Analysis

Hoyan Mok[1]

January 24, 2021

[1]E-mail: victoriesmo@hotmail.com

B

Preface

The latest version: https://github.com/HoyanMok/NotesOnMathematics/tree/master/Analysis
.

Contents

Part I

Mathematical Analysis

Chapter 1

Metric Space and Continuous Mapping

§1 Metric Space

Definition 1.1 (Metric)**.** A function

$$d\colon X^2 \to \mathbb{R}$$

$\forall x, y, z \in X$ satisfying:

a) $d(x,y) = 0 \leftrightarrow x = y$;

b) $d(x,y) = d(y,x)$ (symmetry);

c) $d(x,z) \leq d(x,y) + d(y,z)$ (triangle inequality),

is called a ***metric*** or ***distance*** in X. Such X is said to be equiped with a metric d, (X,d) is called a ***metric space***. If the metric defined over X is definite, we just simply call the X the metric space.

Some examples:

- We can define $\mathbb{R}^n_p := (\mathbb{R}^n, d_p)$, where

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

  while

$$d_\infty(x,y) := \max_{i\in n} \left| x^i - y^i \right|. \tag{1-2}$$

- Similarly we can define metric spaces as $(C[a,b], d_p)$ or simplified $C_p[a,b]$.

$$d_p(f,g) = \left(\int_a^b \left| f - g \right|^p \mathrm{d}x \right)^{1/p}. \tag{1-3}$$

  while $C_\infty[a,b]$ is called a ***Chebyshev metric***, where the metric is defined as $d_\infty(f,g) := \max_{x\in[a,b]} |f(x) - g(x)|$.

- On equivalence class $\tilde{\mathfrak{R}}[a,b]$ over $\mathfrak{R}[a,b]$ similar metric can be defined. Functions are considered equicalent if they are equal up to a null set.

Lemma 1 (Quadruple inequality)**.** *Let (X,d) be a metric space.*

$$\forall a,b,u,v \in X,\ \left|d(a,b) - d(u,v)\right| \leq d(a,u) + d(b,v) \tag{1-4}$$

Proof. Without loss of generality, we assume that $d(a,b) > d(u,v)$. According to the triangle inequality (see def. 1.1), $d(a,b) \leq d(a,u) + d(u,v) + d(v,b)$, which is to prove. □

Definition 1.2 (δ-ball)**.** Let (X,d) be a metric space, and $\delta \in \mathbb{R}_+$, $a \in X$. A set

$$B(a;\delta) = \{x \in X \mid d(a,x) < \delta\}$$

is then called a ***ball*** with a centre at $a \in X$ and a radius of δ, or a ***ball*** of point a.

Definition 1.3. The ***diametre*** of a set $A \subset X$, is defined as:

$$d(A) := \sup\{d(x,y) \mid x,y \in A\}.$$

The distance between a set and a point, and the distance between sets are defined as:

$$d(A,a) := \inf\{d(x,a) \mid x \in A\}, \quad d(A,B) := \inf\{d(x,y) \mid x \in A, y \in B\}.$$

Definition 1.4 (Open set)**.** An ***open set*** $G \in 2^X$ in a metric space (X,d) is a set that satisfies: $\forall x \in G$, $\exists \delta \in \mathbb{R}_+$, s.t. $B(X,\delta) \in 2^G$.

Definition 1.5 (Closed set)**.** A ***closed set*** $F \in 2^X$ in a metric space (X,d) is a set that satisfies: $X - F$ is an open set in (X,d).

A ***closed ball*** $\tilde{B}(X,\delta) := \{x \in X \mid d(a,x) \leq r\}$ is an example of closed sets in (X,d).

Proposition 1. a) An infinite union of open sets is an open set.
b) A definite intersection of open sets is an open set.
c) A definite union of closed sets is a closed set.
d) An infinite intersection of closed sets is a closed set.

Proof. Let $\forall \alpha \in A$, G_α be open sets.

a) $\forall x \in \bigcup_{\alpha\in A} G_\alpha$, $\exists \alpha \in A$ s.t. $x \in G_\alpha$. Since G_α is open, $\exists \delta \in \mathbb{R}_+$ s.t. $B(X,\delta) \subset G_\alpha \subset \bigcup_{\alpha\in A} G_\alpha$.

b) Let G_1, G_2 be open sets in (X,d). $\forall a \in G_1 \cap G_2$, $\exists \delta_1, \delta_2 \in \mathbb{R}_+$ s.t. $B(a;\delta_1) \subset G_1$, $B(a;\delta_2) \subset G_2$. Without loss of generality, let $\delta_1 \geq \delta_2$, therefore $a \in B(a;\delta_1) \cap B(a;\delta_2) = B(a;\delta_2) \subset G_1 \cap G_2$.

c) Just consider $\complement_X \left(\bigcap_{\alpha\in A} F_\alpha\right) = \bigcup_{\alpha\in A} \complement_X(F_\alpha)$ and a).

d) Similarly, $\complement_X (F_1 \cup F_2) = \complement_X(F_1) \cap \complement_X(F_2)$.

□

Definition 1.6 (Neighbourhood)**.** If $x \in X$ is an element of an open set, then such open set is called a ***neighbourhood*** of point x in X, denoted by $U(x)$. The collection of all neighbourhoods of x can be denoted by $\mathscr{U}(x)$.

Definition 1.7 (Interior point)**.** Let $x \in X$, $E \subset X$.

a) If $\exists U(x) \subset E$, x is called an ***interior point*** of E.
b) If $\exists U(x) \subset X - E$, x is called an ***exterior point*** of E.
c) If x isn't an interior point nor exterior point of E, it is called a ***boundary point*** of E. The set of boundary points is called ***boundary***, denoted by ∂E.

Definition 1.8 (Limit point)**.** $a \in X$, $E \subset X$. If $\forall U(a)$, card$\big(E \cap U(a)\big) = \infty$, a is called a ***limit point*** of E.

Definition 1.9 (Closure)**.** The intersections of $E \subset X$ and set of all its limit points is called the ***closure*** of E, denoted by $\overline{E}$.

Theorem 1.1. *Let $F \in 2^X$. F is a closed set in X $\leftrightarrow$ $\overline{F} = F$.*

Proof. $\to$: $\complement_X(F)$ is open, hence its elements are all its interior points. Therefore $\overline{F} - F = \overline{F} \cup \complement_X(F) = \varnothing$, also we know that $F \subset \overline{F}$, hence $F = \overline{F}$.

$\leftarrow$: $F = \overline{F}$ means that $\forall x \in \complement_X(F)$, x is not a boundary of F, which implies that x is an interior point of $X - F$. Therefore $X - F$ is open while F is closed. □

Theorem 1.2. *$\overline{E}$ is always closed.*

Proof. $\forall x \in X - \overline{E}$, since it is not an element of the set E nor its limit points, $\exists U(x)$ s.t. $U(x) \cap \overline{E} = \varnothing$, which implies that x is an extorior point of E, therefore $\overline{E}$ is closed. □

Theorem 1.3. $\overline{E} = \overline{\overline{E}}$.

Proof. Since $\overline{E}$ is closed, its complement is open, which implies that its elements are all exterior points of $\overline{E}$, therefore $\overline{E}$ has contained all of its limit points. □

Definition 1.10. (Metric subspace) We called (X', d') a ***subspace*** of (X, d) when $X' \subset X$ and $\forall x, y \in X'$, $d'(x, y) = d(x, y)$.

§2 Topological Space

Definition 2.1 (Topology)**.** We say X is epuiped with a ***topology*** if we assigned a $\mathscr{T} \subset 2^X$, with the following propoties:

a) $\varnothing \in \mathscr{T}$; $X \in \mathscr{T}$.
b) $\big(\forall \alpha \in A, G_\alpha \in \mathscr{T}\big) \to \bigcup_{\alpha \in A} G_\alpha \in \mathscr{T}$.
c) $\forall G_1, G_2 \in \mathscr{T}$, $G_1 \cap G_2 \in \mathscr{T}$.

We call $(X, \mathscr{T})$ a ***topological space***, and sometimes we might simply call X the topological space.

These conditions is the intrinsic propoties of the open sets we have defined in the metric space[1]. The topology consisting of all the open sets defined in the metric space $(\mathbb{R}; d_2)$ is called the ***standard topology*** of the n-dimension Euclidean space.

Definition 2.2 (Open set)**.** Topology $\mathscr{T}$'s elements are called ***open sets***, and their complements are called ***closed sets***.

[1] See proposition 1

Definition 2.3 (Base)**.** Let $(X, \mathscr{T})$ be a topological space, and $\mathfrak{B} \subset 2^X$. If $\forall G \in \mathscr{T}$, $\exists \{B_\alpha\}_{\alpha \in A} \in 2^{\mathfrak{B}}$ s.t. $\bigcup_{\alpha \in A} B_\alpha = G$, we called $\mathfrak{B}$ a (topological or open) ***base*** of the topology $\mathscr{T}$.

Definition 2.4 (Weight)**.** The smallest possible cardinity of a base of a topology is called the ***weight*** of the topological space.

Definition 2.5 (Neighbourhood)**.** If $x \in U(x)$ and $U(x) \in \mathscr{T}$, then $U(x)$ is a ***neighbourhood*** of x in topological space $(X, \mathscr{T})$. All neighbourhoods of a point x is denoted by $\mathscr{U}(x)$.

If $\mathring{U}(x) := U(x) - \{x\} \neq \varnothing$, then it is a ***deleted neighbourhood***. The collection of deleted neighbourhoods of x is denoted as $\mathring{\mathscr{U}}(x)$.

For example, we define an equivalence relation $\sim$ in $C(\mathbb{R};\mathbb{R})$. If $f, g \in C(\mathbb{R};\mathbb{R})$, at point $a \in \mathbb{R}$:

$$f \sim_a g \;\leftrightarrow\; \exists U(a)\big(\forall x \in U(a),\ f(x) = g(x)\big)\,. \tag{2-1}$$

By collecting all of the continuous functions that are euivalent to f, we call f define a ***germ*** at point a, denoted by f_a. If $f \in C(\mathbb{R};\mathbb{R})$ is defined in $U(a)$, then we can call $\{f_x \mid x \in U(a)\}$ a neighbourhood of germ f_a. Class of neighbourhoods of each f_x constructs a base of topological space $(C(\mathbb{R};\mathbb{R}); \mathscr{T})$, where $\mathscr{T}$ is made of the sets of germs of continuous function in $C(\mathbb{R};\mathbb{R})$.

Definition 2.6 (Hausdorff space)**.** We call a topological space $(X, \mathscr{T})$ a ***Hausdorff space***, ***separated space*** or T_2 ***space***, if $\forall x, y \in X$, $x \neq y \to \big(\exists U(x), U(y)$ s.t. $U(x) \cap U(y) = \varnothing\big)$[2].

Definition 2.7 (Dense set)**.** $E \subset X$ is a ***dense set*** in the topological space $(X, \mathscr{T})$, if $\forall x \in X$, $\forall U(x)$, $U(x) \cap E \neq \varnothing$.

Definition 2.8 (Separable)**.** If there is a *countable* dense set in topological space $(X, \mathscr{T})$, then $(X, \mathscr{T})$ is ***separable***.

We can also define interior points, exterior points, boundary points, limit points in topological space as in metric space.

Definition 2.9 (Topological subspace)**.** Each subset Y of X equiped with topology $\mathscr{T}$ can be given a ***subspace topology*** $\mathscr{T}_Y$ whose elements G_Y are intersections of the subset with an open set G in $(X, \mathscr{T})$ i.e. $\forall G_Y \in \mathscr{T}_Y$, $\exists G \in \mathscr{T}$ s.t. $G_Y = G \cap Y$. Subsets equiped with such topology construct a ***topological subspace*** $(Y, \mathscr{T}_Y)$.

If two topology $\mathscr{T}_1, \mathscr{T}_2$ are defined on the same X, $\mathscr{T}_1$ is said to be ***stronger*** than $\mathscr{T}_2$ if $\mathscr{T}_1 \subsetneq \mathscr{T}_2$.

Definition 2.10 (Direct product)**.** Let $(X_1, \mathscr{T}_1)$ and $(X_2, \mathscr{T}_2)$ be two topological spaces. Their ***direct product*** is defined as $(X_1 \times X_2, \mathscr{T})$, where $\mathscr{T}$ has a basis $\mathscr{B} := \{G_1 \times G_2 \mid G_1 \in \mathscr{T}_1 \wedge G_2 \in \mathscr{T}_2\}$.

§3 Compact Set

Definition 3.1 (Open cover)**.** Let $(X, \mathscr{T})$ be a topological space, $K \in 2^X$ and $\Omega \in 2^{\mathscr{T}}$. We call Ω to be an ***open cover*** over K, if $K \subset \cup\Omega$. If there are two open covers Ω, Ω' over K, and $\Omega' \subset \Omega$, we say that Ω' is a ***subcover*** of Ω.

[2]This definition is also called ***Hausdorff axiom*** or ***separation axiom***.

Definition 3.2 (Compact set)**.** A set $K \in 2^X$ in topological space $(X, \mathscr{T})$ is called a ***compact set*** if each of its open covers has a *finite* subcover.

Specially, $\varnothing$ is compact.

Theorem 3.1. *A set $K \subset X$ is compact in $(X, \mathscr{T})$* **iff** *K is compact in $(K, \mathscr{T}_K)$ itself.*

This theorem tells a truth that whether K is compact or not doesn't dependent on the topological space it's in. This fact can be easily proved: we just need to notice that every open set G_K in $(K, \mathscr{T}_K)$ is an intersection of an open set G in $(X, \mathscr{T})$ and K.

Theorem 3.2 (Compact $\to$ closed (Hausdorff))**.** *If K is compact in a Hausdorff space $(X, \mathscr{T})$*[3]*, then K is a closed set in $(X, \mathscr{T})$.*

Proof. Let x_0 be a limit point of K, which means $\forall U(x_0)$,

$$\operatorname{card} U(x_0) \cap K \notin \mathbb{N}.$$

Assume that $x_0 \notin K$. In a Hausdorff space, $\forall x \in K - \{x_0\}$, $\exists U(x)$ s.t. $U(x) \cap U(x_0) = \varnothing$. Such $U(x)$ construct an open cover $\Omega = \{U(x) | x \in K\} \subset 2^K$. Since K is compact, $\exists \Omega' \subset \Omega$ s.t. $\operatorname{card} \Omega \in \mathbb{N}$.

$$(\cup\Omega') \cap U(x_0) = \left(\bigcup_{k=1}^{n} U_k\right) \cap U(x_0) = \bigcup_{k=1}^{n} (U_k \cap U(x_0)) = \varnothing\,.$$

Since $K \subset \cup\Omega'$, x_0 is an exterior point of K, which leads to a contradiction.

Hence $x_0 \in K$. $\overline{K} = K$. □

Theorem 3.3. *Each decreasing* **nested sequences** *of non-empty compact sets has a non-empty limit, i.e. $\forall (K_n)_{n\in\mathbb{N}} \in \mathscr{P}(X)^{\mathbb{N}}$ s.t. $\forall n \in \mathbb{N}_+$, $K_n \supset K_{n+1} \wedge K_n \neq \varnothing \wedge$ (K_n is compact): $K_n \downarrow K \neq \varnothing$.*

Proof. Assume that $K = \varnothing$. Compact subsets of K_1 are all colsed, while their complements are all open. An open cover Ω can be constructed as $\{K_1 - K_n \mid n \in \mathbb{N}_+\}$. Since K_1 is compact, there would be a finite subcover $\Omega' \subset \Omega$, notice that $(X - K_n)_{n\in\mathbb{N}}$ is also a nested sequence, there must be one single $X - K_{n_0} \in \Omega'$ that covers K_1, which means $K_{n_0} = \varnothing$ contradicting that $\forall n \in \mathbb{N}_+$, K_n is non-empty. □

Theorem 3.4. *A Closed subset F of a compact set K is also compact.*

Proof. If $\Omega_F \subset 2^K$ is an open cover of F. Notice that $K - F$ is open, $\Omega = (\cup\Omega_F) \cap \{K - F\}$ constructs an open cover over over K. Since K is compact there must be a finite cover $\Omega' \subset \Omega$ which obviously also covers over F. □

The following propoties of compact sets are about topological spaces induced from metric spaces.

Definition 3.3 (net)**.** (X, d) is a metric space, $E \in 2^X$. E is called an ***ε-net*** if $\forall x \in X$,$\exists e \in E$, $d(e, x) < \varepsilon$.

Theorem 3.5 (Finite ε-net exists)**.** *If (K, d) is a compact metric space, then $\forall \varepsilon \in \mathbb{R}_+$, $\exists$* finite *ε-net in (K, d).*

[3]See definition 2.6.

Proof. For each point $x \in K$, find it a $B(x, \varepsilon)$, of which an infinite cover Ω over K is made. Since K is compact, there exists a finite subcover $\mathit{\Omega}' = \{B(x_i, \varepsilon)\}_{i \in n}$ ($n \in \mathbb{N}_+$). Therefore $\{x_i\}_{i \in n}$ is a finite ε-net in K. $\square$

Theorem 3.6 (Sequentially compact)**.** *A metric space* (K, d) *is compact* **iff** *it is* **sequentially compact***, that is,* $\forall (x_n)_{n \in \mathbb{N}} \in K^{\mathbb{N}}$*, it has a convergent subsequence* $(x_{k_n})_{n \in \mathbb{N}}$ *(*$k_n \in \mathbb{N}$*;* $k_{n+1} > k_n$*) whose limit* $a \in K$*.*

To prove Theorem 3.6, we need to prove two lemmata first.

Lemma 2. *If* (K, d) *is sequentially compact, then* $\forall \varepsilon \in \mathbb{R}_+$*,* $\exists$ *finite* ε*-net in* (K, d)*.*

Proof. Assume that $\exists \varepsilon_0 \in \mathbb{R}_+$, there were no finite ε_0-net in (K, d). Define such sequence: $(x_n)_{n \in \mathbb{N}}$ s.t. $\forall n \in \mathbb{N}\ \forall k \in n$, $d(x_n, x_k) \geq \varepsilon_0$ (There would always be a next one since there exists no finite ε_0-net or $\{B(x_n; \varepsilon_0)\}_{n \in N}$ gives such). It has no convergent subsequence: if there were a $(x_{k_n})_{n \in \mathbb{N}}$ convergent to $a \in K$, $\exists N, M \in \mathbb{N}_+$, $d(x_N, x_M) \leq d(x_N, a) + d(x_M, a) \leq \varepsilon_0$, which lead to a contradictary. $\square$

Lemma 3. *If* (K, d) *is sequentially compact then every nested sequence of closed non-empty sets* $\{F_n\}_{n \in \mathbb{N}}$ *in* K *have a non-empty intersection.*

Proof. Let $(x_{k_n})_{n \in \mathbb{N}}$ be a convergent subsequence of $(x_n)_{n \in \mathbb{N}}$, where $\forall n \in \mathbb{N}$, $x_n \in F_n$. Let a be the limit of $(x_{k_n})_{n \in \mathbb{N}}$.

Assume that $a \notin \bigcap_{n \in \mathbb{N}} F_n$, in a metric space, $\exists U(a) \in \mathscr{U}(a)$ s.t. $U(a) \cap \left(\bigcap_{n \in \mathbb{N}} F_n\right) = \varnothing$, therefore $U(a) \cap \left(\bigcap_{n \in \mathbb{N}} F_{k_n}\right) = \varnothing$. But this conflict the fact that $\exists N \in \mathbb{N}$, s.t. $n > N \to x_{k_n} \in U(a)$ while $x_{k_n} \in F_{k_n}$. $\square$

Then we get back to the Theorem 3.6.

Proof. $\to$: If $\operatorname{card}\{x_n\}_{n \in \mathbb{N}} \in \mathbb{N}$, it is obvious; Now we let $\operatorname{card}\{x_n\}_{n \in \mathbb{N}} \notin \mathbb{N}$. We can always find finite $1/k$-net $\{B(a_{k,i}, 1/k)\}_{i \in m}$ (Theorem 3.5, $m \in \mathbb{N}$, $a_i \in K$), for all $k \in \mathbb{N}_+$. For each k, there must be at least one $B(a_{k,i_0}; 1/k)$ (for simplication, we denote a_{k,i_0} by a_k) that includes infinite elements in $(x_n)_{n \in \mathbb{N}}$. $\forall n \in \mathbb{N}_+$ (let $k_0 = 0$), select $x_{k_n} \in B(a_{n,0}; 1/n)$, and $\{\overline{B}(x_n; 1/k)\}$ is a nested sequence of a closed non-empty sets in sequentially compact K, (Lemma 3) $\lim_{n \to \infty} x_{k_n} \in K$.

