta \quad\implies\quad \rho(f(x), f(y)) < \varepsilon,$$

we say that f is *uniformly continuous*.

(2) If $\exists \theta > 0$ s.t. for all $x, y \in X$,

$$\rho(f(x), f(y)) \leq \theta d(x, y),$$

we say that f is Lipschitz continuous (θ-Lipschitz).

---

[3] Because $B_\delta^A(a) \subset B_\delta^X(a)$ and $f|_A(B_\delta^A(a)) = f(B_\delta^A(a))$.

Example 1.44. The function $\rho : X \to \mathbb{R}$ in Example 1.52 is Lipschitz continuous.

Proposition 1.45 (Banach Contraction Principle)**.** *Let X be a complete metric space, $f : X \to X$ be a* contraction*, that is, there is $\theta \in (0,1)$, s.t.* p2

$$d(f(x), f(y)) \le \theta d(x,y), \qquad x, y \in X.$$

Then $\exists 1\ x^* \in X$ s.t. $f(x^*) = x^*$ (such x^* is called a fixed point of f).

Proof. Take $x_0 \in X$ and define $x_n = f(x_{n-1})$ for $n \ge 1$, we get a sequence $\{x_n\} \subset X$ with

$$\begin{aligned} d(x_i, x_{i+1}) &= d(f(x_{i-1}), f(x_i)) \\ &\le \theta d(x_{i-1}, x_i) \le \cdots \le \theta^i d(x_0, x_1). \end{aligned}$$

Given $\varepsilon > 0$, since $\theta \in (0,1)$, there is $N \in \mathbb{N}$ such that

$$\frac{\theta^n}{1-\theta} d(x_0, x_1) < \varepsilon \qquad \text{for } n \ge N.$$

If $m > n \ge N$, we have

$$\begin{aligned} d(x_m, x_n) &\le d(x_n, x_{n+1}) + \cdots + d(x_{m-1}, x_m) \\ &\le \sum_{i=n}^{m-1} d(x_i, x_{i+1}) \le d(x_0, x_1) \sum_{i=n}^{m-1} \theta^i \\ &\le \frac{\theta^n}{1-\theta} d(x_0, x_1) < \varepsilon. \end{aligned}$$

So $\{x_n\}$ is Cauchy and $x_n \to x^*$ for some $x^* \in X$. Let $n \to \infty$ in

$$x_n = f(x_{n-1})$$

we get $x^* = f(x^*)$. If f has another fixed point x', we have

$$d(x^*, x') = d(f(x^*), f(x')) \le \theta d(x^*, x').$$

Since $\theta \in (0,1)$ we get $d(x^*, x') = 0$, or $x^* = x'$.

Remark 1.46. Without the completeness of X, we could not get $x_n \to x^*$. Try to construct a counterexample showing that if X is not complete, some contraction $f : X \to X$ could have no fixed point.

Example 1.47 (Picard–Lindelöf). Let $f : [-r, r] \times \overline{B}_r(a) \to \mathbb{R}^n$ be contihuous, $f(t, \cdot)$ be ℓ-Lip. Then for some $h \in (0, r)$, there is a unique $x : (-h, h) \to \mathbb{R}^n$ such that

$$\dot{x} = f(t, x), \qquad x(0) = a. \tag{1.5}$$ ie

Proof. Let

$$M = \sup_{(t,x) \in [-r,r] \times B_r(a)} |f(t,x)|, \qquad h = \min\left\{\frac{r}{M+1}, \frac{1}{\ell+1}\right\},$$

X be the set of all continuous $x : (-h, h) \to \overline{B}_r(a)$ equipped with the metric

$$d(x, y) = \sup_{t \in (-h,h)} |x(t) - y(t)|, \qquad x, y \in X.$$

Then X is complete (Example 1.36). For $x \in X$ we define $Tx : (-h, h) \to \mathbb{R}^n$ via

$$(Tx)(t) = a + \int_0^t f(s, x(s))\, ds. \tag{1.6}$$ c

Because for all $t \in (-h, h)$ we have

$$|(Tx)(t) - a| \le \left| \int_0^t |f(s, x(s))|\, ds \right| \le Mh \le r,$$

That is $(Tx)(t) \in \overline{B}_r(a)$. So $Tx \in X$ and (1.6) defines a map $T : X \to X$.

Given $x, y \in X$, we have

$$\begin{aligned} |(Tx)(t) - (Ty)(t)| &= \left| \int_0^t f(s, x(s))\, ds - \int_0^t f(s, y(s))\, ds \right| \\ &\le \left| \int_0^t |f(s, x(s)) - f(s, y(s))|\, ds \right| \\ &\le \left| \int_0^t \ell\, |x(s) - y(s)|\, ds \right| \le \ell h d(x, y) \end{aligned}$$

for all $t \in (-h, h)$. Consequently

$$d(Tx, Ty) = \sup_{t \in (-h,h)} |(Tx)(t) - (Ty)(t)| \le (\ell h)\, d(x, y).$$

Since $\ell h < 1$, we conclude that T is a contraction and has a unique fixed point $x \in X$, which is the unique solution of the initial value problem (1.5).

Remark 1.48. Central problem in mathematics is *Solving Equations*. Solutions of any equations are fixed points of certain maps[4]. Thus fixed point theory is very useful in proving the existence of solutions.

Proposition 1.45 is the simplest fixed point theorem. Another famous one is the Brouwer fixed point theorem. which says that: If X is a closed ball in $\mathbb{R}^n$, then every continuous map $f : X \to X$ has a fixed point. For an elementary proof, see Liu & Zhang (2017).

p4

Proposition 1.49. *Let X and Y be metric spaces, $f : X \to Y$. Then the following statements are equivalent:*

(1) *f is continuous at $a \in X$.*
(2) *$f(x_n) \to f(a)$ for all $\{x_n\} \subset X$ with $x_n \to a$.*
(3) *For Y-open set V containing $f(a)$, there is X-open set U containing a such that $f(U) \subset V$.*

Proof. (1) $\Rightarrow$ (2). Given $\varepsilon > 0$, there is $\delta > 0$ such that

$$f(B_\delta^X(a)) \subset B_\varepsilon^Y(f(a)).$$

Since $x_n \to a$, $\exists N \in \mathbb{N}$ such that $x_n \in B_\delta^X(a)$ for $n \ge N$. Thus $f(x_n) \in B_\varepsilon^Y(f(a))$. Hence[5] $f(x_n) \to f(a)$.

---

[4] Let $g(x) = x + f(x)$, then solutions of the equation $f(x) = 0$ are fixed points of g.

[5] A crucial point in studying mathematics (and any science) is being able to describe the same thing in different ways. Here "$y_n \to y$" iff "given $\varepsilon > 0$, $d(y_n, y) < \varepsilon$ for $n \gg 1$" iff "given $\varepsilon > 0$, $y_n \in B_\varepsilon(y)$ for $n \gg 1$".

(2) $\Rightarrow$ (3). Otherwise, there is $V \in \mathcal{N}^Y_{f(a)}$, such that $f(U) \not\subset V$ for all $U \in \mathcal{N}^X_a$. In particular,

$$f(B^X_{1/n}(a)) \not\subset V \qquad \text{for all } n \in \mathbb{N}.$$

For each $n \in \mathbb{N}$ we pick $x_n \in B^X_{1/n}(a)$ such that $f(x_n) \notin V$, we get a sequence $\{x_n\} \subset X$ such that $x_n \to a$ but $f(x_n) \not\to f(a)$.

(3) $\Rightarrow$ (1). Given $\varepsilon > 0$, $B^Y_\varepsilon(f(a))$ is Y-open set containing $f(a)$. There is X-open set U contianing a such that $f(U) \subset B^Y_\varepsilon(f(a))$. Take $\delta > 0$ such that $B^X_\delta(a) \subset U$, we conclude

$$f(B^X_\delta(a)) \subset f(U) \subset B^Y_\varepsilon(f(a)).$$

So f is continuous at a.

p3

Proposition 1.50. *$f : X \to Y$ is continuous iff for Y-open set V, $f^{-1}(V)$ is X-open.*

Proof. ($\Rightarrow$). For $a \in f^{-1}(V)$, by Proposition 1.49 there is X-open set U_a containing a, such that $f(U_a) \subset V$. Thus $U_a \subset f^{-1}(V)$,

$$f^{-1}(V) = \bigcup_{a \in f^{-1}(V)} \{a\} \subset \bigcup_{a \in f^{-1}(V)} U_a \subset f^{-1}(V).$$

We see that $f^{-1}(V) = \bigcup_{a \in f^{-1}(V)} U_a$ is open (Compare with the proof of Example 1.29).

($\Leftarrow$). We need to show that given $a \in X$, f is continuous at a. Let V be a Y-open set containing $f(a)$, then $U = f^{-1}(V)$ is an X-open set containing a. By Proposition 1.49, f is continuous at a.

Corollary 1.51. *$f : X \to Y$ is continuous iff for Y-closed set V, $f^{-1}(V)$ is X-closed.*

Proof. Or we can prove via sequences.

($\Rightarrow$). If $\{x_n\} \subset f^{-1}(V)$, $x_n \to a$, then $f(x_n) \in V$, $f(x_n) \to f(a)$. Since V is closed we conclude $f(a) \in V$ or $a \in f^{-1}(V)$. Thus $f^{-1}(V)$ is closed.

Example 1.52. Let $E \subset X$, we define $\rho : X \to \mathbb{R}$ by ex3

$$\rho(x) = \inf_{y \in E} d(x, y). \qquad \text{(the distance from } x \text{ to } E\text{)}$$

Then we have

$$|\rho(x) - \rho(y)| \le d(x, y).$$

In particular, if $x_n \to a$ in X then $\rho(x_n) \to \rho(a)$ in $\mathbb{R}$, thus ρ is continuous. More precisely, ρ is 1-Lipschitz.

Proof. Given $x, y \in X$, take $\{z_n\} \subset E$ such that $d(y, z_n) \to \rho(y)$. Then

$$\rho(x) \le d(x, z_n) \le d(x, y) + d(y, z_n).$$

Letting $n \to \infty$ yields

$$\rho(x) \le d(x, y) + \rho(y), \qquad \rho(x) - \rho(y) \le d(x, y).$$

Similarly we also have $\rho(y) - \rho(x) \le d(x, y)$.

Example 1.53. Let ρ be defined in Example 1.52. For $\varepsilon > 0$ set $E_\varepsilon = \rho^{-1}(-\infty, \varepsilon)$. Then

$$\overline{E} = \bigcap_{n=1}^{\infty} E_{1/n}.$$

Remark 1.54. Note that ρ is continuous, hence E_ε is open. Hence the intersection of infinitely many open sets can be closed.

Proof. Note that $E^\varepsilon = \rho^{-1}(-\infty, \frac{\varepsilon}{2}]$ is closed, $E \subset E^\varepsilon \subset E_\varepsilon$, by Proposition 1.26 we get

$$\overline{E} \subset \bigcap_{n=1}^{\infty} E^{1/n} \subset \bigcap_{n=1}^{\infty} E_{1/n}.$$

If $a \notin \overline{E}$, $B_r(a) \cap E = \emptyset$ for some $r > 0$. If $m^{-1} < r$, then

$$\rho(a) \geq r > \frac{1}{m}.$$

Hence $a \notin E_{1/m}$, we conclude $a \notin \bigcap_{n=1}^{\infty} E_{1/n}$.

Remark 1.55. For $X \neq \emptyset$, $\mathcal{T} \subset 2^X$ is called a topology on X if

(1) $X \in \mathcal{T}, \emptyset \in \mathcal{T}$,
(2) $O_1 \cap O_2 \in \mathcal{T}$ if $O_i \in \mathcal{T}$,
(3) $\bigcup_{\lambda \in \Lambda} O_\lambda \in \mathcal{T}$ if all $O_\lambda \in \mathcal{T}$.

We call $(X, \mathcal{T})$ a topological space, $E \subset X$ is called open if $E \in \mathcal{T}$.

Because of Proposition 1.50, for $f : X \to Y$ between topological spaces, we say that f is continuous if $f^{-1}(V)$ is X-open for all Y-open set V. We don't need a metric!.

Proposition 1.56. *If $f : X \to Y$ is continuous at $a \in X$, $g : Y \to Z$ is continuous at $f(a)$, then $g \circ f : X \to Z$ is continuous at a. Therefore, if f and g are continuous, so is $g \circ f$.*

1.3. Continuity on product, connected, and compact metric spaces. The product space of (Y, d) and (Z, ρ) is $(Y \times Z, h)$, being

$$h((y_1, z_1), (y_2, z_2)) = d(y_1, y_2) + \rho(z_1, z_2) \qquad \text{for } (y_i, z_i) \in Y \times Z. \tag{1.7}$$ ooo

Then it is clear that for $\{(y_n, z_n)\} \subset Y \times Z$,

$$(y_n, z_n) \to (a, b) \qquad \iff \qquad y_n \to a \text{ and } z_n \to b. \tag{1.8}$$ e2

For $f : X \to Y$ and $g : X \to Z$, we define $f \oplus g : X \to Y \times Z$,

$$(f \oplus g)(x) = (f(x), g(x)),$$

sometimes denoted by (f, g).

Proposition 1.57. *$f \oplus g : X \to Y \times Z$ is continuous at $a \in X$ iff f and g are.*

Proof. Using Proposition 1.49 and (1.8).

Proof. Using definition involving balls in X, Y, Z and $Y \times Z$.

Example 1.58. The metric $d : X \times X \to \mathbb{R}$ is continuous.

Proof. If $\{(x_n, y_n)\} \subset X \times X$, $(x_n, y_n) \to (a, b)$, we have $x_n \to a$ and $y_n \to b$. Hence

$$|d(x_n, y_n) - d(a, b)| \leq d(x_n, a) + d(b, y_n) \to 0.$$

Remark 1.59. Similarly, we can consider continuity of maps

$$f : X \to \prod_{i=1}^{n} X_i = X_1 \times \cdots \times X_n,$$

where $\prod_{i=1}^{n}$ is product space of X_i with metric defined similar to (1.7).

Proposition 1.60. *If $f, g : X \to \mathbb{R}^n$ are continuous at $a \in X$, then $f + g$, $f \cdot g$ are also continuous at a. If $n = 1$ and $g(a) \neq 0$, f/g is also continuous at a.*

A metric space (X, d) is disconnected, if $X = V \cup W$ for some disjoint non-empty open sets V and W. If X is not disconnected, then it is connected. A subset $Y \subset X$ is connected, if as a subspace of X it is connected.

Example 1.61. As a subspace of $\mathbb{R}$, $Y = [1, 2] \cup [3, 4]$ is disconnected. How about $(1, 2) \cup (2, 4)$?

Proposition 1.62. *If $X \subset \mathbb{R}$ is connected, then X is an interval.*

Proof. Let $a = \inf X$, $b = \sup X$. We claim that $X = \langle a, b\rangle$[6].

Obviously, $X \subset \langle a, b\rangle$. If $X \neq \langle a, b\rangle$, then $\exists c \in (a, b) \setminus X$. We get two disjoint *non-empty* X-open subsets

$$V = (-\infty, c) \cap X, \qquad W = (c, \infty) \cap X,$$

such that $X = V \cup W$, contradicting the connectedness of X. Hence $X = \langle a, b\rangle$.

Proposition 1.63. *X is disconnected iff $f(X) = \{-1, 1\}$ for* some *continuous function $f : X \to \mathbb{R}$.*

Proof. ($\Rightarrow$) Assume $X = V_+ \cup W_-$ for disjoint non-empty open sets $V_\pm$. Then $f : X \to \mathbb{R}$ given by

$$f(x) = \pm 1 \quad \text{for } x \in V_\pm$$

is continuous and $f(X) = \{-1, 1\}$.

($\Leftarrow$) If there is such a function, then $X = V_+ \cup V_-$ is union of disjoint non-empty *open* sets $V_\pm = f^{-1}(\pm 1)$.

Corollary 1.64. *If $X \subset \mathbb{R}$ is an interval, then X is connected.*

Proof. Otherwise, there is a continuous function $f : X \to \mathbb{R}$ such that $f(X) = \{-1, 1\}$. Since X is an interval, by intermediate value theorem, $f(\xi) = 0$ for some $\xi \in X$, a contradiction.

Proposition 1.65. *If X is connected and $f : X \to Y$ is continuous, then $f(X)$ is connected.*

Proof. If $f(X)$ is disconnected, there are disjoint non-empty $f(X)$-open sets V_i such that

$$f(X) = V_1 \cup V_2.$$

Then there are disjoint non-empty Y-open sets U_i such that $V_i = U_i \cap f(X)$. Since f is continuous, $\Omega_i = f^{-1}(U_i)$ are non-empty X-open sets, such that

$$X = \Omega_1 \cup \Omega_2.$$

We conclude that X is disconnected.

Proof. If $f(X)$ is disconnected, there is continuous function $g : f(X) \to \mathbb{R}$ such that $g(f(X)) = \{-1, 1\}$. Then $h = g \circ f : X \to \mathbb{R}$ is continuous anf $h(X) = \{-1, 1\}$, X is then disconnected.

[6]For example if $a \in X$, $b \notin X$, then $\langle a, b\rangle = [a, b)$.

Corollary 1.66. *If X is connected and $f : X \to \mathbb{R}$ is continuous, then $f(X)$ is an interval. In particular, let $\alpha = \inf_X f$, $\beta = \sup_X f$, if $c \in (\alpha, \beta)$, then there is $\xi \in X$ such that $f(\xi) = c$.*

Definition 1.67. A metric space (X, d) is compact if every $\{x_n\} \subset X$ has convergent subsequence. A subset Y is compact if (Y, d_Y) is compact[7].

Remark 1.68. A sequence $\{x_n\}$ in X converges means that for some $a \in X$, we have $d(x_n, a) \to 0$. The limit a must be in X. For example, as a subspace of $X = \mathbb{R}^n$, B_1 is not compact, because for

$$x_n = \left(1 - \frac{1}{n}, 0, \ldots, 0\right),$$

$\{x_n\}$ has *no* convergent subsequence in B_1, although it converges in $X = \mathbb{R}^n$.

Proposition 1.69. *If X is compact, then X is complete and bounded.*

Proof. Let $\{x_n\} \subset X$ be Cauchy. Then it has a convergent subsequence, thus itself is convergent. Hence X is complete.

If X is unbounded, we construct a sequence $\{x_n\} \subset X$ as follow. Take $x_1 \in X$. Assume that we have chosen $\{x_i\}_{i=1}^n$. Since X is unbounded, for

$$r = 1 + \max_{i \in \overline{n}} d(x_i, x_1),$$

there is $x_{n+1} \in B_r^{c}(x_1)$. Because $d(x_i, x_j) \geq 1$, $\{x_n\}$ has no convergent subsequence.

Corollary 1.70. *If Y is a compact set of X, then Y is closed and bounded.*

Proof. By the proposition, Y is complete subspace of X, thus is closed (Proposition 1.37 (1)). Y is also a bounded subset of Y, thus

$$Y \subset B_R^Y(a) = B_R(a) \cap Y$$

for some $a \in Y$ and $R > 0$. We conclude $Y \subset B_R(a)$.

Remark 1.71. Boundedness of Y also follows from $\operatorname{diam}_X Y = \operatorname{diam}_Y Y < \infty$.

Example 1.72. There are complete and bounded spaces which are not compact. On

$$\ell^2 = \left\{ x = (x_1, x_2, \ldots) \,\middle|\, \sum_{i=1}^{\infty} x_i^2 < \infty \right\}$$

set

$$d(x, y) = \sqrt{\sum_{i=1}^{\infty} (x_i - y_i)^2}.$$

Then the subspace

$$S^\infty = \left\{ x \in \ell^2 \mid d(x, 0) = 1 \right\}$$

is complete and bounded, but not compact because for

$$x^k = \left(\delta_1^k, \delta_2^k, \ldots\right),$$

the sequence $\{x_k\}$ has not convergent subsequence ($d(x_k, x_l) = \sqrt{2}$ if $k \neq l$).

[7] In other words, if $\{y_n\} \subset Y$, there is a subsequence $\{y_{n_i}\}$ such that $y_{n_i} \to y$ for some $y \in Y$.

Proof (ℓ^2 is complete). Let $\{x^k\}$ be a Cauchy sequence. Then for all $i \in \mathbb{N}$,

$$\left|x_i^k - x_i^l\right| \le d(x^k, x^l) \to 0 \qquad \text{as } k, l \to \infty.$$

So $x_i^k \to a_i$. We calim that $x^k \to a$ in ℓ^2. Given $\varepsilon > 0$, there is $K \in \mathbb{N}$ such that $d(x^k, x^K) < \varepsilon$ for $k \ge K$. Take $N \in \mathbb{N}$ such that

$$\sum_{i=N}^{\infty} \left(x_i^K\right)^2 < \varepsilon^2, \qquad \sum_{i=N}^{\infty} a_i^2 < \varepsilon^2. \tag{1.9}$$ N

Then for $k \ge K$,

$$\begin{aligned}
\sum_{i=N}^{\infty} \left(x_i^k\right)^2 &\le \left(\left(\sum_{i=N}^{\infty} \left(x_i^k - x_i^K\right)^2 \right)^{1/2} + \left(\sum_{i=N}^{\infty} (x_i^K)^2 \right)^{1/2} \right)^2 \\
&\le \left(d(x^k, x^K) + \left(\sum_{i=N}^{\infty} (x_i^K)^2 \right)^{1/2} \right)^2 < 4\varepsilon^2.
\end{aligned}$$

Hence

$$\begin{aligned}
d^2(x^k, a) &= \sum_{i=1}^{N} \left(x_i^k - a_i\right)^2 + \sum_{i=N}^{\infty} \left(x_i^k - a_i\right)^2 \\
&\le \sum_{i=1}^{N} \left(x_i^k - a_i\right)^2 + \left(\left(\sum_{i=N}^{\infty} \left(x_i^k\right)^2 \right)^{1/2} + \left(\sum_{i=N}^{\infty} a_i^2 \right)^{1/2} \right)^2 \\
&\le \sum_{i=1}^{N} \left(x_i^k - a_i\right)^2 + 9\varepsilon^2, \qquad (1.10)
\end{aligned}$$ n1

which implies

$$\overline{\lim_{k \to \infty}}\, d(x^k, a) \le 3\varepsilon. \tag{1.11}$$ g

Letting $\varepsilon \to 0$ we get $\lim d(x^k, a) = 0$. Thus $x^k \to a$ in ℓ^2.

Remark 1.73. For every $k \in \mathbb{N}$, $\sum\limits_{i=1}^{\infty} \left(x_i^k\right)^2 < \infty$, thus there is $N \in \mathbb{N}$ such that

$$\sum_{i=N}^{\infty} \left(x_i^k\right)^2 < \varepsilon^2.$$

However, this N depends on k. As a result, we could not get (1.11) by letting $k \to \infty$ in

$$\begin{aligned}
d^2(x^k, a) &= \sum_{i=1}^{N} \left(x_i^k - a_i\right)^2 + \sum_{i=N}^{\infty} \left(x_i^k - a_i\right)^2 \\
&\le \sum_{i=1}^{N} \left(x_i^k - a_i\right)^2 + \left(\left(\sum_{i=N}^{\infty} \left(x_i^k\right)^2 \right)^{1/2} + \left(\sum_{i=N}^{\infty} a_i^2 \right)^{1/2} \right)^2
\end{aligned}$$

$$\leq \sum_{i=1}^{N} \left(x_i^k - a_i \right)^2 + 4\varepsilon^2.$$

The N determined in the (1.9) does not depend on k.

Noting that, *any bounded sequence in* $\mathbb{R}^n$ *has convergent subsequences*, we have

Proposition 1.74. *A subset E of $\mathbb{R}^n$ is compact, iff it is closed and bounded.*

Let $Y \subset X$. A collection of open sets $\{V_\lambda\}_{\lambda \in I}$ satisfying

$$Y \subset \bigcup_{\lambda \in I} V_\lambda$$

is called an open cover of Y (more precisely, X-open cover).

Lemma 1.75 (Lebesgue)**.** *If Y is compact, $\{V_\lambda\}_{\lambda \in I}$ is an open cover of Y, then $\exists \delta > 0$, called the* Lebesgue number *of the open cover, such that for $\forall x \in Y$, $\exists \lambda_x \in I$ such that $B_\delta(x) \subset V_{\lambda_x}$. That is, every δ-balls centering in Y is contained in* some *open set from the cover.*

Proof. Otherwise, $\forall n \in \mathbb{N}$, $\exists x_n \in Y$ such that

$$B_{1/n}(x_n) \not\subset V_\lambda \qquad \text{for all } \lambda \in I. \tag{1.12}$$ e0

Being a sequence in Y, $\{x_n\}$ has a convergent subsequence. Assume $x_{n_i} \to a \in Y$. For some $\lambda' \in I$ we have $a \in V_{\lambda'}$. Since $V_{\lambda'}$ is open, $B_r(a) \subset V_{\lambda'}$ for some $r > 0$.

Since $x_{n_i} \to a$, for $i \gg 1$ we have

$$\frac{1}{n_i} + d(x_{n_i}, a) < r.$$

If $y \in B_{1/n_i}(x_{n_i})$, then

$$\begin{aligned} d(y,a) &\leq d(y, x_{n_i}) + d(x_{n_i}, a) \\ &< \frac{1}{n_i} + d(x_{n_i}, a) < r. \end{aligned}$$

Thus $B_{1/n_i}(x_{n_i}) \subset V_{\lambda'}$, contradicting (1.12).

Theorem 1.76. *If Y is compact, $\{V_\lambda\}_{\lambda \in I}$ is an open cover of Y. Then there is a finite $F \subset I$ such that*

$$Y \subset \bigcup_{\lambda \in F} V_\lambda. \tag{1.13}$$ e4

That is, every open cover of a compact set has a finite subcover.