$\leftarrow$: Assume that there were an open cover $\mathit{\Omega}$ over K having no finite subcover, $\forall n \in \mathbb{N}_+$, $\exists$ finite $1/n$-net (Lemma 3), in which there would be at least one x_n whose $\overline{B}(x_n; \frac{1}{n})$ can't be covered finitely. Then $\overline{B}(x_n; 1/n) \downarrow B = \{a\}$ (Theorem 3.3) can't be finitely covered by any subcover of $\mathit{\Omega}$, which means $\mathit{\Omega}$ can't cover the whole K, leading to the contradiction. $\square$

We now prove a very useful special case for compact sets: compact sets in $\mathbb{R}$.

Lemma 4 (n-dimemsional cuboids are compact)**.** *Let* I *be a cuboid in* $\mathbb{R}^n$ *i.e.*

$$I := \{\boldsymbol{x} \in \mathbb{R}_n \mid a_i \leq x_i \leq b_i, \forall i \in n\}.$$

The cuboid I *is compact.*

Proof. We only need to prove that I is sequentially compact (Theorem 3.6). Let $(\boldsymbol{x}_i)_{i\in\mathbb{N}} \in I^{\mathbb{N}}$.

Denote $S_0 := I$. We divide S_m ($m \in \mathbb{N}$) into 2^n parts by equally dividing every $I_i := \{\boldsymbol{x} \in \mathbb{R}_n \mid a_i \leq x_i \leq b_i\}$ into two. Choose one that contains inifinite points of $(\boldsymbol{x}_i)_{i\in\mathbb{N}}$ as S_{m+1}. Then we get a closed nested sequence $S := (S_i)_{i\in\mathbb{N}}$. Notice that $\forall i \in \mathbb{N}$, S_i can be conceived as a product of n 1-dimension intervals. These intervals are also closed nested sequence, but in $\mathbb{R}$. We have leaned that $\exists! \boldsymbol{\xi} := (\xi_i)_{i\in n}$ s.t. $\{\boldsymbol{\xi}\} := \bigcap S$ from the theory of real numbers.

In every S_k we can find a $\boldsymbol{x}_{i_k}$, which is a convergent subsequence of the arbitary sequence $(\boldsymbol{x}_i)_{i\in\mathbb{N}}$. □

Theorem 3.7 (Compact iff closed and bounded in $\mathbb{R}^n$)**.** *Let $K \in \mathscr{P}(\mathbb{R}^n)$, $n \in \mathbb{N}_+$. The set K is compact* iff *it is closed and bounded.*

Proof. $\to$: We have proved that compact sets are closed in a Hausdorff space (Theorem 3.2). Now we prove that K is also bounded. Let $\boldsymbol{x} \in \mathbb{R}^n$, and we could find an open covers of K:

$$\Omega := \{B(\boldsymbol{x}; n) \mid n \in \mathbb{N}_+\}.$$

Assume that we find a finite subcover $\Omega' := \{B(\boldsymbol{x}; n_k) \mid k \in m\}$, then $d(K) < n_m$.

$\leftarrow$: Since K is bounded, we can find it a n-dimension cuboid I, which we have proved to be compact (Lemma 4). The closed set K in the compact set I is compact (Theorem 3.4).

□

§4 Connected Set

Definition 4.1 (Connected space)**.** Topological space $(X, \mathscr{T})$ is called ***connected*** if there is no ***open-closed set*** (i.e. both open and closed) besides $\varnothing$ and X itself.

Notice that if $A \in 2^X$ is open-closed, its complement $X - A$ is also open-closed, which means a topological space is connected ***iff*** it is not a union of its two open subsets.

Definition 4.2 (Connected set)**.** Let $(X, \mathscr{T})$ be a topological space. Subset C is said to be ***connected*** if subspace $(C, \mathscr{T}_C)$ is connected.

Theorem 4.1. *Let $(X, \mathscr{T})$ be a topological space, and $\{C_\alpha\}_{\alpha\in A}$ be connected subsets of X. If $\bigcap_{\alpha\in A} C_\alpha \neq \varnothing$, then $\bigcup\limits_{\alpha\in A} C_\alpha$ is also connected.*

Proof. Assume that $C = \bigcup_{\alpha\in A} C_\alpha$ were not connected, $\exists E \in 2^C$ s.t. $E \neq \varnothing$, $E \neq C$ and $E, C - E \in \mathscr{T}_C$. For E is not empty there exists a $\beta \in A$ s.t. $E \cap C_\beta \neq \varnothing$.

Now we show that $C_\beta \subset E$. Suppose that $C_\beta \not\subseteq E$, which implies that $(C - E) \cap C_\beta \neq \varnothing$. $E, C - E, C_\beta \in \mathscr{T}_C$, by the definition of the topology, $E \cap C_\beta, (C - E) \cap C_\beta \in \mathscr{T}_C$. This conflicts to the fact that C_β is connected. Therefore $C_\beta \subset E$.

Hence, there exists a $B \subsetneqq A$, $\bigcup_{\beta\in B} C_\beta = A$. Since C_γ, $\gamma \in A - B$ would have a empty intersection with E, which contradicts $\bigcap_{\alpha\in A} C_\alpha \neq \varnothing$. □

Theorem 4.2. *Connected sets have connected closure.*

Proof. □

Theorem 4.3. *$C \subset \mathbb{R}$ is connected* **iff** *$\forall x, z \in C \forall y \in \mathbb{R} (x < y < z) \to y \in C$.*

Proof. $\to$: Assume that there were such $y \in \mathbb{R}$ that $\exists x, z \in C$, $x < y < z$ but $y \notin C$. $\{x \in C \mid x < y\}$ and $\{x \in C \mid x > y\}$ are open in C for they are intersection of open sets in $\mathbb{R}$ and C. Since they're each other's complement, they are both open-closed, which conflicts to the definition of a connected set.

$\leftarrow$: It can be proved that $(\inf C, \sup C) \subset C$. Assume that there were an open-closed proper subset $E \neq \varnothing$ contained in C. Find two points $x \in E$, $z \in C - E$. Without loss of generality, let $x < z$. Since E and $C-E$ are closed, $c_1 = \inf\big(E \cap [a,b]\big) \in E$ while $c_2 = \inf\big((C-E)\cap[a,b]\big) \in C-E$. However $E \cap (C-E) = \varnothing$, hence $c_1 < c_2$, which means $(c_1, c_2) \cap E = \varnothing$. Here's the contradiction. $\square$

Definition 4.3 (Locally connected)**.** A topological space $(X, \mathscr{T})$ is said to be ***locally connected*** if $\forall x \in X$, $\exists U(x)$ s.t. $U(x)$ is connected.

§5 Complete Metric Spaces

We now take a closer look at one of the most important examples of metric spaces: complete spaces.

Definition 5.1 (Cauchy sequence)**.** A sequence $(x_n)_{n\in\mathbb{N}}$ of points in a metric space (X, d) is called a ***fundamental sequence*** or ***Cauchy sequence*** if $\forall \varepsilon \in \mathbb{R}_+$, $\exists N \in \mathbb{N}$ s.t. as long as $m, n > N$, $d(x_n, x_m) < \varepsilon$.

Definition 5.2 (complete space)**.** A metric space (X, d) is ***complete*** if any Cauchy sequence of its points is convergent.

For example, a metric space $C_\infty[a, b]$ is complete while $C_1[a, b]$ isn't. The proof see [2, p. 22].

Theorem 5.1 (Closed subspace of a complete space is complete)**.** *Let (X, d) be a complete space, A is a colsed set of X. The subspace (A, d) is also complete.*

Proof. Let $\langle x_n\rangle_{n\in\mathbb{N}} \in A^{\mathbb{N}}$ be a Cauchy sequence in A. Since X is complete, $\lim_{n\to\infty} x_n = x \in X$. If $x \notin A$, then $\forall U \in \mathscr{U}(x)$, $\operatorname{card}(U \cap A) = \infty$ i.e. x is a limit point of A. By Theorem 1.1, $x \in A$. $\square$

Let us consider an incomplete space $\mathbb{Q}_1$, which is a subspace of the complete space $\mathbb{R}_1$. If $\mathbb{R}_1$ is the smallest complete space containing $\mathbb{Q}_1$, we can say that we have achieved a ***completion*** of $\mathbb{Q}_1$. However, the term "smallest" hasn't been properly defined yet.

Definition 5.3 (completion)**.** If a metric space (X, d) is a subspace of a complete metric space (Y, d) and everywhere dense in it, we call the latter one the ***completion*** of (X, d).

We need to confirm that such completion is the smallest and unique. So we introduce:

Definition 5.4 (isometry)**.** If there exists a ***isometry*** $f\colon X_1 \to X_2$ when (X_1, d_1) and (X_2, d_2) are both metric space, i.e. f is a bijective and $\forall a, b \in X_1$, $d_2\big(f(a), f(b)\big) = d_1(a, b)$, then these two metric spaces are ***isometric***.

This relation is reflexive (id_X), symmetric (f^{-1}), and transitive ($f \circ g$), so it is a equivalence relation, denoted by $\sim$. We shall consider isometric spaces as identical, when only discussing within metric topological topics.

Theorem 5.2. *If metirc spaces (Y_1, d_1) and (Y_2, d_2) are both completions of (X, d), then they are isometric.*

Proof. Between two completions such isometry $f\colon Y_1 \to Y_2$ can be defined: if $x_1, x_2 \in X$,

$$d_2\big(f(x_1), f(x_2)\big) = d\big(f(x_1), f(x_2)\big) = d(x_1, x_2) = d_1(x_1, x_2).$$

For each $y_1 \in Y_1 - X_1$, a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ can be found in the nested sequence of balls centered in y_1. It is obvious that $(x_n)_{n\in\mathbb{N}}$ is also fundamental in Y_2, limitting to $y_2 \in Y_2$.

Differently selected sequences of points $(x'_n)_{n\in\mathbb{N}}$ won't limit to a different y'_2, namely $d(x_n, x'_n)$ shall converge to 0, or the fact that the radii of balls converge to 0 would be violated.

Let $f(y_1) = y_2$.

a) For each $y_2 \in Y_2 - X$, there always exists a Cauchy sequence converging to it, which implies that f is a surjection.

b) On the other hand, we shall notice that $\forall y'_1, y''_1 \in Y_1 - X$,

$$d_1(y'_1, y''_1) = \lim_{n\to\infty} d(x'_n, x''_n) = d_2(y'_2, y''_2)$$

while $(x'_n)_{n\in\mathbb{N}}$ and $(x''_n)_{n\in\mathbb{N}}$ are both Cauchy sequence. This equality proved that f is a injection.

□

Theorem 5.3. *There always exists a completion for every metric space.*

Proof. Let $C_X := \{(x_n)_{n\in\mathbb{N}} \in X^{\mathbb{N}} \mid \forall\varepsilon \in \mathbb{R}_+, \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \in \mathbb{N}(n > N \wedge m > N \to d_X(x_n, x_m) < \varepsilon)\}$, namely the collections of Cauchy sequences in X.

We say two Cauchy sequences $(x_n)_{n\in\mathbb{N}}$, $(x'_n)_{n\in\mathbb{N}}$ are equivalent (or, we shall say in a complete space, that they have a same limit) if $\lim_{n\to\infty} d(x_n, x'_n) = 0$.

It can be easily proved that such relation is a equivalence relation, and it divides C_X into equivalence classes S.

$\forall (x_n)_{n\in\mathbb{N}}, (x'_n)_{n\in\mathbb{N}} \in C_X$, $\forall\varepsilon \in \mathbb{R}_+$, $\exists N \in \mathbb{N}$ s.t. $\forall n, m \in \mathbb{N}$, as long as $n > N$ and $m > N$ (by Lemma 1):

$$|d_X(x_n, x'_n) - d_X(x_m, x'_m)| \le d_X(x_n, x_m) + d_X(x'_n, x'_m) < 2\varepsilon\,.$$

Hence, $(d(x_n, x'_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}_1$. Since $\mathbb{R}_1$ is a complete space, $\lim_{n\to\infty} d(x_n, x'_n)$ always exists. This fact allows us to introduce[4]:

$$d\colon S^2 \to \mathbb{R};\ \big([(x_n)_{n\in\mathbb{N}}], [(x'_n)_{n\in\mathbb{N}}]\big) \mapsto \lim_{n\to\infty} d(x_n, x'_n)$$

A metric space (S_X, d) isometric to any given metric space (X, d_X) can be constructed, where $S_X := \{[(x)_{n\in\mathbb{N}}] \mid x \in X\}$.

Then we shall show that S is the completion of S_X.

Let $\big([(x^i_n)_{n\in\mathbb{N}}]\big)_{i\in\mathbb{N}}$ be a Cauchy sequence in S. By definition, for any $i \in \mathbb{N}_+$, there exists a N that is large enough such that as long as $j > N$, $k > N$, $d_X(x^i_j, x^i_k) < 1/i$. Choose $a^i := x^i_k$ for such $k > N$, so that $d\big([(a^i)_{n\in\mathbb{N}}], [(x^i_n)_{n\in\mathbb{N}}]\big) < 1/i$.

$\forall\varepsilon \in \mathbb{R}_+$, $\exists N \in \mathbb{N}$ (e.g. we can choose $N = \lfloor 4/\varepsilon \rfloor$) s.t. $\forall n, m \in \mathbb{N}$, $p > N \wedge q > N \;\to$

$$d\big([(x^p_n)_{n\in\mathbb{N}}], [(x^q_n)_{n\in\mathbb{N}}]\big) < \frac{\varepsilon}{2} \wedge d\big([(x^p_n)_{n\in\mathbb{N}}], [(a^p)_{n\in\mathbb{N}}]\big) < \frac{1}{p} \wedge d\big([(x^q_n)_{n\in\mathbb{N}}], [(a^q)_{n\in\mathbb{N}}]\big) < \frac{1}{q}\,,$$

[4] We implicitly use the (countable) axiom of choice: we must find a Cauchy sequence for each equivalence class.

therefore when p, q are great enough, (by the triangle inequality)

$$d\big([(a^p)_{n\in\mathbb{N}}], [(a^q)_{n\in\mathbb{N}}]\big) \le \frac{\varepsilon}{2} + \frac{1}{p} + \frac{1}{q} < \varepsilon\,.$$

So, $[(a^n)_{n\in\mathbb{N}}]$ is a Cauchy sequence, therefore it is an element of S.

By $\lim_{i\to\infty} d\big([(x^i_n)_{n\in\mathbb{N}}], [(a^n)_{n\in\mathbb{N}}]\big) = 0$, we found a limit for the arbitary Cauchy sequence $\big([(x^i_n)_{n\in\mathbb{N}}]\big)_{i\in\mathbb{N}}$ in S.

Finally, we have to check that S_X is everywhere dense in S. For any aribitary $[(x_n)_{n\in\mathbb{N}}] \in S$, $\forall\varepsilon$, we can always choose a $N \in \mathbb{N}$ great enough so that $[(x_N)_{n\in\mathbb{N}}] \in S_X \cap B\big([(x_n)_{n\in\mathbb{N}}], \varepsilon\big)$. Since every neighbourhood of $[(x_n)_{n\in\mathbb{N}}]$ contains a ball centred at it, we have proved that $\forall U \in \mathscr{U}([(x_n)_{n\in\mathbb{N}}])\big(U \cap S_X \neq \varnothing\big)$. □

Note: We have already seem such technique when we construct the real numbers from the sequences of rational numbers.

§6 Continuous Mapping

Let's recall the definition of the limitation.

Definition 6.1 (Filter base)**.** A set $\mathscr{B} \subset 2^X$ is called a **(filter) base** in X if the following conditions hold:

a) $\varnothing \notin \mathscr{B}$.
b) $\forall B_1, B_2 \in \mathscr{B}$, $\exists B \in \mathscr{B}$ s.t. $B \subset B_1 \cap B_2 \subset B_2$.

Here is a list of some importants filter bases:

(1) $x \to a$, where $a \in X$, means $\mathring{\mathscr{U}}(a)$;

(2) $x \to \infty$, means $\{V \mid X - V \in \mathscr{U}(a) - \{X\}\}$;

(3) $E \ni x \to a$, means $\{\mathring{U}(a) \cap E \mid \mathring{U}(a) \in \mathring{\mathscr{U}}(a)\}$;

(4) $E \ni x \to \infty$, means $\{E \cap V \mid X - V \in \mathscr{U}(a) - \{X\}\}$.

Introduction of the limits in a topological space is as follows.

Definition 6.2 (Limit)**.** Let $a \in Y$ be the ***limit*** over the base $\mathscr{B} \subset 2^{\mathscr{D}(f)}$ of a mapping $f\colon \mathscr{D}(f) \to Y$, in which Y is epuiped with a topology $\mathscr{T}$.

$$\lim_{\mathscr{B}} f = a \quad := \quad \forall U(a) \in \mathscr{U}(a)\ \exists B \in \mathscr{B}(f(B) \subset U(a)).$$

Such definition is parallel to the definition we have introduced on the limits of real number, hence it basically holds the same propoties, except for:

Theorem 6.1 (Uniqueness of limit in Hausdorff space)**.** *Let Y be a Hausdorff space, $\mathscr{B}$ be a filter base in X, $f \in Y^X$. The limit of f over $\mathscr{B}$ is unique.*

Definition 6.3 (Oscillation)**.** Let X, Y be two topological spaces, $f \in Y^X$, $E \in \mathscr{P}(X)$.

$$\omega(f;E) := \sup\{d_Y(f(x_1), f(x_2)) \mid x_1, x_2 \in E\}$$

is called the ***oscillation*** of the function f in set E. We can also define the ***oscillation*** of f at a point $x \in X$ as

$$\omega(f;x) := \inf\{\omega(f;B) \mid B \in \mathscr{B}\},$$

where $\mathscr{B}$ is a filter base that $\cap\mathscr{B} = \{x\}$.

Theorem 6.2 (Cauchy criterion for existance of limit)**.** *Let $\mathscr{B}$ be a filter base in X, (Y, d) be a complete metric space, and $f \in Y^X$. The mapping f has a limit over base $\mathscr{B}$* **iff** *$\forall\varepsilon \in \mathbb{R}_+$, $\exists B \in \mathscr{B}$ s.t. $\omega(f;B) < \varepsilon$.*

Proof. $\to$: Denote $a := \lim_{\mathscr{B}} f$. $\forall\varepsilon$, $\exists B \in \mathscr{B}$ s.t. $f(B) \subseteq B(a;\varepsilon/2)$

$$\forall x, x' \in B, \quad d(f(x), f(x')) \le d(f(x), a) + d(f(x'), a) < \varepsilon\,.$$

$\leftarrow$: $\forall n \in \mathbb{N}_+$, $\exists B_n \in \mathscr{B}$ s.t. $\omega(f;B_n) < 1/n$. Since $B_n \neq \varnothing$ (the definition of filter base), we can choose[5] $x_n \in B_n$ for any n, so that we get a sequence $\langle f(x_n)\rangle_{n\in\mathbb{N}} \in Y^{\mathbb{N}}$. Let $x \in B_n \cap B_m$ for any m, n that $m > 1/\varepsilon$, $n > 1/2\varepsilon$ for any ε

$$d(f(x_n), f(x_m)) \le d(f(x_n), f(x)) + d(f(x_m), f(x)) < \varepsilon\,,$$

hence $\langle f(x_n)\rangle_{n\in\mathbb{N}}$ is a Cauchy sequence, by the completeness of Y we can find a limit a for it.

Let $m \to \infty$ we get $d(f(x_n), a) \le \varepsilon$. This inequality holds for any ε and n great enough. $\forall x' \in B_n$,

$$d(f(x'), a) \le d(f(x'), f(x_n)) + d(f(x_n), a) < \frac{1}{n} + \varepsilon\,,$$

the right-hand side can be arbitary small, if n is even greater. □

Definition 6.4 (Continuity)**.** A mapping $f\colon X \to Y$, where X, Y is equiped with topology $\mathscr{T}_X$, $\mathscr{T}_Y$, respectively, is said to be ***continuous*** at $x_0 \in X$ (let $y_0 = f(x_0) \in Y$), if $\forall U(y_0)$, $\exists U(x_0)$ s.t. $f(U(x_0)) \subset U(y_0)$. It is ***continuous*** in X if it is continuous at each point $x \in X$.

The set of continuous mappings from X into Y can be denoted by $C(X, Y)$ or $C(X)$ when Y is clear.

Theorem 6.3 (**Criterion for continuity**)**.** *Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be two topological spaces, $f \in Y^X$. The function f is continuous* **iff** *$\forall G_Y \in \mathscr{T}_Y$, $f^{-1}(G_Y) \in \mathscr{T}_X$.*

Proof. $\to$: It is obvious if $f^{-1}(G_Y) = \varnothing$. Hence we assume that $f^{-1}(G_Y) \neq \varnothing$. Let $x_0 \in X$. Since $f \in C(X, Y)$, for G_Y, $\exists U(x_0)$ s.t. $f\big(U(x_0)\big) \subset G_Y$. Also notice that $f\big(U(x_0)\big) \subset G_Y \Rightarrow U(x_0) \subset f^{-1}(G_Y)$, therefore $f^{-1}(G_Y)$ is open.

$\leftarrow$: $\forall x_0 \in X$, let $y_0 = f(x_0)$, $f^{-1}(U(y_0)) \in \mathscr{T}_X$. Notice that $x_0 \in f^{-1}(U(y_0))$, therefore $f \in C(X, Y)$. □

Definition 6.5 (Homeomorphism)**.** Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be two topological spaces. A bijective mapping $f\colon X \to Y$ is a ***homeomorphism*** if $f \in C(X, Y) \wedge f^{-1} \in C(Y, X)$.

[5] I don't know any proof that can avoid using axiom of choices

Definition 6.6 (Homeomorphic spaces)**.** Two topological spaces $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ are said to be ***homeomorphic*** if there exists a homeomorphism $f\colon X \to Y$.

Homeomorphic topological spaces are identical with respect to their topological propoties since the theorem 6.3 has shown that their open sets correspond to each other.

Theorem 6.4 (Continuity of compositions of functions)**.** *Let X, Y, Z be three topological spaces, $E \in \mathscr{P}(X)$. $f \in C(E, Y)$, $g \in C(f(E), Z)$, then*

$$g \circ f \in C(E, Z)\,.$$

Theorem 6.5 (Continuous then locally bounded)**.** *Let $(X, \mathscr{T})$ be a topological space and (Y, d) be a metric space, $f \in Y^X$, $x \in X$. If f is continuous at x, then $\exists U(x) \in \mathscr{U}(x)$ s.t. $U(x)$ is bounded.*

Theorem 6.6 (Continuous iff oscillation is zero)**.** *Let X be a topological space and Y be a metric space, $f \in Y^X$, $x \in X$. The function f is continuous at x* **iff** *$\omega(f; x) = 0$.*

Then we shall introduce some global properties of continuous mappings.

Theorem 6.7 (Conservation of compactness)**.** *Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be two topological spaces. Let $K \subset X$ be a compact set. If $f\colon X \to Y \in C(X, Y)$, then $f(K)$ is compact.*

Proof. For each open cover $\Omega_Y = \{G_Y \in \mathscr{T}_Y\} \subset \mathscr{T}_Y$ over $f(K)$, $f^{-1}(G_Y) \in \mathscr{T}_X$ (Therem 6.3). $f(K) \subset \cup\Omega_Y \Rightarrow K \subset f^{-1}(\cup\Omega_Y) = \cup\Omega_X$, where $\Omega_X = \{f^{-1}(G_Y) \mid G_Y \in \Omega_Y\}$ is an open cover over K. Since K is compact, $\exists \Omega'_X \subset \Omega_X\,(|\Omega'_X| \in \mathbb{N}_+ \wedge K \subset \cup\Omega'_X)$, $f(K) \subset f(\cup\Omega'_X)$. $f(G'_X) \in \Omega_Y$, hence $\Omega'_Y = \{f(G'_X) \mid G'_X \in \Omega'_X\}$ is a finite subcover over $f(K)$. □

Theorem 6.8 (Weierstrass maximum-value theorem)**.** *Let K be a compact topological space, and $f \in C(K, \mathbb{R})$. $\exists x_m, x_M \in K$, s.t. $f(x_m) = m := \inf f(K)$, $f(x_M) = M := \sup f(K)$.*

Proof. By Theorem 6.7, $f(K)$ is also compact, and therefore closed and bounded (Theorem **??**). If $M \notin f(K)$, then open covers $\{B(M; (M-m)/n) - \tilde{B}(M; (M-m)/(n+1)) \mid n \in \mathbb{N}_+\}$ would not have a finite subcover, which is a contradiction to the compactness of $f(K)$. □

Theorem 6.9 (Bijective from compact space to Hausdorff space is homeomorphism)**.** *Let $(K, \mathscr{T}_K)$ be a compact space and $(Y, \mathscr{T}_Y)$ be a Hausdorff space. Let $f \in Y^K$ be a bijective. If $f \in C(K, Y)$, then f is a homeomorphism.*

Proof. $\forall F = K - G$ s.t. $G \in \mathscr{T}_K$ is compact (Theorem 3.4). Hence $f(F)$ is compact (Theorem 6.7), then it is also closed (Theorem 3.2). This fact shows that f^{-1} is continuous (Theorem 6.3). □

Definition 6.7 (Uniformly continuous)**.** Let (X, d_X), (Y, d_Y) be metric spaces, $f \in Y^X$. If $\forall \varepsilon \in \mathbb{R}_+$, $\exists \delta \in \mathbb{R}$, $\forall x \in X$ s.t. $\forall E \in \mathscr{P}(X)$,

$$d_X E < \delta \quad \to \quad \omega(f; E) < \varepsilon\,,$$

then f is said to be a ***uniformly continuous*** mapping.

Theorem 6.10 (Heine-Cantor theorem)**.** *Let (K, d_K) be a compact metric space, and (Y, d_Y) be a metric space. $\forall f \in C(K, Y)$, f is uniformly continuous.*

Proof. $\forall \varepsilon \in \mathbb{R}_+$, we can find it a collections of open balls

$$\Omega = \big\{ B(x;\delta(x)/2) \mid x \in X,\, \omega\big(f; B(x;\delta(x))\big) < \varepsilon \big\}\,,$$

that covers the compact set K, then there exists a finite subcover $\Omega' = \big\{B(x_i;\delta(x_i)/2)\big\}_{i\in n}$. Let $\delta := \min\{\delta(x_i)\}_{i\in n}$.

$\forall x', x'' \in K$, $\exists i \in n$, $x' \in B(x_i;\delta(x_i)/2)$, if $d(x',x'') < \delta$,

$$\delta(x'', x_i) \le \delta(x', x'') + \delta(x', x'') < \delta + \delta(x_i) \le \delta(x_i)\,,$$

therefore $x', x'' \in B(x_i;\delta(x_i))$, we have assume that $\omega(f; B(x_i;\delta(x_i)))$. □

Theorem 6.11 (Cantor (generalised)). *Let K be a compact set, $f \in \mathbb{R}^K$. If $\forall x \in K$, $\omega(f,x) \le \omega_0$, then $\forall \varepsilon \in \mathbb{R}_+$, $\exists \delta \in \mathbb{R}_+$ s.t. $\forall \boldsymbol{x} \in K$, $\omega(f, B_K(\boldsymbol{x};\delta)) < \omega_0 + \varepsilon$.*

Proof. We will get the proof if we repeat the prove of Theorem 6.10, only to replace ε in the definition of Ω by $\omega_0 + \varepsilon$. □

Theorem 6.12 (Conservation of connectedness). *Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be two topological spaces, and $E \subset X$ be a connected set. If $f \in C(X,Y)$, then $f(E)$ is also connected.*

Proof. Only to notice that the open-closed sets in $(f(E), \mathscr{T}_{f(E)})$ have concurrently open-closed pre-images in $(E, \mathscr{T}_E)$. □

Theorem 6.13 (Intermediate-value theorem). *Let $(X, \mathscr{T})$ be a connected topological space, and $f \in C(X, \mathbb{R})$, $f(a) = A$, $f(b) = B$, $A < B$. $\forall C \in [A,B]$, $\exists c \in X$, $f(c) = C$.*

Proof. by Theorem 6.12, $f(X)$ must be a connected set. Hence by Theorem 4.3, we know that $\forall C \in [A,B]$, $C \in f(X)$. □

§7 Contraction

Definition 7.1 (Fixed point). A point $a \in X$ is a ***fixed point*** of a mapping $f\colon X \to X$ if $f(a) = a$.

Definition 7.2 (Contraction). Let (X,d) be a metric space. A mapping $f\colon X \to X$ is called a ***contraction*** if $\exists q \in (0,1) \subset \mathbb{R}$ s.t. $\forall x_1, x_2 \in X$,

$$d(f(x_1), f(x_2)) \le q d(x_1, x_2). \tag{7-1}$$

Lemma 5. *A contraction $f\colon X \to X$ is always continuous.*

Proof. $\forall x \in X$, $\forall \varepsilon \in \mathbb{R}_+$, $\exists \delta < \varepsilon/q$, according to inequality 7-1:

$$f\big(B(x;\delta)\big) \subset B\big(f(x);\varepsilon\big).$$

□

Theorem 7.1 (**Picard-Banach fixed-point principle** or **contraction mapping principle**). *Let (X,d) be a complete metric space. Each contraction $f\colon X \to X$ has a unique fixed point a. Also, $\forall \{x_n\} \subset X$ s.t. $\forall n \in \mathbb{N}\big(f(x_n) = x_{n+1}\big)$ then $\lim_{n\to\infty} x_n = a$, and*

$$d(x_n, a) \le \frac{q^n}{1-q} d(x_1, x_0). \tag{7-2}$$

Proof. By the inequality 7-1:

$$d(x_{n+1}, x_n) \leq q d(x_n, x_{n-1}) \leq \cdots \leq q^n d(x_1, x_0)$$

Therefore, $\forall n, k \in \mathbb{N}$,

$$d(x_{n+k}, x_n) \leq \sum_{i=0}^{k-1} d(x_{n+i+1}, x_{n+i}) \leq \sum_{i=0}^{k-1} q^{n+i} d(x_1, x_0) \leq \frac{q^n}{1-q} d(x_1, x_0)\,, \tag{7-3}$$

which implies that $\langle x_n \rangle_{n \in \mathbb{N}}$ is a Cauchy sequence in a complete space (X, d), hence it converges to a point $a \in X$.

To proof that a is a fixed point of f, since f is continuous (Lemma 5), just notice that

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

If there were another fixed point $a' \in X$ of f, then:

$$0 \leq d(a, a') = d(f(a), f(a')) \leq q d(a, a')$$

which can't be true unless $a = a'$.

By passing to the limit as $k \to \infty$ in the inequality 7-3, we have the inequality 7-2. □

If the factor q is not limited within 1, we obtain:

Definition 7.3 (Lipschitz continuity)**.** Let (X, d_X), (Y, d_Y) be two metric spaces, $f \in Y^X$. If $\exists M \in \mathbb{R}_+$ s.t. $\forall x_1, x_2 \in X$,

$$d_Y(f(x_1), f(x_2)) \leq M d_X(x_1, x_2)\,, \tag{7-4}$$

then f is said to be ***Lipschitz continuous***. Inequality 7-4 is called the ***Libschitz condition***.

It is almost obvious that a Lipschitz continuous mapping is continuous.

Chapter 2

Normed Linear Space and Differential Calculus

§8 Normed Linear Space

Definition 8.1 (Norm)**.** Let V be a linear space over $\mathbb{R}$ or $\mathbb{C}$. A function $\| \,\| \colon X \to \mathbb{R}$ assigning to each vector $\boldsymbol{x} \in X$ a real number $\|\boldsymbol{x}\|$ is called a ***norm*** in the linear space X if:

a) $\|\boldsymbol{x}\| = 0 \;\leftrightarrow\; \boldsymbol{x} = \boldsymbol{0}$ (nondegeneracy);

b) $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$ (homogeneity);

c) $\|\boldsymbol{x}_1 + \boldsymbol{x}_2\| \le \|\boldsymbol{x}_1\| + \|\boldsymbol{x}_2\|$ (the triangle inequality).

A linear space with a norm defined on it is said to be ***normed***.

Over every normed space a distance can be defined as:

$$d(\boldsymbol{x}_1, \boldsymbol{x}_2) = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \tag{8-1}$$

Definition 8.2 (Banach space)**.** Let V be a normed space. If (V, d) is a complete space, where the distance d is defined as Eq. (8-1), then we call V a ***complete normed space*** or ***Banach space***.

Definition 8.3 (Hermitian form)**.** A linear space X on the complex field $\mathbb{C}$ is said to be given a ***Hermitian space*** if there is a mapping $\langle , \rangle \colon X^2 \to \mathbb{C}$ defined, s.t. $\forall \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \in X$, $\forall \lambda \in \mathbb{C}$.

a) $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \overline{\langle \boldsymbol{x}_2, \boldsymbol{x}_1 \rangle}$;

b) $\langle \lambda \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \lambda \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$;

c) $\langle \boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{x}_3 \rangle = \langle \boldsymbol{x}_1, \boldsymbol{x}_3 \rangle + \langle \boldsymbol{x}_2, \boldsymbol{x}_3 \rangle$.

A Hermitian form is said to be ***positive semi-definite***, if $\forall \boldsymbol{x} \in X$, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \ge 0$[1]. A Hermitian form is said to be ***degenerate***, if $\exists \boldsymbol{x} \in X - \{\boldsymbol{0}\}$ s.t. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$. A Hermitian form that is not degenerate is said to be ***non-degenerate***.

Definition 8.4 (Inner product)**.** A non-degenerate positive semi-definite Hermitian form[2]is said to be an ***inner product***. A space equiped with an inner product is said to be a ***inner product space***.

[1] $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \overline{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$, hence $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \in \mathbb{R}$.

[2] Equivalently, a positive definite Hermitian form.

Theorem 8.1 (Cauchy-Bunyakovskii's inequality)**.** *A linear space X on the complex field $\mathbb{C}$ is equiped with an inner product $\langle , \rangle$. $\forall \boldsymbol{x}, \boldsymbol{y} \in X$,*

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{y}, \boldsymbol{y} \rangle \,. \tag{8-2}$$

Proof. The theorem is trivial as $\boldsymbol{y} = \boldsymbol{0}$. Let us assume that $\boldsymbol{y} \neq \boldsymbol{0}$, therefore $\langle \boldsymbol{y}, \boldsymbol{y} \rangle > 0$.

$\forall \lambda \in \mathbb{C}$,

$$0 \leq \langle \boldsymbol{x} + \lambda \boldsymbol{y}, \boldsymbol{x} + \lambda \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle + \lambda \overline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle} + \overline{\lambda} \langle \boldsymbol{x}, \boldsymbol{y} \rangle + |\lambda|^2 \langle \boldsymbol{y}, \boldsymbol{y} \rangle$$

Let $\lambda = -\langle \boldsymbol{x}, \boldsymbol{y} \rangle / \langle \boldsymbol{y}, \boldsymbol{y} \rangle$, we have:

$$0 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle - \frac{|\langle \boldsymbol{x}, \boldsymbol{y} \rangle|^2}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle} \,.$$

□

By the theorem 8.1 we can claim that a linear space on complex number with an inner product $\langle , \rangle$ induces a norm

$$\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} \,, \tag{8-3}$$

and a metric

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| \,. \tag{8-4}$$

Theorem 8.2 (Continuity of norm)**.** *Let X be a normed space with a norm $\| * \|$. The mapping $\| * \| \in \mathbb{R}^X$ is continuous in X.*

Proof. $\forall \boldsymbol{x} \in X$, $\forall \varepsilon \in \mathbb{R}_+$, if $\|\Delta \boldsymbol{x}\| < \varepsilon$, then

$$\|\boldsymbol{x} + \Delta \boldsymbol{x}\| \leq \|\boldsymbol{x}\| + \|\Delta \boldsymbol{x}\| < \|\boldsymbol{x}\| + \varepsilon \,.$$

□

Definition 8.5 (Hilbert space)**.** If a linear space is equiped with an inner poduct, and together with its induced metric constructs a complete metric space, we call it a ***Hilbert space***. If the induced metric space is not complete, we shall call it a ***pre-Hilbert space***.

§9 Linear Operators

Definition 9.1 (Norm)**.** Let $\mathscr{A}$ be a n-multilinear operator space over normed space $(\boldsymbol{X}_i)_{i \in n}$ to a normed space Y i.e. $\mathscr{A} \in \mathcal{L}(X_0, X_1, \cdots, X_{n-1}; Y)$. We define the norm $\|\mathscr{A}\|$ as:

$$\|\mathscr{A}\| := \sup \left\{ \frac{\|\mathscr{A}(\boldsymbol{x}_i)_{i \in n}\|_Y}{\prod_{i \in n} \|\boldsymbol{x}_i\|_{X_i}} \middle| \forall i \in n,\ \boldsymbol{x}_i \in X_i - \{\boldsymbol{0}\} \right\} \,, \tag{9-1}$$

where the subscripts denote which spaces the norms are defined in.

The following theorem gives an equivalent definition:

Theorem 9.1. *Let $\mathscr{A} \in \mathcal{L}(X_0, X_1, \cdots, X_{n-1}; Y)$.*

$$\|\mathscr{A}\| = \{\|\mathscr{A}(\boldsymbol{e}_i)_{i \in n}\|_Y \mid \forall i \in n,\ \boldsymbol{e}_i \in X_i \wedge \|\boldsymbol{e}_i\|_{X_i} = 1\} \,. \tag{9-2}$$

Theorem 9.2. *Let $\mathscr{A} \in \mathcal{L}(X_0, X_1, \cdots, X_{n-1}; Y)$, and let $\|\mathscr{A}\| < \infty$.*

$$\|\mathscr{A}(\boldsymbol{x}_i)_{i\in n}\|_Y \le \|\mathscr{A}\| \prod_{i\in n} \|\boldsymbol{x}_i\|_{X_i} \,. \tag{9-3}$$

Definition 9.2 (Bounded linear operators)**.** Let $\mathscr{A} \in \mathcal{L}(X_0, X_1, \cdots, X_{n-1}; Y)$. If $\|\mathscr{A}\| < \infty$, then $\mathscr{A}$ is said to be ***bounded***.

Theorem 9.3 (Continuous at zero iff bounded)**.** *Let $\mathscr{A} \in \mathcal{L}(X_0, X_1, \cdots, X_{n-1}; Y)$. Denote $\prod_{i\in n} X_i$ by X. The operator $\mathscr{A}$ is continuous at $\mathbf{0} \in X$[3] **iff** it is bounded.*

Proof. First assume that $\mathscr{A}$ is bounded.

When $\|\mathscr{A}\| = 0$ it is trivial. Hence we assume that $\|\mathscr{A}\| > 0$.

$\forall \varepsilon \in \mathbb{R}_+$, if $\Delta\boldsymbol{x} := (\Delta\boldsymbol{x}_i)_{i\in n} \in X$ meets the condition that $\forall i \in n$, $\|\Delta\boldsymbol{x}_i\|_{X_i} < \sqrt[n]{\varepsilon/\|\mathscr{A}\|}$ then

$$\begin{aligned} d_Y(\mathscr{A}(\mathbf{0} + \Delta\boldsymbol{x}), \mathscr{A}(\mathbf{0})) &= d_Y(\mathscr{A}(\Delta\boldsymbol{x}), \mathbf{0}) = \|\mathscr{A}(\Delta\boldsymbol{x})\|_Y \\ &\le \|\mathscr{A}\| \prod_{i\in n} \|\Delta\boldsymbol{x}_i\|_{X_i} < \varepsilon \,. \end{aligned}$$

Then we assume that $\mathscr{A}$ is continuous at $\mathbf{0}$.

Set any positive $\varepsilon \in \mathbb{R}_+$, $\exists \delta \in \mathbb{R}_+$, when $\forall i \in n$, $\boldsymbol{x}_i \in X_i - \{\mathbf{0}\}$ and $\|\boldsymbol{x}_i\|_{X_i} \le \delta$, $\|\mathscr{A}(\boldsymbol{x})\| \le \varepsilon$.

Since every unit vector $\boldsymbol{e}_i$ can be written as $\delta\boldsymbol{e}_i/\delta$, where $\delta\boldsymbol{e}_i \in X_i - \{\mathbf{0}\}$ and $\|\delta\boldsymbol{e}_i\|_{X_i} = \delta$, then

$$\|\mathscr{A}(\boldsymbol{e}_i)_{i\in n}\|_Y = \frac{1}{\delta^n} \|\mathscr{A}(\delta e_i)_{i\in n}\|_Y \le \frac{\varepsilon}{\delta^n} \,,$$

which implies that the operator $\mathscr{A}$ is bounded. □

Theorem 9.4 (Continuous at zero then at everywhere)**.** *Let $\mathscr{A} \in \mathcal{L}(X_0, X_1, \cdots, X_{n-1}; Y)$. Denote $\prod_{i\in n} X_i$ by X. If the operator is continuous at $\mathbf{0} \in X$, then it is continuous in X.*

Proof. By theorem 9.3, we have learned that an operator continuous at $\mathbf{0}$ is bounded.