Proof. Assume that the open cover has no finite subcover. Let $\delta > 0$ be the Lebesgue number of the open cover $\{V_\lambda\}_{\lambda \in I}$. Take $x_1 \in F$.

(1) If $Y \subset B_\delta(x_1)$, then $Y \subset V_{\lambda_{x_1}}$ and $F = \{\lambda_{x_1}\}$ fufills the requirement.

(2) If $Y \not\subset B_\delta(x_1)$, then $\exists x_2 \in Y \setminus B_\delta(x_1)$. If

$$Y \subset B_\delta(x_1) \cup B_\delta(x_2),$$

we are done ($F = \{\lambda_{x_1}, \lambda_{x_2}\}$). Otherwise we can take $x_3 \in Y \setminus \bigcup_{i=1}^{2} B_\delta(x_i)$.

(3) Repeating this procedure, if

$$Y \not\subset \bigcup_{i=1}^{n} B_\delta(x_i), \qquad \text{we take } x_{n+1} \in Y \setminus \bigcup_{i=1}^{n} B_\delta(x_i).$$

This procedure must stop in finite steps[(8)]: for some $\ell \in \mathbb{N}$ we will have

$$Y \subset \bigcup_{i=1}^{\ell} B_\delta(x_i)$$

and (1.13) is true for $F = \{\lambda_{x_i}\}_{i=1}^{\ell}$.

Remark 1.77. The converse is also true. If Y is not compact, some sequence $\{x_n\}$ in Y has no comnvergent subsequence (in Y). In other words, for $x \in Y$, $\exists r_x > 0$ such that $B^Y_{r_x}(x)$ contains only finite many term of $\{x_n\}$ (this is not the same as $B^Y_{r_x}(x) \cap \{x_n\}$ is finite set). Suppose $B^Y_{r_x}(x) = B_{r_x}(x) \cap Y$, then $\{B_{r_x}(x)\}_{x\in Y}$ is an X-open cover of Y without finite subcover.

Proposition 1.78. *If X is compact and $f : X \to Y$ is continuous, then $f(X)$ is compact.*

Proof. Let $\{V_\lambda\}_{\lambda\in I}$ be Y-open cover of $f(X)$, $U_\lambda = f^{-1}(V_\lambda)$. Then $\{U_\lambda\}_{\lambda\in I}$ is X-open cover of X, there is finite $F \subset I$ such that

$$X = \bigcup_{\lambda\in F} U_\lambda \qquad \Longrightarrow \qquad f(X) = \bigcup_{\lambda\in F} f(U_\lambda) \subset \bigcup_{\lambda\in F} V_\lambda.$$

Proof. Let $\{y_n\}$ be a sequence in $f(X)$. Then $y_n = f(x_n)$ for $x_n \in X$. Assume $x_{n_i} \to a$, we deduce $y_{n_i} \to f(a) \in f(X)$.

Remark 1.79. If $f : X \to Y$ is cotinuous and $K \subset X$ is compact, then $f|_K : K \to Y$ is continuous. By Proposition 1.78 we see that $f(K)$ is compact.

Corollary 1.80. *If X is compact and $f : X \to \mathbb{R}$ is continuous, $\alpha = \inf_X f$, $\beta = \sup_X f$. Then $\alpha \in f(X)$, $\beta \in f(X)$.*

Example 1.81. If $A \in 2^{\mathbb{R}^n} \setminus \{\emptyset, \mathbb{R}^n\}$ is open, then A is not closed. Thus, $\mathbb{R}^n$ is connected.

Proof (S. Liu). Take $a \in \mathbb{R}^n \setminus A$. Since A is closed, $\exists x \in A$ such that

$$|x - a| = \inf_{y\in A} |y - a| . \tag{1.14}$$

But A is open, $\exists r \in (0, |x - a|)$ such that $B_r(x) \subset A$. Let

$$x' = x - \frac{r}{2\,|x-a|}\,(x-a)\,,$$

then it can be checked that $x' \in B_r(a)$, hence $x' \in A$; but

$$\begin{aligned} \left|x' - a\right| &= \left|(x-a) - \frac{r}{2\,|x-a|}\,(x-a)\right| \\ &= \left|1 - \frac{r}{2\,|x-a|}\right| |x-a| < |x-a|\,, \end{aligned}$$

violating (1.14).

[(8)]Otherwise, since $d(x_i, x_j) \geq \delta$ we obtain a sequence $\{x_n\} \subset Y$ with no convergent subsequence.

Proposition 1.82. *If X is compact and $f : X \to Y$ is continuous. Then f is uniformly continuous.*

Proof. Let $\varepsilon > 0$ be given. For $a \in X$, $\exists r_a > 0$ such that

$$\rho(f(x), f(a)) < \frac{\varepsilon}{2} \qquad \text{for } x \in B_{r_a}(a).$$

Then $\{B_{r_a}(a)\}_{a\in X}$ is open cover of X. Let $\delta > 0$ be the Lebesgue number.

Let $x, y \in X$ satisfying $d(x, y) < \delta$. There is $a \in X$ such that

$$B_\delta(x) \subset B_{r_a}(a).$$

That is, $x, y \in B_{r_a}(a)$, and we have

$$\begin{aligned}\rho(f(x), f(y)) &\le \rho(f(x), f(a)) + \rho(f(a), f(y))\\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Proof. If f is not uniformly continuous[9], $\exists\varepsilon > 0$, for $\forall\delta > 0$, $\exists x, y \in X$,

$$d(x, y) < \delta \qquad \text{but} \qquad \rho(f(x), f(y)) \ge \varepsilon.$$

Take $\delta = 1/n$, we get sequences $\{x_n\}$ and $\{y_n\}$ in X,

$$d(x_n, y_n) < \frac{1}{n} \qquad \text{but} \qquad \rho(f(x_n).f(y_n)) \ge \varepsilon. \tag{1.15}$$ e5

Since X is compact, we have $x_{n_i} \to a$ for a subsequence $\{x_{n_i}\}$. Then also $y_{n_i} \to a$. But f is continuous at a, we get

$$\rho(f(x_{n_i}).f(y_{n_i})) \le \rho(f(x_{n_i}), f(a)) + \rho(f(a), f(y_{n_i})) \to 0,$$

contradicting (1.15).

Proof. Let $\varepsilon > 0$ be given. For $a \in X$, $\exists\delta_a > 0$ such that

$$f(B_{\delta_a}(a)) \subset B_{\varepsilon/2}(f(a)). \tag{1.16}$$ eB

Then $\left\{B_{\delta_a/2}(a)\right\}_{a\in X}$ is an open cover of X. Since X is compact, there is a finite subcover $\left\{B_{\delta_i/2}(a_i)\right\}_{i=1}^n$, here for simplicity we have denoted δ_{a_i} by δ_i.

Set $\delta = 2^{-1}\min\limits_{j\in\overline{n}}\delta_j$. Let $x, y \in X$ satisfying $d(x, y) < \delta$. Since

$$X = \bigcup_{i=1}^n B_{\delta_i/2}(a_i),$$

we have $x \in B_{\delta_i/2}(a_i)$ for some $i \in \overline{n}$. Because

$$d(y, a_i) \le d(y, x) + d(x, a_i) < \delta + \frac{\delta_i}{2} \le \delta_i,$$

we see that $x, y \in B_{\delta_i}(a)$. Then (1.16) implies

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

[9]The negation of "f is uniformly continuous".

Example 1.83. Assume $f : \mathbb{R}^m \to \mathbb{R}^n$ is continuous,

$$\lim_{|x|\to\infty} f(x) = 0, \tag{1.17}$$ z

then f is uniformly continuous.

Proof. Otherwise, there are $\varepsilon > 0$ and $\{x_k\} \subset \mathbb{R}^m$, $\{y_k\} \subset \mathbb{R}^m$ such that

$$|x_k - y_k| < \frac{1}{k} \qquad \text{but} \qquad |f(x_k) - f(y_k)| \geq \varepsilon. \tag{1.18}$$ eR

Because of (1.17), $\exists R > 0$ such that $|f(x)| < \frac{\varepsilon}{2}$ for $x \in B_R^c$. From (1.18) we deduce

$$|x_k| \leq R + 1, \qquad |y_k| \leq R + 1,$$

Otherwise

$$|f(x_k) - f(y_k)| \leq |f(x_k)| + |f(y_k)| < \varepsilon.$$

Since $\{x_k\}$ and $\{y_k\}$ are bounded, from the first inequality in (1.18), there are $a \in \mathbb{R}^m$ and subsequences $\{x_{k_i}\}$ and $\{y_{k_i}\}$ such that $x_{k_i} \to a$, $y_{k_i} \to a$. Hence

$$|f(x_{k_i}) - f(y_{k_i})| \to |f(a) - f(a)| = 0,$$

contradicting the second inequality in (1.18).

Proof. Given $\varepsilon > 0$, $\exists R > 0$ such that $|f(x)| < \frac{\varepsilon}{2}$ for $x \in B_R^c$. Since $D = \{|x| \leq R + 1\}$ is compact, f is uniformly continuous on D, there is $\delta \in (0, 1)$ such that

$$|x - y| < \delta \text{ and } x, y \in D \qquad \Longrightarrow \qquad |f(x) - f(y)| < \varepsilon.$$

For $x, y \in \mathbb{R}^m$ with $|x - y| < \delta$,

(1) if both x and y are in D, then $|f(x) - f(y)| < \varepsilon$.

(2) if one of x and y is not in D, then since $\delta < 1$, both of them are in B_R^c. Hence

$$|f(x) - f(y)| \leq |f(x)| + |f(y)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

2. Uniform convergence

2.1. Pointwise and uniform convergence.

Definition 2.1. Let X, Y be metric spaces, $E \subset X$, $f : E \to Y$ be a map, $a \in \overline{E}$, $b \in Y$. We say that $f(x)$ converges to b (or b is the limit of $f(x)$) as $x \to a$, write

$$\lim_{x\to a} f(x) = b, \tag{2.1}$$ e7

if for any $\varepsilon > 0$, $\exists \delta > 0$, such that[10]

$$f(E \cap B_\delta^X(a)) \subset B_\varepsilon^Y(b). \tag{2.2}$$ e8

Remark 2.2. We need $a \in \overline{E}$ (otherwise the limit of f at $a \notin \overline{E}$ can be *any* element in Y). If $a \in E$ and (2.1) holds, then $b = f(a)$. Using limit, f is continuous at a iff

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

Proposition 2.3. *(2.1) holds iff $f(x_n) \to b$ for all $\{x_n\} \subset E$ with $x_n \to a$.*

---

[10] Instead of (2.2), some authors require $f(E \cap (B_\delta^X(a) \backslash a)) \subset B_\varepsilon^Y(b)$.

Example 2.4. Let $f : [0,1) \to \mathbb{R}$, $f(0) = 0$, $f(x) = 1 + x^2$ for $x \in (0,1)$. Then f does *not* converge to 1 as $x \to 0$, but

$$\lim_{x\to 0} \left(1 + x^2\right) = 1.$$

Consider a sequence of maps $f_n : X \to Y$, where X is a set, Y is a metric space. Given $x \in X$, $\{f_n(x)\}$ is a sequence in Y. Thus it makes sense to consider the convergence of $\{f_n(x)\}$. If it converges, the limit should depend on x, denoted by $f(x)$. If $\{f_n(x)\}$ converges for all $x \in X$, we get a new map $f : X \to Y$ via

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

This map f is called pointwise limit of the sequence $\{f_n\}$, denoted by $f_n \to f$ on X.

If X is also a metric space and all f_n are continuous, is the limit function f continuous?

ex

Example 2.5. Consider $f_n : [0,1] \to \mathbb{R}$ defined by $f_n(x) = x^n$. It is easy to see that

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

We see that each f_n is continuous but the limit f is discontinuous at $x = 1$.

ex4

Example 2.6. Let $f_n = n\chi^{(0,n^{-1}]} : [0,1] \to \mathbb{R}$, then

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

but

$$\int_0^1 f_n = 1 \not\to 0 = \int_0^1 f. \tag{2.3}$$

e10

For $f, g : X \to Y$, we set

$$d_\infty(f,g) = \sup_{x\in X} \rho(f(x), g(x)). \tag{2.4}$$

ed

Note that for some f and g, one may have $d_\infty(f,g) = +\infty$. When $Y = \mathbb{R}$, we denote

$$|f|_\infty = \sup_{x\in X} |f(x)|.$$

Note that $|f|_\infty < \infty$ iff f is bounded. Using this notation, $d_\infty(f,g)$ reduces to

$$|f - g|_\infty = \sup_{x\in X} |f(x) - g(x)|.$$

Definition 2.7. Let $f_n, f : X \to Y$ We say that[11] f_n converges to f uniformly on X, write $f_n \rightrightarrows f$ on X, if $d_\infty(f_n, f) \to 0$ (*In case* $Y = \mathbb{R}$, *this reduces to* $|f_n - f|_\infty \to 0$).

Remark 2.8. $f_n \to f$ on X means, given $x \in X$ we have $f_n(x) \to f(x)$. That is, for $\forall\varepsilon > 0$, $\exists N$ s.t.

$$\rho(f_n(x), f(a)) < \varepsilon \qquad \text{for all } n \geq N.$$

However, this N depends on both ε and x. For the same ε, deferent x requires different N. While $f_n \rightrightarrows f$ means that N depends only on ε, it works for all $x \in X$. Thus, uniformly convergence is a stronger concept.

---

[11]If $A \subset X$ and $f_n|_A \rightrightarrows f|_A$ on A, that is

$$\sup_{x\in A} \rho(f_n(x), f(x)) \to 0,$$

we say that f_n converges to f uniformly on A.

Remark 2.9. If $A \subset B \subset X$ and $f_n \rightrightarrows f$ on B, then $f_n \rightrightarrows f$ on A.

In Example 2.5,

$$f_n(x) - f(x) = \begin{cases} x^n & x \in [0,1), \\ 0 & x = 1. \end{cases}$$

Thus $f_n \not\rightrightarrows f$ because

$$d_\infty(f_n, f) = \sup_{x \in [0,1]} |f_n(x) - f(x)| = \sup_{x \in [0,1)} x^n = 1 \not\to 0.$$

Example 2.10. Let $f_n : [0,1] \to \mathbb{R}$ be given by $f_n(x) = (1-x)\,x^n$. It is easy to see that 211
$f_n \to \mathbf{0}$ on $[0,1]$. From

$$[(1-x)\,x^n]' = x^{n-1}\,[n - (n+1)\,x] = 0$$

we get $x = \frac{n}{n+1}$. Thus

$$\begin{aligned} d_\infty(f_n, \mathbf{0}) &= \sup_{x \in [0,1]} |f_n(x) - 0| = \sup_{x \in [0,1]} (1-x)\,x^n \\ &= [(1-x)\,x^n]_{x = \frac{n}{n+1}} \\ &= \left(1 - \frac{n}{n+1}\right)\left(\frac{n}{n+1}\right)^n \to 0. \end{aligned}$$

in conclusion, $f_n \rightrightarrows \mathbf{0}$.

A more interesting example is the sequence $\{f_n\}$ given by $f_n(x) = \left(1 + \frac{x}{n}\right)^n$. It turns out that for $f(x) = e^x$ and $\forall a > 0$,

$$f_n \rightrightarrows f \text{ on } [0,a], \qquad \text{but} \qquad f_n \not\rightrightarrows f \text{ on } [0,\infty).$$

The following proposition may be useful to prove the above statements.

Proposition 2.11. *$f_n \rightrightarrows f$ iff for any $\{x_n\} \subset X$, $\rho_n = \rho(f_n(x_n), f(x_n)) \to 0$.* pe

Proof. ($\Leftarrow$) If $f_n \not\rightrightarrows f$, then

$$d_\infty(f_n, f) = \sup_{x \in X} \rho(f_n(x_n), f(x_n)) \not\to 0.$$

There is $\varepsilon > 0$ and $n_k \nearrow \infty$ such that

$$d_\infty(f_{n_k}, f) = \sup_{x \in X} \rho(f_{n_k}(x), f(x)) \geq 2\varepsilon.$$

Hence

$$\rho(f_{n_k}(y_k), f(y_k)) \geq \varepsilon$$

for some $y_k \in X$. Choose $a \in X$ and define

$$x_n = \begin{cases} y_k & \text{if } n = n_k, \\ a & \text{if } n \notin \{n_k\}_{k=1}^\infty. \end{cases}$$

We see that $\{\rho_n\}$ has a subsequence $\{\rho_{n_k}\}$ such that

$$\rho_{n_k} = \rho(f_{n_k}(x_{n_k}), f(x_{n_k})) = \rho(f_{n_k}(y_k), f(y_k)) \geq \varepsilon$$

for all k. Hence $\rho_n \not\to 0$.

After studying the next example, you are invited to solve Examples 2.5 and 2.10 using Proposition 2.11.

Example 2.12. Consider a sequence of functions $f_n(x) = \left(1 + \frac{x}{n}\right)^n$. Let $f(x) = e^x$, then $f_n \to f$ on $\mathbb{R}$.

(1) Given $a > 0$, $f_n \rightrightarrows f$ on $[0, a]$.

(2) $f_n \not\rightrightarrows f$ on $[0, \infty)$.

Remark 2.13. The right hand side of

$$d_\infty(f_n, f) = \sup_x \left| \left(1 + \frac{x}{n}\right)^n - e^x \right|$$

is difficult to handle. So it is not convenient to prove the results using definition.

Proof. (a) Take $\{x_n\} \subset [0, a]$. Then because $|x_n| \le a$ and

$$\ln(1+t) = t - \tfrac{1}{2}t^2 + o(t^2) \qquad \text{as } t \to 0, \tag{2.5}$$ xx

we deduce

$$\begin{aligned} f_n(x_n) - f(x_n) &= e^{n \ln\left(1 + \frac{x_n}{n}\right)} - e^{x_n} \\ &= e^{x_n}\left(e^{n \ln\left(1+\frac{x_n}{n}\right) - x_n} - 1\right) \\ &= e^{x_n}\left(e^{n\left(\frac{x_n}{n} - \frac{1}{2}\left(\frac{x_n}{n}\right)^2 + o\left(\left(\frac{x_n}{n}\right)^2\right)\right) - x_n} - 1\right) \\ &= e^{x_n}\left(e^{-\frac{1}{2}\frac{x_n^2}{n} + o\left(\frac{x_n^2}{n}\right)} - 1\right) \to 0. \end{aligned} \tag{2.6}$$ y7

By Proposition 2.11, $f_n \rightrightarrows f$ on $[0, a]$.

(b) Take $x_n = n$. The result follows from $\{x_n\} \subset [0, \infty)$ and

$$f_n(x_n) - f(x_n) = 2^n - e^n \not\to 0.$$

Remark 2.14. If you don't feel comfortable with the Landau notation $o(t)$, (2.5) should be written as

$$\ln(1+t) = t - \tfrac{1}{2}t^2 + \eta(t), \qquad \text{where } \lim_{t\to 0} \frac{\eta(t)}{t^2} = 0.$$

Therefore

$$\ln\left(1 + \frac{x_n}{n}\right) = \frac{x_n}{n} - \frac{1}{2}\left(\frac{x_n}{n}\right)^2 + \eta\left(\frac{x_n}{n}\right)$$

with

$$n\eta\left(\frac{x_n}{n}\right) = \frac{x_n^2}{n} \frac{\eta\left(\frac{x_n}{n}\right)}{\left(\frac{x_n}{n}\right)^2} \to 0 \qquad \text{as } n \to \infty.$$

Hence

$$\begin{aligned} e^{n \ln\left(1+\frac{x_n}{n}\right) - x_n} &= e^{n\left(\frac{x_n}{n} - \frac{1}{2}\left(\frac{x_n}{n}\right)^2 + \eta\left(\frac{x_n}{n}\right)\right) - x_n} \\ &= e^{-\frac{x_n^2}{2n} + n\eta\left(\frac{x_n}{n}\right)} \to 1, \end{aligned}$$

and

$$f_n(x_n) - f(x_n) = e^{x_n}\left(e^{n \ln\left(1+\frac{x_n}{n}\right) - x_n} - 1\right) \to 0.$$

Given a sequence of maps $f_n : X \to Y$, how can we know whether $\{f_n\}$ converges to some $f : X \to Y$ uniformly[12]?

[12] Without knowing f. All the above criteria need to know f.

pu

Proposition 2.15. *Let Y be complete, then $f_n \rightrightarrows f$ for some $f : X \to Y$, iff it is Cauchy, i.e., $\forall \varepsilon > 0$, $\exists N$, $d_\infty(f_m, f_n) < \varepsilon$ for all $m, n \geq N$.*

Proof. ($\Rightarrow$) is easy and does not depend on the completeness of Y.

($\Leftarrow$) Firstly we need to construct a possible limit function $f : X \to Y$. For $x \in X$,

$$\rho(f_m(x), f_n(x)) \leq d_\infty(f_m, f_n).$$

Hence $\{f_n(x)\}$ is a Cauchy sequence in Y. We define $f : X \to Y$ via

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

Given $\varepsilon > 0$, $\exists N$ such that for $m, n \geq N$ and $x \in X$ we have

$$\rho(f_m(x), f_n(x)) \leq d_\infty(f_m, f_n) < \varepsilon.$$

Let $m \to \infty$, by the continuity of metric function we get

$$\rho(f(x), f_n(x)) \leq \varepsilon \qquad \text{for all } n \geq N \text{ and } x \in X.$$

Thus

$$d_\infty(f_n, f) = \sup_{x\in X} \rho(f(x), f_n(x)) \leq \varepsilon,$$

we get $f_n \rightrightarrows f$.

pc

Proposition 2.16. *Assume that $f_n : X \to Y$ are continuous at $a \in X$, $f_n \rightrightarrows f$, then f is also continuous at a. Hence, if $f_n \in C(X, Y)$ and $f_n \rightrightarrows f$, then $f \in C(X, Y)$.*

Proof. Given $\varepsilon > 0$, since $f_n \rightrightarrows f$, we take n such that

$$d_\infty(f_n, f) < \frac{\varepsilon}{3}.$$

Because f_n is continuous at a, $\exists \delta > 0$ such that for all $x \in B_\delta(a)$ we have

$$\rho(f_n(x), f_n(a)) < \frac{\varepsilon}{3}.$$

Consequently

$$\begin{aligned}\rho(f(x), f(a)) &\leq \rho(f(x), f_n(x)) + \rho(f_n(x), f_n(a)) + \rho(f_n(a), f(a))\\ &\leq 2d_\infty(f_n, f) + \rho(f_n(x), f_n(a)) < \varepsilon \qquad \text{for all } x \in B_\delta(a),\end{aligned}$$

and we deduce that f is continuous at a.

Proof. Let $\{x_k\} \subset X$, $x_k \to a$. Because $f_n \rightrightarrows f$, given $\varepsilon > 0$, there is $n \in \mathbb{N}$ such that $d_\infty(f, f_n) < \varepsilon$. Thus

$$\begin{aligned}\rho(f(x_k), f(a)) &\leq \rho(f(x_k), f_n(x_k)) + \rho(f_n(x_k), f_n(a)) + \rho(f_n(a), f(a))\\ &\leq 2d_\infty(f, f_n) + \rho(f_n(x_k), f_n(a))\\ &< 2\varepsilon + \rho(f_n(x_k), f_n(a)).\end{aligned}$$

Noting that f_n is continuous at a, we get

$$\varlimsup_{k\to\infty} \rho(f(x_k), f(a)) \leq 2\varepsilon.$$

Since ε is arbitrary, the limsup is zero, and we deduce $f(x_k) \to f(a)$.

Remark 2.17. From both proofs, we see that if $f_n \rightrightarrows f$ and there is a subsequence $\{f_{n_k}\}$ such that each f_{n_k} is continuous at a, then f is continuous at a.

cp

Proposition 2.18. *Let $E \subset X$, $f_n : E \to Y$, $f_n \rightrightarrows f$ on E. If Y is complete, $a \in \overline{E}$ and*

$$b_n = \lim_{x \to a} f_n(x).$$

Then the limits below exist and are equal

$$\lim_{x \to a} f(x) = \lim_{n \to \infty} b_n.$$

In other words,

$$\lim_{x \to a} \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \lim_{x \to a} f_n(x).$$

Note that Proposition 2.16 is a direct consequence of Proposition 2.18.

Proof. Given $\varepsilon > 0$, by Proposition 2.15, $\exists N$, for all $m, n \geq N$ we have

$$\rho(f_m(x), f_n(x)) < \varepsilon, \qquad \forall x \in X.$$

Let $x \to a$, we get $\rho(b_m, b_n) \leq \varepsilon$. Thus $\{b_n\}$ is a Cauchy sequence in Y, let

$$b = \lim_{n \to \infty} b_n. \tag{2.7}$$ be

It remains to prove

$$\lim_{x \to a} f(x) = b. \tag{2.8}$$ e9

Take $\{x_k\} \subset X$, $x_k \to a$. For $\varepsilon > 0$, take n such that

$$d_\infty(f_n, f) < \varepsilon, \qquad \rho(b_n, b) < \varepsilon.$$

Then

$$\begin{aligned} \rho(f(x_k), b) &\leq \rho(f(x_k), f_n(x_k)) + \rho(f_n(x_k), b) \\ &\leq \varepsilon + \rho(f_n(x_k), b). \end{aligned}$$

Because $f_n(x_k) \to b_n$ as $k \to \infty$, we get

$$\begin{aligned} \varlimsup_{k \to \infty} \rho(f(x_k), b) &\leq \varlimsup_{k \to \infty} (\varepsilon + \rho(f_n(x_k), b)) \\ &= \varepsilon + \rho(b_n, b) < 2\varepsilon. \end{aligned}$$

Since ε is arbitrary, we deduce $f(x_k) \to b$, and (2.8) is proved.

Remark 2.19. Alternatively, after getting (2.7) as above, we prove (2.8) using ε–δ. Given $\varepsilon > 0$, take n such that

$$d_\infty(f_n, f) < \varepsilon, \qquad \rho(b_n, b) < \varepsilon. \tag{2.9}$$ fe

Because

$$\lim_{x \to a} f_n(x) = b_n,$$

there is $\delta > 0$ such that (the second inclusion is by (2.9))

$$f_n(B^X_\delta(a) \cap E) \subset B^Y_\varepsilon(b_n) \subset B^Y_{2\varepsilon}(b).$$

Now, for $x \in B^X_\delta(a) \cap E$ we deduce (note that $f_n(x) \in B^Y_{2\varepsilon}(b)$)

$$\begin{aligned} \rho(f(x), b) &\leq \rho(f(x), f_n(x)) + \rho(f_n(x), b) \\ &\leq d_\infty(f_n, f) + 2\varepsilon < 3\varepsilon. \end{aligned}$$

This proves (2.8).

Proposition 2.20. *Let $f_n \in C(X, Y)$. If $f_n \rightrightarrows f$ and $x_n \to a$ in X, then $f_n(x_n) \to f(a)$.*

Proof. Given $\varepsilon > 0$, take N such that $d_\infty(f_n, f) \le \varepsilon$ for $n \ge N$ We have

$$\begin{aligned}\rho(f_n(x_n), f(a)) &\le \rho(f_n(x_n), f(x_n)) + \rho(f(x_n), f(a)) \\ &\le d_\infty(f_n, f) + \rho(f(x_n), f(a)) \\ &\le \varepsilon + \rho(f(x_n), f(a)).\end{aligned}$$

By the continuity of f, as $n \to \infty$ we get $\rho(f(x_n), f(a)) \to 0$. Hence

$$\varlimsup_{n\to\infty} \rho(f_n(x_n), f(a)) \le \varlimsup_{n\to\infty} (\varepsilon + \rho(f(x_n), f(a))) = \varepsilon.$$

Proposition 2.21. *Let $f, f_n : X \to Y$. If $f_n \rightrightarrows f$ and each f_n is bounded (meaning $f_n(X)$ is bdd subset of Y), then f is also bounded.*

2.2. Uniform convergence with integration and differentiation. A partition of $[a, b]$ is a finite subset P with $a, b \in P$. We may assume that $P = \{x_i\}_{i=0}^n$, where $a = x_0 < \cdots < x_n = b$. Given $f : [a, b] \to \mathbb{R}$, set $\Delta x_i = x_i - x_{i-1}$

$$m_i = \inf_{[x_{i-1}, x_i]} f, \qquad M_i = \sup_{[x_{i-1}, x_i]} f, \qquad \omega_i = M_i - m_i$$

for $i \in \overline{n}$, we define the Darboux sums

$$s(P) = \sum_{i=1}^n m_i \Delta x_i, \qquad S(P) = \sum_{i=1}^n M_i \Delta x_i$$

and amplitude area

$$\Omega(P) = S(P) - s(P) = \sum_{i=1}^n \omega_i \Delta x_i.$$

Proposition 2.22. *$f : [a, b] \to \mathbb{R}$ is Riemannian integrable (we write $f \in R[a, b]$), iff given $\varepsilon > 0$, $\Omega(P) < \varepsilon$ for some partition P.*

p7

Proposition 2.23. *If $f_n \in R[a, b]$, $f_n \rightrightarrows f$, then $f \in R[a, b]$ and $\int_a^b f_n \to \int_a^b f$, i.e.,*

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b \lim_{n\to\infty} f_n.$$

Proof. Since $|f_n - f|_\infty \to 0$, given $\varepsilon > 0$, $\exists n$ such that $|f_n - f|_\infty < \varepsilon$. Since $f_n \in R[a, b]$, $\Omega_{f_n}(P) < \varepsilon$ for some partition $P = \{x_i\}_{i=0}^n$. For $\xi, \eta \in [x_{i-1}, x_i]$, we have

$$\begin{aligned}|f(\xi) - f(\eta)| &\le |f(\xi) - f_n(\xi)| + |f_n(\xi) - f_n(\eta)| + |f_n(\eta) - f(\eta)| \\ &\le 2\,|f - f_n|_\infty + \omega_i^{f_n} < 2\varepsilon + \omega_i^{f_n}.\end{aligned}$$

Hence

$$\omega_i^f = \sup_{\xi,\eta\in[x_{i-1},x_i]} |f(\xi) - f(\eta)| \le 2\varepsilon + \omega_i^{f_n}$$

and $f \in R[a, b]$, because

$$\begin{aligned}\Omega_f(P) &= \sum_{i=1}^n \omega_i^f \Delta x_i \le \sum_{i=1}^n \left(2\varepsilon + \omega_i^{f_n}\right) \Delta x_i \\ &= 2\varepsilon\,(b-a) + \Omega_{f_n}(P) < [2\,(b-a) + 1]\,\varepsilon.\end{aligned}$$

Observing

$$\left|\int_a^b f_n - \int_a^b f\right| = \left|\int_a^b (f_n - f)\right| \le \int_a^b |f_n - f|$$
$$\le \int_a^b |f_n - f|_\infty = (b-a)\,|f_n - f|_\infty \to 0,$$

we get $\int_a^b f_n \to \int_a^b f$.

Example 2.24. From (2.3), we know that in Example 2.6 $f_n \not\rightrightarrows f$. On the other hand, in Example 2.5, $f_n \not\rightrightarrows f$ but $\int_0^1 f_n \to \int_0^1 f$.

Proposition 2.25. *If $f_n \in C^1[a,b]$, $f_n' \rightrightarrows g$. If $f_n(c) \to \alpha$ for some $c \in [a,b]$, then there is $f \in C^1[a,b]$ such that $f_n \rightrightarrows f$ and $f' = g$, i.e.*

$$\left(\lim_{n\to\infty} f_n\right)' = \lim_{n\to\infty} f_n'.$$

Proof. For $x \in [a,b]$, by Proposition 2.23

$$f_n(x) = f_n(c) + \int_c^x f_n' \to \alpha + \int_c^x g =: f(x),$$

we see that $f_n \to f$ on $[a,b]$. Moreover, $f' = g \in C[a,b]$, thus $f \in C^1[a,b]$.

Since

$$|f_n(x) - f(x)| = \left|\left(f_n(c) + \int_c^x f_n'\right) - \left(\alpha + \int_c^x g\right)\right|$$
$$\le |f_n(c) - \alpha| + \left|\int_c^x (f_n' - g)\right|$$
$$\le |f_n(c) - \alpha| + \int_a^b |f_n' - g|$$
$$\le |f_n(c) - \alpha| + (b-a)\left|f_n' - g\right|_\infty,$$

we get $f_n \rightrightarrows f$ because

$$|f_n - f|_\infty \le |f_n(c) - \alpha| + (b-a)\left|f_n' - g\right|_\infty \to 0.$$

2.3. Series of functions. Given $f_n : X \to \mathbb{R}$ ($n \in \mathbb{N}$), for $m \in \mathbb{N}$, we define $S_m = \sum_{n=1}^m f_n : X \to \mathbb{R}$ via

$$S_m(x) = \sum_{n=1}^m f_n(x).$$

If $\lim\limits_{m\to\infty} S_m(x)$ exists for $\forall x \in X$, call the limit $S(x)$, we get a function $S : X \to \mathbb{R}$ and we have $S_m \to S$ on X. Therefore

$$S(x) = \lim_{m\to\infty} S_m(x) = \lim_{m\to\infty} \sum_{n=1}^m f_n(x) =: \sum_{n=1}^\infty f_n(x),$$

and we denote $S = \sum_{m=1}^\infty f_m$. In general, we call the formal infinite sum $\sum_{m=1}^\infty f_m$ a series of functions, even if it does not *converge* (in that case it is simply a symbol without mathematical meaning).

Because $S = \sum_{m=1}^{\infty} f_m$ is the pointwise limit of the partial sum S_m, we say that the series converges to S point-wise. If $S_m \rightrightarrows S$, we say that the series converges uniformly, and write $S = \sum_{m=1}^{\infty} f_m$ uniformly on X. Because

$$\begin{aligned} |f_n|_\infty &= |S_n - S_{n-1}|_\infty = |(S_n - S) + (S - S_{n-1})|_\infty \\ &\le |S_n - S|_\infty + |S_{n-1} - S|_\infty \to 0, \end{aligned}$$

we have: 26

Proposition 2.26. *If $\sum_{n=1}^{\infty} f_n$ converges uniformly, then $f_n \rightrightarrows 0$.*

Thus, if $f_n \not\rightrightarrows 0$, then $\sum_{n=1}^{\infty} f_n$ does not converge uniformly. The converse of Proposition 2.26 is not true. Can you find a counterexample?

Proposition 2.27. *If $|f_n|_\infty \le a_n$ and the numerical series $\sum_{n=1}^{\infty} a_n$ converges, then the serie of functions $\sum_{n=1}^{\infty} f_n$ converges uniformly to its sum S.*

Proof. We need to show that $\{S_m\}$ converges uniformly. Given $\varepsilon > 0$, $\exists N$ such that

$$\sum_{i=n}^{m} a_i < \varepsilon \qquad \text{for } m \ge n \ge N.$$

Because $|f_n|_\infty \le a_n$, we deduce

$$\begin{aligned} |S_m - S_n|_\infty &= |f_{n+1} + \cdots + f_m|_\infty \\ &\le |f_{n+1}|_\infty + \cdots + |f_m|_\infty \\ &\le \sum_{i=n}^{m} a_i < \varepsilon. \end{aligned}$$

The desired result follows from Proposition 2.15.

Theorem 2.28. *Suppose $\sum_{n=1}^{\infty} f_n$ uniformly converges to S on $[a,b]$.*

(1) *If f_n is continuous at $x_0 \in [a,b]$, then S is continuous at x_0. If $f_n \in C[a,b]$, then $S \in C[a,b]$.*
(2) *If $f_n \in R[a,b]$, then $S \in R[a,b]$ and*

$$\int_a^b S = \int_a^b \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int_a^b f_n.$$

Example 2.29. If $f_n \in C[a,b]$, $\sum_{n=1}^{\infty} f_n(a)$ does not converge. Then $\sum_{n=1}^{\infty} f_n$ does not converge uniformly on (a,b). Thus, $\sum_{n=1}^{\infty} n^{-x}$ converges point-wise on $(1,\infty)$ but not uniformly.

Theorem 2.30. *If $f_n \in C^1[a,b]$, $\sum_{n=1}^{\infty} f_n(a)$ converges, $\sum_{n=1}^{\infty} f_n'$ uniformly converges to g on $[a,b]$, then $\sum_{n=1}^{\infty} f_n$ uniformly converges to some $G \in C^1[a,b]$ on $[a,b]$, moreover $G' = g$, that is*

$$\left(\sum_{n=1}^{\infty} f_n\right)' = \sum_{n=1}^{\infty} f_n'.$$

Integrating (differentiating) term by term is powerful to find the sum of some series.

Example 2.31. For $x \in (-\pi,\pi)$, find $S(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} \tan \frac{x}{2^n}$.

Proof. Firstly $S(0) = 0$. For $x \in (-\pi, \pi) \setminus 0$, the series converges *uniformly* on $[0, x]$. Integrating term by term, we get

$$\begin{aligned}
\int_0^x S(t)dt &= \sum_{n=1}^{\infty} \int_0^x \frac{1}{2^n} \tan \frac{t}{2^n}\, dt \\
&= -\sum_{n=1}^{\infty} \ln \cos \frac{x}{2^n} = -\lim_{N\to\infty} \sum_{n=1}^{N} \ln \cos \frac{x}{2^n} \\
&= -\lim_{N\to\infty} \ln \left(\cos \frac{x}{2} \cos \frac{x}{2^2} \cdots \cos \frac{x}{2^N} \right) \\
&= -\lim_{N\to\infty} \ln \frac{\sin x}{2^N \sin \frac{x}{2^N}} = -\ln \frac{\sin x}{x}.
\end{aligned}$$

Thus

$$S(x) = \left(-\ln \frac{\sin x}{x} \right)' = \frac{1}{x} - \cot x.$$

Example 2.32. Find $S(x) = \sum\limits_{n=1}^{\infty} n\,(n+1)\,x^n$.

Proof. For $x \in (-1, 1)$, the domain of S, we perform formal computation (by nice properties of power series, the uniform convergence needed is valid):

$$\begin{aligned}
S(x) &= \sum_{n=1}^{\infty} n\,(n+1)\,x^n = \sum_{n=1}^{\infty} \left(n x^{n+1} \right)' \\
&= \left(\sum_{n=1}^{\infty} n x^{n+1} \right)' = \left(x^2 \sum_{n=1}^{\infty} n x^{n-1} \right)' \\
&= \left(x^2 \sum_{n=0}^{\infty} (x^n)' \right)' = \left(x^2 \left(\sum_{n=0}^{\infty} x^n \right)' \right)' \\
&= \left(x^2 \left(\frac{1}{1-x} \right)' \right)' = \left(\frac{x^2}{(1-x)^2} \right)' = \frac{2x}{(1-x)^3}.
\end{aligned}$$

exq

Example 2.33. Find $S(x) = \sum\limits_{n=0}^{\infty} \dfrac{x^n}{n!}$.

Proof. The series converges for all $x \in \mathbb{R}$. For $x \in \mathbb{R}$, differentiating term by term (can we?) we get

$$\begin{aligned}
S'(x) &= \left(\sum_{n=0}^{\infty} \frac{x^n}{n!} \right)' = \sum_{n=0}^{\infty} \left(\frac{x^n}{n!} \right)' \\
&= \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty} \frac{x^n}{n!} = S(x).
\end{aligned}$$

Thus

$$\begin{aligned}
[e^{-x} S(x)]' &= e^{-x} \left[S'(x) - S(x) \right] = 0, \\
e^{-x} S(x) &= e^{-0} S(0) = 1.
\end{aligned}$$

Consequently $S(x) = e^x$.

Example 2.34. As exercise, find

$$S(x) = \sum_{n=1}^{\infty} n(n+2)x^n, \qquad s(x) = \sum_{n=0}^{\infty} \frac{(-1)^{n+1}x^{2n+1}}{(2n+1)!}.$$

3. Multivariable differential calculus

3.1. Partial derivative, differentiability. In single variable calculus, the derivative of a function f at a is defined as

$$f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}. \tag{3.1}$$

de

If f is m-variable, then both a and h are points in $\mathbb{R}^m$, it makes no sense to divide $f(a+h)-f(a)$ by h. Thus derivative of multivariable functions must be defined differently. We start with partial derivative.

Let $a = (a^1, \ldots, a^m) \in \mathbb{R}^m$, $r > 0$. We consider an m-variable function[13]

$$f : B_r(a) \to \mathbb{R}, \qquad f(x) = f(x^1, \ldots, x^m).$$

For each $i \in \overline{m}$ we have a single variable function $\varphi_i : (-r, r) \to \mathbb{R}$,

$$\varphi_i(t) = f(a + te_i) = f(a^1, \ldots, a^i + t, \ldots, a^m),$$

where $e_i = (0, \ldots, 1, \ldots, 0)$. The partial derivetive of f with respect to x^i at a is defined by the first equality below

$$\left.\frac{\partial f}{\partial x^i}\right|_a = \varphi_i'(0) = \lim_{t\to 0} \frac{\varphi_i(t)-\varphi_i(0)}{t} = \lim_{t\to 0}\frac{f(a+te_i)-f(a)}{t},$$

which is also denoted by $\partial_i f(a)$, $f_{x^i}(a)$ or $f_i(a)$.

From the definiton, we see that partial derivative is defined via derivative of single variable function. It is clear that $\partial_i f(a)$ is the rate of change of f at a with respect to the i^{th} variable x^i. What is the geometric interpretation of $\partial_i f(a)$?

If $\partial_i f(a)$ exists for all $i \in \overline{m}$, we call

$$\nabla f(a) = (\partial_1 f(a), \ldots, \partial_m f(a))$$

the gradient of f at a, which can also be denoted by $\mathrm{grad}\, f(a)$.

xd

Example 3.1. Consider $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = \sqrt[3]{xy}$. Do $\partial_1 f(0,0)$ and $\partial_2 f(0,0)$ exist? How about $\partial_1 f(1,0)$ and $\partial_2 f(1,0)$?

Proof. To investigate $\partial_1 f(0,0)$, we consider

$$\varphi(t) = f((0,0) + t(1,0)) = f(t,0).$$

By the definition of f, we see that $\varphi(t) \equiv 0$. Thus

$$\partial_1 f(0,0) = \varphi'(0) = 0.$$

Similarly $\partial_2 f(0,0) = 0$. Therefore $\nabla f(0,0) = (0,0)$.

Remark 3.2. Unlike single variable functions, f can be discontinuous at a even if $\partial_i f(a)$ exists for all $i \in \overline{m}$.

[13] In differential calculus we are interested in the local behavior of f near interior points of its domain. Therefore, for simplicity we may assume that f is defined on some ball $B_r(a)$.

Let $f : B_r(a) \to \mathbb{R}$. We say that f is differentiable at a, if there is $\lambda \in \mathbb{R}^m$ such that[14]

$$f(a+h) - f(a) - \lambda \cdot h = o(|h|) \qquad \text{as } h \to 0, \tag{3.2}$$

that is

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \lambda \cdot h}{|h|} = 0. \tag{3.2}$$ wee

From this definition, f is differentiable at a means that the change of f at a can be approximated by the linear function $h \mapsto \lambda \cdot h$ of h (the change of input), the error is higher order infinitesimal with respect to $|h|$.

In lower dimensional case $m = 2$ or $m = 3$, we can use x, y, z to denote independent variables. For example, 2-variable function $(x, y) \mapsto f(x, y)$ is differentiable at $(a, b) \in \mathbb{R}^2$ means there are $\lambda, \mu \in \mathbb{R}$ such that

$$\lim_{\rho \to 0} \frac{f(a+h, b+k) - f(a,b) - (\lambda k + \mu h)}{\rho} = 0.$$

where $\rho = \sqrt{h^2 + k^2}$.

Remark 3.3. Let $\lambda = (\lambda_1, \ldots, \lambda_m)$ satisfies (3.2), we call

$$df_a : h \mapsto \lambda \cdot h$$

the differential of f at a. If $|h| \ll 1$, we write $h = (dx^1, \ldots, dx^m)$, thus the differential

$$df = \lambda_1 dx^1 + \cdots + \lambda_m dx^m,$$

actually, this is the value of the differential at the h. From (3.2) we see that when $|h| \ll 1$

$$f(a+h) - f(a) \approx df, \tag{Newton's approximation}$$

thus the differential of f at a is a very good approximation of the change of f at a. 5t

Theorem 3.4. *If f is differentiable at a, i.e., f satisfies (3.2), then*

(1) *f is continuous at a,*
(2) *for $i \in \overline{m}$ we have $\partial_i f(a) = \lambda_i$, thus $\lambda = \nabla f(a)$.*

Proof. (1) From (3.2) we have

$$\lim_{|h| \to 0} f(a+h) = f(a),$$

thus f is continuous at a.

(2) Note that (3.2) implies

$$\lim_{t \to 0} \frac{f(a+te_i) - f(a) - \lambda \cdot (te_i)}{|te_i|} = 0,$$

hence

$$\begin{aligned} \partial_i f(a) &= \lim_{t \to 0} \frac{f(a+te_i) - f(a)}{t} \\ &= \lim_{t \to 0} \left(\frac{|te_i|}{t} \frac{f(a+te_i) - f(a) - \lambda \cdot (te_i)}{|te_i|} + \lambda \cdot e_i \right) \\ &= \lambda \cdot e_i = \lambda_i. \end{aligned}$$

[14]When $m = 1$ this is equivalent to (3.1), however, (3.1) makes no sense for $m > 1$. The equivalent form (3.2) resolves this difficulty.

pf

Proposition 3.5 (Fermat). *Let $U \subset \mathbb{R}^m$, $a \in U^\circ$ be a local extreme point of $f : U \to \mathbb{R}$. If $\partial_i f(a)$ exists then $\partial_i f(a) = 0$.*

Proof. Assume $B_r(a) \subset U$, then $t = 0$ is local extreme point of φ_i. Hence
$$\partial_i f(a) = \varphi_i'(0) = 0.$$

Let Ω be open subset in $\mathbb{R}^m$, $f : \Omega \to \mathbb{R}$. If f has partial derivative with respect to x^i at all $x \in \Omega$, then we have the partial derivative function (also called partial derivative) $\partial_i f : \Omega \to \mathbb{R}$,
$$x \mapsto \left.\frac{\partial f}{\partial x^i}\right|_x .$$
We say that f is continuously differentiable, write $f \in C^1(\Omega)$, if $\partial_i f \in C(\Omega)$ for all $i \in \overline{m}$.

Theorem 3.6. *Let $f : B_r(a) \to \mathbb{R}$. If $\partial_i f : B_r(a) \to \mathbb{R}$ is continuous at a for all $i \in \overline{m}$, then f is differentiable at a.*

Proof. Given $h \in B_r$, to investigate the limit (3.2), let $p_0 = a$,
$$p_k = a + \sum_{i=1}^{k} h^i e_i .$$
Applying the Lagrange mean value theorem to the single-variable function
$$t \mapsto f(a^1 + h^1, \dots, a^{k-1} + h^{k-1}, t, a^{k+1}, \dots, a^m)$$
on $[a^k, a^k + h^k]$, we have
$$f(p_k) - f(p_{k-1}) = \partial_k f(\xi_k) h^k,$$
for some $\xi_k \in (p_{k-1}, p_k)$. Thus
$$\begin{aligned}
\frac{|f(a+h) - f(a) - \nabla f(a)\cdot h|}{|h|} &= \frac{\left|\sum_{k=1}^m \big((f(p_k) - f(p_{k-1})) - \partial_k f(a) h^k\big)\right|}{|h|} \\
&\le \frac{1}{|h|}\sum_{k=1}^m |\partial_k f(\xi_k) - \partial_k f(a)| \left|h^k\right| \\
&\le \sum_{k=1}^m |\partial_k f(\xi_k) - \partial_k f(a)| \to 0 \qquad \text{as } h \to 0,
\end{aligned}$$
because $\partial_k f$ are continuous at a and $\xi_k \to a$ for all $k \in \overline{m}$ as $h \to 0$.

Example 3.7. Consider $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = \sqrt[3]{xy}$. Is f differentiable at $(0,0)$?

Proof. From Example 3.1 we have $\nabla f(0,0) = (0,0)$. For the differentiability of f at $(0,0)$, we consider the left hand side[15] of (3.2)
$$f((0,0) + (h,k)) - f(0,0) - \nabla f(0,0)\cdot(h,k) = f(h,k) = \sqrt[3]{hk}.$$

[15] If f is differentiable at $(0,0)$, by Theorem 3.4 (2), the λ on the left hand side of (3.2) must be $\nabla f(0,0)$.

Since

$$\lim_{(h,k)\to 0} \frac{\sqrt[3]{hk}}{\sqrt{h^2+k^2}} = 0$$

is not true, we conclude that f is not differentiable at $(0,0)$.

Now consider vector-valued function $f = (f^1, \ldots, f^n) : B_r(a) \to \mathbb{R}^n$. The partial derivative of f with respect to x^i at a is defined by

$$\partial_i f(a) = \left.\frac{\partial f}{\partial x^i}\right|_a = \varphi_i'(0) = \lim_{t\to 0} \frac{f(a+te_i)-f(a)}{t},$$

where $\varphi_i : (-r,r) \to \mathbb{R}^n$, $\varphi_i(t) = f(a+t)$. It is clear that

$$\partial_i f(a) = \left(\partial_i f^1(a), \ldots, \partial_i f^n(a)\right).$$

If there is $n \times m$ matrix[16] A such that

$$f(a+h) - f(a) - Ah = o(|h|) \qquad \text{as } h \to 0, \tag{3.3}$$ fd

that is

$$\lim_{h\to 0} \frac{f(a+h)-f(a)-Ah}{|h|} = 0, \tag{3.3}$$

we say that f is differentiable at a. It turns out that such A is unque, we call it the derivative of f at a and denote it by $f'(a)$.

u8

Remark 3.8. Since (3.3) involves matrix multiplication, we shall consider h as column vector. In what follows we often consider vector-valued functions $f : x \mapsto y$ as maps between column vectors.

From (3.3) we see that for small h, the linear map $A : \mathbb{R}^m \to \mathbb{R}^n$ is a very good linear approximation of the nonlinear map[17] $h \mapsto f(a+h) - f(a)$. We expect to get *local* properties of f near a through inverstigating $A = f'(a)$. This is the fundamental idea of differential calculus.

Let A^i be the rows of A, then

$$A = \begin{pmatrix} A^1 \\ \vdots \\ A^n \end{pmatrix},$$

Using

$$\begin{aligned} \left|f^i(a+h) - f^i(a) - A^i \cdot h\right| &\le |f(a+h) - f(a) - Ah| \\ &\le \sum_{i=1}^n \left|f^i(a+h) - f^i(a) - A^i \cdot h\right| \end{aligned}$$

we can easily prove:

---

[16]Here we view h as a column vector. Viewing h as a row vector, A should be $m \times n$ matrix and (3.3) should be

$$f(a+h) - f(a) - hA = o(|h|).$$

[17]called the increment of f at a.

Theorem 3.9. *The map $f : B_r(a) \to \mathbb{R}^n$ is differntiable at a iff all its components f^i are differentiable at a. In this case*

$$f'(a) = \begin{pmatrix} \dfrac{\partial f^1}{\partial x^1} & \cdots & \dfrac{\partial f^1}{\partial x^m} \\ \vdots & & \vdots \\ \dfrac{\partial f^n}{\partial x^1} & \cdots & \dfrac{\partial f^n}{\partial x^m} \end{pmatrix}_a = \begin{pmatrix} \nabla f^1 \\ \vdots \\ \nabla f^n \end{pmatrix} = (\partial_1 f, \ldots, \partial_m f).$$

If $\partial_i f^j(a)$ exist for all $i \in \overline{m}$ and $j \in \overline{n}$, we have the Jacobian matrix of f at a

$$\left(\frac{\partial f^i}{\partial x^j}\right)_a = \begin{pmatrix} \partial_1 f^1 & \cdots & \partial_m f^1 \\ \vdots & & \vdots \\ \partial_1 f^m & \cdots & \partial_m f^m \end{pmatrix}_a$$

even if f is not differentiable at a (in this case this matrix could not be denoted by $f'(a)$). When $m = n$, its determinant

$$J_f(a) = \det\left(\frac{\partial f^i}{\partial x^j}\right)_a = \left.\frac{\partial(f^1, \ldots, f^m)}{\partial(x^1, \ldots, x^m)}\right|_a$$

is call the Jacobian determinant of f at a.