$\forall \boldsymbol{x}, \Delta\boldsymbol{x} \in X$,

$$\begin{aligned} & d_Y(\mathscr{A}(\boldsymbol{x} + \Delta\boldsymbol{x}), \mathscr{A}(\boldsymbol{x})) = \|\mathscr{A}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \mathscr{A}(\boldsymbol{x})\|_Y \\ & \quad = \Big\| \mathscr{A}(\Delta\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \mathscr{A}(\boldsymbol{x}_1, \Delta\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \Delta\boldsymbol{x}_{n-1}) \\ & \qquad + \mathscr{A}(\Delta\boldsymbol{x}_0, \Delta\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\boldsymbol{x}_0, \cdots, \Delta\boldsymbol{x}_{n-2}, \Delta\boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\Delta\boldsymbol{x}) \Big\|_Y \\ & \quad \le \|\mathscr{A}(\Delta\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1})\|_Y + \cdots + \|\mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \Delta\boldsymbol{x}_{n-1})\|_Y \\ & \qquad + \cdots + \|\mathscr{A}(\Delta\boldsymbol{x})\|_Y \\ & \quad \le \|\mathscr{A}\| \sum_{S\in\mathscr{P}(n)-\{\varnothing\}} \prod_{i\in n-S} \|\boldsymbol{x}_i\|_{X_i} \prod_{j\in S} \|\Delta\boldsymbol{x}_j\|_{X_j} \,. \end{aligned}$$

By setting $\max\{\|x_i\|_{X_i} \mid i \in n\} < \varepsilon \max\left\{ \sqrt[n]{\prod_{i\in n-S} \|\boldsymbol{x}_i\|_{X_i}} \mid S \in \mathscr{P}(n) - \{\varnothing\}\right\}/(2^n - 1)\|\mathscr{A}\|$ we have $d_Y(\mathscr{A}(\boldsymbol{x} + \Delta\boldsymbol{x}), \mathscr{A}(\boldsymbol{x})) < \varepsilon$ for any $\varepsilon \in \mathbb{R}_+$. □

[3] Be reminiscent of the Defintion 2.10

Theorem 9.3 and Theorem 9.4 show the equivalence for linear operators of being bounded and being continuous. We shall denote the space of all the bounded n-multilinear operators from X_0, $\cdots$, X_{n-1} to Y by $\mathcal{B}(X_0, \cdots, X_{n-1}; Y)$.

Corollary 1 (Linear operators from finite dimensional space are continuous)**.** *If* $\forall i \in n$, $\dim X_i < \infty$, *then*

$$\mathcal{L}(X_0, \cdots, X_{n-1}; Y) = \mathcal{B}(X_0, \cdots, X_{n-1}; Y)\,.$$

Corollary 2 (Continuous at a point then at everywhere)**.** *Let* $\mathscr{A} \in \mathcal{L}(X_0, X_1, \cdots, X_{n-1}; Y)$. *Denote* $\prod_{i\in n} X_i$ *by* X, *and Let* $\boldsymbol{x} = (\boldsymbol{x}_i)_{i\in n} \in X$. *If the operator is continuous at* $\boldsymbol{x}$, *then it is continuous in* X.

Proof. □

Definition 9.3 (Isomorphism)**.** Two normed space are ***isomorphic*** if their exists an ***isomorphism*** f between them, s.t. f is a isomorphsm between two linear space, and f and f^-1 are continuous.

Theorem 9.5. *If two normed spaces have the same finite dimension, they are isomorphic.*

Theorem 9.6 (Space of bounded linear operators is normed linear space)**.** $\mathcal{B}(X_0, \cdots, X_{n-1}; Y)$ *is a normed linear space, the norm is defined as in Eq.* (9-1).

Theorem 9.7 (Norm of operator composition)**.** *Let* X, Y, Z *be three normed spaces, and* $\mathscr{A} \in \mathcal{B}(X; Y)$, $\mathscr{B} \in \mathcal{B}(Y; Z)$.

$$\|\mathscr{B}\mathscr{A}\| \le \|\mathscr{B}\|\|\mathscr{A}\|\,.$$ [4]

Proof.

$$\begin{aligned}\|\mathscr{B}\mathscr{A}\| &= \sup\left\{\|\mathscr{B}\mathscr{A}\boldsymbol{x}\|_Z/\|\boldsymbol{x}\|_X \mid \boldsymbol{x} \in X - \{\boldsymbol{0}\}\right\}\\ &\le \|\mathscr{B}\| \sup\left\{\|\mathscr{A}\boldsymbol{x}\|_Y/\|\boldsymbol{x}\|_X \mid \boldsymbol{x} \in X - \{\boldsymbol{0}\}\right\} = \|\mathscr{B}\|\|\mathscr{A}\|\,.\end{aligned}$$

□

Theorem 9.8 (completeness)**.** *If* Y *is a Banach space, so is* $\mathcal{B}(X_0, \cdots, X_{n-1}; Y)$.

Proof. Let $(\mathscr{A}_i)_{i\in\mathbb{N}} \in \mathcal{B}(X_0, \cdots, X_{n-1}; Y)^{\mathbb{N}}$ be a Cauchy sequence. $\forall \boldsymbol{x} := (\boldsymbol{x}_i)_{i\in n} \in X := \prod_{i\in n} X_i$,

$$\|\mathscr{A}_\ell \boldsymbol{x} - \mathscr{A}_m \boldsymbol{x}\|_Y = \|(\mathscr{A}_\ell - \mathscr{A}_m)\boldsymbol{x}\|_Y \le \|\mathscr{A}_\ell - \mathscr{A}_m\| \prod_{i\in n} \|\boldsymbol{x}_i\|_{X_i}\,,$$

therefore $(\mathscr{A}_i \boldsymbol{x})_{i\in\mathbb{N}} \in Y^{\mathbb{N}}$ is also a Cauchy sequence.

Since Y is a Banach space, we denote the limit of the Cauchy sequence $(\mathscr{A}_i\boldsymbol{x})_{i\in n}$ by $\mathscr{A}\boldsymbol{x}$. We need to prove that $\mathscr{A} \in \mathcal{B}(X_0, \cdots, X_{n-1}; Y)$.

It is obvious that $\mathscr{A} \in \mathcal{L}(X_0, \cdots, X_{n-1}; Y)$, therefore we only need to show that $\|\mathscr{A}\| < \infty$.

Let $\boldsymbol{e} := (\boldsymbol{e}_i)_{i\in n} \in X$, where $\forall i \in n$, $\|\boldsymbol{e}_i\|_{X_i} = 1$. $\forall \varepsilon \in \mathbb{R}_+$, $\exists N \in \mathbb{N}$, if $\ell > N$, then

$$0 \le \|\mathscr{A}\boldsymbol{e}\|_Y \le \|\mathscr{A}_\ell \boldsymbol{e}\|_Y + \varepsilon \le \|\mathscr{A}_\ell\| + \varepsilon\,,$$

Since $\{\|\mathscr{A}_i\| \mid i \in \mathbb{N}\}$ is bounded, we claim that $\left\{\|\mathscr{A}\boldsymbol{e}\| \mid \boldsymbol{e} = (\boldsymbol{e}_i)_{i\in n} \in X \wedge \forall i \in n\big(\|\boldsymbol{e}_i\|_{X_i} = 1\big)\right\}$ is also bounded. □

[4] By convention, we denote $\mathscr{B} \circ \mathscr{A}$ by $\mathscr{B}\mathscr{A}$, and $(\mathscr{B}\mathscr{A})(\boldsymbol{x})$ by $\mathscr{B}\mathscr{A}\boldsymbol{x}$ (since the compositions of the operator is associative).

Theorem 9.9. *$\forall m \in n$,*

$$\exists f \in \mathcal{B}(X_0, \cdots, X_{n-1}; Y)^{\mathcal{B}(X_0,\cdots,X_{m-1};B(X_m,\cdots,X_{n-1};Y))}$$

s.t. f is a isomorphism between two linear spaces and it conserves the norm structure i.e.

$$\|f(\mathscr{B})\| = \|\mathscr{B}\|.$$

Proof. $\forall \mathscr{B} \in \mathcal{B}\big(X_0, \cdots, X_{m-1}; B(X_m, \cdots, X_{n-1}; Y)\big)$, $\forall \boldsymbol{x} := (\boldsymbol{x}_i)_{i\in n} \in X := \prod_{i\in n} X_i$, $f(\mathscr{B})\boldsymbol{x} := \mathscr{B}(\boldsymbol{x}_i)_{i\in n}(\boldsymbol{x}_j)_{j\in n\backslash m}$.

Obviously $f \in \mathcal{L}\big(\mathcal{B}(X_0, \cdots, X_{m-1}; B(X_m, \cdots, X_{n-1}; Y)); \mathcal{B}(X_0, \cdots, X_{n-1}; Y)\big)$. If $f(\mathscr{B}) = \mathscr{O}_X$, $\mathscr{B} = \mathscr{O}_{\prod_{i\in m} X_m}$, therefore $\ker f = \{\mathscr{O}_{\prod_{i\in m} X_m}\}$, which implies that f is a isomorphism between two linear spaces.

$$\begin{aligned}
\|\mathscr{B}\| &= \sup\left\{\frac{\|\mathscr{B}(\boldsymbol{x}_i)_{i\in m}\|}{\prod_{i\in m}\|\boldsymbol{x}_i\|_{X_i}}\middle|\forall i \in m,\ \boldsymbol{x}_i \in X_i \wedge \boldsymbol{x}_i \neq \mathbf{0}\right\} \\
&= \sup\left\{\frac{\sup\left\{\frac{\|f(\mathscr{B})(\boldsymbol{x})\|_Y}{\prod_{i\in n\backslash m}\|\boldsymbol{x}_i\|_{X_i}}\middle|\forall i \in n\backslash m,\ \boldsymbol{x}_i \in X_i \wedge \boldsymbol{x}_i \neq \mathbf{0}\right\}}{\prod_{i\in m}\|\boldsymbol{x}_i\|_{X_i}}\middle|\forall i \in m,\ \boldsymbol{x}_i \in X_i \wedge \boldsymbol{x}_i \neq \mathbf{0}\right\} \\
&= \sup\left\{\frac{\|f(\mathscr{B})(\boldsymbol{x})\|_Y}{\prod_{i\in n}\|\boldsymbol{x}_i\|_{X_i}}\middle|\forall i \in n,\ \boldsymbol{x}_i \in X_i \wedge \boldsymbol{x}_i \neq \mathbf{0}\right\} = \|f(\mathscr{B})\|
\end{aligned}$$

□

Corollary 3. *$\mathcal{B}(X_0; \mathcal{B}(X_1; \cdots; \mathcal{B}(X_{n-1}; Y)\cdots))$ and $\mathcal{B}(X_0, \cdots, X_{n-1}; Y)$ are isomorphic.*

§10 Differentiation

Definition 10.1 (Differentiation)**.** Let X, Y be two normed spaces. A mapping f from $D \in \mathscr{P}(X)$ to Y is said to be ***differentiable*** at an interior point $\boldsymbol{x} \in D$ if $\exists \mathscr{L}(\boldsymbol{x}) \in \mathcal{B}(X; Y)$[5] s.t. $\forall \Delta\boldsymbol{x} \in X$ $\big(\boldsymbol{x} + \Delta\boldsymbol{x} \in D\big)$,

$$f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x}) = \mathscr{L}(\boldsymbol{x})\Delta\boldsymbol{x} + \alpha(\boldsymbol{x}; \Delta\boldsymbol{x}), \tag{10-1}$$

where $\alpha(\boldsymbol{x}; \Delta\boldsymbol{x}) = o(\Delta\boldsymbol{x})$ as $\Delta\boldsymbol{x} \to 0$, i.e. $\lim_{\Delta\boldsymbol{x}\to\mathbf{0}} \|\alpha(\boldsymbol{x}; \Delta\boldsymbol{x})\|_Y / \|\Delta\boldsymbol{x}\|_X = 0$.

Such $\mathscr{L}|_{\boldsymbol{x}}$ is called the ***differential*** of f at $\boldsymbol{x}$[6], denoted by $\mathrm{d}f(\boldsymbol{x})$ or $f'(\boldsymbol{x})$.

Theorem 10.1 (Uniqueness)**.** *Let X and Y be two normed spaces. If a mapping $f \in Y^D$ where $D \in \mathscr{P}(X)$ is differentiable at $\boldsymbol{x}$ which is an interior point of D, then the differential of f at $\boldsymbol{x}$ is unique.*

Proof. Let their be two differentials $\mathscr{L}_1(\boldsymbol{x})$, $\mathscr{L}_2(\boldsymbol{x})$, by the definition (10-1), we have:

$$\big(\mathscr{L}_1(\boldsymbol{x}) - \mathscr{L}_2(\boldsymbol{x})\big)\Delta\boldsymbol{x} = o(\Delta\boldsymbol{x}),$$

[5] $\boldsymbol{x}$ here is an argument.

[6] Alternatively, ***tangent mapping*** or ***derivative***.

hence $\|(\mathscr{L}_1(\boldsymbol{x}) - \mathscr{L}_2(\boldsymbol{x}))\Delta\boldsymbol{x}\|_Y = o(\|\Delta\boldsymbol{x}\|_X)$, therefore

$$\lim_{\|\Delta\boldsymbol{x}\|_X \to 0} \left\|(\mathscr{L}_1(\boldsymbol{x}) - \mathscr{L}_2(\boldsymbol{x}))\frac{\Delta\boldsymbol{x}}{\|\Delta\boldsymbol{x}\|_X}\right\|_Y = 0\,,$$

This means that whatever the direction of unit vector $\Delta\boldsymbol{x}/\|\Delta\boldsymbol{x}\|_X$ is, the norm of $\left\|(\mathscr{L}_1(\boldsymbol{x}) - \mathscr{L}_2(\boldsymbol{x}))\Delta\boldsymbol{x}/\|\Delta\boldsymbol{x}\|_X\right\|_Y$ is always zero, therefore $\|\mathscr{L}_1(\boldsymbol{x}) - \mathscr{L}_2(\boldsymbol{x})\|$. By the definition of norms, this means that $\mathscr{L}_1(\boldsymbol{x}) - \mathscr{L}_2(\boldsymbol{x}) = \mathscr{O}$, or $\mathscr{L}_1(\boldsymbol{x}) = \mathscr{L}_2(\boldsymbol{x})$. □

Theorem 10.1 gives us the right to define:

Definition 10.2 (Derivative mapping)**.** Let X, Y be two normed spaces, $D \in \mathscr{P}(X)$, $f \in Y^D$, $\Delta(f) := \{\boldsymbol{x} \in X \mid f \text{ is differentiable at } \boldsymbol{x}\}$.

$$f' \colon \Delta(f) \to \mathscr{B}(X, Y);\ \boldsymbol{x} \mapsto \mathrm{d}f(\boldsymbol{x})$$

is called the ***derivative mapping*** of f.

Warning: We use $f'(\boldsymbol{x})$ to denote the linear operator on X instead of a point in Y (when $X = Y = \mathbb{R}$, they are the isomorphic). It is obvious that $\forall \mathscr{A} \in \mathcal{B}(X;Y)$, $\forall \boldsymbol{x} \in X$, $\mathrm{d}\mathscr{A}(\boldsymbol{x}) = \mathscr{A}$, which is different from the usual notations that writes $f(x) = \mathrm{e}^x \to f'(x) = \mathrm{e}^x = f(x)$ and $f(x) = ax \to f'(x) = a$.

To make it clear, we must remember: $f \in Y^X$, $f' \in \mathscr{B}(X;Y)^X$, $f'(\boldsymbol{x}) \in \mathscr{B}(X;Y)$, $f'(\boldsymbol{x})\Delta\boldsymbol{x} \in Y$.

It is always convenient to define such notation:

Definition 10.3. Let X_i, $i \in n$ be normed spaces, and $X := \prod_{i \in n} X_i$. We define $\mathrm{d}\boldsymbol{x}_i$ as:

$$\mathrm{d}\boldsymbol{x}_i \Delta\boldsymbol{x} = \Delta\boldsymbol{x}_i\,,$$

for any $\Delta\boldsymbol{x} := (\Delta\boldsymbol{x}_i)_{i \in n} \in X$.

Actually, $\mathrm{d}\boldsymbol{x}_i$ can be conceive as the differential of the projective operator $X \to X_i$. If $n = 1$, $\mathrm{d}\boldsymbol{x} = \mathrm{id}_X$, therefore we can write:

$$\mathrm{d}f(\boldsymbol{x}) = f'(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,,$$

which is the notation we have been very familiar with.

Theorem 10.2 (Differentiable then continuous)**.** *Let X and Y be two normed spaces. If a mapping $f \in Y^D$ where $D \in \mathscr{P}(X)$ is differentiable at $\boldsymbol{x}$ which is an interior point of D, then f is continuous at $\boldsymbol{x}$.*

Proof. as $\|\Delta\boldsymbol{x}\| \to 0$

$$\|f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x})\|_Y \leq \|\mathscr{L}(\boldsymbol{x})\Delta\boldsymbol{x}\|_Y + \|\alpha(\boldsymbol{x};\Delta\boldsymbol{x})\|_Y \leq \|\mathscr{L}(\boldsymbol{x})\|\|\Delta\boldsymbol{x}\|_X + \|\alpha(\boldsymbol{x};\Delta\boldsymbol{x})\|_Y \to 0\,.$$

□

Theorem 10.3 (Linearity of differentiation)**.** *Let X, Y be two normed space on $\mathbb{F}$ ($\mathbb{C}$ or $\mathbb{R}$), $\boldsymbol{x} \in X$ is an interior point. The space of all mappings differentiable at $\boldsymbol{x}$ is also a linear space on $\mathbb{F}$.*

Theorem 10.4 (Chain rule)**.** *Let X, Y, Z be three normed spaces, $D \in \mathscr{P}(X)$, $f \in Y^D$, $g \in Z^{f(D)}$, and f be differentiable at $\boldsymbol{x} \in D$, g be differentiable at $\boldsymbol{y} := f(\boldsymbol{x}) \in f(D)$.*

$$(g \circ f)'(\boldsymbol{x}) = g'(\boldsymbol{y}) f'(\boldsymbol{x})^{7} .$$

For example, $(\mathscr{A} \circ f)'(\boldsymbol{x}) = \mathscr{A} f'(\boldsymbol{x})$, since $\mathscr{A}'(\boldsymbol{y}) = \mathscr{A}$.

Theorem 10.5 (Differentiation of inverse mappings)**.** *Let X, Y be two normed spaces, $D \in \mathscr{P}(X)$, bijective $f \in X^D$, and f be differentiable at $\boldsymbol{x} \in D$, and there be an inverse $[f'(\boldsymbol{x})]^{-1}$ for $f'(\boldsymbol{x})$. Then, f^{-1} is also differentiable at $\boldsymbol{y} := f(\boldsymbol{x})$, and*

$$(f^{-1})'(\boldsymbol{y}) = [f'(\boldsymbol{x})]^{-1} .$$

Consider a mappings $f \colon X \to Y$, where $Y := \prod_{i \in n} Y_i$, normed with $\|\boldsymbol{y}\|_Y := \sqrt[p]{\sum_{i \in n} \|\boldsymbol{y}_i\|_{Y_i}^p}$. By writing f as $(f_i)_{i \in n}$ such that $f(\boldsymbol{x}) = (f_i(\boldsymbol{x}))_{i \in n}$, and

$$f'(\boldsymbol{x})\Delta\boldsymbol{x} = (f_i'(\boldsymbol{x})\Delta\boldsymbol{x})_{i \in n} ,$$

we can conclude that f is differentiable at $\boldsymbol{x} \in X$ ***iff*** for each $f_i \colon X \to Y_i$, $i \in n$, is differentiable at $\boldsymbol{x}$.