Example 3.10. Let $A = (a^i_j)_{n\times m}$, $f : \mathbb{R}^m \to \mathbb{R}^n$ is defined by $f(x) = Ax$. For $a \in \mathbb{R}^m$, find $f'(a)$. xe

Proof. It is clear that

$$f(a + h) - f(a) - Ah = 0,$$

from the definition (3.3) it is clear that $f'(a) = A$.

To study the operations of differential maps, we need the norm of matrixs. Let A be an $n \times m$ matrx, then the function $h \mapsto |Ah|$ is continuous on $\mathbb{R}^m$ (why?), thus is bounded on ∂B^m_1. We define the (operator) norm of A by

$$\|A\| = \sup_{|h|=1} |Ah|.$$

Its geometric meaning is the maximal stretch ratio of $A : \mathbb{R}^m \to \mathbb{R}^n$ along all directions. Obviously

(1) For all $x \in \mathbb{R}^m$ we have $|Ax| \le \|A\|\,|x|$.
(2) Given $\ell \times m$ B, then $\|BA\| \le \|B\|\,\|A\|$.

Proposition 3.11 (derivative rule)**.** *If $f, g : B_r(a) \to \mathbb{R}^n$ are differentiable at a, $\lambda \in \mathbb{R}$, then* u9

(1) *$f + \lambda g$ is differentiable at a, $(f + \lambda g)'(a) = f'(a) + \lambda g'(a)$;*
(2) *$f \cdot g$ is differentiable at a and $(f \cdot g)'(a) = f^{\mathrm{T}}(a)g'(a) + g^{\mathrm{T}}(a)f'(a)$.*

Proof. 2) As $h \to 0$ we have

$$f(a + h) = f(a) + f'(a)h + o(h), \qquad g(a + h) = g(a) + g'(a)h + o(h).$$

Because[18] $o(h) + o(h) = o(h)$ and

$$f(a) \cdot o(h) = o(h), \qquad f'(a)h \cdot g'(a)h = o(h), \qquad f'(a)h \cdot o(h) = o(h),$$

[18] Equalities like this mean that: if $\varphi(h) = o(h)$ and $\psi(h) = o(h)$, then $\varphi(h) + \psi(h) = o(h)$.

we deduce

$$
\begin{aligned}
(f\cdot g)(a+h) &= \big(f(a)+f'(a)h+o(h)\big)\cdot\big(g(a)+g'(a)h+o(h)\big)\\
&= f(a)\cdot g(a) + f(a)\cdot g'(a)h + f'(a)h\cdot g(a) + o(h)\\
&= (f\cdot g)(a) + f^{\mathrm{T}}(a)g'(a)h + g^{\mathrm{T}}(a)f'(a)h + o(h)\\
&= (f\cdot g)(a) + \big(f^{\mathrm{T}}(a)g'(a) + g^{\mathrm{T}}(a)f'(a)\big)\,h + o(h).
\end{aligned}
$$

Hence $f\cdot g$ is differentiable at a and $(f\cdot g)'(a) = f^{\mathrm{T}}(a)g'(a) + g^{\mathrm{T}}(a)f'(a)$.

Remark 3.12. In the above proof, both $f(a)$ and $g'(a)h$ are column vector (Remark 3.8), hence their dot product

$$f(a)\cdot g'(a)h = f^{\mathrm{T}}(a)g'(a)h.$$

Example 3.13. Let $A = \big(a_{ij}\big)_{n\times n}$, $f:\mathbb{R}^n\to\mathbb{R}$ is defined by $f(x) = Ax\cdot x$, that is

$$f(x) = \sum_{i,j=1}^{n} a_{ij}x^i x^j.$$

For $a\in\mathbb{R}^n$ find $\nabla f(a)$.

Proof. Since $f(x) = Ax\cdot x$, using Proposition 3.11 (2) and Example 3.10

$$
\begin{aligned}
\nabla f(a) = f'(a) &= (Ax)^{\mathrm{T}}_{x=a}\,(x)'_{x=a} + (x)^{\mathrm{T}}_{x=a}\,(Ax)'_{x=a}\\
&= a^{\mathrm{T}}A^{\mathrm{T}}I_n + a^{\mathrm{T}}A = a^{\mathrm{T}}\left(A^{\mathrm{T}}+A\right).
\end{aligned}
$$

In particular, if A is symetric, then $\nabla f(a) = 2a^{\mathrm{T}}A$.

Proof. Since f is polynomial, it is differentiable. To find $\nabla f(a)$, it suffices to find

$$
\begin{aligned}
\partial_k f(a) = \partial_k|_{x=a}\left(\sum_{i,j=1}^{n} a_{ij}x^i x^j\right) &= \sum_{i,j=1}^{n}\partial_k|_{x=a}\left(a_{ij}x^ix^j\right)\\
&= \sum_{i,j=1}^{n} a_{ij}\,\partial_k|_{x=a}\left(x^ix^j\right)\\
&= \sum_{i,j=1}^{n} a_{ij}\left(a^i\partial_k|_{x=a}x^j + a^j\partial_k|_{x=a}x^i\right)\\
&= \sum_{i,j=1}^{n} a_{ij}\left(a^i\delta^j_k + a^j\delta^i_k\right) = \sum_{i=1}^{n} a_{ik}a^i + \sum_{j=1}^{n} a_{kj}a^j\\
&= \left(a^{\mathrm{T}}\left(A^{\mathrm{T}}+A\right)\right)_k.
\end{aligned}
$$

Thus $\nabla f(a) = a^{\mathrm{T}}\left(A^{\mathrm{T}}+A\right)$.

Example 3.14. If $A = (a_{ij})_{n\times n}$ is positive symmetric matrix, $f = \nabla F$ for some $F\in C^1(\mathbb{R}^n)$ satisfying

$$\lim_{|x|\to\infty}\frac{F(x)}{|x|^2} = 0. \tag{3.4}$$

Then the nonlinear algebraic equation $Ax = f^{\mathrm{T}}(x)$, in component form

$$\sum_{j=1}^{n} a_{ij}x^j = f_i(x^1, \dots, x^n), \qquad i \in \overline{n},$$

has a solution.

Proof. Let $\lambda_1 > 0$ be the smallest eigenvalue of A, then

$$Ax \cdot x \geq \lambda_1 |x|^2 \qquad \text{for all } x \in \mathbb{R}^n.$$

Consider the C^1-function $\Phi : \mathbb{R}^n \to \mathbb{R}$,

$$\Phi(x) = \frac{1}{2}Ax \cdot x - F(x).$$

As $|x| \to \infty$ we have

$$\frac{\Phi(x)}{|x|^2} \geq \frac{\frac{1}{2}\lambda_1 |x|^2 - F(x)}{|x|^2} \to \frac{1}{2}\lambda_1.$$

Which implies

$$\lim_{|x|\to\infty} \Phi(x) = +\infty.$$

Hence there is $\xi \in \mathbb{R}^n$ such that $\Phi(\xi) = \inf_{\mathbb{R}^n} \Phi$. By Proposition 3.5 we deduce

$$0 = \nabla\Phi(\xi) = \xi^{\mathrm{T}}A - \nabla F(\xi) = \xi^{\mathrm{T}}A - f(\xi).$$

That is $A\xi = f^{\mathrm{T}}(\xi)$.

Remark 3.15. The condition (3.4) can be weaken as

$$\overline{\lim_{|x|\to\infty}} \frac{F(x)}{|x|^2} < \frac{\lambda_1}{2}.$$

3.2. Chain rule. The chain rule is very useful for differentiating multivariable functions. Recall that $f : B_r^m(a) \to \mathbb{R}^n$ is differentiable at a means there is an $n \times m$ matrix A suc that

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Ah}{|h|} = 0.$$

Theorem 3.16 (Chain rule)**.** *If $g : B_r(a) \to \mathbb{R}^n$ is differentiable at a, U is open set in $\mathbb{R}^n$ containning $g(B_r(a))$, and $f : U \to \mathbb{R}^\ell$ is differentiable at $b = g(a)$, then $f \circ g : B_r(a) \to \mathbb{R}^\ell$ is differentiable at a and* chn

$$(f \circ g)'(a) = f'(b)g'(a).$$

The conclusion of the theorem says that the Jacobian metrix of $f \circ g$ at a is the product of the Jacobian matrix of f at $b = g(a)$ and the Jacobian metrix of g at a. That is, if $g : x \mapsto u$ is differentiable at a, $f : u \mapsto y$ is differentiable at $b = g(a)$, then $f \circ g : x \mapsto y$ is differentiable at a and

$$\begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \cdots & \frac{\partial y^1}{\partial x^m} \\ \vdots & & \vdots \\ \frac{\partial y^\ell}{\partial x^1} & \cdots & \frac{\partial y^\ell}{\partial x^m} \end{pmatrix}_a = \begin{pmatrix} \frac{\partial y^1}{\partial u^1} & \cdots & \frac{\partial y^1}{\partial u^n} \\ \vdots & & \vdots \\ \frac{\partial y^\ell}{\partial u^1} & \cdots & \frac{\partial y^\ell}{\partial u^n} \end{pmatrix}_b \begin{pmatrix} \frac{\partial u^1}{\partial x^1} & \cdots & \frac{\partial u^1}{\partial x^m} \\ \vdots & & \vdots \\ \frac{\partial u^n}{\partial x^1} & \cdots & \frac{\partial u^n}{\partial x^m} \end{pmatrix}_a .$$

Or equivalently,

$$\left.\frac{\partial y^k}{\partial x^i}\right|_a = \sum_{j=1}^{n} \left.\frac{\partial y^k}{\partial u^j}\right|_b \cdot \left.\frac{\partial u^j}{\partial x^i}\right|_a \qquad \text{for } i \in \overline{m}, k \in \overline{\ell}.$$

Proof (Theorem 3.16). Let $A = f'(b)$, $B = g'(a)$. Since g is continuous at a, we may choose $\varepsilon > 0$ and $\delta \in (0, r)$ such that $B^n_\varepsilon(b) \subset U$, $g(B^m_\delta(a)) \subset B^n_\varepsilon(b)$.

For $\eta : B^m_\delta(0) \to \mathbb{R}^n$ and $\lambda : B^n_\varepsilon(0) \to \mathbb{R}^\ell$ determined by

$$f(b+k) - f(b) = Ak + \lambda(k), \tag{3.5}$$ w

$$g(a+h) - g(a) = Bh + \eta(h), \tag{3.6}$$ w5

we have

$$\lim_{|k|\to 0} \frac{\lambda(k)}{|k|} = 0, \qquad \lim_{|h|\to 0} \frac{\eta(h)}{|h|} = 0. \tag{3.7}$$ w7

Decreasing δ if necessary, we may assume $|\eta(h)| \le |h|$.

Let $k = Bh + \eta(h)$. Then (3.6) becomes $g(a+h) = b + k$. Thus by (3.5) we have

$$\begin{aligned}(f\circ g)(a+h) - (f\circ g)(a) &= f(b+k) - f(b)\\ &= A(Bh+\eta(h)) + \lambda(Bh+\eta(h))\\ &= (AB)h + [A\eta(h) + \lambda(Bh+\eta(h))].\end{aligned}$$

From $|A\eta(h)| \le \|A\|\,|\eta(h)|$ we have

$$\lim_{h\to 0} \frac{A\eta(h)}{|h|} = 0.$$

Therefore, it remains to show

$$\lim_{h\to 0} \frac{\lambda(Bh+\eta(h))}{|h|} = 0. \tag{3.8}$$ w9

Given $\varepsilon > 0$, from (3.7) and $\lambda(0) = 0$, there is $\rho > 0$ such that

$$|\lambda(k)| \le \frac{\varepsilon}{1+\|B\|}\,|k| \qquad \text{for } k \in B^n_\rho.$$

If $|h| < \min\left\{\delta, (\|B\|+1)^{-1}\rho\right\}$, then

$$\begin{aligned}|Bh+\eta(h)| &\le |Bh| + |\eta(h)|\\ &\le \|B\|\,|h| + |h| < \rho.\end{aligned}$$

Consequently

$$\begin{aligned}|\lambda(Bh+\eta(h))| &\le \frac{\varepsilon}{1+\|B\|}\,|Bh+\eta(h)|\\ &\le \frac{\varepsilon}{1+\|B\|}\,(\|B\|\,|h| + |h|) = \varepsilon\,|h|,\end{aligned}$$

and (3.8) is proved.

wt0

Corollary 3.17. *Let $g : B_r(a) \to \mathbb{R}^n$. If $\partial_i g(a)$ exists and f is differentiable at $b = g(a)$, then $f\circ g : B_r(a) \to \mathbb{R}^\ell$ has partial derivative with respect to x^i at a and*

$$\partial_i(f\circ g)(a) = f'(b)\partial_i g(a). \tag{3.9}$$ wch

Proof. Because $\partial_i g(a)$ exsits, $\varphi : (-r,r) \to \mathbb{R}^n$, $\varphi(t) = g(a + te_i)$ is differentiable at $t = 0$[19]. Applying chain rule to

$$(-r,r) \xrightarrow{\varphi} U \xrightarrow{f} \mathbb{R}^\ell,$$

yields the desired conclusion.

Remark 3.18. Let $f : B_r(a) \to \mathbb{R}^n$ be differentiable at a, $h \in \mathbb{R}^m$. For $g : t \mapsto a + th$, applying chain rule to the composition w4

$$(-\varepsilon,\varepsilon) \xrightarrow{g} B_r(a) \xrightarrow{f} \mathbb{R}^n$$

yields

$$f'(a)h = \frac{d}{dt}\bigg|_{t=0} f(a+th).$$

Example 3.19. Let $g : \mathbb{R}^2 \to \mathbb{R}^2$, $g(x,y) = (x+2y, ye^x)$, $f : \mathbb{R} \to \mathbb{R}^2$ is differentiable with

$$f(0) = (0,1)\,, \qquad f'(0) = (-1,2)\,.$$

Find $(g \circ f)'(0)$.

Proof. By the chain rule

$$\begin{aligned}
(g \circ f)'(0) &= g'(f(0))f'(0) \\
&= \begin{pmatrix} \partial_x g^1 & \partial_y g^1 \\ \partial_x g^2 & \partial_y g^2 \end{pmatrix}_{(0,1)} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ ye^x & e^x \end{pmatrix}_{(0,1)} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.
\end{aligned}$$

Theorem 3.20 (Lagrange)**.** *If $f : \Omega \to \mathbb{R}$ is continuous on $[a,b] \subset \Omega^\circ$, and differentiable in (a,b), then $\exists \xi \in (a,b)$ such that* wt1

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

Remark 3.21. In many books on the topic, f is required to be differentiable over the whole Ω, this prevents applications to some problems such as Example 3.24.

Proof. We convert the multivariable problem into single variably one by restricing the variable on a direction. Let $\varphi : [0,1] \to \mathbb{R}$, $\varphi(t) = f(a + t(b-a))$. By 3.16, φ is continuous on $[0,1]$ and differentiable in $(0,1)$, and

$$\varphi'(t) = f'(a + t(b-a))(b-a).$$

Applying the Lagrange mean value theorem to φ on $[0,1]$, $\exists \tau \in (0,1)$ such that

$$\begin{aligned}
f(b) - f(a) = \varphi(1) - \varphi(0) &= \varphi'(\tau) \\
&= f'(a + \tau(b-a))(b-a).
\end{aligned}$$

We see that $\xi = a + \tau(b-a)$ satisfies the requirment.

[19]For single variable functions, differentiability is equivalent to existence of derivative.

Remark 3.22. For real-valued function f, we have

$$f'(x)h = (\partial_1 f(x), \dots, \partial_m f(x)) \begin{pmatrix} h^1 \\ \vdots \\ h^m \end{pmatrix} = \nabla f(x) \cdot h.$$

Therefore, the conclusion of Theorem 3.20 can also be written as

$$f(b) - f(a) = \nabla f(\xi) \cdot (b - a).$$

Example 3.23. Let $u : \mathbb{R}^2 \to \mathbb{R}$ be differentiable and satisfies

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = 0, \qquad u(x, 0) = 0.$$

Show that $u \equiv 0$.

Proof. For $(x_0, y_0) \in \mathbb{R}^2$, we consider $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\varphi(t) = u(x_0 + t, y_0 + t).$$

It is clear that φ is differentiable on $\mathbb{R}$, for $t \in \mathbb{R}$ we have

$$\dot{\varphi}(t) = \frac{\partial u}{\partial x}\bigg|_{(x_0+t, y_0+t)} + \frac{\partial u}{\partial u}\bigg|_{(x_0+t, y_0+t)} = 0.$$

Thus φ is constant function. Hence

$$u(x_0, y_0) = \varphi(0) = \varphi(-y_0) = u(x_0 - y_0, 0) = 0.$$

sx

Example 3.24. Let $f : B_\delta(a) \to \mathbb{R}$ be continuous, and differentiable on $B_\delta(a) \setminus \{a\}$;

$$(x - a) \cdot \nabla f(x) < 0 \qquad \text{for} x \in B_\delta(a) \setminus \{a\}.$$

Then a is maximizer of f.

Proof. For $\forall x \in B_\delta(a) \setminus \{a\}$, f is continuous on $[a, x]$, and differentiable on (a, x). By Theorem 3.20, there is

$$\xi = a + \tau(x - a) \in (a, b), \qquad \tau \in (0, 1),$$

such that

$$f(x) - f(a) = \nabla f(\xi) \cdot (x - a) = \frac{1}{\tau} \nabla f(\xi) \cdot (\xi - a) < 0.$$

We see that a is the maximizer of f.

Theorem 3.20 is *not true* for vector-valued functions, but we have a weaker result.

tmv

Theorem 3.25 (Meanvalue inequality)**.** *If $f : \Omega \to \mathbb{R}^n$ is continuous on $[a, b] \subset \Omega$ and differentiable in (a, b), then $\exists \xi \in (a, b)$ such that*

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

Proof. The idea is converting vector-valued function into scale function via dot product. Consider $\varphi : \Omega \to \mathbb{R}$,

$$\varphi(x) = (f(b) - f(a)) \cdot f(x).$$

By Proposition 3.11, $\varphi \in C^1(\Omega)$ and

$$\varphi'(x) = (f(b) - f(a))^{\mathrm{T}} f'(x).$$

By the Lagrange mean value theorem, $\exists \xi \in (a, b)$ such that

$$
\begin{aligned}
|f(b) - f(a)|^2 = \varphi(b) - \varphi(a) &= \varphi'(\xi)(b-a) \\
&= (f(b) - f(a))^{\mathrm{T}}\, f'(\xi)\,(b-a) \\
&= (f(b) - f(a)) \cdot \big(f'(\xi)(b-a)\big) \\
&\le |(f(b) - f(a))|\, |f'(\xi)(b-a)| \\
&\le |f(b) - f(a)|\, \|f'(\xi)\|\, |b-a|\,.
\end{aligned}
$$

3.3. Directional derivative and gradient. The directional derivative of $f : B_r(a) \to \mathbb{R}$ at a in the direction $\ell \in \mathbb{R}^m$ is defined by

$$
\left.\frac{\partial f}{\partial \ell}\right|_a = \varphi_\ell'(0) = \left.\frac{d}{dt}\right|_{t=0} f(a + t\ell) = \lim_{t\to 0} \frac{f(a+t\ell) - f(a)}{t},
$$

it is also denoted by $\nabla_\ell f(a)$, where $\varphi_\ell : (-r, r) \to \mathbb{R}$, $\varphi_\ell(t) = f(a + t\ell)$.

The directional derivative $\nabla_\ell f(a)$ is the rate of change of f at a in the direction ℓ. Obviously $\nabla_{e_i} f(a) = \partial_i f(a)$.

Remark 3.26. We may also define one-side derectional derivative

$$
\nabla_\ell^{\pm} f(a) = (\varphi_\ell)'_{\pm}(0) = \lim_{t\to 0\pm} \frac{f(a+t\ell) - f(a)}{t}.
$$

Then, $\nabla_\ell f(a)$ exists iff both $\nabla_\ell^{\pm} f(a)$ exists and $\nabla_\ell^{+} f(a) = \nabla_\ell^{-} f(a)$. We need such one-side derivative if a is a boundary point of the domain of f.

w0

Theorem 3.27. *If f is differentiable at a, then $\nabla_\ell f(a) = \ell \cdot \nabla f(a)$ for all $\ell \in \mathbb{R}^m$.*

Let θ be the angle between ℓ and $\nabla f(a)$, then

$$
\nabla_\ell f(a) = |\ell|\, |\nabla f(a)| \cos\theta.
$$

Thus, $\nabla f(a)$ is the direction along which f grows most rapidly.

Informally, because f is differentiable at a,

$$
f(a+h) - f(a) = \nabla f(a) \cdot h + o(|h|) \qquad \text{as } h \to 0.
$$

Let $h = t\ell$, *then* $o(|h|) = o(t)$. Hence

$$
\frac{f(a+t\ell) - f(a)}{t} = \ell \cdot \nabla f(a) + \frac{o(t)}{t} \to \ell \cdot \nabla f(a) \qquad \text{as } t \to 0.
$$

Proof. As $t \to 0$,

$$
\begin{aligned}
\frac{f(a+t\ell) - f(a)}{t} &= \frac{|t\ell|}{t} \left(\frac{f(a+t\ell) - f(a) - \nabla f(a)\cdot(t\ell)}{|t\ell|} + \frac{t\nabla f(a)\cdot \ell}{|t\ell|} \right) \\
&= \frac{|t\ell|}{t}\, \frac{f(a+t\ell) - f(a) - \nabla f(a)\cdot(t\ell)}{|t\ell|} + \nabla f(a)\cdot \ell \\
&\to \nabla f(a)\cdot \ell, \qquad (3.10)
\end{aligned}
$$

edd

this implies $\varphi_\ell'(0) = \nabla f(a) \cdot \ell$. Here, the first term in the second line of (3.10) goes to zero because f is differentiable at a.

Remark 3.28. Theorem 3.27 can also be proved via Remark 3.18.

Example 3.29. Consider $f : \mathbb{R}^2 \to \mathbb{R}$,

$$f(x,y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

As $x \to 0$,

$$f(x, x^2) \to \frac{1}{2} \neq f(0,0),$$

thus f is not continuous, hence not differentiable, at $(0,0)$. For $\ell = (h,k)$, define $\varphi : t \mapsto f(th, tk)$. If $k \neq 0$,

$$\begin{aligned} \nabla_\ell f(0,0) &= \lim_{t\to 0} \frac{\varphi(t) - \varphi(0)}{t} \\ &= \lim_{t \to 0} \frac{1}{t} \frac{(th)^2 tk}{(th)^4 + (tk)^2} = \frac{h^2}{k}. \end{aligned}$$

If $k = 0$, then $\varphi(t) = f(th, 0) = 0$, hence

$$\nabla_\ell f(0,0) = \dot{\varphi}(0) = 0.$$

Thus, along any direction ℓ, the directional derivative $\nabla_\ell f(0,0)$ exists, but f is not differentiable at $(0,0)$.

Theorem 3.27 reveals the meaning of gradient for scalar functions. We can also define divergence for vector fields on $\mathbb{R}^m$ and curl for vector fields on $\mathbb{R}^3$. To explain their meaning, we need integrals of multivariable functions.

Let Ω be an open subset of $\mathbb{R}^m$. A map $F = \left(F^1, \dots, F^m\right) : \Omega \to \mathbb{R}^m$ is called a vector field. The divergence of F at $a \in \Omega$ is defined by

$$\operatorname{div} F(a) = (\nabla \cdot F)(a) = \sum_{i=1}^m \left.\frac{\partial F^i}{\partial x^i}\right|_a$$

If $\operatorname{div} F(a)$ exist for all $a \in \Omega$, we get a new scalar function $\operatorname{div} F$ from the vector field F:

$$\operatorname{div} F = \nabla \cdot F : \Omega \to \mathbb{R}, \qquad x \mapsto \operatorname{div}(x).$$

When $m = 3$ and F is C^1, we can also produce a new vector field $\operatorname{rot} F = \nabla \times F : \Omega \to \mathbb{R}^3$,

$$\begin{aligned} \operatorname{rot} F(x) = (\nabla \times F)(x) &= \det \begin{pmatrix} e_1 & e_2 & e_3 \\ \partial_1 & \partial_2 & \partial_3 \\ F^1 & F^2 & F^3 \end{pmatrix} \\ &= \left(\partial_2 F^3 - \partial_3 F^2, \partial_3 F^1 - \partial_1 F^3, \partial_1 F^2 - \partial_2 F^1\right), \end{aligned}$$

called the curl of F.

From the definition, similar to gradient, both divergence and curl are first order differential operators. From the rules of partial derivative we can esily obtain the rules for these operators.

div

Proposition 3.30. *If $g \in C^1(\Omega)$, $F \in C^1(\Omega, \mathbb{R}^m)$, then*

$$\operatorname{div}(gF) = g \operatorname{div} F + \nabla g \cdot F.$$

Proof. Compute directly:

$$\operatorname{div}(gF) = \sum_{i=1}^{m} \partial_i (gF^i) = \sum_{i=1}^{m} \left(g(\partial_i F^i) + (\partial_i g) F^i \right) = g \operatorname{div} F + \nabla g \cdot F.$$

3.4. Inverse function theorem. Let $a \in \mathbb{R}^m$, an open set containing a is called an open neighbouhood of a. The collection of all open neighbouhoods of a is denoted by $\mathcal{N}_a$ (or $\mathcal{N}_a^m$ if we need to specify the dimension).

Let U and V be open sets of $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, $f : U \to V$. If f is bijective and both f and $f^{-1} : V \to U$ are C^k, then f is called a C^k-diffeomorphism (then we must have $m = n$). If $a \in U$ and there are $A \in \mathcal{N}_a$ and $B \in \mathcal{N}_{f(a)}$ such that $f|_A : A \to B$ is a C^k-diffeomorphism, then we called f a local C^k-diffeomorphism at a.

wti

Theorem 3.31 (Inverse function theorem)**.** *Let Ω be open subset of $\mathbb{R}^m$, $f \in C^k(\Omega, \mathbb{R}^m)$, $a \in \Omega$. If $\det f'(a) \neq 0$, then f is a local C^k-diffeomorphism at a.*

w3

Lemma 3.32. *Let Ω be open subset in $\mathbb{R}^m$, $f \in C^1(\Omega, \mathbb{R}^m)$, $a \in \Omega$. If $\det f'(a) \neq 0$, then $\exists \varepsilon > 0$, such that $B_\varepsilon[a] \subset \Omega$ and*

$$|f(x) - f(y)| \geq \varepsilon |x - y| \qquad \text{for } x, y \in B_\varepsilon[a]. \tag{3.11}$$ zz

Proof (Method 1). Otherwise, for $\forall n$, there are $x_n, y_n \in B_{1/n}(a)$, such that[20]

$$\begin{aligned} \frac{1}{n} |x_n - y_n| &> |f(x_n) - f(y_n)| \\ &= \left| \begin{pmatrix} \nabla f^1(\xi_n^1)(x_n - y_n) \\ \vdots \\ \nabla f^m(\xi_n^m)(x_n - y_n) \end{pmatrix} \right|, \end{aligned} \tag{3.12}$$ eh

where $\xi_n^i \in [x_n, y_n]$ is obtained by applying Theorem 3.20 to f^i.

We may assume

$$h_n = \frac{x_n - y_n}{|x_n - y_n|} \to h,$$

then $h \neq 0$. Let $n \to \infty$ after dividing both sides of (3.12) by $|x_n - y_n|$, noticing $\xi_n^i \to a$ for all $i \in \overline{m}$ we get $f'(a)h = 0$, contradicting to $\det f'(a) \neq 0$.

Proof (Method 2). Let $A = f'(a)$. Because A is invertible, $\exists \delta > 0$ such that

$$|Ax| \geq 2\delta |x|, \qquad \forall x \in \mathbb{R}^m.$$

Consider the C^1-map $\varphi : \Omega \to \mathbb{R}^m$, $\varphi(x) = Ax - f(x)$. We have

$$\varphi'(a) = A - f'(a) = 0_m,$$

i.e., $\|\varphi'(a)\| = 0$. By the *continuity* of $x \mapsto \|\varphi'(x)\|$, $\exists \varepsilon > 0$ such that $\|\varphi'(x)\| \leq \delta$ for $x \in B_\varepsilon(a)$.

For $x, y \in B_\varepsilon(a)$, by the meanvalue inequality (Theorem 3.25), $\exists \xi \in (x, y)$, such that

$$\delta |x - y| \geq \|\varphi'(\xi)\| \, |x - y| \geq |\varphi(x) - \varphi(y)|$$

---

[20]For vector-valued functions, the relation between f and its derivative is the inequality

$$|f(x_n) - f(y_n)| \leq \|f'(\xi_n)\| |x_n - y_n|.$$

Unfortunately, the inequality is on the wrong direction: we could not link it with the left hand side of (3.12). Observing that for scalar functions, the relation is an equality, in the second step of (3.12) we apply Theorem 3.20 to the components of f.

$$
\begin{aligned}
&= |A(x-y) - (f(x)-f(y))| \\
&\geq |A(x-y)| - |f(x)-f(y)| \\
&\geq 2\delta |x-y| - |f(x)-f(y)| .
\end{aligned}
$$

Now (3.11) follows.

Remark 3.33. The second proof does not relay on the local compactness of $\mathbb{R}^m$, so it can be generalized to infinite dimensional spaces (in such spaces, bounded sequences need not have convergent subsequences).

w6

Lemma 3.34. *Let G be an open subset of $\mathbb{R}^m$, $f : G \to \mathbb{R}^m$ is C^1. If $\det f'(x) \neq 0$ for $\forall x \in G$, then $f(G)$ is an open subset of $\mathbb{R}^m$.*

Proof. Let $b \in f(G)$, we show that $b \in [f(G)]^\circ$. Take $a \in f^{-1}(b)$. By Lemma 3.32, there is $\varepsilon > 0$ such that $f : B_\varepsilon[a] \to \mathbb{R}^m$ is injective. Thus $f(x) \neq b$ for $\forall x \in \partial B_\varepsilon(a)$. Hence

$$\mu = \inf_{x \in \partial B_\varepsilon(a)} |f(x) - b| > 0.$$

Given $y \in B_{\mu/2}(b)$, consider the function $\psi : B_\varepsilon[a] \to \mathbb{R}$,

$$\psi(x) = |f(x) - y|^2 .$$

For $x \in \partial B_\varepsilon(a)$ we have

$$
\begin{aligned}
\psi(x) = |f(x)-y|^2 &\geq \{|f(x)-b| - |b-y|\}^2 \\
&> \left\{\mu - \frac{\mu}{2}\right\}^2 = \frac{\mu^2}{4} > |b-y|^2 = \psi(a).
\end{aligned}
$$

Therefore ψ takes its minimum at some $\xi \in B_\varepsilon(a)$, and we have

$$0 = \psi'(\xi) = f'(\xi)(f(\xi) - y).$$

From $\det f'(\xi) \neq 0$ we have $y = f(\xi)$, namely $y \in f(G)$. Thus $B_{\mu/2}(b) \subset f(G)$ and $b \in [f(G)]^\circ$.

Proof (Theorem 3.31). Since $\det f'(a) \neq 0$, by the continuity of $x \mapsto \det f'(x)$ and Lemma 3.32, there are $\varepsilon > 0$, such that $B_\varepsilon(a) \subset \Omega$, $\det f'(x) \neq 0$ for $x \in B_\varepsilon(a)$, and

$$|f(x_1) - f(x_2)| \geq \varepsilon |x_1 - x_2| , \qquad \forall x_1, x_2 \in B_\varepsilon(a). \tag{3.13}$$ e3e

By Lemma 3.34, $V = f(B_\varepsilon(a))$ is an open neighbourhood of $b = f(a)$. Obviously $f : B_\varepsilon(a) \to V$ is bijective, let $\varphi : V \to B_\varepsilon(a)$ be its inverse. From (3.13) we get

$$|\varphi(y_1) - \varphi(y_2)| \leq \frac{1}{\varepsilon} |y_1 - y_2| . \tag{3.14}$$ 3e5

So $\varphi : V \to B_\varepsilon(a)$ is continuous.

For $y \in V$, we prove that φ is differentiable at y. For $k \in \mathbb{R}^m \backslash 0$ small, Let

$$x = \varphi(y), \qquad h = \varphi(y+k) - \varphi(y).$$

Then

$$y + k = f(\varphi(y+k)) = f(x+h), \qquad |h| = |\varphi(y+k) - \varphi(y)| \leq \frac{1}{\varepsilon} |k| .$$

Since $k \neq 0$ and φ is injective, we have $h \neq 0$. Moreover, as $k \to 0$ we have $h \to 0$. From

$$\begin{aligned}
\frac{\left|\varphi(y+k)-\varphi(y)-[f'(x)]^{-1}k\right|}{|k|} &= \frac{\left|h-[f'(x)]^{-1}k\right|}{|k|} \\
&= \frac{\left|[f'(x)]^{-1}\left(f'(x)h-(f(x+h)-f(x))\right)\right|}{|k|} \\
&\leq \frac{\|[f'(x)]^{-1}\|\,|f'(x)h-(f(x+h)-f(x))|}{|h|}\frac{|h|}{|k|} \\
&\leq \frac{\|[f'(x)]^{-1}\|}{\varepsilon}\frac{|f(x+h)-f(x)-f'(x)h|}{|h|}
\end{aligned}$$

and the differentiability of f at x, we get

$$\lim_{k\to 0}\frac{\left|\varphi(y+k)-\varphi(y)-[f'(x)]^{-1}k\right|}{|k|} = 0.$$

Thus, φ is differentiable at y and $\varphi'(y) = [f'(x)]^{-1}$, that is

$$(f^{-1})'(y) = \left[f'(x)\right]^{-1} = \left[f'(f^{-1}(y))\right]^{-1}.$$

By the formula for inverse matrix and continuity of f' and f^{-1}, we see that f^{-1} is C^1.

Remark 3.35. Our proof of Theorem 3.31 relies on Lemma 3.34, whose proof in turn relies on the local compactness of $\mathbb{R}^m$ (thus is not valid if $\mathbb{R}^m$ is replaced by an infinite dimensional Banach space; although the conclusion remains true). Theorem 3.31 can also be proved via Banach's contraction principle (Proposition 1.45); this approach does not rely on the local compactness.

The inverse function theorem says that for $f : \Omega \to \mathbb{R}^m$, if the linerization $f'(a) : \mathbb{R}^m \to \mathbb{R}^m$ is invertible, then locally f is invertible near a. In the same spirit, for $f : \Omega \to \mathbb{R}^n$, if $f'(a) : \mathbb{R}^m \to \mathbb{R}^n$ is suejective (injective), we expect f to be locally surjective (injective). top

Theorem 3.36. *Let Ω be open subset in $\mathbb{R}^m$, $a \in \Omega$, $f : \Omega \to \mathbb{R}^n$ is C^1, $f(a) = b$. If* $\operatorname{rank} f'(a) = n$ (this means $f'(a) : \mathbb{R}^m \to \mathbb{R}^n$ is suejective)*, then $b \in [f(\Omega)]^\circ$.*

Remark 3.37. That $b \in [f(\Omega)]^\circ$ means that all points near b are contained in the image of f. For this reason we say that f is locally surjective at a.

In particular, If for $\forall x \in \Omega$ we have $\operatorname{rank} f'(x) = n$, then $f(\Omega)$ is open subset of $\mathbb{R}^m$. Thus Lemma 3.34 is a special case of Theorem 3.36. re

Remark 3.38. Let Ω be open subset of $\mathbb{R}^m$, $f : \Omega \to \mathbb{R}^n$ is a C^1-map, $a \in \Omega$. If

$$\operatorname{rank} f'(a) < n,$$

we say that a is a critical point of f. Thus, Theorem 3.36 says that if a is not a critical point of f, then f is locally surjective at a.

Proof. Let $f = \left(f^1, \ldots, f^n\right)$. We may assume

$$\det\left(\partial_i f^j(a)\right)_{i,j\in\overline{n}} \neq 0,$$

Define $\Phi : \mathbb{R}^m \to \mathbb{R}^m$,

$$\Phi(x) = \left(f(x), x^{n+1}-a^{n+1}, \ldots, x^m - a^m\right).$$

Then $\Phi(a) = (b, 0)$,

$$\Phi'(a) = \begin{pmatrix} \partial_1 f^1 & \cdots & \partial_n f^1 & \partial_{n+1} f^1 & \cdots & \partial_m f^1 \\ \vdots & & \vdots & \vdots & & \vdots \\ \partial_1 f^n & \cdots & \partial_n f^n & \partial_{n+1} f^n & \cdots & \partial_m f^n \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix}_a$$

is invertible. By Theorem 3.31, there are $U \in \mathcal{N}_a^m$ and $V \in \mathcal{N}_{(b,0)}^m$ such that $\Phi : U \to V$ is diffeomorphism.

Hence, for some $\varepsilon > 0$ we have

$$B_\varepsilon^m(b, 0) \subset V = \Phi(U) \subset \Phi(\Omega).$$

By the definition of Φ we see $B_\varepsilon^n(b) \subset f(\Omega)$. Indeed, if $y \in B_\varepsilon^n(b)$ then $(y, 0) \in B_\varepsilon^m(b, 0)$, so there is $x \in \Omega$ such that

$$(y, 0) = \Phi(x) = \left(f(x), x^{n+1} - a^{n+1}, \dots, x^m - a^m\right),$$

That is $y = f(x) \in f(\Omega)$.

Remark 3.39. As we have seen, for C^1-map $f : \Omega \to \mathbb{R}^n$ and $a \in \Omega$,

(1) if $f'(a) : \mathbb{R}^m \to \mathbb{R}^n$ is invertible, then f is locally invertible (Theorem 3.31);
(2) if $f'(a) : \mathbb{R}^m \to \mathbb{R}^n$ is surjective, then f is locally surjective (Theorem 3.36);
(3) if $f'(a) : \mathbb{R}^m \to \mathbb{R}^n$ is injective, then f is locally injective (please write down the precise statement and prove it. This is an extra credit problem).

That is, f locally inherits the properties of the linear map $f'(a)$, which is much easy to study. That is why the derivative $f'(a)$ is so important. All these results (and the implicit function theorem in the next section) are corollaries of the inverse function theorem. This justifies to say that the inverse function theorem is *the fundamental theorem of differential calculus*.

se

Example 3.40. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be C^1, $\det f'(x) \neq 0$ for all $x \in \mathbb{R}^n$. If

$$\lim_{|x| \to \infty} |f(x)| = +\infty, \tag{3.15}$$ ew

then $f(\mathbb{R}^n) = \mathbb{R}^n$.

Remark 3.41. (1) This means for $\forall b \in \mathbb{R}^n$, then nonlinear algebraic equation $f(x) = b$ has a solution. (2) Actually f is also injective, thus it is a diffeomorphism; see Katriel (1994) for a proof via Mountain Pass Theorem Ambrosetti & Rabinowitz (1973).

Proof. From (3.15) we know that $f(\mathbb{R}^n)$ is closed. From Remark 3.38 we known that $f(\mathbb{R}^n)$ is open. Using Example 1.81 we deduce $f(\mathbb{R}^n) = \mathbb{R}^n$.

Proof. Given $b \in \mathbb{R}^n$, the function $\varphi : \mathbb{R}^n \to \mathbb{R}$ given by

$$\varphi(x) = \frac{1}{2}|f(x) - b|^2$$

attains its minmum at some $\xi \in \mathbb{R}^n$. Since $f'(\xi)$ is invertible, $f(\xi) = b$ follows from

$$0 = \nabla\varphi(\xi) = (f(\xi) - b)^{\mathrm{T}} f'(\xi).$$

Proposition 3.42 (Liu & Liu (2018))**.** *Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be C^1-map with $n \geq 2$,* rank $f'(x) < n$ *for at most finitely many $x \in \mathbb{R}^m$. If $f(\mathbb{R}^m)$ is closed, then $f(\mathbb{R}^m) = \mathbb{R}^n$.*

Remark 3.43. Easy example shows that the result is not true if $n = 1$.

This is a generalization of Example 3.40. Using this proposition we deduce the fundamental theorem of algebra, see Liu & Liu (2018) for the details.

3.5. Implicit function theorem. Let U and V be open subset of $\mathbb{R}^m$ and $\mathbb{R}^n$, $F : U \times V \to \mathbb{R}^p$, $(a, b) \in U \times V$. Then we have a map $F_2 : V \to \mathbb{R}^p$, $y \mapsto F(a, y)$. We define

$$\partial_y F(a, b) = F_2'(b).$$

Similarly we define $\partial_x F(a, b)$. Then $\partial_x F$ and $\partial_y F$ are linear maps from $\mathbb{R}^m$ and $\mathbb{R}^n$ to $\mathbb{R}^p$ respectively, with the matrices

$$\partial_x F(a, b) = \begin{pmatrix} \partial_{x^1} F^1 & \cdots & \partial_{x^m} F^1 \\ \vdots & & \vdots \\ \partial_{x^1} F^p & \cdots & \partial_{x^m} F^p \end{pmatrix},$$

$$\partial_y F(a, b) = \begin{pmatrix} \partial_{y^1} F^1 & \cdots & \partial_{y^n} F^1 \\ \vdots & & \vdots \\ \partial_{y^1} F^p & \cdots & \partial_{y^n} F^p \end{pmatrix}.$$

Proposition 3.44. *Suppose $F : U \times V \to \mathbb{R}^p$, $(a, b) \in U \times V$.*

(1) *If F is differentiable at (a, b), then $F_1 : x \mapsto F(x, b)$ is differentiable at a, $F_2 : y \mapsto F(a, y)$ is differentiable at b, and we have*

$$F'(a, b)(h, k) = \partial_x F(a, b)h + \partial_y F(a, b)k, \quad (h, k) \in \mathbb{R}^m \times \mathbb{R}^n. \tag{3.16}$$ par

(2) *If $\partial_x F$ and $\partial_y F$ are continuous at (a, b), then F is differentiable at (a, b) and we have (3.16).*

By considering the components of F, the proof is easy. Note that if we consider $F'(a, b)$, $\partial_x F(a, b)$ and $\partial_y F(a, b)$ as matrices, (3.16) should be written as

$$F'(a, b) \begin{pmatrix} h \\ k \end{pmatrix} = \partial_x F(a, b)h + \partial_y F(a, b)k$$

and we have the block decomposition $F'(a, b) = \big(\partial_x F(a, b), \partial_y F(a, b)\big)$.

tif

Theorem 3.45 (Implicit function theorem)**.** *Let U and V be open sets in $\mathbb{R}^m$ and $\mathbb{R}^n$, $F \in C^1(U \times V, \mathbb{R}^n)$, $(a, b) \in U \times V$. If*

$$F(a, b) = 0, \qquad \det\big[\partial_y F(a, b)\big] \neq 0,$$

then there are $r > 0$ and a C^1-map $\varphi : B_r^m(a) \to V$ such that $B_r^m(a) \subset U$ and

(1) *$\varphi(a) = b$,*
(2) *for $\forall x \in B_r^m(a)$ we have $F(x, \varphi(x)) = 0$.*
(3) *if $(x, y) \in B_r^m(a) \times B_r^n(b)$ is such that $F(x, y) = 0$, then $y = \varphi(x)$.*

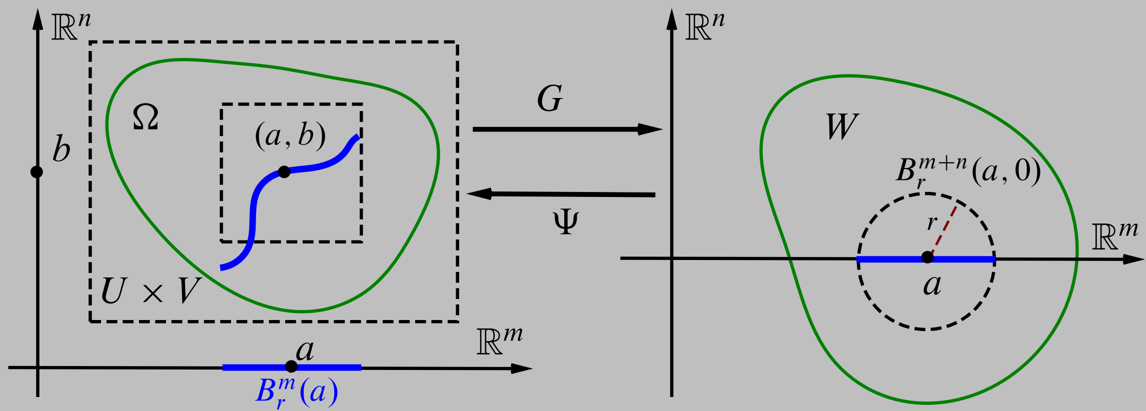
Remark 3.46. Because of (2), we call φ the *implicite function* defined by $F(x, y) = 0$.

Proof. Define $G : U \times V \to \mathbb{R}^m \times \mathbb{R}^n$, $G(x, y) = (x, F(x, y))$. Then $G \in C^1$,

$$G'(a,b) = \begin{pmatrix} I_m & 0 \\ \partial_x F(a,b) & \partial_y F(a,b) \end{pmatrix}.$$

Obviously $\det G'(a,b) \neq 0$, $G(a,b) = (a,0)$. By inverse function theorem, there are $\Omega \in \mathcal{N}^{m+n}_{(a,b)}$ and $W \in \mathcal{N}^{m+n}_{(a,0)}$ such that $G : \Omega \to W$ is diffeomorphism. Let $\Psi : W \to \Omega$ be the local inverse of G. From the definition of G, for $(x, z) \in W$ we have $\Psi^1(x, z) = x$, thus

$$\Psi(x, z) = (x, \Psi^2(x, z)).$$



Take $r > 0$ such that $B_r^{m+n}(a,0) \subset W$. Then for $x \in B_r^m(a)$ we have $(x,0) \in W$. This enables us to define a C^1-map $\varphi : B_r^m(a) \to \mathbb{R}^n$ by

$$\varphi(x) = \Psi^2(x,0),$$

with $\varphi(a) = \Psi^2(a,0) = b$. For $x \in B_r^m(a)$ we have $F(x, \varphi(x)) = 0$, because

$$\begin{aligned}(x, F(x,\varphi(x))) &= G(x,\varphi(x)) = G(x, \Psi^2(x,0)) \\ &= G(\Psi(x,0)) = (G \circ \Psi)(x,0) = (x,0).\end{aligned}$$

By decreasing r we may assume that $B_r^m(a) \times B_r^n(b) \subset \Omega$. If $(x, y) \in B_r^m(a) \times B_r^n(b)$ is such that $F(x, y) = 0$, then

$$G(x, y) = (x,0) = G(\Psi(x,0)) = G(x, \Psi^2(x,0)) = G(x, \varphi(x)).$$

From this we get $y = \varphi(x)$, because G is injective in Ω, $(x, y) \in \Omega$ and

$$(x, \varphi(x)) = \big(x, \Psi^2(x,0)\big) = \Psi(x,0) \in \Omega.$$

How to compute the derivative of $y = \varphi(x)$? Let $\Phi : x \mapsto F(x, \varphi(x))$, it is the composition of $g : x \mapsto (x, \varphi(x))$ and F. Since for $\forall x \in O$ we have $\Phi(x) = 0$, we deduce

$$\begin{aligned} 0 = \Phi'(x) &= F'(x,\varphi(x))g'(x) \\ &= \big(\partial_x F(x,\varphi(x)), \partial_y F(x,\varphi(x))\big) \begin{pmatrix} I_m \\ \varphi'(x) \end{pmatrix} \\ &= \partial_x F(x,\varphi(x)) + \partial_y F(x,\varphi(x))\varphi'(x), \end{aligned}$$

Note that

$$\partial_y F(a,b) = \partial_y F(a, \varphi(a))$$

is invertible, by continuity, for smaller O we may assume that $\partial_y F(x,\varphi(x))$ is invertible for $x \in O$. For such x, multiplying $\left[\partial_y F(x,\varphi(x))\right]^{-1}$ to both sides of the above equality we get

$$\begin{aligned}\varphi'(x) &= -\left[\partial_y F(x,\varphi(x))\right]^{-1} \partial_x F(x,\varphi(x)) \\ &= -\left[\partial_y F(x,y)\right]^{-1} \partial_x F(x,y).\end{aligned} \tag{3.17}$$

et

In practical computation, we take derivative with respect to x^k on both sides of

$$F^i(x^1,\dots,x^m,y^1,\dots,y^n) = 0, \qquad i = 1,\dots,n$$

to get

$$\frac{\partial F^i}{\partial x^k} + \sum_{j=1}^n \frac{\partial F^i}{\partial y^j}\frac{\partial y^j}{\partial x^k} = 0, \qquad i = 1,\dots,n,$$

then solve for $\partial y^j/\partial x^k$ using Cramer rule.

Example 3.47. Where does the equation

$$-3 + x^2 + 2ye^x + z + e^{x^2y^2z} = 0 \tag{3.18}$$

eg

define a function $z = g(x,y)$ implicitly? Compute $\partial_x g(0,1)$.

Proof. Denote the left hand side by $F(x,y,z)$. Since

$$\partial_z F = 1 + e^{x^2y^2z}\partial_z\left(x^2y^2z\right) = 1 + x^2y^2e^{x^2y^2z} > 0,$$

by Theorem 3.45 the equation *locally* defines a function $z = g(x,y)$ near every point $(x,y,z) \in F^{-1}(0)$. Actually g is defined globally because given $(x,y) \in \mathbb{R}^2$ there is a unique $z \in \mathbb{R}$ such that $F(x,y,z) = 0$.