Theorem 10.6 (Differentiation of multilinear operators)**.** *Let $X_0, \cdots, X_{n-1}$, Y be normed spaces, $\mathscr{A} \in \mathcal{B}(X_0, \cdots, X_{n-1}; Y)$. Let $X := \prod_{i \in n} X_i$ be normed space with a norm defined as:*

$$\forall \boldsymbol{x} := (\boldsymbol{x}_i)_{i \in n} \in X, \quad \|\boldsymbol{x}\|_X := \Big(\sum_{i \in n} \|\boldsymbol{x}_i\|_{X_i}^p \Big)^{1/p} . \tag{10-2}$$

Then, $\mathscr{A}$ is differentiable at all interior point $\boldsymbol{x} \in X$, and

$$\mathrm{d}\mathscr{A}(\boldsymbol{x}) = \mathscr{A}(\mathrm{d}\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{n-2}, \mathrm{d}\boldsymbol{x}_{n-1}) .$$

Proof. By Eq. (10-2), we have $\forall i \in n$,

$$\|\boldsymbol{x}_i\|_{X_i} \le \|\boldsymbol{x}\|_X \le \sum_{j \in n} \|\boldsymbol{x}_j\|_{X_j} .$$

Therefore $\forall i, j \in n$,

$$\frac{\|\Delta\boldsymbol{x}_i\|_{X_i} \|\Delta\boldsymbol{x}_j\|_{X_j}}{\|\Delta\boldsymbol{x}\|_X} \le \frac{\|\Delta\boldsymbol{x}_i\|_{X_i} \|\Delta\boldsymbol{x}_j\|_{X_j}}{\|\Delta\boldsymbol{x}_i\|_{X_i}} = \|\Delta\boldsymbol{x}_j\|_{X_j} \le \|\Delta\boldsymbol{x}\|_X ,$$

or $\|\Delta\boldsymbol{x}_i\|_{X_i} \|\Delta\boldsymbol{x}_j\|_{X_j} = o(\boldsymbol{x}; \Delta\boldsymbol{x})$.

$$\begin{aligned}
&\mathscr{A}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \mathscr{A}(\boldsymbol{x}) \\
&\qquad = \mathscr{A}(\Delta\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \mathscr{A}(\boldsymbol{x}_1, \Delta\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \Delta\boldsymbol{x}_{n-1}) \\
&\qquad\quad + \mathscr{A}(\Delta\boldsymbol{x}_0, \Delta\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\boldsymbol{x}_0, \cdots, \Delta\boldsymbol{x}_{n-2}, \Delta\boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\Delta\boldsymbol{x}) \\
&\qquad = \mathscr{A}(\Delta\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{n-2}, \Delta\boldsymbol{x}_{n-1}) + o(\boldsymbol{x}; \Delta\boldsymbol{x}) ,
\end{aligned}$$

[7] Remember, we write the composition of two linear operators omitting the "$\circ$" in the middle.

where we utilize the fact that

$$\|\mathscr{A}(\Delta\boldsymbol{x}_0, \Delta\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1})\|_Y \le \|\mathscr{A}\|\|\Delta\boldsymbol{x}_0\|_{X_0}\|\Delta\boldsymbol{x}_1\|_{X_1}\prod_{i\in n\setminus 2}\|\boldsymbol{x}_i\|_{X_i} = o(\boldsymbol{x};\Delta\boldsymbol{x}), \cdots$$

Therefore

$$\mathrm{d}\mathscr{A}(\boldsymbol{x})\Delta\boldsymbol{x} = \mathscr{A}(\Delta\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{n-2}, \Delta\boldsymbol{x}_{n-1})$$

or

$$\mathrm{d}\mathscr{A}(\boldsymbol{x}) = \mathscr{A}(\mathrm{d}\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}) + \cdots + \mathscr{A}(\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{n-2}, \mathrm{d}\boldsymbol{x}_{n-1}).$$

□

Let $\mathcal{U}(X;Y)$ be the set of ***reversible operators*** in $\mathcal{B}(X;Y)$ i.e. $\forall\mathscr{A}\in\mathscr{U}(X;Y)$, $\exists\mathscr{A}^{-1}\in\mathscr{B}(Y;X)$ s.t.

$$\mathscr{A}\mathscr{A}^{-1} = \mathrm{id}_Y; \quad \mathscr{A}^{-1}\mathscr{A} = \mathrm{id}_X.$$

Theorem 10.7 (Differential of reversion)**.** *Let X be a complete normed space, and Y be a normed space. $\mathscr{A}\in\mathcal{U}(X;Y)$, $\delta\mathscr{A}\in\mathcal{B}(X;Y)$. If $\|\delta\mathscr{A}\| < \|\mathscr{A}^{-1}\|^{-1}$, then $\mathscr{A}+\delta\mathscr{A}\in\mathcal{U}(X;Y)$,*

$$(\mathscr{A}+\delta\mathscr{A})^{-1} = \mathscr{A}^{-1} - \mathscr{A}^{-1}\delta\mathscr{A}\mathscr{A}^{-1} + o(\delta\mathscr{A}),$$

as $\delta\mathscr{A}\to\mathscr{O}$.

Proof. Since X is complte, by Theorem 9.8, we know $\mathcal{B}(X;X)$ is complete. Notice $-\mathscr{A}^{-1}\delta\mathscr{A}\in\mathcal{B}(X;X)$, and by Theorem 9.7,

$$\|-\mathscr{A}^{-1}\delta\mathscr{A}\| \le \|\mathscr{A}^{-1}\|\|\delta\mathscr{A}\| < \|\mathscr{A}^{-1}\|\|\mathscr{A}^{-1}\|^{-1} = 1,$$

$\forall\varepsilon\in\mathbb{R}_+$, let

$$N > \log_{\|\mathscr{A}^{-1}\delta\mathscr{A}\|}\frac{\varepsilon(1-\|\mathscr{A}^{-1}\delta\mathscr{A}\|)}{\|\mathscr{A}^{-1}\delta\mathscr{A}\|}$$

(we assume that $\mathscr{A}^{-1}\delta\mathscr{A}\ne\mathscr{O}$, or the inequality is trivial), $m>n>N$, then

$$\begin{aligned}\left\|\sum_{k=n+1}^{m}(-\mathscr{A}^{-1}\delta\mathscr{A})^k\right\| &\le \sum_{k=n+1}^{m}\|\mathscr{A}^{-1}\delta\mathscr{A}\|^k = \frac{1-\|\mathscr{A}^{-1}\delta\mathscr{A}\|^{m-n}}{1-\|\mathscr{A}^{-1}\delta\mathscr{A}\|}\|\mathscr{A}^{-1}\delta\mathscr{A}\|^{n+1}\\ &\le \frac{\|\mathscr{A}^{-1}\delta\mathscr{A}\|^{n+1}}{1-\|\mathscr{A}^{-1}\delta\mathscr{A}\|} < \varepsilon,\end{aligned}$$

hence $\sum_{k\in n}(-\mathscr{A}^{-1}\delta\mathscr{A})^k$ is a Cauchy sequence, therefore convergent i.e. $\sum_{k\in\mathbb{N}}(-\mathscr{A}^{-1}\delta\mathscr{A})^k$.

We can verify $\sum_{k\in\mathbb{N}}(-\mathscr{A}^{-1}\delta\mathscr{A})^k = (\mathrm{id}_X + \mathscr{A}^{-1}\delta\mathscr{A})^{-1}$.

Since $\mathscr{A}+\delta\mathscr{A} = \mathscr{A}(\mathrm{id}_X + \mathscr{A}^{-1}\delta\mathscr{A})$, we conclude

$$(\mathscr{A}+\delta\mathscr{A})^{-1} = \sum_{k\in\mathbb{N}}(-\mathscr{A}^{-1}\delta\mathscr{A})^k\mathscr{A}^{-1},$$

and

$$\begin{aligned}\|(\mathscr{A}+\delta\mathscr{A})^{-1}-\mathscr{A}^{-1}+\mathscr{A}^{-1}\delta\mathscr{A}\mathscr{A}^{-1}\| &= \left\|\sum_{k=2}^{\infty}(-\mathscr{A}^{-1}\delta\mathscr{A})^k\mathscr{A}^{-1}\right\| \\ &\le \sum_{k=2}^{\infty}\|\mathscr{A}^{-1}\delta\mathscr{A}\|^k\|\mathscr{A}^{-1}\| = \frac{\|\mathscr{A}^{-1}\|\|\mathscr{A}^{-1}\delta\mathscr{A}\|^2}{1-\|\mathscr{A}^{-1}\delta\mathscr{A}\|} = o(\|\delta\mathscr{A}\|)\,.\end{aligned}$$

□

Let $f \in Y^X$ where $X := \prod_{i\in n} X_i$. We define a mapping

$$\varphi_i\colon X_i \to X;\ \boldsymbol{x}_i \mapsto (\boldsymbol{a}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_{i-1}, \boldsymbol{x}_i, \boldsymbol{a}_{i+1}, \cdots, \boldsymbol{a}_{n-1})\,, \tag{10-3}$$

so that $f \circ \varphi_i$ means the mapping of alone $\boldsymbol{x}_i$, leaving other variables unchanged.

Definition 10.4 (Partial derivative)**.** Let $f \in Y^X$ where $X := \prod_{i\in n} X_i$ be the product of normed spaces, Y be a normed space. $\forall i \in n$, φ_i is defined as Eq. (10-3). If $f \circ \varphi_i$ is differentiable at an interior point $\boldsymbol{a}_i \in X_i$, we call its derivative at this point the ***partial derivative*** of f with respect to $\boldsymbol{x}_i$ at $\boldsymbol{a} := (\boldsymbol{a}_i)_{i\in n}$, denoted by $\partial_i f(\boldsymbol{a})$ or $\frac{\partial f}{\partial \boldsymbol{x}_i}(\boldsymbol{a})$.

Theorem 10.8 (Differentiable then partial derivative exists)**.** *Let $X_1, \cdots, X_{n-1}$ and Y be normed spaces, $X := \prod_{i\in n} X_i$, $f \in Y^X$, $\boldsymbol{a} \in X$. If f is differentiable at $\boldsymbol{a}$, then $\forall i \in n$, $f\circ\varphi_i$ is differentiable $\boldsymbol{a}_i \in X_i$, and*

$$\mathrm{d}f(\boldsymbol{a}) = \sum_{i\in n} \partial_i f(\boldsymbol{a})\, \mathrm{d}\boldsymbol{x}_i\,. \tag{10-4}$$

Definition 10.5 (Continuously differentiable)**.** Let $f \in Y^X$ and differentiable at $\boldsymbol{x} \in X$. If the derivative mapping $f' \in \mathscr{B}(X;Y)^X$ is continuous at $\boldsymbol{x}$, we say that f is ***continuously differentiable*** at point $\boldsymbol{x}$.

We can denote all continuously differentiable mappings from an open set X to Y by $C^{(1)}(X,Y)$[8].

By Theorem 10.2 we know that $C^{(1)}(X,Y) \subset C(X,Y)$.

Theorem 10.9 (Continuously differentiable iff partial derivative is continuous (differentiable mapping))**.** *Let $X_0, \cdots, X_{n-1}$, Y be normed spaces, $X := \prod_{i\in n} X_i$, $\boldsymbol{x} \in X$, $f \in Y^X$ is differentiable at $\boldsymbol{x}$. f is continuouly differentiable at $\boldsymbol{x}$* iff *$\forall i \in n$, $\partial_i f \in \mathcal{B}(X_i;Y)^X$.*

Proof.

$$\begin{aligned}\|\partial_i f(\boldsymbol{x}+\Delta\boldsymbol{x}) - \partial_i f(\boldsymbol{x})\| &\le \left\|\sum_{j\in n}\big(\partial_j f(\boldsymbol{x}+\Delta\boldsymbol{x}) - \partial_i f(\boldsymbol{x})\big)\right\| = \|\,\mathrm{d}f(\boldsymbol{x}+\Delta\boldsymbol{x}) - \mathrm{d}f(\boldsymbol{x})\| \\ &\le \sum_{j\in n}\|\partial_j f(\boldsymbol{x}+\Delta\boldsymbol{x}) - \partial_j f(\boldsymbol{x})\|\end{aligned}$$

□

[8] or $C^{(1)}(X)$ if you are sure about what Y is.

Definition 10.6 (Derivative with respect to a vector)**.** Let X and Y be two normed space over $\mathbb{R}$ or $\mathbb{C}$, U be an open set in X, $f \in Y^U$, $\boldsymbol{x} \in U$. The derivative of f with respect to a vector $\boldsymbol{\ell}$ is defined as:

$$\frac{\partial f}{\partial \boldsymbol{\ell}}(\boldsymbol{x}) := \lim_{t \to 0} \frac{1}{t}[f(\boldsymbol{x} + t\boldsymbol{\ell}) - f(\boldsymbol{x})] \,.$$

Theorem 10.10 (Derivative with respect to a vector when differentiable)**.** *Let X and Y be two normed space over $\mathbb{R}$ or $\mathbb{C}$, U be an open set in X, $f \in Y^U$, $\boldsymbol{x} \in U$. If f is differentiable at $\boldsymbol{x}$, then $\forall \boldsymbol{\ell} \in X$, the derivative of f with respect to $\boldsymbol{\ell}$ exists, and*

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\ell}}(\boldsymbol{x}) = f'(\boldsymbol{x})\boldsymbol{\ell} \,.$$

Proof.

$$\lim_{t \to 0} \frac{1}{t}[f(\boldsymbol{x} + t\boldsymbol{\ell}) - f(\boldsymbol{x})] = \lim_{t \to 0} \frac{1}{t}[f'(\boldsymbol{x})t\boldsymbol{\ell} + o(t\boldsymbol{\ell})] = f'(\boldsymbol{x})\boldsymbol{\ell} \,.$$

□

§11 Finite-Increment Theorem

We now study the generalisation of the Lagrangian mean value theorem, or the finite-increment theorem.

Let us recall and generalised the definition of interval:

Definition 11.1. Let X be a linear space over a field $\mathbb{F}$ which contains $\mathbb{R}$, $\boldsymbol{a}, \boldsymbol{b} \in X$. The ***closed*** and ***open interval*** is defined as:

$$\begin{aligned} [\boldsymbol{x}, \boldsymbol{y}] &:= \{\boldsymbol{x} + \theta(\boldsymbol{y} - \boldsymbol{x}) \mid 0 \le \theta \le 1\} \,, \\ (\boldsymbol{x}, \boldsymbol{y}) &:= \{\boldsymbol{x} + \theta(\boldsymbol{y} - \boldsymbol{x}) \mid 0 < \theta < 1\} \,. \end{aligned}$$

Similarly we can define $[\boldsymbol{x}, \boldsymbol{y})$, $(\boldsymbol{x}, \boldsymbol{y}]$.

Theorem 11.1 (Finite-increment theorem)**.** *Let X and Y be two normed spaces, $G \in \mathscr{T}_X$, where $\mathscr{T}_X$ is the topology induced by the norm $\| * \|_X$. Let $f \in C(G, Y)$, $[\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta \boldsymbol{x}] \subset G$. If $\forall \boldsymbol{x} \in (\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta \boldsymbol{x})$, f is differentiable at $\boldsymbol{x}$, then*

$$\|f(\boldsymbol{x}_0 + \Delta x) - f(\boldsymbol{x}_0)\|_Y \le \sup\{\|f'(\boldsymbol{\xi})\| \mid \boldsymbol{\xi} \in (\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta \boldsymbol{x})\} \|\Delta \boldsymbol{x}\|_X \,.$$

Proof. First we assume that f is differentiable in closed interval $[\boldsymbol{x}, \boldsymbol{x} + \Delta \boldsymbol{x}]$ (later we would return to the more generalised situation).

Let us denote $M_{[t_1,t_2]} := \sup\{\|f'(\boldsymbol{x}_0 + t\Delta \boldsymbol{x})\| \mid t \in [t_1, t_2]\}$. If there exists $\varepsilon_0 \in \mathbb{R}_+$, $\|f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) - f(\boldsymbol{x})\|_Y > (M_{[0,1]} + \varepsilon_0)\|\Delta \boldsymbol{x}\|_X$, since

$$\|f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) - f(\boldsymbol{x})\|_Y \le \|f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) - f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}/2)\|_Y + \|f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}/2) - f(\boldsymbol{x})\| \,,$$

and $M_{[0,1/2]} \le M_{[0,1]}$, $M_{[1/2,1]} \le M_{[0,1]}$, the following two inequality *cannot* be both true:

$$\begin{aligned} \|f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) - f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}/2)\|_Y &\le (M_{[1/2,1]} + \varepsilon_0)\|\Delta \boldsymbol{x}\|_X/2 \,; \\ \|f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}/2) - f(\boldsymbol{x})\| &\le (M_{[0,1/2]} + \varepsilon_0)\|\Delta \boldsymbol{x}\|_X/2 \,. \end{aligned}$$

We would repeatedly divide the interval which does not satisfies the finite-increment theorem into two, and finally we would have a collections of closed intervals $\langle [a_i, b_i] \rangle_{i\in\mathbb{N}}$ s.t. $a_i \le a_{i+1} < b_{i+1} \le b_i$, $\forall i \in \mathbb{N}$, over which the inequality

$$\|f(\boldsymbol{x}_0 + b_i\Delta\boldsymbol{x}) - f(\boldsymbol{x}_0 + a_i\Delta\boldsymbol{x})\|_Y > (M_{[a_i,b_i]} + \varepsilon_0)|b_i - a_i|\|\Delta\boldsymbol{x}\|_X$$

holds.

Since $[0,1]$ is a compact set in $\mathbb{R}$, and $|b_i - a_i| = 2^{-i}$, $\exists c \in [0,1]$ s.t. $\bigcap_{i\in\mathbb{N}}[a_i, b_i] = \{c\}$.

Because we can say c divides all $[a_i, b_i]$ into two, we shall always choose one of $\{a_i, b_i\}$ as c_i s.t.

$$\|f(\boldsymbol{x}_0 + c\Delta\boldsymbol{x}) - f(\boldsymbol{x}_0 + c_i\Delta\boldsymbol{x})\|_Y > (M_{[c,c_i]} + \varepsilon_0)|c_i - c|\|\Delta\boldsymbol{x}\|_X \,. \tag{11-1}$$

However, by the differentiability of f at $\boldsymbol{x}_0 + c\Delta\boldsymbol{x}$, $\forall\varepsilon \in \mathbb{R}_+$, there exists an $N \in \mathbb{N}$, as long as $i > N$

$$\begin{aligned}\|f(\boldsymbol{x}_0 + c\Delta\boldsymbol{x}) - f(\boldsymbol{x}_0 + c_i\Delta\boldsymbol{x})\|_Y &\le \|f'(\boldsymbol{x}_0 + c\Delta\boldsymbol{x})\||c_i - c|\|\Delta\boldsymbol{x}\|_X + o(|c_i - c|)\|\Delta\boldsymbol{x}\|_X \\ &\le (M_{[c,c_i]} + \varepsilon)|c_i - c|\|\Delta\boldsymbol{x}\|_X \,.\end{aligned}$$

Letting $\varepsilon = \varepsilon_0$ we would find a contradiction.

Now if the function f is only differentiable in $(\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta\boldsymbol{x})$, we have proved that $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in (\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta\boldsymbol{x})$,

$$\|f(\boldsymbol{x}_2) - f(\boldsymbol{x}_1)\| \le M_{[t_1,t_2]}\|\boldsymbol{x}_1, \boldsymbol{x}_2\|_X \,.$$

where $\boldsymbol{x}_1 = \boldsymbol{x}_0 + t_1\Delta\boldsymbol{x}$, $\boldsymbol{x}_2 = \boldsymbol{x}_0 + t_2\Delta\boldsymbol{x}$.

Since both $\| * \|$ and f is continuous (Theorem 8.2 and Theorem 10.2), we shall pass $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ to $\boldsymbol{x}_0$ and $\boldsymbol{x}_0 + \Delta\boldsymbol{x}$, and get

$$\|f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - f(\boldsymbol{x}_0)\|_Y \le \sup\{\|f'(\boldsymbol{\xi})\| \mid \boldsymbol{\xi} \in (\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta\boldsymbol{x})\}\|\Delta\boldsymbol{x}\|_X \,.$$

□

Corollary 4. *Let X and Y be two normed spaces, $G \in \mathscr{T}_X$, where $\mathscr{T}_X$ is the topology induced by the norm $\| * \|_X$. Let $f \in C(G, Y)$, $[\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta\boldsymbol{x}] \subset G$. $\forall \mathscr{A} \in \mathcal{B}(X, Y)$,*

$$\|f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x}) - \mathscr{A}\Delta\boldsymbol{x}\|_Y \le \sup\left\{\|f'(\boldsymbol{\xi}) - \mathscr{A}\|\|\Delta x\|_X \mid \boldsymbol{\xi} \in [\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta\boldsymbol{x}]\right\}.$$

Proof. Define:

$$F\colon [0,1] \to Y;\ t \mapsto f(\boldsymbol{x} + t\Delta\boldsymbol{x}) - \mathscr{A}t\Delta\boldsymbol{x} \,.$$

By the finite-increment theorem 11.1,

$$\begin{aligned}\|F(1) - F(0)\|_Y &= \|f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x}) - \mathscr{A}\Delta\boldsymbol{x}\|_Y \\ &\le \sup\left\{\|F'(\xi)\| \mid \xi \in [0,1]\right\}|1 - 0| = \sup\left\{\|f'(\boldsymbol{x} + \xi\Delta\boldsymbol{x})\Delta\boldsymbol{x} - \mathscr{A}\Delta\boldsymbol{x}\| \mid \xi \in [0,1]\right\} \\ &\le \sup\left\{\|f'(\boldsymbol{x} + \xi\Delta\boldsymbol{x}) - \mathscr{A}\| \mid \xi \in [0,1]\right\}\|\Delta\boldsymbol{x}\|_X \,.\end{aligned}$$

□

Theorem 11.2 (Continuously differentiable then Lipschitz continuous)**.** *Let K be a convex*[9] *compact set in a normed space X, and Y be a normed space, $f \in Y^K$. If $f \in C^{(1)}(K, Y)$, then f is Lipschitz continuous.*

[9] a ***convex set*** is a set that contains all points on the straight segment jointing any two points i.e. $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in C$, $[\boldsymbol{x}_1, \boldsymbol{x}_2] \subset C$.

Proof. $f' \in C(K; \mathcal{B}(X;Y))$, $\| * \|_Y \in C(Y;\mathbb{R})$, hence the composition $g\colon K \to \mathbb{R};\ \boldsymbol{x} \mapsto \|f'(\boldsymbol{x})\|_Y$ is also continuous. Recall Theorem 6.8, we conclude that $\exists M$, $\forall \boldsymbol{x} \in K$, $g(\boldsymbol{x}) \le M$.