To compute $\partial_x g(0,1)$, differentiating (3.18) having in mind that z is function of (x,y), we get

$$\begin{aligned}0 &= 2x + 2ye^x + z_x + e^{x^2y^2z}\partial_x\left(x^2y^2z\right) \\ &= 2x + 2ye^x + z_x + e^{x^2y^2z}y^2\left(2xz + x^2z_x\right), \\ z_x &= -\frac{2x + 2ye^x + 2xzy^2e^{x^2y^2z}}{1 + x^2y^2e^{x^2y^2z}}.\end{aligned}$$

From (3.18) we see that when $(x,y) = (0,1)$ we have $z = 0$. Hence

$$\partial_x g(0,1) = \left[-\frac{2x + 2ye^x + 2xzy^2e^{x^2y^2z}}{1 + x^2y^2e^{x^2y^2z}}\right]_{(0,1,0)} = -2.$$

Example 3.48. Consider $F = (F_1, F_2) : \mathbb{R}^4 \to \mathbb{R}^2$,

$$\begin{aligned}F^1(x,y,u,v) &= xv + yu^3 + u^4, \\ F^2(x,y,u,v) &= xy + u + v^3 + v.\end{aligned}$$

The point $P\ (1,1,-1,0)$ is a solution of the system

$$\begin{cases} F^1(x,y,z,u,v) = 0, \\ F^2(x,y,z,u,v) = 0. \end{cases} \tag{3.19}$$

eF

We have

$$
\begin{aligned}
F'(P) &= \begin{pmatrix} \partial_x F^1 & \partial_y F^1 & \partial_u F^1 & \partial_v F^1 \\ \partial_x F^2 & \partial_y F^2 & \partial_u F^2 & \partial_v F^2 \end{pmatrix}_P \\
&= \begin{pmatrix} v & u^3 & 3u^2 y + 4u^3 & x \\ y & x & 1 & 1+3v^2 \end{pmatrix}_P \\
&= \begin{pmatrix} 0 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.
\end{aligned}
$$

Since

$$\det \partial_{(u,v)} F(P) = \det \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \neq 0,$$

we see that near P the system (3.19) determines a map

$$\varphi = (f, g) : (x, y) \mapsto (u, v), \qquad \varphi(1,1) = (-1, 0).$$

To find $\partial_x f$, having in mind that u and v are functions of (x, y), we differentiating (3.19) with respect to x:

$$\begin{cases} v + x v_x + 3yu^2 u_x + 4u^3 u_x = 0, \\ y + u_x + 3v^2 v_x + v_x = 0. \end{cases} \qquad \begin{cases} \left(3yu^2 + 4u^3\right) u_x + x v_x = -v, \\ u_x + \left(3v^2 + 1\right) v_x = -y. \end{cases}$$

From this we get

$$
\begin{aligned}
u_x &= \frac{1}{\det \begin{pmatrix} 3yu^2 + 4u^3 & x \\ 1 & 3v^2 + 1 \end{pmatrix}} \det \begin{pmatrix} -v & x \\ -y & 3v^2 + 1 \end{pmatrix} \\
&= \frac{xy - v\left(3v^2 + 1\right)}{(3yu^2 + 4u^3)(3v^2 + 1) - x}.
\end{aligned} \tag{3.20}
$$

Let's compute $\varphi'(1,1)$. We may compute u_y, v_x and v_y as above, then

$$\varphi'(1,1) = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}_{(1,1,-1,0)}.$$

Alternatively, we can also apply (3.17) to get

$$
\begin{aligned}
\varphi'(1,1) &= -\left[\partial_{(u,v)} F(1,1,-1,0)\right]^{-1} \partial_{(x,y)} F(1,1,-1,0) \\
&= -\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & -1 \\ -\frac{1}{2} & 0 \end{pmatrix}.
\end{aligned}
$$

In particular, $u_x(1,1) = -\frac{1}{2}$, coincides with the result given in (3.20).

Now we look back to surfaces in $\mathbb{R}^n$. For surface, we mean subset of $\mathbb{R}^n$ which is locally a graph $G_f = \{(z, \varphi(z))\}$ of smooth function $\varphi : \mathbb{R}^{n-1} \to \mathbb{R}$.

Example 3.49. Let $F : \mathbb{R}^n \to \mathbb{R}$ be a C^1-function such that $M = F^{-1}(0)$ is not empty[21], $\nabla F(x) \neq 0$ for $x \in M$. Consider $a \in M$, we may assume $\partial_n F(a) \neq 0$, then by implicit function theorem, from

$$F(x^1, \ldots, x^n) = 0$$

we may locally express x^n via $\left(x^1, \ldots, x^{n-1}\right)$,

$$x^n = \varphi(x^1, \ldots, x^{n-1}),$$

[21] We call $F^{-1}(c)$ the level set of F at c.

where φ is a C^1-function. Near the point a, $x \in M$ iff x lies on the graph of φ. Thus M is a surface.

Let $\gamma : (-\varepsilon, \varepsilon) \to M$ be a smooth curve on M, $\gamma(0) = a$. Then $F(\gamma(t)) = 0$ hence

$$0 = (F \circ \gamma)'(0) = \nabla F(a) \cdot \dot{\gamma}(0).$$

This means that $\nabla F(a)$ is orthogonal to curves on M passing a. Thus $\nabla F(a)$ is a normal vector of M at a.

Remark 3.50. The converse is also true: If $h \perp \nabla F(a)$, then $h = \dot{\gamma}(0)$ for some $\gamma : (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = a$. This can be prove via the implicit function theorem. An interesting proof via ODE can be found in (Thorpe, 1994, Chapter 3).

Example 3.51. Let U be open subset of $\mathbb{R}^{n-1}$, $x : U \to \mathbb{R}^n$ is a C^1-map. If for all $u \in U$,

$$\operatorname{rank} x'(u) = n - 1,$$

then $S = x(U)$ is a surface in $\mathbb{R}^n$.

For $a = x(\alpha) \in S$, where $\alpha \in U$, since

$$\operatorname{rank} x'(\alpha) = n - 1,$$

we may assume that

$$\left.\frac{\partial(x^1, \dots, x^{n-1})}{\partial(u^1, \dots, u^{n-1})}\right|_\alpha \neq 0.$$

By inverse function theorem, near $\left(a^1, \dots, a^{n-1}\right)$ and α, the map

$$\left(u^1, \dots, u^{n-1}\right) \mapsto \left(x^1, \dots, , x^{n-1}\right)$$

is invertible, that is, we can express u^i by $z = \left(x^1, \dots, , x^{n-1}\right)$,

$$u^i = u^i(z) = u^i(x^1, \dots, , x^{n-1}).$$

Consequently, near a, S is graph of the C^1-function

$$\begin{aligned} x^n &= x^n(u^1, \dots, u^{n-1}) \\ &= x^n(u^1(z), \dots, u^{n-1}(z)) \\ &= \varphi(z) = \varphi(x^1, \dots, , x^{n-1}). \end{aligned}$$

So S is a smooth surface. We also know that the normal vector of S at $a = x(\alpha)$ is

$$N = \left(\frac{\partial(x^2, \dots, x^n)}{\partial(u^1, \dots, u^{n-1})}, \dots, (-1)^{n+1} \frac{\partial(x^1, \dots, x^{n-1})}{\partial(u^1, \dots, u^{n-1})} \right)_\alpha .$$

4. Lebesgue measure and integrals

Let $f : [a, b] \to \mathbb{R}_+$ be integrable, then

$$I = \int_a^b f$$

is the area of the planar region bounded by the graph of f and x-axis. Thus integral is closely related to area, volume and their higher dimensional analogies, called measure.

4.1. Lebesgue measure. We will define a class $\mathcal{M}$ of measurable subsets on $\mathbb{R}^n$ and a measure function $m : \mathcal{M} \to [0, \infty]$, such that

(1) if $A \in \mathcal{M}$, then $A^c \in \mathcal{M}$, and $m(a + A) = m(A)$ for $a \in \mathbb{R}^n$;
(2) if A is open, then $A \in \mathcal{M}$ (thus true for A closed), $m(\emptyset) = 0$, $m([0, 1]^n) = 1$;
(3) if $\{A_k\}_{k=1}^\infty \subset \mathcal{M}$, then $\bigcup_{k=1}^\infty A_k \in \mathcal{M}$ and

$$m\left(\bigcup_{k=1}^\infty A_k\right) \le \sum_{k=1}^\infty m(A_k), \qquad \text{(sub-aditivity)}$$

"=" holds if A_i are disjoint (this is called *countable additivity*).

As a consequence we also have

- for $A, B \in \mathcal{M}$, $A \subset B$ implies $m(A) \le m(B)$.

We start with outer measure. Given $\Omega \subset \mathbb{R}^n$, a natural method to measure its size is to define the *outer measure* of Ω as

$$m^*(\Omega) = \inf\left\{ \sum_{k=1}^\infty |I_k| \,\middle|\, I_k \text{ are boxes in } \mathbb{R}^n \text{ such that } \Omega \subset \bigcup_{k=1}^\infty I_k \right\},$$

where for box $I = \prod_{i=1}^n (a_i, b_i)$, its volume is defined as

$$|I| = \prod_{i=1}^n (b_i - a_i).$$

By definition boxes I are open, their closure $\overline{I} = \prod_{i=1}^n [a_i, b_i]$ are called closed boxes.

Proposition 4.1. *The outer measure has the following properties:*

(1) $m^*(\emptyset) = 0$, *$A \subset B$ implies* $m^*(A) \le m^*(B)$;
(2) $m^*(a + A) = m^*(A)$;
(3) *for* $\{A_k\}_{k=1}^\infty \subset 2^{\mathbb{R}^n}$,

$$m^*\left(\bigcup_{k=1}^\infty A_k\right) \le \sum_{k=1}^\infty m^*(A_k).$$

Proof. Given $\varepsilon > 0$, there are boxes $\{I_k^\ell\}$ such that for all ℓ,

$$\sum_{\ell=1}^\infty |I_k^\ell| < m^*(A_k) + \frac{\varepsilon}{2^k}.$$

Since the boxes $\{I_k^\ell\}$ form a cover of $\bigcup_{k=1}^\infty A_k$, by definition of m^* we have

$$\begin{aligned} m^*\left(\bigcup_{k=1}^\infty A_k\right) &\le \sum_{k=1}^\infty \sum_{\ell=1}^\infty |I_k^\ell| \le \sum_{k=1}^\infty \left(m^*(A_k) + \frac{\varepsilon}{2^k}\right) \\ &= \sum_{k=1}^\infty m^*(A_k) + \sum_{k=1}^\infty \frac{\varepsilon}{2^k} = \sum_{k=1}^\infty m^*(A_k) + \varepsilon. \end{aligned}$$

Letting $\varepsilon \to 0$ ends the proof.

Proposition 4.2. *For* $I = \prod_{i=1}^n (a_i, b_i)$, $m^*(\overline{I}) = |I|$. *Thus,* $m^*([0, 1]^n) = 1$.

Proof. Given $\varepsilon > 0$, $\overline{I}$ is covered by $\prod_{i=1}^{n}(a_i - \varepsilon, b_i + \varepsilon)$, so

$$m^*(\overline{I}) \le \left|\prod_{i=1}^{n}(a_i - \varepsilon, b_i + \varepsilon)\right| = \prod_{i=1}^{n}(b_i - a_i + 2\varepsilon) \to \prod_{i=1}^{n}(b_i - a_i) = |I|$$

as $\varepsilon \to 0$, thus $m^*(\overline{I}) \le |I|$. Let $\{I_k\}$ be a box-cover of $\overline{I}$ such that

$$\sum_{k=1}^{\infty}|I_k| \le m^*(\overline{I}) + \varepsilon$$

since $\overline{I}$ is compact, $I \subset \overline{I} \subset \bigcup_{k=1}^{\ell} I_k$ for some ℓ. Thus

$$|I| \le \sum_{k=1}^{\ell}|I_k| \le \sum_{k=1}^{\infty}|I_k| \le m^*(\overline{I}) + \varepsilon.$$

Let $\varepsilon \to 0$ we get $|I| \le m^*(\overline{I})$.

Corollary 4.3. *For a box I, $m^*(I) = |I|$.*

Example 4.4. Since $\mathbb{Q} = \{q_k\}_{k=1}^{\infty}$ and $m^*(\{q\}) = 0$,

$$m^*(\mathbb{Q}) \le \sum_{k=1}^{\infty} m^*(q_k) = \sum_{k=0}^{\infty} 0 = 0,$$

$m^*([0,1]\setminus\mathbb{Q}) = 1$ because

$$\begin{aligned} 1 = m^*([0,1]) &\le m^*([0,1]\cap\mathbb{Q}) + m^*([0,1]\setminus\mathbb{Q}) \\ &= m^*([0,1]\setminus\mathbb{Q}) \le m^*([0,1]) = 1. \end{aligned}$$

If $A \cap B = \emptyset$, we expect

$$m^*(A \cup B) = m^*(A) + m^*(B). \tag{4.1}$$ e90

Unfortunately, this is not true, althought (4.1) is true if

$$\operatorname{dist}(A, B) = \inf_{x\in A, y\in B}|x - y| > 0.$$

To have (countable) additivity, we have to restrict to a subclass $\mathcal{M} \subset 2^{\mathbb{R}^n}$ called measurable sets.

Definition 4.5 (Carathéodory)**.** A subset $E \subset \mathbb{R}^n$ is measurable, if

$$m^*(T) \ge m^*(T \cap E) + m^*(T \setminus E) \quad \text{for all } T \subset \mathbb{R}^n, \tag{4.2}$$ e91

we then call $m(E) = m^*(E)$ the (Lebesgue) measure of E. The calss of measurable sets is denoted by $\mathcal{M}$.

rk1

Remark 4.6. (4.2) is acturally an equality because "$\le$" is autamatically true. If $E_1 \in \mathcal{M}$, $E_1 \cap E_2 = \emptyset$, testing $T \cap (E_1 \cup E_2)$ via the measurability of E_1 we get

$$m^*(T \cap (E_1 \cup E_2)) = m^*(T \cap E_1) + m^*(T \cap E_2).$$

Using mathematical induction and Proposition 4.9 (3), if $\{E_k\}_{k=1}^{m} \in \mathcal{M}$ are disjoint then

$$m^*\left(T \cap \bigcup_{k=1}^{m} E_k\right) = \sum_{k=1}^{m} m^*(T \cap E_k).$$

Remark 4.7. Given $E \subset \mathbb{R}^n$, if

$$m^*(I) \geq m^*(I \cap E) + m^*(I \setminus E) \tag{4.3}$$

for all box I, then (4.2) holds and $E \in \mathcal{M}$.

To see this, let $\varepsilon > 0$ and take boxes $\{I_k\}$ covering T such that

$$\begin{aligned}
\varepsilon + m^*(T) &\geq \sum_{k=1}^{\infty} |I_k| = \sum_{k=1}^{\infty} m^*(I_k) \\
&\geq \sum_{k=1}^{\infty} [m^*(I_k \cap E) + m^*(I_k \cap E^{\mathrm{c}})] \\
&\geq m^*\left(\left(\bigcup_{k=1}^{\infty} I_k\right) \cap E\right) + m^*\left(\left(\bigcup_{k=1}^{\infty} I_k\right) \cap E^{\mathrm{c}}\right) \\
&\geq m^*(T \cap E) + m^*(T \cap E^{\mathrm{c}}).
\end{aligned}$$

Letting $\varepsilon \to 0$ gives (4.2).

Proposition 4.8. *Half space $H = \{x_n > 0\}$ is measurable.*

Proof. Given a box I, if $I \cap H = \emptyset$, then $I \setminus H = \emptyset$ and (4.3) holds. If $I \cap H \neq \emptyset$ then both $I_1 = I \cap H$ and $I_2 = I \setminus H$ are boxes (I_2 may be empty), and $I = I_1 \cup I_2$, we get

$$\begin{aligned}
m^*(I) &= |I| = |I_1| + |I_2| \\
&= m^*(I_1) + m^*(I_2) = m^*(I \cap H) + m^*(I \setminus H).
\end{aligned}$$

Proposition 4.9. *Properties of measurable sets.*

(1) *$E \in \mathcal{M}$ implies $E^{\mathrm{c}} \in \mathcal{M}$.*
(2) *$E \in \mathcal{M}$ if $m^*(E) = 0$.*
(3) *$E_1, E_2 \in \mathcal{M}$, then $E_1 \cup E_2 \in \mathcal{M}$.*
(4) *$\{E_k\}_{k=1}^{\infty} \subset \mathcal{M}$ implies $\bigcup_{k=1}^{\infty} E_k \in \mathcal{M}$ and $\bigcap_{k=1}^{\infty} E_k \in \mathcal{M}$.*
(5) *if $\{E_k\}_{k=1}^{\infty} \subset \mathcal{M}$ are disjoint, then for $T \subset \mathbb{R}^n$,*

$$m^*\left(T \cap \bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} m^*(T \cap E_k). \tag{4.4}$$

In particular, take $T = \mathbb{R}^n$ we get

$$m\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} m(E_k).$$

Proof. (1) is clear. If $m^*(E) = 0$ then $m^*(T \cap E) = 0$ and (4.2) follows, thus (2) is true.

(3) Given $T \subset \mathbb{R}^n$, using the measurability of E_1 to test T and then using that of E_2 to test $T \cap E_1$ and $T \setminus E_1$, we get

$$\begin{aligned}
m^*(T) &\geq m^*(T \cap E_1) + m^*(T \setminus E_1) \\
&\geq m^*((T \cap E_1) \cap E_2) + m^*((T \cap E_1) \setminus E_2) + m^*((T \setminus E_1) \cap E_2) \qquad (4.5) \\
&\qquad + m^*((T \setminus E_1) \setminus E_2) \\
&\geq m^*(T \cap (E_1 \cup E_2)) + m^*(T \setminus (E_1 \cup E_2)).
\end{aligned}$$

Note that the union of the three sets in (4.5) is precisely $T \cap (E_1 \cup E_2)$, and we have used the sub-aditivity of m^* in the last step.

(4) Firstly we assume that $\{E_k\}$ are disjoint. Set

$$S = \bigcup_{k=1}^{\infty} E_k, \qquad S_m = \bigcup_{k=1}^{m} E_k.$$

Then $S_m \in \mathcal{M}$, thus for $T \subset \mathbb{R}^n$ we have (see Remark 4.6)

$$\begin{aligned} m^*(T) &= m^*(T \cap S_m) + m^*(T \backslash S_m) \\ &= \sum_{k=1}^{m} m^*(T \cap E_k) + m^*(T \backslash S_m) \\ &\geq \sum_{k=1}^{m} m^*(T \cap E_k) + m^*(T \backslash S). \end{aligned}$$

Let $m \to \infty$ we get

$$\begin{aligned} m^*(T) &\geq \sum_{k=1}^{\infty} m^*(T \cap E_k) + m^*(T \backslash S) \qquad (4.6) \\ &\geq m^*(T \cap S) + m^*(T \backslash S). \end{aligned}$$

e20

So $S \in \mathcal{M}$. Replacing T by $T \cap S$ in (4.6) we get (4.4).

For the general case that $\{E_k\}$ are not disjoint, we set

$$E^1 = E_1, \qquad E^k = E_k \backslash \bigcup_{j=1}^{k-1} E_k.$$

Then $\{E^k\}$ are disjoint and $\bigcup_{k=1}^{\infty} E_k = \bigcup_{k=1}^{\infty} E^k$ are measurable.

Corollary 4.10. *If $E, F \in \mathcal{M}$, $E \subset F$, $m(E) < \infty$, then $F \backslash E \in \mathcal{M}$ and*

$$m(F \backslash E) = m(F) - m(E).$$

cl

Corollary 4.11. *If $E_k \in \mathcal{M}$, $E_1 \subset E_2 \subset \cdots$, then*

$$m\left(\bigcup_{k=1}^{\infty} E_k\right) = \lim_{k \to \infty} m(E_k).$$

Proof. If $m(E_\ell) = \infty$ for some $\ell \in \mathbb{N}$, both sides are ∞ and the result is true. Thus we assume $m(E_k) < \infty$ for all k. Let $E^0 = \emptyset$, $E^k = E_k \backslash E_{k-1}$. Then

$$\bigcup_{k=1}^{\infty} E_k = \bigcup_{k=1}^{\infty} E^k, \qquad m(E^k) = m(E_k) - m(E_{k-1}).$$

Since E^k are disjoint,

$$\begin{aligned} m\left(\bigcup_{k=1}^{\infty} E_k\right) &= m\left(\bigcup_{k=1}^{\infty} E^k\right) \\ &= \sum_{k=1}^{\infty} m(E^k) = \lim_{N \to \infty} \sum_{k=1}^{N} (m(E_k) - m(E_{k-1})) \end{aligned}$$

$$= \lim_{N\to\infty} m(E_N).$$

Corollary 4.12. *If I is a box, then $I \in \mathcal{M}$. If E is open (or closed), then $E \in \mathcal{M}$. All Broel sets are measurable.*

Proof. Boxes are finite intersection of half spaces, and open sets are countable union of boxes (see the lemma below).

Lemma 4.13. *Let Ω be an open set in $\mathbb{R}^n$, then Ω is a countable union of boxes.*

Proof. For $a \in \Omega$, there is a box

$$I_r(\tilde{a}) = \prod_{i=1}^{n} \left(\tilde{a}^i - r, \tilde{a}^i + r\right)$$

with $r \in \mathbb{Q}$ and $\tilde{a} \in \mathbb{Q}^n$ such that

$$a \in I_r(\tilde{a}) \subset \Omega. \tag{4.7}$$

Let $\mathcal{I}$ be the collection of all these boxes, then $\mathcal{I}$ is countable, and $\Omega = \bigcup_{I\in\mathcal{I}} I$.

The box $I_r(\tilde{a})$ in (4.7) can be chosen as follow. Take $\delta > 0$ such that $B_\delta(a) \subset \Omega$, then take $r \in \mathbb{Q}$ and $\tilde{a} \in \mathbb{Q}^n$ such that

$$0 < r < \frac{\delta}{2\sqrt{n}}, \qquad |\tilde{a}^i - a^i| < r.$$

Then clearly $a \in I_r(\tilde{a})$. If $y \in I_r(\tilde{a})$ then $|y^i - \tilde{a}^i| < r$, hence

$$\begin{aligned} |y - a| &\le |y - \tilde{a}| + |\tilde{a} - a| \\ &= \sqrt{\sum_{i=1}^{n} |y^i - \tilde{a}^i|^2} + \sqrt{\sum_{i=1}^{n} |\tilde{a}^i - a^i|^2} \\ &< \sqrt{nr^2} + \sqrt{nr^2} = 2\sqrt{n}r < \delta. \end{aligned}$$

We see that $y \in B_r(a)$, hence $I_r(\tilde{a}) \subset \Omega$.

4.2. Measurable functions. Let $\Omega \in \mathcal{M}$, $f : \Omega \to \mathbb{R}^\ell$ is measurable if $f^{-1}(V) \in \mathcal{M}$ for all open set $V \subset \mathbb{R}^\ell$. We use $\mathcal{M}(\Omega, \mathbb{R}^\ell)$ to denote the set of such f, and denote $\mathcal{M}(\Omega) = \mathcal{M}(\Omega, \mathbb{R})$.

Remark 4.14. Since open sets are countable union of boxes, for $f \in \mathcal{M}(\Omega, \mathbb{R}^\ell)$, it suffices to require $f^{-1}(I) \in \mathcal{M}$ for every box $I \subset \mathbb{R}^\ell$.

Lemma 4.15. *Let $\Omega \in \mathcal{M}$, $f : \Omega \to \mathbb{R}^\ell$ be continuous, then $f \in \mathcal{M}(\Omega, \mathbb{R}^\ell)$.*

Proof. For $V \subset \mathbb{R}^\ell$ open, $f^{-1}(V)$ is Ω-open. Thus

$$f^{-1}(V) = U \cap \Omega$$

for some open set $U \subset \mathbb{R}^n$. It follows that $f^{-1}(V) \in \mathcal{M}$.

Lemma 4.16. *If $f : \Omega \to W$ is measurable, $g : W \to \mathbb{R}^k$ is continuous, then $g \circ f \in \mathcal{M}(\Omega)$.*

Proof. For open $V \subset \mathbb{R}^k$, $g^{-1}(V)$ is W-open thus

$$g^{-1}(V) = W \cap U$$

for some open $U \subset \mathbb{R}^\ell$. Consequently

$$(g \circ f)^{-1}(V) = f^{-1}(g^{-1}(V)) = f^{-1}(W \cap U) = f^{-1}(U) \in \mathcal{M}.$$

c2

Corollary 4.17. *$f = (f_1, \dots, f_\ell) \in \mathcal{M}(\Omega, \mathbb{R}^\ell)$ iff $f_i \in \mathcal{M}(\Omega)$ for all $i \in \overline{\ell}$.*

Proof. ($\Rightarrow$) Let $\pi_i : \mathbb{R}^\ell \to \mathbb{R}$ be the projection, then π_i is continuous and

$$f_i = \pi_i \circ f \in \mathcal{M}(\Omega).$$

($\Leftarrow$) For box $I = \prod_{i=1}^{\ell}(a^i, b^i)$, $f_i^{-1}(a^i, b^i) \in \mathcal{M}$ for all $i \in \overline{\ell}$. Thus

$$f^{-1}(I) = \bigcap_{i=1}^{\ell} f_i^{-1}(a^i, b^i) \in \mathcal{M}.$$

cc

Corollary 4.18. *If $f, g \in \mathcal{M}(\Omega)$, then $f \pm g$, fg, $\max\{f, g\}$, $\min\{f, g\}$ are all measurable, and $f/g \in \mathcal{M}(\Omega^*)$ with $\Omega^* = \Omega \backslash g^{-1}(0)$.*

Proof. Let $\varphi : \Omega^* \to \mathbb{R}^2$, $\psi : \mathbb{R} \times \mathbb{R} \backslash 0 \to \mathbb{R}$ be given by

$$\varphi(x) = (f(x), g(x)), \qquad \psi(u, v) = u/v.$$

Then φ is measurable, ψ is continuous. Thus $f/g = \psi \circ \varphi \in \mathcal{M}(\Omega^*)$.

Remark 4.19. Note that $\Omega^* \in \mathcal{M}$ (prove this!), thus it makes sense to talk about measurable functions on Ω^*.

Corollary 4.20. *If $f \in \mathcal{M}(\Omega)$, then $|f|$ and $f^\pm = \dfrac{|f| \pm f}{2} \in \mathcal{M}(\Omega)$.*

Proof. Because $g : u \mapsto |u|$ is continuous, it follows $|f| = g \circ f \in \mathcal{M}(\Omega)$.