Since K is convex, $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in K$, $[\boldsymbol{x}_1, \boldsymbol{x}_2] \subset K$. By finite-increment theorem 11.1, we have:

$$\|f(\boldsymbol{x}_2) - f(\boldsymbol{x}_1)\|_Y \le \sup\left\{\|f'(\boldsymbol{x})\| \mid \boldsymbol{x} \in [\boldsymbol{x}_1, \boldsymbol{x}_2]\right\}\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_X \le M\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_X\,.$$

□

Theorem 11.3. *Let K be a convex compact set in a normed space X, and Y be a normed space, $f \in C^{(1)}(K, Y)$. $\exists \omega \in \mathbb{R}^{\mathbb{R}}$ s.t. $\lim_{x \to +0} \omega(x) = 0$, and $\forall \boldsymbol{x} \in X$, if $\Delta \boldsymbol{x} \in K \cap B(\boldsymbol{x}; \delta)$, then*

$$\|f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x}) - f'(\boldsymbol{x})\Delta\boldsymbol{x}\|_Y \le \omega(\delta)\|\Delta\boldsymbol{x}\|_X\,,$$

for some $\delta \in \mathbb{R}_+$.

Proof. By Corollary 4,

$$\|f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x}) - f'(\boldsymbol{x})\Delta\boldsymbol{x}\|_Y \le \sup\left\{\|f'(\boldsymbol{\xi}) - f'(\boldsymbol{x})\| \mid \boldsymbol{\xi} \in [\boldsymbol{x}_0, \boldsymbol{x}_0 + \Delta\boldsymbol{x}]\right\}\|\Delta x\|_X\,.$$

Let

$$\omega(\delta) = \sup\left\{\|f'(\boldsymbol{x}_2) - f'(\boldsymbol{x}_1)\| \mid \boldsymbol{x}_1, \boldsymbol{x}_2 \in K \wedge d_X(\boldsymbol{x}_1, \boldsymbol{x}_2) < \delta\right\}.$$

□

With the finite-increment theorem, we can generalised Theorem 10.9 to any mappings, instead of differentiable mappings alone.

Theorem 11.4 (Continuously differentiable iff partial differential is continuous)**.** *Let $X_0, \cdots, X_{n-1}$, Y be normed spaces, $X := \prod_{i \in n} X_i$, $G \in \mathscr{T}_X$, $f \in Y^G$.*

$f \in C^{(1)}(G, Y) \ \leftrightarrow\ \forall i \in n$, $\partial_i f \in C(G, \mathcal{B}(X;Y))$.

Proof. $\to$: We have proved that if the mapping f is continuously differentiable in G, $\forall i \in n$, $\partial_i f$ is continuous. (Theorem 10.9).

$\leftarrow$: Denote

$$\mathscr{L} := \sum_{i \in n} \partial_i f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_i\,,$$

and we shall show that $\mathscr{L}$ is the differential of f at $\boldsymbol{x} \in G$.

Let us introduce a notation,

$$\Delta_i f(\boldsymbol{a}) := f(\boldsymbol{a}_0, \cdots, \boldsymbol{a}_{i-1}, \boldsymbol{a}_i + \Delta\boldsymbol{x}_i, \boldsymbol{a}_{i+1}, \cdots, \boldsymbol{a}_{n-1}) - f(\boldsymbol{a})\,.$$

Then

$$\begin{aligned}
&f(\boldsymbol{x} + \Delta\boldsymbol{x}) - f(\boldsymbol{x}) - \mathscr{L}\Delta\boldsymbol{x}\\
&= \Delta_0 f(\boldsymbol{x}_0, \boldsymbol{x}_1 + \Delta\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1} + \Delta\boldsymbol{x}_{n-1}) - \partial_0 f(\boldsymbol{x})\Delta\boldsymbol{x}_0\\
&\quad + \Delta_1 f(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2 + \Delta\boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n-1} + \Delta\boldsymbol{x}_{n-1}) - \partial_1 f(\boldsymbol{x})\Delta\boldsymbol{x}_2\\
&\quad + \cdots + \Delta_{n-1} f(\boldsymbol{x})\Delta\boldsymbol{x}_{n-1} - \partial_{n-1} f(\boldsymbol{x})\Delta\boldsymbol{x}_{n-1}\,.
\end{aligned}$$

By Corollary 4, we have:

$$
\begin{aligned}
&\|f(\boldsymbol{x}+\Delta\boldsymbol{x})-f(\boldsymbol{x})-\mathscr{L}\Delta\boldsymbol{x}\|_Y \\
&\leq \|\Delta_0 f(\boldsymbol{x}_0,\boldsymbol{x}_1+\Delta\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n-1}+\Delta\boldsymbol{x}_{n-1})-\partial_0 f(\boldsymbol{x})\Delta\boldsymbol{x}_0\|_Y \\
&\quad +\cdots+\|\Delta_{n-1}f(\boldsymbol{x})\Delta\boldsymbol{x}_{n-1}-\partial_{n-1}f(\boldsymbol{x})\Delta\boldsymbol{x}_{n-1}\|_Y \\
&\leq \sup\Big\{\big\|\partial_0 f(\boldsymbol{\xi}_0,\boldsymbol{x}_1+\Delta\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n-1}+\Delta\boldsymbol{x}_{n-1}) \\
&\qquad\qquad -\partial_0 f(\boldsymbol{x}_0,\boldsymbol{x}_1+\Delta\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{n-1}+\Delta\boldsymbol{x}_{n-1})\big\|_Y \mid \boldsymbol{\xi}_0\in[\boldsymbol{x}_0,\boldsymbol{x}_0+\Delta\boldsymbol{x}_0]\Big\}\|\Delta\boldsymbol{x}_0\|_{X_0} \\
&\quad +\cdots+\sup\Big\{\big\|\partial_{n-1}f(\boldsymbol{x}_0,\cdots,\boldsymbol{\xi}_{n-1})-\partial_{n-1}f(\boldsymbol{x})\big\|_Y \mid \boldsymbol{\xi}_{n-1}\in[\boldsymbol{x}_0,\boldsymbol{x}_0+\Delta\boldsymbol{x}_0]\Big\}\|\Delta\boldsymbol{x}_{n-1}\|_{X_{n-1}}\,.
\end{aligned}
$$

Since $\partial_i f\in C(X_i,Y)$, we know

$$
\begin{aligned}
&\lim_{\Delta\boldsymbol{x}_i\to\boldsymbol{0}}\sup\Big\{\big\|\partial_0 f(\boldsymbol{x}_0,\cdots,\boldsymbol{\xi}_i,\cdots,\boldsymbol{x}_{n-1}+\Delta\boldsymbol{x}_{n-1}) \\
&\qquad\qquad -\partial_0 f(\boldsymbol{x}_0,\cdots,\boldsymbol{x}_i,\cdots,\boldsymbol{x}_{n-1}+\Delta\boldsymbol{x}_{n-1})\big\|_Y \mid \boldsymbol{\xi}_i\in[\boldsymbol{x}_0,\boldsymbol{x}_0+\Delta\boldsymbol{x}_0]\Big\} \\
&=0\,.
\end{aligned}
$$

Since $\max\{\|\Delta\boldsymbol{x}_i\|_{X_i}\}_{i\in n}\leq\|\Delta\boldsymbol{x}\|_X$ (check Eq. (10-2)), we know that

$$
f(\boldsymbol{x}+\Delta\boldsymbol{x})-f(\boldsymbol{x})-\mathscr{L}\Delta\boldsymbol{x}=o(\Delta\boldsymbol{x})\,,
$$

which means $\mathrm{d}f(\boldsymbol{x})=\mathscr{L}$. □

Then we shall use finite-increment theorem 11.1 to prove some useful theorems.

Theorem 11.5 (Derivative functions doesn't have removable discontinuity)**.** *Let X, Y be two normed spaces, $\boldsymbol{x}_0\in X$, $U\in\mathscr{U}(\boldsymbol{x}_0)$, $f\in Y^U$. If f is differentiable in $\mathring{U}:=U-\{\boldsymbol{x}_0\}$, and*

$$
\lim_{\boldsymbol{x}\to\boldsymbol{x}_0}f'(\boldsymbol{x})=\mathscr{L}\in\mathcal{B}(X;Y)\,,
$$

then f is differentiable at $\boldsymbol{x}_0$ and $f'(\boldsymbol{x}_0)=\mathscr{L}$.

Proof. Find a $\Delta\boldsymbol{x}$ that satisfies $[\boldsymbol{x},\boldsymbol{x}+\Delta\boldsymbol{x}]\subset U$. By Corollary 4, as $\Delta\boldsymbol{x}\to 0$, we have

$$
\|f(\boldsymbol{x}_0+\Delta\boldsymbol{x})-f(\boldsymbol{x}_0)-\mathscr{L}\Delta\boldsymbol{x}\|_Y\leq\sup\{\|f'(\boldsymbol{\xi})-\mathscr{L}\|\,|\,\boldsymbol{\xi}\in(\boldsymbol{x}_0,\boldsymbol{x}_0+\Delta\boldsymbol{x})\}\,\|\Delta\boldsymbol{x}\|_X=o(1)\|\Delta\boldsymbol{x}\|_X=o(\Delta\boldsymbol{x})\,.
$$

By the definition of differetial, we know $f'(\boldsymbol{x}_0)=\mathscr{L}$. □

Theorem 11.6 (Constant if derivative is zero in a convex open set)**.** *Let X, Y be normed spaces, U be a convex open set in X, $f\in Y^U$. If $\forall\boldsymbol{x}\in U$, f is differentiable at $\boldsymbol{x}$, and $f'(\boldsymbol{x})=\mathscr{O}$, then f is a constant function from U i.e. $\exists\boldsymbol{y}_0\in Y$, $\forall\boldsymbol{x}\in U$, $f(\boldsymbol{x})=\boldsymbol{y}_0$.*

Proof. Let $\boldsymbol{x}_0\in U$. $\forall\boldsymbol{x}\in U$, since U is convex, $[\boldsymbol{x}_0,\boldsymbol{x}]\subset U$. The finite-increment theorem 11.1 therefore yields:

$$
\|f(\boldsymbol{x})-f(\boldsymbol{x}_0)\|_Y\leq\sup\{\|f'(\boldsymbol{\xi})\|\mid\boldsymbol{\xi}\in[\boldsymbol{x}_0,\boldsymbol{x}]\}\|\boldsymbol{x}-\boldsymbol{x}_0\|_X=0\,.
$$

In the normed space Y this implies that $f(\boldsymbol{x}_0)=f(\boldsymbol{x})$. □

Theorem 11.7 (Constant if derivative is zero in a connected open set)**.** *Let X, Y be normed spaces, U be a connected open set in X, $f \in Y^U$. If $\forall \boldsymbol{x} \in U$, f is differentiable at $\boldsymbol{x}$, and $f'(\boldsymbol{x}) = \mathscr{O}$, then f is a constant function from U.*

Proof. Let $\boldsymbol{x}_0 \in U$. Consider a set $E := \{\boldsymbol{x} \in U \mid f(\boldsymbol{x}) = f(\boldsymbol{x}_0)\}$.

First, E is open. $\forall \boldsymbol{x} \in E$, $\exists B(\boldsymbol{x};\delta) \subset U$. Since $\forall \boldsymbol{x}' \in B(\boldsymbol{x};\delta)$, $[\boldsymbol{x}, \boldsymbol{x}'] \subset B(\boldsymbol{x};\delta)$, f is constant in $B(\boldsymbol{x};\delta)$ and therefore $B(\boldsymbol{x};\delta) \subset E$. In conclusion, all points in E are interior.

Then, $U - E$ is also open in the topological subspace U, with the same reason.

Since E is not empty, ($\boldsymbol{x}_0 \in E$), the only choice for a open-closed set in a connected set U is U itself, i.e. $\forall \boldsymbol{x} \in U$, $f(\boldsymbol{x}) = f(\boldsymbol{x}_0)$. □

§12 Higher-Order Derivative

We denote the zeroth and first differential of $f \in Y^U$, where U is an open set in a normed space X, by $f^{(0)} := f$, $f^{(1)} := f'$.

Definition 12.1 (n-th differentiation)**.** Let X and Y be normed spaces, with induced topologies $\mathscr{T}_X$ and $\mathscr{T}_Y$. For brevity, we define $Y_0 := Y$, and $Y_{n+1} := \mathcal{B}(X; Y_n)$.

The definition of ***n*-th differential** is introduced below recursively: We have already defined the zeroth and the first defferentiation. If the n-th differential $f^{(n)} \in {Y_n}^U$ is differentiable in $U \in \mathscr{T}_X$ [10], we can define the $(n+1)$-th differential $f^{(n)}(\boldsymbol{x})$ by:

$$f^{(n+1)} = (f^{(n)})' .$$

Theorem 12.1 (Higher-oder differentiation operates on vectors)**.** *Let X and Y be normed spaces, with induced topologies $\mathscr{T}_X$ and $\mathscr{T}_Y$, $U \in \mathscr{T}_X$, $\boldsymbol{x} \in U$, $(\boldsymbol{\ell}_i)_{i\in n} \in X^n$. If $f \in Y^U$ has n-th differential $f^{(n)}$ in U,*

$$((f^{(n)}(\boldsymbol{x})\boldsymbol{\ell}_0)\cdots\boldsymbol{\ell}_{n-1}) = \frac{\partial}{\partial\boldsymbol{\ell}_0}\cdots\frac{\partial}{\partial\boldsymbol{\ell}_{n-1}} f(\boldsymbol{x}) . \tag{12-1}$$

Proof. See Theorem 10.10. □

Theorem 12.2 (Symmetry of higher-order differentiation)**.** *Let $\sigma \in S_n$ where S_n is the symmetric group*[11] *on n. Let X and Y be normed spaces, with induced topologies $\mathscr{T}_X$ and $\mathscr{T}_Y$, $U \in \mathscr{T}_X$, $\boldsymbol{x} \in U$, $(\boldsymbol{\ell}_i)_{i\in n} \in X^n$. If $f \in Y^U$ has n-th differential $f^{(n)}$ in U, then*

$$\frac{\partial}{\partial\boldsymbol{\ell}_{\sigma(0)}}\cdots\frac{\partial}{\partial\boldsymbol{\ell}_{\sigma(n-1)}} f(\boldsymbol{x}) = \frac{\partial}{\partial\boldsymbol{\ell}_0}\cdots\frac{\partial}{\partial\boldsymbol{\ell}_{n-1}} f(\boldsymbol{x}) .$$

Proof. We shall only prove the case when $n = 2$.

The second differetial $f''(\boldsymbol{x})$ exists implies that the first differential $f'(\boldsymbol{x})$ also exists. Since U is open, there exists an open ball $B(0;\delta) \subset U$, where $\delta \in \mathbb{R}_+$.

Let

$$\begin{aligned}\Delta(t) &:= f(\boldsymbol{x} + t\boldsymbol{\ell}_0 + t\boldsymbol{\ell}_1) - f(\boldsymbol{x} + t\boldsymbol{\ell}_0) - f(\boldsymbol{x} + t\boldsymbol{\ell}_1) + f(\boldsymbol{x}) ,\\ D(t, t') &:= f(\boldsymbol{x} + t\boldsymbol{\ell}_0 + t'\boldsymbol{\ell}_1) - f(\boldsymbol{x} + t'\boldsymbol{\ell}_1) ,\end{aligned}$$

[10] Y_n is also a normed space.
[11] Or permutation

where $t \in [0, \delta)$, $t' \in [0, t]$.

It is obvious that $\Delta(t) = D(t,t) - D(t,0)$. By the finite-increment theorem 11.1,

$$\begin{aligned}
\|\Delta(t) - t^2[f''(\boldsymbol{x})\boldsymbol{\ell}_0]\boldsymbol{\ell}_1\|_Y &= \|D(t,t) - D(t,0) - t^2[f''(\boldsymbol{x})\boldsymbol{\ell}_0]\boldsymbol{\ell}_1\|_Y \\
&\le t \sup\left\{ \left\| \frac{\partial D}{\partial t'}(t,\theta) - t\theta[f''(\boldsymbol{x})\boldsymbol{\ell}_0]\boldsymbol{\ell}_1 \right\|_Y \middle| \theta \in (0,t) \right\} \\
&\le t\|\boldsymbol{\ell}_1\|_X \sup\left\{ \|f'(\boldsymbol{x} + t\boldsymbol{\ell}_0 + \theta\boldsymbol{\ell}_1) - f'(\boldsymbol{x} + \theta\boldsymbol{\ell}_1) - t\theta f''(\boldsymbol{x})\boldsymbol{\ell}_0\| \middle| \theta \in (0,t) \right\} \\
&= t\|\boldsymbol{\ell}_1\|_X \sup\left\{ \|\theta f''(\boldsymbol{x})(t\boldsymbol{\ell}_0 + \theta\boldsymbol{\ell}_1 - \theta\boldsymbol{\ell}_1) - t\theta f''(\boldsymbol{x})\boldsymbol{\ell}_0 + o(t)\| \middle| \theta \in (0,t) \right\} \\
&= o(t^2)\,.
\end{aligned}$$

Hence,

$$[f''(\boldsymbol{x})\boldsymbol{\ell}_0]\boldsymbol{\ell}_1 = \lim_{t\to 0} \frac{\Delta(t)}{t^2}\,.$$

Substituting $(\boldsymbol{\ell}_0, \boldsymbol{\ell}_1)$ by $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_0)$ in the definition of $\Delta(t)$ doesn't change its value, hence we have proved the theorem in the case when $n = 2$. □

Theorem 12.2 implies that the n-th derivative $f^{(n)}(\boldsymbol{x})$ corresponds to a n-symmetric multilinear operator in $\mathcal{B}(X, \cdots, X; Y)$[12], and we shall denote:

$$f^{(n)}(\boldsymbol{x})(\boldsymbol{\ell}_i)_{i\in n} := ((f^{(n)}(\boldsymbol{x})\boldsymbol{\ell}_0)\cdots)\boldsymbol{\ell}_{n-1}\,, \tag{12-2}$$

and

$$f^{(n)}(\boldsymbol{x})\boldsymbol{\ell}^n := f^{(n)}(\boldsymbol{\ell}, \cdots, \boldsymbol{\ell})\,. \tag{12-3}$$

Theorem 12.3. *Let $X_0, \cdots, X_{m-1}, Y$ be normed spaces, and $X := \prod_{i\in m} X_i$. Let $f \in Y^U$ where U is an open set in X. If $\forall (i_k)_{k\in n} \in m^n$, $\forall \boldsymbol{x} \in U$, n-th partial derivative*

$$\partial_{i_0} \cdots \partial_{i_{m-1}} f(\boldsymbol{x})$$

exists and continuous (with respect to $\boldsymbol{x}$), then f is n-th differentiable at $\boldsymbol{x}$ i.e. $f^{(n)}$ exists, and is also continous.

Further more,

$$f \in C^{(n)}(U) \;\leftrightarrow\; \forall (i_k)_{k\in n} \in m^n,\; \partial_{i_0} \cdots \partial_{i_{m-1}} f \in C\,,$$

where we denote the set of n-th differentiable functions on U by $C^{(n)}(U;Y)$ ($C^{(n)}(U)$, alternatively).

§13 Applications of Differentiation

13.1 Taylor's Formula

Theorem 13.1 (Taylor's formula)**.** *Let X and Y be two normed spaces, $\boldsymbol{x} \in X$, $U \in \mathscr{U}(x)$, $f \in Y^U$. If f is $(n-1)$-th differentiable in U, and n-th differentiable at point $\boldsymbol{x}$, then as $\Delta\boldsymbol{x} \to \boldsymbol{0}$ ($\boldsymbol{x} + \Delta\boldsymbol{x} \in U$),*

$$f(\boldsymbol{x} + \Delta\boldsymbol{x}) = \sum_{k\in n+1} f^{(k)}(\boldsymbol{x}) \frac{\Delta\boldsymbol{x}^k}{k!} + o(\|\Delta\boldsymbol{x}\|_X^n)\,, \tag{13-1}$$

where we have made use of the notation we introduced at Eq. (12-3).