Because of Corollary 4.17, we may focus on scalar functions $f : \Omega \to \mathbb{R}$.

ol

Lemma 4.21. *Let $f : \Omega \to \mathbb{R}$, then $f \in \mathcal{M}(\Omega)$ iff $\{f > c\} \in \mathcal{M}$ for all $c \in \mathbb{R}$.*

Proof. ($\Rightarrow$) This follows from

$$\{f > c\} = \bigcup_{i=0}^{\infty} f^{-1}(c + i, c + i + 2).$$

($\Leftarrow$) We havc $\{f \le c\} \in \mathcal{M}$, hence $\{f < c\} \in \mathcal{M}$ becuase

$$\{f < c\} = \bigcup_{k=1}^{\infty} \left\{ f \le c - \frac{1}{k} \right\}.$$

Now, for a box (α, β) in $\mathbb{R}$, we have $\{f > \alpha\} \in \mathcal{M}$, $\{f < \beta\} \in \mathcal{M}$. hence

$$f^{-1}(\alpha, \beta) = \{f > \alpha\} \cap \{f < \beta\} \in \mathcal{M}.$$

For scalar functions we may allow them to take values in $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$. This enable us to write $\sup_k f_k$ (which is a function $\sup_k f_k : \Omega \to \overline{\mathbb{R}}$ whose value at x is $\sup_k f_k(x)$) for a sequence of functions $f_k : \Omega \to \mathbb{R}$ (otherwise if $\{f_k(x)\}_{k=1}^{\infty}$ is unbounded then $(\sup_k f_k)(x)$ makes no sense).

Motivated by Lemma 4.21, we say that $f : \Omega \to \overline{\mathbb{R}}$ is measurable, if $\{f > c\} \in \mathcal{M}$ for all $c \in \mathbb{R}$. When f is $\mathbb{R}$-valued, this coincides with the prevision definition. We still use $\mathcal{M}(\Omega)$ to denote the set of $\overline{\mathbb{R}}$-valued measurable functions.

Example 4.22. If $f : \Omega \to \overline{\mathbb{R}}$ is measurable, then the $\mathbb{R}$-valued function $f_* : \Omega_* \to \mathbb{R}$ is measurable, where $\Omega_* = \{x \in \Omega \mid |f(x)| < \infty\}$.

Proof. Because

$$\begin{aligned}\Omega_* &= \Omega \backslash (\{|f| = \infty\}) \\ &= \Omega \backslash \left(\bigcap_{k=1}^{\infty} (\{f > k\} \cup \{f < -k\}) \right)\end{aligned}$$

we see that $\Omega_* \in \mathcal{M}$. Thus, given $c \in \mathbb{R}$,

$$\Omega_*(f_* > c) = \Omega_* \cap \Omega(f > c)$$

is measurable. Hence f_* is measurable.

xpq

Proposition 4.23. *If $\{f_k\}_{k=1}^{\infty} \subset \mathcal{M}(\Omega)$, then $\sup_{k \geq 1} f_k$, $\inf_{k \geq 1} f_k$, $\overline{\lim} f_k$, $\underline{\lim} f_k$ are all measurable on Ω. In particular, if $f_k \to f$ pointwise on Ω, then f is also measurable.*

Proof. Given $c \in \mathbb{R}$, $\{f_k > c\}$ is measurable. Thus

$$\{\sup f_k > c\} = \bigcup_{k=1}^{\infty} \{f_k > c\}$$

is measurable. We deduce $\sup_k f_k \in \mathcal{M}(\Omega)$. Similarry $\inf_k f_k \in \mathcal{M}(\Omega)$. Consequently,

$$\overline{\lim_{k \to \infty}} f_k = \inf_{m \geq 1} \sup_{k \geq m} f_k$$

is also measurable. If exists, $\lim f_k = \overline{\lim} f_k$ is also measurable.

Now we generalize Corollary 4.18 to $\overline{\mathbb{R}}$-valued functions.

Corollary 4.24. *If $f, g \in \mathcal{M}(\Omega)$, then $f \pm g$, fg, $\max\{f, g\}$, $\min\{f, g\}$ are all measurable, and $f/g \in \mathcal{M}(\Omega^*)$ with $\Omega^* = \Omega \backslash g^{-1}(0)$.*

Proof. For $k \in \mathbb{N}$ we define $f_k, g_k : \Omega \to \mathbb{R}$ via

$$f_k(x) = \begin{cases} f(x) & \text{if } |f(x)| \leq k, \\ k & \text{if } f(x) > k, \\ -k & \text{if } f(x) < -k, \end{cases} \qquad g_k(x) = \begin{cases} g(x) & \text{if } |g(x)| \leq k, \\ k & \text{if } g(x) > k, \\ -k & \text{if } g(x) < -k. \end{cases}$$

Then f_k and g_k are measurable $\mathbb{R}$-valued functions[(22)]. By Corollary 4.18, $f_k g_k \in \mathcal{M}(\Omega)$. Since $f_k g_k \to fg$, Proposition 4.23 yields $fg \in \mathcal{M}(\Omega)$.

Let $P(x)$ is a statement involving $x \in \Omega$. We say that $P(x)$ holds for almost every $x \in \Omega$ (a.e. $x \in \Omega$ for short), if $P(x)$ is true for all $x \in \Omega \backslash e$ for some $e \subset \Omega$ with $m(e) = 0$. For example, let D be the Dirichlet function, then $D(x) = a$ a.e. $x \in \Omega$.

[(22)] Given $c \in \mathbb{R}$, we have

$$\{f_k > c\} = \begin{cases} \emptyset & \text{if } c \geq k, \\ \{f > c\} & \text{if } c \in [-k, k), \\ \Omega & \text{if } c < -k. \end{cases}$$

4.3. Lebesgue integration for nonnegative functions. The indicator function of a subset $A \subset \mathbb{R}^n$ is $\chi^A : \mathbb{R}^n \to \mathbb{R}$ defined by

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

Given $\Omega \in \mathcal{M}$, the fuction $f : \Omega \to \mathbb{R}$ given by

$$f = \sum_{i=1}^{\ell} c_i \chi^{E_i}$$

is called simple function, where $\Omega = \bigcup_{i=1}^{\ell} E_i$ with $E_i \in \mathcal{M}$ disjoint, and $\{c_i\}_{i=1}^{\ell} \subset \mathbb{R}$. Obviously $f \in \mathcal{M}(\Omega)$.

The integral of the above simple function is

$$\int_\Omega f = \sum_{i=1}^{\ell} c_i m(E_i). \tag{4.8}$$ uo

Example 4.25. The Dirichlet function D is simple and we have $\int_{\mathbb{R}} D = 0$. ta

Lemma 4.26. *If $f, g : \Omega \to \mathbb{R}$ are simple, then*

(1) *$f + g$ and cf ($c \in \mathbb{R}$) are also simple, and*

$$\int_\Omega (f+g) = \int_\Omega f + \int_\Omega g, \qquad \int_\Omega cf = c \int_\Omega f.$$

(2) *if $f \le g$ then $\int_\Omega f \le \int_\Omega g$ (exercise).*

Proof. Assume

$$f = \sum_{i=1}^{\ell} c_i \chi^{E_i}, \qquad g = \sum_{j=1}^{k} d_i \chi^{F_j}.$$

Then Ω has disjoint partitions

$$\Omega = \bigcup_{i=1}^{\ell} E_i = \bigcup_{i=1}^{\ell} \left(E_i \cap \left(\bigcup_{j=1}^{k} F_j \right) \right) = \bigcup_{i=1}^{\ell} \bigcup_{j=1}^{k} \Omega_{ij},$$

where $\Omega_{ij} = E_i \cap F_j$.

(1) It is clear that $f + g$ is simple, because

$$f + g = \sum_{i=1}^{\ell} \sum_{j=1}^{k} (c_i + d_j) \chi^{\Omega_{ij}}.$$

Noting that

$$m(E_i) = m\left(E_i \cap \left(\bigcup_{j=1}^{k} F_j \right) \right) = \sum_{i=1}^{k} m(E_i \cap F_j) = \sum_{j=1}^{k} m(\Omega_{ij})$$

and similarly for $m(F_j)$, we deduce

$$\int_\Omega (f+g) = \sum_{i=1}^{\ell} \sum_{j=1}^{k} (c_i + d_j) m(\Omega_{ij})$$

$$= \sum_{i=1}^{\ell} c_i \sum_{j=1}^{k} m(\Omega_{ij}) + \sum_{j=1}^{k} d_j \sum_{i=1}^{\ell} m(\Omega_{ij})$$

$$= \sum_{i=1}^{\ell} c_i m(E_i) + \sum_{j=1}^{k} d_j m(F_j) = \int_\Omega f + \int_\Omega g.$$

(2) With respect to the partition $\{\Omega_{ij}\}_{i\in\bar{\ell}, j\in\bar{k}}$,

$$f = \sum_{i,j} \alpha_{ij} \chi^{\Omega_{ij}}, \qquad g = \sum_{i,j} \beta_{ij} \chi^{\Omega_{ij}}.$$

Given a pair of indices (i, j). If $\Omega_{ij} \neq \emptyset$, take $x \in \Omega_{ij}$, we have

$$\alpha_{ij} = f(x) \leq g(x) = \beta_{ij}.$$

Hence

$$\int_\Omega f = \sum_{i,j} \alpha_{ij} m(\Omega_{ij}) \leq \sum_{i,j} \beta_{ij} m(\Omega_{ij}) = \int_\Omega g.$$

Lemma 4.27. *Let $f : \Omega \to [0, \infty)$ be simple, $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$ with $\Omega_k \in \mathcal{M}$, $\Omega_k \subset \Omega_{k+1}$ for all k. Then* 11

$$\int_\Omega f = \lim_{k\to\infty} \int_{\Omega_k} f.$$

Proof. Assume

$$f = \sum_{i=1}^{\ell} c_i \chi^{E_i}, \qquad \text{then } f|_{\Omega_k} = \sum_{i=1}^{\ell} c_i \chi^{E_i \cap \Omega_k}.$$

Since (see Corollary 4.11)

$$m(E_i \cap \Omega_k) \to m(E_i \cap \Omega) = m(E_i),$$

as $k \to \infty$, we deduce

$$\int_{\Omega_k} f = \sum_{i=1}^{\ell} c_i m(E_i \cap \Omega_k) \to \sum_{i=1}^{\ell} c_i m(E_i) = \int_\Omega f.$$

Let $f : \Omega \to [0, \infty]$ be measurable, its Lebesgue integral is defined by

$$\int_\Omega f = \sup_{\varphi \in S_f} \int_\Omega \varphi,$$

where S_f is the set of all simple functions $\varphi : \Omega \to [0, \infty)$ satisfying $\varphi \leq f$. When f is simple this *reduces* to the integral of simple functions defined earlier[(23)]. Clearly

$$0 \leq \int_\Omega f \leq \infty,$$

one should note that $\int_\Omega f = \infty$ is possible. If $E \subset \Omega$ is measurable, instead of $\int_E f|_E$ we write $\int_E f$.

[(23)] If f is simple, let I be the integral of f in the sense of (4.8). Since $f \in S_f$ we have $I \leq \sup_{\varphi\in S_f} \int_\Omega \varphi$. On the other hand, if $\varphi \in S_f$ then $\varphi \leq f$. By Lemma 4.26 (2) we have $\int_\Omega \varphi \leq I$. Hence $\sup_{\varphi\in S_f} \int_\Omega \varphi \leq I$. We conclude $I = \sup_{\varphi\in S_f} \int_\Omega \varphi$.

lc

Lemma 4.28. *If $E \subset \Omega$ is measurable, then*

$$\int_E f = \int_\Omega f\chi^E.$$

Proof. Given $\varphi \in S_{f|_E}$, we define $\tilde{\varphi} : \Omega \to [0, \infty)$ by

$$\tilde{\varphi}(x) = \begin{cases} \varphi(x) & \text{if } x \in E, \\ 0 & \text{if } x \in \Omega\backslash E. \end{cases}$$

Then $\tilde{\varphi} \in S_{f\chi^E}$,

$$\int_E \varphi = \int_\Omega \tilde{\varphi} \le \int_\Omega f\chi^E, \qquad \text{thus } \int_E f \le \int_\Omega f\chi^E.$$

Given $\psi \in S_{f\chi^E}$, it is clear that $\psi|_E \in S_{f|_E}$. Hence

$$\int_\Omega \psi = \int_E \psi|_E \le \int_E f, \qquad \text{thus } \int_\Omega f\chi^E \le \int_E f.$$

Proposition 4.29. *Let $f, g : \Omega \to [0, \infty]$ be measurable,*

(1) $\int_\Omega f = 0$ *iff* $f = 0$ *a.e.* Ω.
(2) $f \le g$ *implies* $\int_\Omega f \le \int_\Omega g$.

Proof. (1) ($\Leftarrow$) If $f = 0$ a.e., then $\varphi = 0$ a.e. for all $\varphi \in S_f$. Thus $\int_\Omega \varphi = 0$ and

$$\int_\Omega f = \sup_{\varphi \in S_f} \int_\Omega \varphi = \sup_{\varphi \in S_f} 0 = 0.$$

($\Rightarrow$) We may assume $m(\Omega) > 0$. If $\int_\Omega f = 0$, then $m(\{f > k^{-1}\}) = 0$ for all $k \in \mathbb{N}$. Otherwise $\varphi = k^{-1}\chi^{\{f > k^{-1}\}} \in S_f$ for some k, and we have

$$\int_\Omega f \ge \int_\Omega \varphi = k^{-1} m(\{f > k^{-1}\}) > 0.$$

Now $f = 0$ a.e. follows from

$$\{f > 0\} = \bigcup_{k=1}^\infty \left\{ f > \frac{1}{k} \right\}.$$

le

Corollary 4.30. *If $E \subset \Omega$ is measurable, then*

$$\int_E f \le \int_\Omega f. \tag{4.9}$$ el

Proof. Since $f\chi^E \le f$, Lemma 4.28 yields

$$\int_E f = \int_\Omega f\chi^E \le \int_\Omega f.$$

Theorem 4.31 (Levi). *Let $f_k : \Omega \to [0, \infty]$ be measurable, $f_k \le f_{k+1}$ for all k, $f = \lim f_k$, then*

$$\int_\Omega f_k \to \int_\Omega f. \tag{4.10}$$ er

Proof. From $f_k \le f$ we have $\int_\Omega f_k \le \int_\Omega f$. Thus

$$\lim \int_\Omega f_k \le \int_\Omega f. \tag{4.11}$$ e12

Given $h \in S_f$, take $c \in (0,1)$ and set $\Omega_k = \{f_k \ge ch\}$. Then $\Omega_k \subset \Omega_{k+1}$ for all k,

$$\Omega = \bigcup_{k=1}^{\infty} \Omega_k.$$

To see this, let $x \in \Omega$. If $f(x) = 0$ then $x \in \Omega_k$ for all k because $h(x) = 0$; if $f(x) > 0$ then $f_k(x) > ch(x)$ for $k \gg 1$ because $f_k(x) \to f(x)$ and $f(x) > ch(x)$, hence $x \in \Omega_k$ for $k \gg 1$.

By Corollary 4.30 we get

$$\int_\Omega f_k \ge \int_{\Omega_k} f_k \ge \int_{\Omega_k} ch = c\int_{\Omega_k} h.$$

Now Lemma 4.27 yields

$$\lim_{k\to\infty} \int_\Omega f_k \ge c \lim_{k\to\infty} \int_{\Omega_k} h = c\int_\Omega h.$$

Let $c \to 1$ we deduce

$$\lim_{k\to\infty} \int_\Omega f_k \ge \int_\Omega h, \qquad \text{hence } \lim_{k\to\infty} \int_\Omega f_k \ge \int_\Omega f.$$

This and (4.11) give (4.10).

Proposition 4.32. *Let $f : \Omega \to [0,\infty]$ be measurable, then there is a sequence of simple functions $\{\varphi_k\}$ such that $\varphi_k \nearrow f$.*

Proof. For $k \in \mathbb{N}$, let $E_k = \{f \ge k\}$,

$$E_{k,j} = \left\{ \frac{j-1}{2^k} \le f < \frac{j}{2^k} \right\}, \qquad j \in \overline{k \cdot 2^k}.$$

Now define $\varphi_k : \Omega \to [0,\infty)$,

$$\varphi_k = k\chi^{E_k} + \sum_{j=1}^{k\cdot 2k} \frac{j-1}{2^k} \chi^{\Omega_{k,j}}.$$

Then $\varphi_k \le f$. Moreover: (1) $\varphi_k \le \varphi_{k+1}$; (2) $\varphi_k \to f$.

(1) Given $x \in \Omega$. If $x \in E_k$ then $x \in E_{k+1}$ or $x \in E_{k+1,\ell}$ with $\ell \ge k \cdot 2^{k+1} + 1$. In both cases

$$\varphi_{k+1}(x) \ge k = \varphi_k(x).$$

If $x \in E_{k,j}$ for some $j \in \overline{k \cdot 2^k}$, then

$$\frac{j-1}{2^k} \le f(x) < \frac{j}{2^k}, \qquad \frac{(2j-1)-1}{2^{k+1}} \le f(x) < \frac{2j}{2^{k+1}}.$$

We see that $x \in E_{k+1,\ell}$ for some $\ell \ge 2j-1$. Thus

$$\varphi_{k+1}(x) = \frac{\ell-1}{2^{k+1}} \ge \frac{j-1}{2^k} = \varphi_k(x).$$

(2) Given $x \in \Omega$. If $f(x) = \infty$, then $\varphi_k(x) = k$ for all k; if $f(x) \le A$ then for $k > A$ there is $j \in \overline{k \cdot 2^k}$ such that

$$\frac{j-1}{2^k} \le f(x) < \frac{j}{2^k},$$

hence $0 \le f(x) - \varphi_k(x) \le 2^{-k}$. In both case we have $\varphi_k(x) \to f(x)$.

add

Proposition 4.33. *Let $f, g : \Omega \to [0, \infty]$ be measurable, then*

$$\int_\Omega (f+g) = \int_\Omega f + \int_\Omega g. \tag{4.12}$$

ea

Proof. Take two sequence of simple functions $\varphi_k \nearrow f$, $\psi_k \nearrow g$. Then

$$\varphi_k + \psi_k \nearrow f + g,$$

and since $\varphi_k + \psi_k$ are simple, Lemma 4.26 yields

$$\int_\Omega (\varphi_k + \psi_k) = \int_\Omega \varphi_k + \int_\Omega \psi_k.$$

Now (4.12) follows from this and Levi.

Corollary 4.34. *Let $f_k : \Omega \to [0, \infty]$ be measurable, then*

$$\int_\Omega \sum_{k=1}^\infty f_k = \sum_{k=1}^\infty \int_\Omega f_k.$$

Remark 4.35. In Riemann integral, the right hand side still makes sense, but $\sum_{k=1}^\infty f_k$ maybe not integrable.

Proof. Let $F_\ell = \sum_{k=1}^\ell f_k$, then $F_\ell \le F_{\ell+1}$,

$$\sum_{k=1}^\ell \int_\Omega f_k = \int_\Omega \sum_{k=1}^\ell f_k = \int_\Omega F_\ell.$$

By Levi we have

$$\sum_{k=1}^\infty \int_\Omega f_k = \lim_{\ell\to\infty} \sum_{k=1}^\ell \int_\Omega f_k = \lim_{\ell\to\infty} \int_\Omega F_\ell = \int_\Omega \lim_{\ell\to\infty} F_\ell = \int_\Omega \sum_{k=1}^\infty f_k.$$

Remark 4.36. If $\Omega = \Omega_1 \sqcup \Omega_2$, $\Omega_i \in \mathcal{M}$, then

$$\int_\Omega f = \int_\Omega f\left(\chi^{\Omega_1} + \chi^{\Omega_2}\right) = \int_\Omega f\chi^{\Omega_1} + \int_\Omega \chi^{\Omega_2} = \int_{\Omega_1} f + \int_{\Omega_2} f.$$

Corollary 4.37. *If $f = g$ a.e. Ω, then $\int_\Omega f = \int_\Omega g$.*

Proof. Let $E = \{f \ne g\}$, then $m(E) = 0$, $\int_E f = 0$,

$$\int_\Omega f = \int_E f + \int_{\Omega\setminus E} f = \int_{\Omega\setminus E} g = \int_\Omega g.$$

Lemma 4.38 (Fatou)**.** *Let $f_k : \Omega \to [0, \infty]$ be measurable, we have*

$$\int_\Omega \varliminf_{k\to\infty} f_k \le \varliminf_{k\to\infty} \int_\Omega f_k.$$

Proof. Let $g_\ell = \inf_{k\ge \ell} f_k$, then

$$g_\ell \le g_{\ell+1}, \qquad g_\ell \le f_\ell, \qquad \lim_{k\to\infty} g_k = \varliminf_{k\to\infty} f_k.$$

for all ℓ. Levi yields

$$\begin{aligned}\int_\Omega \varliminf_{k\to\infty} f_k &= \int_\Omega \lim_{k\to\infty} g_k = \lim_{k\to\infty}\int_\Omega g_k \\ &= \varliminf_{k\to\infty}\int_\Omega g_k \le \varliminf_{k\to\infty}\int_\Omega f_k.\end{aligned}$$

po

Proposition 4.39. *Let $f : \Omega \to [0,\infty]$ be measurable, $\int_\Omega f < \infty$, then*

(1) $m(\{f = \infty\}) = 0.$

(2) *for* $\forall \varepsilon > 0$, $\exists \delta > 0$ *such that*

$$\int_E f < \varepsilon.$$

for $E \subset \Omega$ with $m(E) < \delta$. (the absolute continuity of Lebesgue integral*)*

Proof. (1) Let $A = \{f = \infty\}$, then $A \in \mathcal{M}$ because

$$A = \bigcap_{k=1}^{\infty} \{f > k\}.$$

For all $k \in \mathbb{N}$ we have $k\chi^A \le f$,

$$km(A) = \int_\Omega k\chi^A \le \int_\Omega f < \infty, \qquad m(A) \le \frac{1}{k}\int_\Omega f.$$

Thus $m(A) = 0$.

(2) Given $\varepsilon > 0$, take $\varphi \in S_f$ such that (equality follows from Proposition 4.33)

$$\int_\Omega (f-\varphi) = \int_\Omega f - \int_\Omega \varphi < \frac{\varepsilon}{2}.$$

Let $\delta = \varepsilon/\left(2(1+|\varphi|_\infty)\right)$. If $E \subset \Omega$, $m(E) < \delta$, then

$$\begin{aligned}\int_E f &= \left(\int_E f - \int_E \varphi\right) + \int_E \varphi \\ &\le \int_\Omega (f-\varphi) + \int_E \varphi \le \frac{\varepsilon}{2} + |\varphi|_\infty m(E) < \varepsilon.\end{aligned}$$

Example 4.40. Let $f : \Omega \to [0,\infty]$ be measurable, $\int_\Omega f < \infty$. Then $F : (0,\infty) \to \mathbb{R}$ defined below is continuous:

$$F(r) = \int_{\Omega\cap B_r} f.$$

Proof (via Absolute Continuity). Let $r_0 \in (0,\infty)$, we prove that F is continuous at r_0. Given $\varepsilon > 0$, by Proposition 4.39, there is $\eta > 0$ such that

$$\int_E f < \varepsilon$$

for all $E \subset \Omega$ satisfying $m(E) < \eta$. Take $\delta > 0$ such that

$$m(B_r \backslash B_{r_0}) < \eta \qquad \text{if } r \in (r_0, r_0+\delta).$$

Then for $r \in (r_0, r_0 + \delta)$ we have

$$|F(r) - F(r_0)| = \int_{\Omega \cap B_r} f - \int_{\Omega \cap B_{r_0}} f = \int_{\Omega \cap (B_r \setminus B_{r_0})} f < \varepsilon$$

because $m(\Omega \cap (B_r \setminus B_{r_0})) < \eta$. This proves that F is right-continuous at r_0:

$$\lim_{r \to r_0+} F(r) = F(r_0).$$

Similarly we can prove that F is left-continuous at r_0.

Proof (via Levi). Let $r_0 \in (0, \infty)$ and $r_n \nearrow r_0$. Then

$$f_n = f \chi^{\Omega \cap B_{r_n}} \nearrow f \chi^{\Omega \cap B_{r_0}}.$$

By Levi,

$$F(r_n) = \int_{\Omega \cap B_{r_n}} f = \int_{\Omega} f_n \to \int_{\Omega} f \chi^{\Omega \cap B_{r_0}} = \int_{\Omega \cap B_{r_0}} f = F(r_0).$$

We still need to prove $F(r_n) \to F(r_0)$ for $r_n \searrow r_0$ (exercise).

Remark 4.41. More genetral result is true: $G : (0, \infty) \times \mathbb{R}^n \to \mathbb{R}$ given below is continuous:

$$G(r, x) = \int_{\Omega \cap B_r(x)} f.$$

Lemma 4.42 (Borel-Cantelli)**.** *Let* $\Omega_k \in \mathcal{M}$, $\sum_{k=1}^{\infty} m(\Omega_k) < \infty$, *then* $m(\overline{\lim}\, \Omega_k) = 0$.

Remark 4.43. Given a sequence of sets A_k, we define

$$\overline{\lim_{k \to \infty}}\, \Omega_k = \{x \mid x \in \Omega_i \text{ for infinitely many } i\} = \bigcap_{i=1}^{\infty} \bigcup_{k=i}^{\infty} \Omega_k,$$

$$\underline{\lim}_{k \to \infty}\, \Omega_k = \{x \mid x \notin \Omega_i \text{ for at most finitely many } i\} = \bigcup_{i=1}^{\infty} \bigcap_{k=i}^{\infty} \Omega_k.$$

Borel-Cantelli lemma is frequently used in probability.

Proof. Let $f_k = \chi^{\Omega_k}$, then

$$\int \sum_{k=1}^{\infty} f_k = \sum_{k=1}^{\infty} \int f_k = \sum_{k=1}^{\infty} m(\Omega_k) < \infty.$$

Hence

$$m(\overline{\lim}\, \Omega_k) = m\left(\sum_{k=1}^{\infty} f_k = \infty \right) = 0,$$

because: x belongs to infinitetely many Ω_k iff $\sum_{k=1}^{\infty} f_k(x) = \infty$.