[12]By Corollary 3, these two spaces are isomorphic

Proof. If we consider each term of the Taylor's formula as a function of $\Delta\boldsymbol{x}$, we can find them to be differentiable (with respect to $\Delta\boldsymbol{x}$), since $f^{(k)}(\boldsymbol{x}) \in \mathcal{B}(X, \cdots, X; Y)$. The derivative of the symmetric k-linear operator

$$T_k(\Delta\boldsymbol{x}) := \frac{1}{k!} f^{(k)}(\boldsymbol{x})\Delta\boldsymbol{x}^k$$

with respect to $\Delta\boldsymbol{x}$ is[13]:

$$T_k'(\Delta\boldsymbol{x})\boldsymbol{\ell} = \frac{1}{(k-1)!} f^{(k)}(\boldsymbol{x})\Delta\boldsymbol{x}^{k-1}\boldsymbol{\ell}\,.$$

Hence, if we assume that the Eq. (13-1) holds for $n-1$, by the finite-increment theorem 11.1, we conclude:

$$\begin{aligned}
&\left\| f(\boldsymbol{x}+\Delta\boldsymbol{x}) - \sum_{k\in n+1} T_k(\Delta\boldsymbol{x}) \right\|_Y \\
&\quad \le \sup\left\{ \left\| f'(\boldsymbol{x}+\boldsymbol{\xi}) - \sum_{k\in n} \frac{1}{k!} f^{(k+1)}(\boldsymbol{x})\boldsymbol{\xi}^k \right\|_Y \,\middle|\, \boldsymbol{\xi} \in [\boldsymbol{0}, \Delta\boldsymbol{x}] \right\} \|\Delta\boldsymbol{x}\|_X \\
&\quad = o(\boldsymbol{\xi}^{n-1})\|\Delta\boldsymbol{x}\|_X = o(\Delta\boldsymbol{x}^n)\,.
\end{aligned}$$

□

Theorem 13.2. *Let X, Y be two normed spaces, U be an open set in X, $f \in C^{(n)}(X;Y)$. Let $[\boldsymbol{x}, \boldsymbol{x}+\Delta\boldsymbol{x}] \subset U$, and f be $(n+1)$-th differentiable in $(\boldsymbol{x}, \boldsymbol{x}+\Delta\boldsymbol{x})$.*

If $\forall\boldsymbol{\xi} \in (\boldsymbol{x}, \boldsymbol{x}+\Delta\boldsymbol{x})$, $\|f^{(n+1)}(\boldsymbol{\xi})\| \le M$, then

$$\left\| f(\boldsymbol{x}+\Delta\boldsymbol{x}) - \sum_{k\in n+1} \frac{1}{k!} f^{(k)}(\boldsymbol{x})\Delta\boldsymbol{x}^k \right\|_Y \le \frac{M}{(n+1)!}\|\Delta\boldsymbol{x}\|_X^{n+1}\,.$$

Proof. Define a function $g \in Y^{[0,1]}$:

$$g(t) := f(\boldsymbol{x}+\Delta\boldsymbol{x}) - \sum_{k\in n+1} \frac{(1-t)^k}{k!} f^{(k)}(\boldsymbol{x}+t\Delta\boldsymbol{x})\Delta\boldsymbol{x}^k\,,$$

Notice the derivative of $(1-t)^k f^{(k)}(\boldsymbol{x}+t\Delta\boldsymbol{x})/k!$ with respect to t is:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{(1-t)^k}{k!} f^{(k)}(\boldsymbol{x}+t\Delta\boldsymbol{x})\right) = \frac{(1-t)^k}{k!} f^{(k+1)}(\boldsymbol{x}+t\Delta\boldsymbol{x})\Delta\boldsymbol{x} - \frac{k(1-t)^{k-1}}{k!} f^{(k)}(\boldsymbol{x}+t\Delta\boldsymbol{x})\,,$$

We have:

$$g'(t) = -\frac{(1-t)^n}{n!} f^{(n+1)}(\boldsymbol{x}+t\Delta\boldsymbol{x})\Delta\boldsymbol{x}^{n+1}\,,$$

therefore

$$\|g'(t)\| \le \frac{|1-t|^n}{n!}\|f^{(n+1)}(\boldsymbol{x}+t\Delta\boldsymbol{x})\|\|\Delta\boldsymbol{x}\|_X^{n+1} \le \frac{M(1-t)^n}{n!}\|\Delta\boldsymbol{x}\|_X^{n+1}\,.$$

[13] cf. Theorem 10.6

Making use of $[-(1-t)^{n+1}]' = (n+1)(1-t)^n$ and the definition of differentiation, $\forall \varepsilon \in \mathbb{R}_+$, $\exists \delta \in \mathbb{R}_+$, if $1-t \le \delta$, then:

$$\|g(t)\|_Y - \frac{\varepsilon}{2}(1-t) \le \|g'(t)\|(1-t) \le \frac{M(1-t)^n}{n!}(1-t)\|\Delta \boldsymbol{x}\|_X^{n+1} \le \frac{M(1-t)^{n+1}}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1} + \frac{\varepsilon}{2}(1-t)\,,$$

or

$$\|g(t)\|_Y \le \frac{M(1-t)^{n+1}}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1} + \varepsilon(1-t)\,.$$

Since such δ exists, for ε, we define δ' as the supremum of the δs[1, p. 64], i.e.

$$\delta' := \sup\left\{\delta \in \mathbb{R}_+ \middle| 1-t \le \delta \ \rightarrow\ \|g(t)\|_Y \le \frac{M(1-t)^{n+1}}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1} + \varepsilon(1-t)\,.\right\}$$

If $\delta' \ne 1$, then for $t < 1-\delta'$, again we make use of the definition of differentiation, starting at δ', $\exists \eta \in \mathbb{R}_+$, if $\delta' - t \le \eta$, then

$$\|g(t) - g(\delta')\|_Y \le \frac{M[(1-\delta')^{n+1} - (1-t)^{n+1}]}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1} + \varepsilon(\delta' - t)\,,$$

and

$$\begin{aligned}
\|g(t)\|_Y &\le \|g(t) - g(\delta')\|_Y + \|g(\delta')\|_Y \\
&\le \frac{M[(1-\delta')^{n+1} - (1-t)^{n+1}]}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1} + \varepsilon(\delta' - t) + \frac{M(1-\delta')^{n+1}}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1} + \varepsilon(1-\delta') \\
&= \frac{M(1-t)^{n+1}}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1} + \varepsilon(1-t)\,,
\end{aligned}$$

which contradicts to the definition of δ'.

Hence $\delta' = 1$, or:

$$\|g(0)\|_Y \le \frac{M}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1} + \varepsilon\,,$$

which holds for any $\varepsilon \in \mathbb{R}_+$, hence:

$$\|g(0)\|_Y \le \frac{M}{(n+1)!}\|\Delta \boldsymbol{x}\|_X^{n+1}\,. \tag{13-2}$$

Eq. (13-2) is to prove. □

Lemma 6. *Let X, Y be a linear space, $\mathscr{A} \in \mathcal{B}(X, \cdots, X; Y)$ i.e. $\mathscr{A}$ is an n-linear operators from X, ..., X to Y. If $\forall \boldsymbol{x} \in X$, $\mathscr{A}\boldsymbol{x}^n = \boldsymbol{0}$, then $\forall (\boldsymbol{x}_i)_{i\in n} \in X^n$, $\mathscr{A}(\boldsymbol{x}_i)_{i\in n} = \boldsymbol{0}$.*

Proof.

$$\begin{aligned}
2\mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_1) &= \mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_1) + \mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_2) \\
&= \mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_0) + \mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_1 - \boldsymbol{x}_0) + \mathscr{A}(\boldsymbol{x}_1, \boldsymbol{x}_0 - \boldsymbol{x}_1) + \mathscr{A}(\boldsymbol{x}_1, \boldsymbol{x}_1) \\
&= \mathscr{A}(\boldsymbol{x}_0, \boldsymbol{x}_0) + \mathscr{A}(\boldsymbol{x}_1, \boldsymbol{x}_1) - \mathscr{A}(\boldsymbol{x}_1 - \boldsymbol{x}_0, \boldsymbol{x}_1 - \boldsymbol{x}_0)\,.
\end{aligned}$$

□

Theorem 13.3 (The uniqueness of Taylor's finite expansion)**.** *Let X, Y be normed spaces, $f \in Y^U$ where U is an open set in X. If f is n-th differentiable at point $\boldsymbol{x} \in U$, and $\forall k \in n+1$, exists k-linear operators $\mathscr{L}_k$ s.t.*

$$f(\boldsymbol{x} + \Delta\boldsymbol{x}) = \sum_{k \in n+1} \mathscr{L}_k \Delta\boldsymbol{x}^k + o(\|\Delta\boldsymbol{x}\|_X^n)$$

as $\Delta\boldsymbol{x} \to \mathbf{0}$, then, $\mathscr{L}_k = f^{(k)}(\boldsymbol{x})$.

Proof. It is obvious that $\mathscr{L}_0 = f^{(0)}(\boldsymbol{x}) = f(\boldsymbol{x})$. Assume that $\forall i \in k$, $f^{(i)}(\boldsymbol{x}) = \mathscr{L}_i$, then

$$\sum_{i \in k+1} \frac{1}{i!} f^{(i)}(\boldsymbol{x}) \Delta\boldsymbol{x}^i + o(\|\Delta\boldsymbol{x}\|_X^k) = \sum_{i \in k+1} \frac{1}{i!} \mathscr{L}_i \Delta\boldsymbol{x}^i + o(\|\Delta\boldsymbol{x}\|_X^k)\,,$$

hence:

$$[f^{(k)}(\boldsymbol{x}) - \mathscr{L}_k] \Delta\boldsymbol{x}^k = o(\|\Delta\boldsymbol{x}\|_X^k)\,.$$

Divides each sides by $\|\Delta\boldsymbol{x}\|_X^k$ and passing the limit $\Delta\boldsymbol{x} \to 0$, we have:

$$\lim_{\Delta\boldsymbol{x} \to 0} [f^{(k)}(\boldsymbol{x}) - \mathscr{L}_k] \left(\frac{\Delta\boldsymbol{x}}{\|\Delta\boldsymbol{x}\|_X} \right)^k = \lim_{\Delta\boldsymbol{x} \to 0} o(1) = \mathbf{0}\,,$$

which means $\forall \hat{\boldsymbol{e}} \in X$ s.t. $\|\hat{\boldsymbol{e}}\|_X = 1$, $[f^{(k)}(\boldsymbol{x}) - \mathscr{L}_k] \hat{\boldsymbol{e}}^k = \mathbf{0}$. This means $f^{(k)}(\boldsymbol{x}) - \mathscr{L}_k = \mathscr{O}$, by Lemma 6. □

13.2 Interior Extrema

Definition 13.1 (Extremum)**.** Let X be a normed space, and $f \in \mathbb{R}^X$. If $\boldsymbol{x} \in X$ satisfies: $\exists U \in \mathscr{U}(\boldsymbol{x})$ s.t. $\forall \boldsymbol{x}' \in U - \{\boldsymbol{x}\}$, $f(\boldsymbol{x}) > f(\boldsymbol{x}')$, then $\boldsymbol{x}$ is a ***locally maximum point*** of f. Similarily, we can define ***locally minimun point***. Both locally maximum point and minimum point are called ***extrmum point***.

Theorem 13.4. *Let X be a normed space, U is an open set in X, and $f \in \mathbb{R}^U$. The mapping f is n-th differentiable in U, and $(n+1)$-th differentiable at $\boldsymbol{x} \in U$, where $n \in \mathbb{N}_+$. $\forall k \in n+1$, $f^{(k)}(\boldsymbol{x}) = \mathscr{O}$, and $f^{(n+1)}(\boldsymbol{x}) \neq \mathscr{O}$.*

If f reach its extremum at $\boldsymbol{x}$, then $n+1 \in 2\mathbb{Z}$ and $f^{(n+1)}(\boldsymbol{x})$ is semidefinite, i.e. $\nexists \Delta\boldsymbol{x}, \Delta\boldsymbol{x}' \in X$ s.t. $f^{(n+1)}(\boldsymbol{x}) \Delta\boldsymbol{x}^{n+1} f^{(n+1)}(\boldsymbol{x}) \Delta\boldsymbol{x}'^{n+1} < 0$.

Proof. $\exists \Delta\boldsymbol{x} \in X$, $f^{(n+1)}(\boldsymbol{x}) \Delta\boldsymbol{x}^{n+1} \neq 0$ since $f^{(n+1)}(\boldsymbol{x}) \neq \mathscr{O}$. $\exists \delta \in \mathbb{R}_+$, as $t \in (-\delta, \delta)$,

$$o(1) = \frac{1}{t^{n+1}} o((t\Delta\boldsymbol{x})^n) > -\frac{1}{(n+1)!} f^{(n+1)}(\boldsymbol{x}) \Delta\boldsymbol{x}^{n+1}\,,$$

hence

$$f(\boldsymbol{x} + t\Delta\boldsymbol{x}) - f(\boldsymbol{x}) = \left(\frac{1}{(n+1)!} f^{(n+1)}(\boldsymbol{x}) \Delta\boldsymbol{x}^{n+1} + o(1) \right) t^{n+1}\,.$$

If the difference remians its sign, then $n+1$ must be an even number. □

Theorem 13.5. *Let X be a normed space, U is an open set in X, and $f \in \mathbb{R}^U$. The mapping f is n-th differentiable in U, and $(n+1)$-th differentiable at $\boldsymbol{x} \in U$, where $n \in \mathbb{N}_+$. $\forall k \in n+1$, $f^{(k)}(\boldsymbol{x}) = \mathscr{O}$, and $f^{(n+1)}(\boldsymbol{x}) \neq \mathscr{O}$.*

If $\exists \delta \in \mathbb{R}_+$, $\forall \hat{\mathbf{e}} \in X$ s.t. $\|\hat{\mathbf{e}}\|_X = 1$, $|f^{(n+1)}(\boldsymbol{x})\hat{\mathbf{e}}^{n+1}| \geq \delta$, then f reaches its extremum. If $f^{(n+1)}(\boldsymbol{x})\hat{\mathbf{e}}^{n+1} > 0$, then $\boldsymbol{x}$ is a local maximum point; If $f^{(n+1)}(\boldsymbol{x})\hat{\mathbf{e}}^{n+1} < 0$, then $\boldsymbol{x}$ is a local minimum point.

Proof. Assume that $f^{(n+1)}(\boldsymbol{x})\Delta\boldsymbol{x}^{n+1} > 0$.

$$\begin{aligned}
f(\boldsymbol{x} - \Delta\boldsymbol{x}) - f(\boldsymbol{x}) &= \frac{1}{k!} f^{(n+1)}(\boldsymbol{x})\Delta\boldsymbol{x}^{n+1} + o(\Delta\boldsymbol{x}^{n+1}) \\
&= \|\Delta\boldsymbol{x}\|_X^{n+1} \left(\frac{1}{k!} f^{(n+1)}(\boldsymbol{x}) \left(\frac{\Delta\boldsymbol{x}}{\|\Delta\boldsymbol{x}\|_X} \right)^{n+1} + o(1) \right) \\
&\geq \|\Delta\boldsymbol{x}\|_X^{n+1} \left(\frac{\delta}{k!} + o(1) \right) \to \|\Delta\boldsymbol{x}\|_X^{n+1} \frac{\delta}{k!} > 0\,.
\end{aligned}$$

□

§14 Implicit Function Theorem

Theorem 14.1 (Implicit function theorem)**.** *Let X, Z be normed spaces, and Y be a Banach space. $\boldsymbol{x}_0 \in X$, $\boldsymbol{y}_0 \in Y$. Denote*

$$W := B(\boldsymbol{x}_0; \alpha) \times B(\boldsymbol{y}_0; \beta)\,,$$

where $\alpha, \beta \in \mathbb{R}_+$. If $F \in Z^W$ satisfies:

a) $F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \mathbf{0}$;
b) F is continuous at $(\boldsymbol{x}_0, \boldsymbol{y}_0)$;
c) There exists the partial derivative of $F(\boldsymbol{x}, \boldsymbol{y})$ with respect to $\boldsymbol{y} \in Y$: $\partial_{\boldsymbol{y}} F(\boldsymbol{x}, \boldsymbol{y})$ in W, and $\partial_{\boldsymbol{y}} F$ is continuous at point $(\boldsymbol{x}_0, \boldsymbol{y}_0)$;
d) $\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{B}(Y; Z)$ is reversible i.e. $\exists [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \in \mathcal{B}(Z; Y)$ s.t.

$$\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \circ [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} = [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \circ \partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \mathrm{id}_Y\,,$$

then, $\exists U \in \mathscr{U}(\boldsymbol{x}_0)$, $\exists V \in \mathscr{U}(\boldsymbol{y}_0)$, $\exists f \in V^U$ s.t. f is continuous at $\boldsymbol{x}_0$, $U \times V \subset W$ and $\forall \boldsymbol{x} \in U$, $\forall \boldsymbol{y} \in V$,

$$F(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0} \;\leftrightarrow\; f(\boldsymbol{x}) = \boldsymbol{y}\,.$$

Before our proof of the theorem, some explanation to it might be necessary. Given a $\boldsymbol{x} \in B(\boldsymbol{x}_0; \alpha)$, we want to find a $f(\boldsymbol{x}) \in B(\boldsymbol{y}_0; \beta)$ that satisfies $F[\boldsymbol{x}, f(\boldsymbol{x})] = \mathbf{0}$. If we have made an guess $\boldsymbol{y}$, the error shall be $\Delta = f(\boldsymbol{x}) - \boldsymbol{y}$, of course since we don't know exactly what $f(\boldsymbol{x})$ is, we shall estimate it.

Then we made an approximation. We assume that the behaviour of $F(\boldsymbol{x}, \boldsymbol{y})$ is linear with respect to $\boldsymbol{y}$ around $(\boldsymbol{x}, f(\boldsymbol{x}))$, i.e.

$$F(\boldsymbol{x}, \boldsymbol{y}) \approx \partial_{\boldsymbol{y}} F(\boldsymbol{x}, f(\boldsymbol{x}))(\boldsymbol{y} - f(\boldsymbol{x})) \approx \partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)(\boldsymbol{y} - f(\boldsymbol{x}))$$

If we find that $F(\boldsymbol{x},\boldsymbol{y}) \neq \boldsymbol{0}$, we know that $\boldsymbol{y}$ is not the $f(\boldsymbol{x})$ we are searching for, and by our approximation, it is about:

$$\Delta \approx \tilde{\Delta} = [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0)]^{-1} \circ F(\boldsymbol{x},\boldsymbol{y}) \,.$$

Making use of our estimate of the error to correct $\boldsymbol{y}$, we get $\boldsymbol{y}' = \boldsymbol{y} - \tilde{\Delta}$. However, since we made a approximation (which is too much!), $\boldsymbol{y}'$ is also not $f(\boldsymbol{x})$. So we repeat the procedure, which is estimate the error, correct it, and estimate the error again ...

But wait, would we finally get what we want? In analysis this is a bad question – maybe we shall ask: as we repeat the procedure, would the result gets closed enough to the answer? The proof below would answer.

Proof. Consider a function from $B(\boldsymbol{y}_0;\beta)$ to Y:

$$\Delta_{\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{y} - [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0)]^{-1} \circ F(\boldsymbol{x},\boldsymbol{y}) \,,$$

Obviously, if $\boldsymbol{y} = f(\boldsymbol{x}) \leftrightarrow F(\boldsymbol{x},\boldsymbol{y})$ then $\Delta_{\boldsymbol{x}}(f(\boldsymbol{x})) = f(\boldsymbol{x})$ i.e. $f(\boldsymbol{x})$ is a fix-point of the function $\Delta_{\boldsymbol{x}}(\boldsymbol{y})$ of $\boldsymbol{y}$ with fixed $\boldsymbol{x}$. Now we need to prove such fix-point exists.

The function $F(\boldsymbol{x},\boldsymbol{y})$ is differentiable with respect to $\boldsymbol{y}$ in W, so is $\Delta_{\boldsymbol{x}}(\boldsymbol{y})$:

$$\Delta'_{\boldsymbol{x}}(\boldsymbol{y}) = \mathrm{id}_Y - [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{y}} F(\boldsymbol{x},\boldsymbol{y}) = [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0)]^{-1} [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0) - \partial_{\boldsymbol{y}} F(\boldsymbol{x},\boldsymbol{y})] \,.$$

Take the norm of each side and by Theorem 9.7, we have:

$$\|\Delta'_{\boldsymbol{x}}(\boldsymbol{y})\| \leq \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0)]^{-1}\| \|\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0) - \partial_{\boldsymbol{y}} F(\boldsymbol{x},\boldsymbol{y})\| \,.$$

Since $\partial_{\boldsymbol{y}} F$ is continuous at $(\boldsymbol{x}_0,\boldsymbol{y}_0)$, $\forall \varepsilon \in (0,1)$, if γ is small enough, $\forall \boldsymbol{x} \in B(\boldsymbol{x}_0;\gamma/2)$, $\forall \boldsymbol{y} \in B(\boldsymbol{y}_0;\gamma/2)$[14] ,

$$\|\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0) - \partial_{\boldsymbol{y}} F(\boldsymbol{x},\boldsymbol{y})\| < \frac{\varepsilon}{\|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0)]^{-1}\|} \,.$$

By the finite-increment theorem 11.1, $\forall \boldsymbol{y},\boldsymbol{y}' \in B(\boldsymbol{y}_0;\gamma/2)$,

$$\begin{aligned}
\|\Delta_{\boldsymbol{x}}(\boldsymbol{y}') - \Delta_{\boldsymbol{x}}(\boldsymbol{y})\|_Y &\leq \sup\{\|\Delta'_{\boldsymbol{x}}(\boldsymbol{\xi})\| \mid \boldsymbol{\xi} \in [\boldsymbol{y},\boldsymbol{y}']\} \|\boldsymbol{y} - \boldsymbol{y}'\|_Y \\
&\leq \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0)]^{-1}\| \|\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0,\boldsymbol{y}_0) - \partial_{\boldsymbol{y}} F(\boldsymbol{x},\boldsymbol{y})\| \|\boldsymbol{y} - \boldsymbol{y}'\|_Y \\
&< \varepsilon \|\boldsymbol{y} - \boldsymbol{y}'\|_Y \,.
\end{aligned}$$

In another word, $\Delta_{\boldsymbol{x}}(\boldsymbol{y})$ is a ε-contraction, from $B(\boldsymbol{y}_0;\gamma/2)$ to $B(\boldsymbol{y}_0;\gamma/2)$.