4.4. Absolutely integrable functions. Sign-changing measurable functions $f : \Omega \to \overline{\mathbb{R}}$ is absolutely integrable if $\int_\Omega |f| < \infty$. The set of all such functions is denoted by $L^1(\Omega)$ or simply $L(\Omega)$. If $f \in L(\Omega)$, its Lebesgue integral is

$$\int_\Omega f = \int_\Omega f^+ - \int_\Omega f^-.$$

Proposition 4.44. *For $f, g \in L(\Omega)$, $c \in \mathbb{R}$,*

(1) $\int_\Omega cf = c \int_\Omega f$,

(2) $f + g \in L(\Omega)$ *and*

$$\int_\Omega (f+g) = \int_\Omega f + \int_\Omega g. \tag{4.13}$$

(3) $\int_\Omega f \le \int_\Omega g$ *if* $f \le g$ *a.e.* Ω.

Proof. Since $|f+g| \le |f| + |g|$, we get $f + g \in L(\Omega)$. To get (4.13), we may assume that instead of being $\overline{\mathbb{R}}$-valued, f and g are $\mathbb{R}$-valued. In fact, since $\int_\Omega |f| < \infty$ and $\int_\Omega |g| < \infty$, the measure of

$$E = \{|f| = \infty\} \cup \{|g| = \infty\}$$

is zero. Define $\mathbb{R}$-valued functions $\tilde{f}, \tilde{g} : \Omega \to \mathbb{R}$ via

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \Omega \backslash E, \\ 0 & x \in E, \end{cases} \qquad \tilde{g}(x) = \begin{cases} g(x) & x \in \Omega \backslash E, \\ 0 & x \in E. \end{cases}$$

Then $\tilde{f} = f$ a.e., $\tilde{g} = g$ a.e., and $\tilde{f} + \tilde{g} = f + g$ a.e.. Hence

$$\int_\Omega \tilde{f} = \int_\Omega f, \qquad \int_\Omega \tilde{g} = \int_\Omega g, \qquad \int_\Omega \left(\tilde{f} + \tilde{g}\right) = \int_\Omega (f+g).$$

From this the aditivity law (4.13) for $\overline{\mathbb{R}}$-valued follows from that law for $\mathbb{R}$-valued functions.

Having this remark in mind, from

$$(f+g)^+ - (f+g)^- = f + g = f^+ - f^- + g^+ - g^-,$$

we deduce[24]

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+.$$

Integrating both sides, using the aditivity of integtrals of nonegative functions, we have

$$\begin{aligned} \int_\Omega (f+g)^+ + \int_\Omega f^- + \int_\Omega g^- &= \int_\Omega \left((f+g)^+ + f^- + g^-\right) \\ &= \int_\Omega \left((f+g)^- + f^+ + g^+\right) \\ &= \int_\Omega (f+g)^- + \int_\Omega f^+ + \int_\Omega g^+. \end{aligned}$$

Since all integrals are finite, we get

$$\int_\Omega (f+g) = \int_\Omega (f+g)^+ - \int_\Omega (f+g)^-$$

[24] Adding both sides by f^-, g^- and then $(f+g)^-$, this is valid because all these are finite (it make no sense to add $+\infty$ to both sides of an equality).

$$= \left(\int_\Omega f^+ - \int_\Omega f^-\right) + \left(\int_\Omega g^+ - \int_\Omega g^-\right)$$

$$= \int_\Omega f + \int_\Omega g.$$

Theorem 4.45 (Lebesgue dominated theorem)**.** *Let $f_k : \Omega \to \overline{\mathbb{R}}$ be measurable, $|f_k| \le g$ for some $g \in L(\Omega)$. If $f_k \to f$ on Ω, then*

$$\int_\Omega |f_k - f| \to 0, \qquad \textit{in particular } \int_\Omega f_k \to \int_\Omega f.$$

Proof. Let $g_k = |f_k - f|$, then

$$h_k := 2g - g_k \ge 0, \qquad h_k \to 2g \text{ a.e. } \Omega.$$

By Fatou,

$$\int_\Omega 2g \le \varliminf_{k\to\infty} \int_\Omega h_k = \varliminf_{k\to\infty} \left(\int_\Omega 2g - \int_\Omega g_k\right)$$

$$= \int_\Omega 2g - \varlimsup_{k\to\infty} \int_\Omega g_k.$$

It follows that

$$\varlimsup_{k\to\infty} \int_\Omega g_k = 0, \qquad \text{that is } \int_\Omega g_k \to 0.$$

Example 4.46. Find

$$I = \lim_{n\to\infty} \int_{\mathbb{R}} \frac{\sin(x/n)}{1+x^2} dx.$$

Proof. Let $g, f_n : \mathbb{R} \to \mathbb{R}$,

$$f_n(x) = \frac{\sin(x/n)}{1+x^2}, \qquad g(x) = \frac{1}{1+x^2}.$$

Then $g \in L(\mathbb{R})$, $f_n \to 0$ a.e. $\mathbb{R}$. Therefore

$$\lim_{n\to\infty} \int_{\mathbb{R}} \frac{\sin(x/n)}{1+x^2} dx = \int_{\mathbb{R}} \lim_{n\to\infty} \frac{\sin(x/n)}{1+x^2} dx = \int_{\mathbb{R}} \dot{0} dx = 0.$$

pd

Proposition 4.47. *Let $f : \Omega \times (a,b) \to \overline{\mathbb{R}}$, $f(\cdot,t) \in L(\Omega)$ for all $t \in (a,b)$, $f(x,\cdot)$ is differentiable. If there is $g \in L(\Omega)$ such that $|\partial_t f(x,t)| \le g(x)$ for all $(x,t) \in \Omega \times (a,b)$, then the function $\varphi : (a,b) \to \mathbb{R}$ given by*

$$\varphi(t) = \int_\Omega f(x,t)\, dx$$

is differentiable,

$$\varphi'(t) = \frac{d}{dt} \int_\Omega f(x,t)\, dx = \int_\Omega \frac{\partial f(x,t)}{\partial t} dx.$$

Proof. Given $t_0 \in (a,b)$ and $t_n \to t_0$, define $f_n : \Omega \to \mathbb{R}$,

$$f_n(x) = \frac{f(x,t_n) - f(x,t_0)}{t_n - t_0}.$$

Then $f_n \to \partial_t f(\cdot, t_0)$ on Ω, and by the mean value theorem, for $x \in \Omega$ we have

$$|f_n(x)| = \left|\frac{f(x,t_n) - f(x,t_0)}{t_n - t_0}\right| = |\partial_t(x, \xi_n)| \le g(x),$$

where $\xi_n \in (t_0, t_n)$ may depend on x. Using Lebesgue dominated theorem,

$$\begin{aligned}
\varphi'(t_0) &= \lim_{n\to\infty} \frac{\varphi(t_n) - \varphi(t_0)}{t_n - t_0} \\
&= \lim_{n\to\infty} \int_\Omega f_n(x)\,dx = \int_\Omega \partial_t f(x, t_0)\,dx.
\end{aligned}$$

Example 4.48. Compute

$$\varphi(t) = \int_{-\infty}^{\infty} e^{-x^2/2} \cos(tx)\ \mathrm{d}x.$$

Proof. Let $f(x,t) = e^{-x^2/2}\cos(tx)$, then

$$|\partial_t f(x,t)| = \left|xe^{-x^2/2}\sin(tx)\right| \le |x|\,e^{-x^2/2} =: g(x).$$

Since $g \in: L(\mathbb{R})$, Proposition 4.47 applies, and we have

$$\begin{aligned}
\dot\varphi(t) &= \int_{-\infty}^{\infty} \partial_t\left(e^{-x^2/2}\cos(tx)\right)\,\mathrm{d}x = -\int_{-\infty}^{\infty} xe^{-x^2/2}\sin(tx)\ \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \sin(tx)\ \mathrm{d}e^{-x^2/2} = \left[e^{-x^2/2}\sin(tx)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}\left(\sin(tx)\right) \\
&= -\int_{-\infty}^{\infty} e^{-x^2/2} t\cos(tx)\ dx = -t\varphi(t).
\end{aligned}$$

We deduce

$$\dot\varphi(t) + t\varphi(t) = 0, \qquad \varphi(0) = \int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}.$$

Solving this ODE, we get

$$\int_{-\infty}^{\infty} e^{-x^2/2}\cos(tx)\ \mathrm{d}x = \sqrt{2\pi}e^{-t^2/2}.$$

Proposition 4.49. *Let $f_k : \Omega \to \overline{\mathbb{R}}$ be measurable. If $\sum_i \int_\Omega |f_i| < \infty$, then $\sum_i f_i = S$ a.e. on Ω for some $S \in L(\Omega)$, and*

$$\int_\Omega S = \sum_{i=1}^{\infty} \int_\Omega f_i.$$

Proof. By Levi,

$$\int_\Omega \sum_i |f_i| = \sum_i \int_\Omega |f_i| < \infty, \tag{4.14}$$

hence $F := \sum_i |f_i| < \infty$ a.e. on Ω. Thus $\sum_i f_i = S$ a.e. on Ω for some measurable $S : \Omega \to \overline{\mathbb{R}}$. Since $|S| \le F$, we see from (4.14) that $S \in L(\Omega)$. Let $S_k = \sum_{i=1}^k f_i$, then $S_k \to S$, $|S_k| \le F$. Applying Lebesgue we get

$$\sum_{i=1}^{k} \int_\Omega f_i = \int_\Omega S_k \to \int_\Omega S.$$

4.5. Relation with Riemann integral. Lebesgue integral extends Riemann integral.

Theorem 4.50. *Let $f : [a, b] \to \mathbb{R}$ be bounded, D is the set of discontinuous points. Then $f \in R[a, b]$ iff $m(D) = 0$. In this case $f \in L[a, b]$ and*

$$\int_a^b f = \int_{[a,b]} f.$$

Proof. For a partition $P = \{x_i\}_{i=0}^n$ of $[a, b]$, let

$$\varphi = \sum_{i=1}^n m_i \chi^{(x_{i-1},x_i]}, \qquad \psi = \sum_{i=1}^n M_i \chi^{(x_{i-1},x_i]}, \tag{4.15}$$ ep

where

$$m_i = \inf_{[x_{i-1},x_i]} f, \qquad M_i = \sup_{[x_{i-1},x_i]} f.$$

We have

$$s(P) = \int_{[a,b]} \varphi, \qquad S(P) = \int_{[a,b]} \psi.$$

Let P_n be a sequence of partition of $[a, b]$ such that $|P_n| \to 0$, $P_n \subset P_{n+1}$. Then

$$\varphi_1 \le \varphi_2 \le \cdots \le f \le \cdots \le \psi_2 \le \psi_1,$$

where φ_n and ψ_n are the simple functions in (4.15) for the partition P_n. Obviously

$$\varphi = \sup_n \varphi_n, \qquad \psi = \inf_n \psi_n$$

are bounded and measurable, thus in $L[a, b]$.

Let $Q = \bigcup_{n=1}^\infty P_n$, since $|P_n| \to 0$ we have (verifying pointwise[25])

$$\varphi \le f \le \psi, \qquad \{\varphi < \psi\} \subset D \subset \{\varphi < \psi\} \cup Q.$$

Because $m(Q) = 0$, we get $m(D) = m(\{\varphi < \psi\})$.

By Lebesgue dominated theorem,

$$\int_{[a,b]} \varphi = \lim_n \int_{[a,b]} \varphi_n = \lim_n s(P_n), \qquad \int_{[a,b]} \psi = \lim_n S(P_n).$$

Thus

$$\omega := \lim_n [S(P_n) - s(P_n)] = \int_{[a,b]} (\psi - \varphi).$$

We conclude (noting $\varphi \le \psi$)

$$f \in R[a,b] \quad \Leftrightarrow \quad \omega = 0 \quad \Leftrightarrow \quad \psi = \varphi \text{ a.e} \quad \Leftrightarrow \quad m(D) = 0.$$

---

[25] If $\varphi(x) < \psi(x)$, then

$$\inf_n (\psi_n(x) - \varphi_n(x)) = \psi(x) - \varphi(x) =: \varepsilon > 0.$$

This *means* that for all n, the amplitude of f on the subinterval(s) of P_n containing x is not less than ε. So f is not continuous at x.

If $\varphi(x) = \psi(x)$ and $x \notin Q$, then for all n there is a *unique* subinterval $[x_{i-1}^n, x_i^n]$ containing x and the amplitude of f on $[x_{i-1}^n, x_i^n]$, which equals $\psi_n(x) - \varphi_n(x)$, goes to 0 as $n \to \infty$. Thus f is continuous at x.

In this case, $f = \varphi$ a.e., thus[26] $f \in L[a,b]$ and

$$\int_a^b f = \lim_n s(P_n) = \lim_n \int_{[a,b]} \varphi_n = \int_{[a,b]} \varphi = \int_{[a,b]} f.$$

4.6. Fubini theorem. To comput higher dimensional integrals we convert them into iterated lower dimensional ones.

Theorem 4.51 (Tonelli). *If $f : \mathbb{R}^m \times \mathbb{R}^n \to [0,\infty]$ is measurable, then*

(1) *for a.e. $x \in \mathbb{R}^m$, $f(x,\cdot) : \mathbb{R}^n \to [0,\infty]$ is measurable.*
(2) *$F_f : \mathbb{R}^m \to [0,\infty]$ defined below is measurable:*

$$F_f(x) = \int_{\mathbb{R}^n} f(x,y)\,dy. \tag{4.16}$$ ef

(3) *we have*

$$\int_{\mathbb{R}^m\times\mathbb{R}^n} f(x,y)\,dxdy = \int_{\mathbb{R}^m} F_f(x)\,dx = \int_{\mathbb{R}^m} dx \int_{\mathbb{R}^n} f(x,y)\,dy.$$

Theorem 4.52 (Fubini). *If $f \in L(\mathbb{R}^m \times \mathbb{R}^n)$, then*

(1) *for a.e. $x \in \mathbb{R}^m$, $f(x,\cdot) \in L(\mathbb{R}^n)$.*
(2) *then function $F_f \in L(\mathbb{R}^m)$, and*

$$\int_{\mathbb{R}^m\times\mathbb{R}^n} f(x,y)\,dxdy = \int_{\mathbb{R}^m} dx \int_{\mathbb{R}^n} f(x,y)\,dy.$$

In particular, if $f \in L(\mathbb{R}^m \times \mathbb{R}^n)$, then

$$\int_{\mathbb{R}^m} dx \int_{\mathbb{R}^n} f(x,y)\,dy = \int_{\mathbb{R}^n} dy \int_{\mathbb{R}^m} f(x,y)\,dx.$$

Example 4.53. Since[27]

$$\int_0^1 dx \int_0^1 \frac{y^2-x^2}{(x^2+y^2)^2}\,dy \neq \int_0^1 dy \int_0^1 \frac{y^2-x^2}{(x^2+y^2)^2}\,dx,$$

we conclude that if $f : [0,1]\times[0,1] \to \mathbb{R}$ is the integrand, then $f \notin L([0,1]\times[0,1])$.

Example 4.54. Let $f : \Omega \to [0,\infty]$ be measurable,

$$V_f = \{(x,y) \mid x \in \Omega, 0 \le y \le f(x)\}.$$

Then

$$m(V_f) = \int_\Omega f.$$

[26] That $f \in \mathcal{M}[a,b]$ also follows from its a.e. continuity.

[27] Since $\int \frac{y^2-x^2}{(x^2+y^2)^2}\,dy = -\frac{y}{x^2+y^2}$,

$$\int_0^1 dx \int_0^1 \frac{y^2-x^2}{(x^2+y^2)^2}\,dy = \int_0^1 \left[-\frac{y}{x^2+y^2}\right]_{y=0}^{y=1} dx = \int_0^1 \frac{-1}{1+x^2}dx = -\frac{1}{4}\pi.$$

Proof. We omit the verification that V_f is measurable. Becsause

$$\chi^{V_f}(x, y) = \chi^{\Omega}(x)\chi^{[0,f(x)]}(y),$$

we have

$$\begin{aligned} m(V_f) &= \int_{\mathbb{R}^{n+1}} \chi^{V_f}(x, y)\, dxdy = \int_{\mathbb{R}^n} dx \int_{\mathbb{R}} \chi^{V_f}(x, y)\, dy \\ &= \int_{\mathbb{R}^n} dx \int_{\mathbb{R}} \chi^{\Omega}(x)\chi^{[0,f(x)]}(y)\, dy \\ &= \int_{\mathbb{R}^n} \chi^{\Omega}(x) \left(\int \chi^{[0,f(x)]}(y)\, dy \right) dx \\ &= \int_{\mathbb{R}^n} \chi^{\Omega}(x) f(x)\, dx = \int_{\Omega} f(x)\, dx. \end{aligned}$$

5. Appendix

5.1. Logic and quantifiers. A proposition is a statement that is TRUE or FALSE. The negative of p is denoted by $\neg p$. A compound proposition is a proposition that involves the assembly of multiple statements.

https://en.wikiversity.org/wiki/Compound_Propositions_and_Useful_Rules

Example 5.1. Suppose p is false, then "if p then q ($p \to q$)" is always true (even q is false).

Example 5.2. $p \vee \neg q \to r$ means p or $\neg q$ implies r. That is, either p or $\neg q$ is true, r would be true.

Example 5.3. "$p \to q$" is equivalent to "$\neg q \to \neg p$". Thus, to prove "if p then q", it suffices to show that "if q is not true, then p is not true". This is *proof by contradiction*.

Some propositions depend on x, we write $p(x)$. In analysis and many branchs of mathematics, we will encounter

(1) there is x such that $p(x)$ ($\exists x$, $p(x)$),
(2) for all x we have $p(x)$ ($\forall x$, $p(x)$).

Example 5.4. For a sequence of real numbers a_n, $a_n \to a$ means

$$\forall \varepsilon > 0, \exists N, \text{if } n \geq N \text{ then } |a_n - a| < \varepsilon.$$

$a_n \not\to a$ means

$$\exists \varepsilon > 0, \forall N, \exists n \geq N \text{ such that } |a_n - a| \geq \varepsilon.$$

5.2. Sets and functions. We will not define what a set is.

(1) $x \in A$, $x \notin A$.
(2) $A \subset B$, $B \supset A$ (we will not use $A \subseteq B$), proper subset.

Example 5.5. $A = \{1, 2, a\}$, $a \in A$, $3 \notin A$.

Example 5.6. $\{x \in S \mid P(x)\}$ is the set of $x \in S$ such that P is true.

Example 5.7. $\emptyset$, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$.

Set operations:

(1) $A \cap B$, $A \cup B$, $A \backslash B$

(2) For a family of sets A_λ ($\lambda \in \Lambda$),

$$\bigcup_{\lambda\in\Lambda} A_\lambda = \{x \mid x \in A_\lambda \text{ for } \textit{some } \lambda \in \Lambda\},$$

$$\bigcap_{\lambda\in\Lambda} A_\lambda = \{x \mid x \in A_\lambda \text{ for } \textit{all } \lambda \in \Lambda\}.$$

If $\Lambda = \mathbb{N}$, instead of $\bigcup_{\lambda\in\Lambda} A_\lambda$ we write

$$\bigcup_{\lambda=1}^{\infty} A_\lambda = \bigcup_{n=1}^{\infty} A_n$$

for $\bigcup_{\lambda\in\Lambda} A_\lambda$. We have

$$X\backslash \bigcup_{\lambda\in\Lambda} A_\lambda = \bigcap_{\lambda\in\Lambda} (X\backslash A_\lambda), \qquad X\backslash \bigcap_{\lambda\in\Lambda} A_\lambda = \bigcup_{\lambda\in\Lambda} (X\backslash A_\lambda).$$

(3) $A \times B$. For example,

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) \mid x \in \mathbb{R}, y \in \mathbb{R}\}.$$

Viewing (x, y) as coordinate of point on a plan, we regard $\mathbb{R}^2$ as the plane.

(4) $\prod_{i=1}^n A_i = A_1 \times \cdots \times A_n = \left\{\left(x^1, \ldots, x^n\right) \mid x^i \in A_i \text{ for } i \in \overline{n}\right\}$.

Given nonempty sets A and B. A map $f : A \to B$ is a rule that assigns each $a \in A$ a unique element $b \in B$. Here b depends on a, called the image of a, and denoted by $f(a)$. But what is a rule?

Definition 5.8. Given nonempty sets A and B. A map $f : A \to B$ (with domain $D_f = A$ and target B) is a subset of $A \times B$ such that: for $\forall a \in A$, $\exists 1\ b \in B$ such that $(a, b) \in f$; we write $b = f(a)$. When $B = \mathbb{R}$, we call f a real function on A.

Remark 5.9. We can think of f as a machine, inputing $a \in A$, it produces the output $f(a)$.

The image of $E \subset A$ is

$$f(E) = \{f(a) \mid a \in E\}.$$

$R_f = f(A)$ is the range of f. The preimage of $F \subset B$ is

$$f^{-1}(F) = \{a \in A \mid f(a) \in F\}.$$

Example 5.10. The rule $x \mapsto x^2$ is a map $f : \mathbb{R} \to \mathbb{R}$. Here $D_f = \mathbb{R}$, $R_f = [0, \infty)$.

$$f[-1, 2) = [0, 4), \qquad f^{-1}[-1, 2) = f^{-1}[0, 2) = \left(-\sqrt{2}, \sqrt{2}\right).$$

Example 5.11. Given $f : X \to Y$, it is easy to prove:

(1) $f(A \cup B) = f(A) \cup f(B)$, $f(A \cap B) \subset f(A) \cap f(B)$.
(2) $f^{-1}(E \cup F) = f^{-1}(E) \cup f^{-1}(F)$, $f^{-1}(E \cap F) \subset f^{-1}(E) \cap f^{-1}(F)$.

Similar results are also true for infinite union or intersection.

The map $f : A \to B$ is

(1) injective: if $\# f^{-1}(b) \leq 1$ for all $b \in B$,
(2) surjective: if $f(A) = B$,
(3) bijective: if f is both injective and surjective.

Remark 5.12. $f : A \to B$ is surjective means that for $\forall b \in B$, the equation

$$f(x) = b$$

always has a solution in A.

If $f : A \to B$ is bijective, then the map

$$f^{-1} = \{(b, a) \mid (a, b) \in f\}$$

is call the inverse (map) of f. Namely $f^{-1} : B \to A$,

$$f^{-1}(b) = a \qquad \text{iff} \qquad f(a) = b.$$

If $f : A \to B$, $g : B \to C$, then the coposition $g \circ f : A \to C$ is defined by

$$(g \circ f)(x) = g(f(x)), \qquad \forall x \in A.$$

We have:

(1) $(g \circ f)^{-1}(E) = f^{-1}(g^{-1}(E))$ for $E \subset C$.
(2) $(h \circ g) \circ f = h \circ (g \circ f)$ for $h : C \to D$.

Given $f : A \to B$ and $E \subset A$, we have a new map

$$f|_E : E \to B, \qquad f|_E(x) = f(x) \quad \text{for } \forall x \in E,$$

called the restriction of f to E.

Given $f : A \to B$, if there is $F : X \to B$ for some $X \supset A$ such that $f = F|_A$, then F is an extension of f.

5.3. Backup. Proposition 1.49: (2) $\Rightarrow$ (1). If f is not continuous at a, $\exists \varepsilon > 0$ such that

$$f(B^X_{1/n}(a)) \not\subset B^Y_\varepsilon(f(a)) \qquad \text{for all } n \in \mathbb{N}.$$

For each n we pick $x_n \in B^X_{1/n}(a)$ such that $f(x_n) \notin B^Y_\varepsilon(f(a))$, we get a sequence $\{x_n\} \subset X$ such that $x_n \to a$ but $f(x_n) \not\to f(a)$.

(1) $\Rightarrow$ (3). Take $\varepsilon > 0$ such that $B^Y_\varepsilon(f(a)) \subset V$, then take $\delta > 0$ such that

$$f(B^X_\delta(a)) \subset B^Y_\varepsilon(f(a)).$$

The X-open set $U = B^X_\delta(a)$ satisfies $f(U) \subset V$ and $a \in U$.

Proposition 1.50:

Proof (Without using Proposition 1.49). ($\Rightarrow$). For $a \in f^{-1}(V)$, we have $f(a) \in V$. Thus $\exists \varepsilon > 0$ such that $B^Y_\varepsilon(f(a)) \subset V$. Since f is continuous at a, $\exists \delta > 0$ s.t.

$$f(B^X_\delta(a)) \subset B^Y_\varepsilon(f(a)) \subset V.$$

That is $B^X_\delta(a) \subset f^{-1}(V)$, $a \in \left(f^{-1}(V)\right)^\circ$. So $f^{-1}(V) = \left(f^{-1}(V)\right)^\circ$ and $f^{-1}(V)$ is X-open.

($\Leftarrow$). We need to show that given $a \in X$, f is continuous at a. Given $\varepsilon > 0$, $B^Y_\varepsilon(f(a))$ is a Y-open set containing $f(a)$, then $f^{-1}(B^Y_\varepsilon(f(a)))$ is an X-open set containing a. There is $\delta > 0$ such that

$$B^X_\delta(a) \subset f^{-1}(B^Y_\varepsilon(f(a))),$$

which implies $f(B^X_\delta(a)) \subset B^Y_\varepsilon(f(a))$, f is continuous at a.

References

Ambrosetti A, Rabinowitz PH (1973). *Dual variational methods in critical point theory and applications*. J. Functional Analysis, 14 349–381.

Katriel G (1994). *Mountain pass theorems and global homeomorphism theorems*. Ann. Inst. H. Poincaré C Anal. Non Linéaire, 11(2) 189–209.

Liu P, Liu S (2018). *On the surjectivity of smooth maps into Euclidean spaces and the fundamental theorem of algebra*. Amer. Math. Monthly, 125(10) 941–943.

Liu S, Zhang Y (2017). *On the change of variables formula for multiple integrals*. J. Math. Study, 50(3) 268–276.

Thorpe JA (1994). *Elementary topics in differential geometry*. Undergraduate Texts in Mathematics. Springer-Verlag, New York. Corrected reprint of the 1979 original.