To apply the Picard-Banach fixed-point principle 7.1, we need to find a closed metric subspace $\big(\tilde{B}(\boldsymbol{y}_0;\delta), d_Y\big)$, where $\delta \leq \gamma/2$. By Theorem 5.1, $\tilde{B}(\boldsymbol{y}_0;\delta)$ is also complete. But we don't know if $\Delta_{\boldsymbol{x}}\big(\tilde{B}(\boldsymbol{y}_0;\delta)\big) \subset \tilde{B}(\boldsymbol{y}_0;\delta)$ yet. To satisfy this, we find a $\zeta \in (0,\gamma/2)$, s.t. $\|\Delta_{\boldsymbol{x}}(\boldsymbol{y}_0) - \boldsymbol{y}_0\|_Y < \delta(1-\varepsilon)$ if $d_X(\boldsymbol{x},\boldsymbol{x}_0) < \zeta$ so that

$$\|\Delta_{\boldsymbol{x}}(\boldsymbol{y}) - \boldsymbol{y}_0\|_Y \leq \|\Delta_{\boldsymbol{x}}(\boldsymbol{y}) - \Delta_{\boldsymbol{x}}(\boldsymbol{y}_0)\|_Y + \|\Delta_{\boldsymbol{x}}(\boldsymbol{y}_0) - \boldsymbol{y}_0\|_Y < \varepsilon\|\boldsymbol{y} - \boldsymbol{y}_0\|_Y + (\varepsilon-1)\delta < \varepsilon\varepsilon + (\varepsilon-1)\delta = \delta \,.$$

Hence, there exists the unique fixed point for $\Delta_{\boldsymbol{x}}(\boldsymbol{y}) \in \tilde{B}(\boldsymbol{y};\delta)$ for each $\boldsymbol{x} \in U := B(\boldsymbol{x}_0;\zeta)$, which is the $f(\boldsymbol{x})$ we have been searching for.

Finally we check if $f\colon U \to V$ is continuous at $\boldsymbol{x}_0$. For any $\delta' \in (0,\delta)$, we can find another $\zeta' \in (0,\zeta)$ s.t. $\|\Delta_{\boldsymbol{x}}(\boldsymbol{y}_0) - \boldsymbol{y}_0\|_Y < \delta'(1-\varepsilon)$ if $d_X(\boldsymbol{x},\boldsymbol{x}_0) < \zeta'$, so that $\|\Delta_{\boldsymbol{x}}(\boldsymbol{y}) - \boldsymbol{y}_0\|_Y < \delta'$. □

[14] so that $d_{X\times Y}\big((\boldsymbol{x},\boldsymbol{y}),(\boldsymbol{x}_0,\boldsymbol{y}_0)\big) = \sqrt[p]{\|\boldsymbol{x}-\boldsymbol{x}_0\|_X^p + \|\boldsymbol{y}-\boldsymbol{y}_0\|_Y^p} \leq d_X(\boldsymbol{x},\boldsymbol{x}_0) + d_Y(\boldsymbol{y},\boldsymbol{y}_0) < \gamma$

Theorem 14.2 (Continuity of implicit function)**.** *Let X, Z be normed spaces, and Y be a Banach space. $\boldsymbol{x}_0 \in X$, $\boldsymbol{y}_0 \in Y$. Denote*

$$W := B(\boldsymbol{x}_0; \alpha) \times B(\boldsymbol{y}_0; \beta)\,,$$

where $\alpha, \beta \in \mathbb{R}_+$. If $F \in Z^W$ satisfies:

a) $F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \mathbf{0}$;

b) $F \in C(W; Z)$;

c) There exists the partial derivative of $F(\boldsymbol{x}, \boldsymbol{y})$ with respect to $\boldsymbol{y} \in Y$: $\partial_{\boldsymbol{y}} F(\boldsymbol{x}, \boldsymbol{y})$ in W, and $\partial_{\boldsymbol{y}} F$ is continuous at point $(\boldsymbol{x}_0, \boldsymbol{y}_0)$;

d) $\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{B}(Y; Z)$ is reversible,

then, $\exists U \in \mathscr{U}(\boldsymbol{x}_0)$, $\exists V \in \mathscr{U}(\boldsymbol{y}_0)$, $\exists f \in C(U; Y)$ s.t. $U \times V \subset W$ and $\forall \boldsymbol{x} \in U$, $\forall \boldsymbol{y} \in V$,

$$F(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0} \;\leftrightarrow\; f(\boldsymbol{x}) = \boldsymbol{y}\,.$$

Proof. By Theorem 10.7, $\|\partial F_{\boldsymbol{y}}(\boldsymbol{x}, \boldsymbol{y})^{-1}\|$ is continuous in some neighbourhoods. Hence, the conditions of implicit function theorem are also satisfied in these neighbourhoods. □

Theorem 14.3 (Differentiability of implicit function)**.** *Let X, Z be normed spaces, and Y be a Banach space. $\boldsymbol{x}_0 \in X$, $\boldsymbol{y}_0 \in Y$. Denote*

$$W := B(\boldsymbol{x}_0; \alpha) \times B(\boldsymbol{y}_0; \beta)\,,$$

where $\alpha, \beta \in \mathbb{R}_+$. If $F \in Z^W$ satisfies:

a) $F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \mathbf{0}$;

b) F is continuous at $\boldsymbol{x}_0, \boldsymbol{y}_0$;

c) There exist the partial derivatives of $F(\boldsymbol{x}, \boldsymbol{y})$ with respect to $\boldsymbol{y} \in Y$: $\partial_{\boldsymbol{y}} F(\boldsymbol{x}, \boldsymbol{y})$ and with respect to $\boldsymbol{x}$: $\partial_{\boldsymbol{x}} F(\boldsymbol{x}, \boldsymbol{y})$, in W, and $\partial_{\boldsymbol{y}} F$, $\partial_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{y})$ are continuous at point $(\boldsymbol{x}_0, \boldsymbol{y}_0)$;

d) $\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{B}(Y; Z)$ is reversible,

then, $\exists U \in \mathscr{U}(\boldsymbol{x}_0)$, $\exists V \in \mathscr{U}(\boldsymbol{y}_0)$, $\exists f \in V^U$ s.t. $U \times V \subset W$ and $\forall \boldsymbol{x} \in U$, $\forall \boldsymbol{y} \in V$,

$$F(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0} \;\leftrightarrow\; f(\boldsymbol{x}) = \boldsymbol{y}\,,$$

and, f is differentiable at $\boldsymbol{x}_0$:

$$f'(\boldsymbol{x}_0) = -[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)\,. \tag{14-1}$$

Proof. Let's vertify that the right-hand side of Eq. (14-1) is the differential of f at $\boldsymbol{x}_0$. Find a $\boldsymbol{x} + \Delta \boldsymbol{x}$ within U[15],

$$\begin{aligned}
&\big\|f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - \boldsymbol{y}_0 + [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x}\big\|_Y \\
&\quad = \big\|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \big(\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)[f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - \boldsymbol{y}_0] + \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x}\big)\big\|_Y \\
&\quad \le \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1}\| \big\|\big(\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)[f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - \boldsymbol{y}_0] + \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x}\big) \\
&\qquad\qquad + \big(F(\boldsymbol{x}, f(\boldsymbol{x})) - F(\boldsymbol{x}_0, \boldsymbol{y}_0)\big)\big\|_Y\,.
\end{aligned}$$

[15] notice that $F(\boldsymbol{x}, f(\boldsymbol{x})) = F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \mathbf{0}$.

Since $F'_{\boldsymbol{x}}$, $F'_{\boldsymbol{y}}$ are continuous at $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, F is differentiable at $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ (Theorem 11.4). As $(\boldsymbol{x}_0 + \Delta\boldsymbol{x}, f(\boldsymbol{x}_0 + \Delta\boldsymbol{x})) \to (\boldsymbol{x}_0, \boldsymbol{y}_0)$:

$$\begin{aligned}
&\big\| f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - \boldsymbol{y}_0 + [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x} \big\|_Y \\
&\quad \le \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1}\| o\big(\Delta\boldsymbol{x}, f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - f(\boldsymbol{x}_0)\big) \\
&\quad = \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1}\| o(1)(\|\Delta\boldsymbol{x}\|_X + \|f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - f(\boldsymbol{x}_0)\|_Y)\,.
\end{aligned}$$

However,

$$\begin{aligned}
&\|f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - f(\boldsymbol{x}_0)\|_Y \\
&\quad = \big\| f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - \boldsymbol{y}_0 + [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x} - [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x} \big\|_Y \\
&\quad \le \big\| f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - \boldsymbol{y}_0 + [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x} \big\|_Y \\
&\qquad\quad + \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)\| \|\Delta\boldsymbol{x}\|_X\,,
\end{aligned}$$

hence we have:

$$\begin{aligned}
&\big\| f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - \boldsymbol{y}_0 + [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x} \big\|_Y \\
&\quad \le \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1}\| \Big[(1 + \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)\|) \|\Delta\boldsymbol{x}\|_X \\
&\qquad\qquad + \big\| f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - f(\boldsymbol{x}_0) + [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x} \big\|_Y \Big] o(1)\,,
\end{aligned}$$

or,

$$\begin{aligned}
&\big\| f(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) - \boldsymbol{y}_0 + [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \Delta\boldsymbol{x} \big\|_Y \\
&\quad \le \frac{\|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1}\| (\|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)\| + 1)}{1 - \|[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1}\| o(1)} o(1) \|\Delta\boldsymbol{x}\|_X\,.
\end{aligned}$$

By the continuity of f at $\boldsymbol{x}_0$, as $\Delta\boldsymbol{x} \to \boldsymbol{0}$, $o(1) \to 0$ as well, hence we have proved that:

$$f'(\boldsymbol{x}_0) = -[\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \partial_{\boldsymbol{x}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)\,.$$

□

Theorem 14.4 (Continuous differentiability of implicit function)**.** *Let X, Z be normed spaces, and Y be a Banach space. $\boldsymbol{x}_0 \in X$, $\boldsymbol{y}_0 \in Y$. Denote*

$$W := B(\boldsymbol{x}_0; \alpha) \times B(\boldsymbol{y}_0; \beta)\,,$$

where $\alpha, \beta \in \mathbb{R}_+$. If $F \in Z^W$ satisfies:
a) $F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \boldsymbol{0}$;
b) $F \in C^{(1)}(W; Z)$;
c) $\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{B}(Y; Z)$ is reversible i.e. $\exists [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \in \mathcal{B}(Z; Y)$ s.t.

$$\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \circ [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} = [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \circ \partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \mathrm{id}_Y\,,$$

then, $\exists U \in \mathscr{U}(\boldsymbol{x}_0)$, $\exists V \in \mathscr{U}(\boldsymbol{y}_0)$, $\exists f \in C^{(1)}(U; Y)$ s.t. $U \times V \subset W$ and $\forall \boldsymbol{x} \in U$, $\forall \boldsymbol{y} \in V$,

$$F(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0} \;\leftrightarrow\; f(\boldsymbol{x}) = \boldsymbol{y}\,.$$

Proof. By Theorem 11.4, we know $\partial_{\boldsymbol{x}} F$ and $\partial_{\boldsymbol{y}} F$ are continuous in U, V. By Theorem 10.7, $[\partial_{\boldsymbol{y}} F]^{-1}$ is also continuous, hence $f'(bvx)$, being the composition of continuous mapping (given by Eq. 14-1), is also continuous. $\square$

Recursively we can prove:

Theorem 14.5 (n-th continuous differentiability of implicit function)**.** *Let X, Z be normed spaces, and Y be a Banach space. $\boldsymbol{x}_0 \in X$, $\boldsymbol{y}_0 \in Y$. Denote*

$$W := B(\boldsymbol{x}_0; \alpha) \times B(\boldsymbol{y}_0; \beta)\,,$$

where $\alpha, \beta \in \mathbb{R}_+$. If $F \in Z^W$ satisfies:
a) $F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \boldsymbol{0}$;
b) $F \in C^{(k)}(W; Z)$;
c) $\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{B}(Y; Z)$ is reversible i.e. $\exists [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \in \mathcal{B}(Z; Y)$ s.t.

$$\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) \circ [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} = [\partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0)]^{-1} \circ \partial_{\boldsymbol{y}} F(\boldsymbol{x}_0, \boldsymbol{y}_0) = \mathrm{id}_Y\,,$$

then, $\exists U \in \mathscr{U}(\boldsymbol{x}_0)$, $\exists V \in \mathscr{U}(\boldsymbol{y}_0)$, $\exists f \in C^{(k)}(U; Y)$ s.t. $U \times V \subset W$ and $\forall \boldsymbol{x} \in U$, $\forall \boldsymbol{y} \in V$,

$$F(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0} \;\leftrightarrow\; f(\boldsymbol{x}) = \boldsymbol{y}\,.$$

Chapter 3

Integration

§15 Lebesgue Measure

We now generalise the concepts of 'length', 'area' and 'volume', that is, we want to *measure* the subset of a normed space.

Definition 15.1 (Cuboid)**.** Let X_i, $i \in n$ be 1D normed spaces. The ***cuboids*** $I_{\boldsymbol{a},\boldsymbol{b}}$ in $X := \prod_{i\in n} X_i$, where $\boldsymbol{a}, \boldsymbol{b} \in X$, are defined as:

$$I_{\boldsymbol{a},\boldsymbol{b}} := \{\boldsymbol{x} \in X \mid x_i \in [a_i, b_i], \forall i \in n\}\,.$$

Before our definition of volume of subsets of X, we discuss on the volume of cuboids. The volume, or, the measure of the cuboids shall be like:

$$\mu(I_{\boldsymbol{a},\boldsymbol{b}}) = \prod_{i\in n} \|a_i - b_i\|_i \tag{15-1}$$

If a (countable) collection of cuboids are pairwise disjoint i.e. in which each two cuboids are disjoint, we shall expect their union has a volume:

$$\mu\left(\bigcup_{i\in\mathbb{N}} I_i\right) = \sum_{i\in\mathbb{N}} \mu(I_i)\,,$$

where the right hand side could be finite or ∞. Moreover, if they have no common interior point pairwisely, the equation still holds.

If there are a collections of cuboids $\{I_i\}_{i\in n}$ that covers the given cuboid I, we shall see:

$$\mu(I) \le \sum_{i\in n} \mu(I_k)\,.$$

We shall expect the measue of the subsets of X has the same properties. But we must limit our discussion on *some* subsets of X, and we may study the reason in real analysis later.

Definition 15.2 (σ-algebra)**.** Let $\mathscr{F} \in 2^X$ be a collection of subsets of a set X. If $\mathscr{F}$ satisfies:

1) $\varnothing \in \mathscr{F}$;
2) $\forall A \in \mathscr{F}$, $X - A \in \mathscr{F}$ (closed under complementation);
3) $\forall \langle A_i \rangle_{i\in\mathbb{N}} \in \mathscr{F}^{\mathbb{N}}$, $\bigcup_{i\in\mathbb{N}} A_i \in \mathscr{F}$ (closed under counterable unions),

then $\mathscr{F}$ is said to be a ***σ-algebra***.

As an example, the σ-algebra closure of cuboids (The intersections of all σ-algeras containing all cuboids) is called the ***Borel sets***.

Definition 15.3 (Measure)**.** Let $\mathscr{F}$ be a σ-algebra over X, $\mu \in (\{0\} \cup \mathbb{R}_+)^{\mathscr{F}}$. If the function μ satisfies:

1) $\mu(\varnothing) = 0$;
2) (Countable additivity) If $\langle A_i \rangle_{i\in\mathbb{N}} \in \mathscr{F}^{\mathbb{N}}$ are pair wise disjoint, then

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i)\,,$$

then μ is called a ***measure function***.

The pair $(X, \mathscr{F}, \mu)$ is called a ***measurable space***, and the sets in $\mathscr{F}$ are called ***measurable sets***. The image of a set in $\mathscr{F}$ under μ is called the measure of the set.

We shall study one of the most import measures: Lebesgue measure.

Definition 15.4 (Lebesgue outer measure)**.** The ***Lebsgue outer measure*** λ^* is a function from 2^X to $[0, \infty] \subset \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, and is defined as:

$$\lambda^*(A) := \inf\left\{\sum_{i\in\mathbb{N}} \mu(I_i) \,\middle|\, A \subseteq \bigcup_{i\in\mathbb{N}} I_i\right\}\,,$$

where the volume of the cuboids μ is defined as Eq. (15-1).

Theorem 15.1 (Monotone of Lebesgue outer measure)**.** *If $A \subseteq B$, then $\lambda^*(A) \le \lambda^*(B)$.*

Proof. If $\{I_i\}_{i\in\mathbb{N}}$ covers B, then they must cover A. □

Theorem 15.2 (Countable subadditivity of Lebesgue outer measure)**.** $\forall \langle A_k \rangle_{k\in\mathbb{N}} \in (2^X)^{\mathbb{N}}$,

$$\lambda^*\left(\bigcup_{k\in\mathbb{N}} A_k\right) \le \sum_{k\in\mathbb{N}} \lambda^*(A_k)\,.$$

Proof. $\forall \varepsilon \in \mathbb{R}_+$, by the definition of infimum, for each $k \in \mathbb{N}$, find a sequence of cuboids $\langle I_i^{(k)} \rangle_{i\in n}$ that covers A_k, and:

$$\lambda^*(A_k) + \frac{\varepsilon}{2^k} > \sum_{i\in\mathbb{N}} \mu(I_i^{(k)})\,.$$

Summing the equation over k, we have:

$$\sum_{k\in\mathbb{N}} \sum_{i\in\mathbb{N}} \mu(I_i^{(k)}) \le \sum_{k\in\mathbb{N}} \lambda^*(A_k) + \varepsilon\,.$$

As $\langle I_i^{(k)} \rangle_{i,k\in\mathbb{N}}$ covers $\bigcup_{k\in\mathbb{N}} A_k$, we have:

$$\sum_{k\in\mathbb{N}} \lambda^*(A_k) + \varepsilon \geq \sum_{k\in\mathbb{N}} \sum_{i\in\mathbb{N}} \mu(I_i^{(k)}) \geq \lambda^* \left(\bigcup_{k\in\mathbb{N}} A_k \right) .$$

The inequality holds for any positive real number ε, hence:

$$\sum_{k\in\mathbb{N}} \lambda^*(A_k) \geq \lambda^* \left(\bigcup_{k\in\mathbb{N}} A_k \right) .$$

□

Definition 15.5 (Carathéodory criterion)**.** If $E \in 2^X$ satisfies that $\forall A \in 2^X$:

$$\lambda^*(A) = \lambda^*(A \cap E) + \lambda^*(A - E) ,$$

then we say that E is ***Lebesgue measurable***, and we can say $\lambda(E) := \lambda^*(E)$.

We shall denote the collection of Lebesgue measurable sets in X by $\mathscr{F}$.

Theorem 15.3 (Lebesgue measurable sets is closed under finite unions)**.** *Let E_1, E_2 be two Lebesgue measuable sets. $E_1 \cup E_2 \in \mathscr{F}$.*

Proof. $\forall A \in 2^X$,

$$\begin{aligned}\lambda^*(A) &= \lambda^*(A \cap E_1) + \lambda^*(A - E_1) \\ &= \lambda^*(A \cap E_1 \cap E_2) + \lambda^*(A \cap E_1 - E_2) + \lambda^*((A - E_1) \cap E_2) + \lambda^*(A - E_1 - E_2) .\end{aligned}$$

It is easy to verify that:

$$(A \cap E_1 \cap E_2) \cup ((A \cap E_1) - E_2) \cup ((A - E_1) \cap E_2) = A \cap (E_1 \cup E_2) ,$$

therefore by Theorem 15.2:

$$\lambda^*(A) \geq \lambda^*(A \cap (E_1 \cup E_2)) + \lambda^*(A - (E_1 \cup E_2)) .$$

But $(A \cap (E_1 \cup E_2)) \cup (A - (E_1 \cup E_2)) = A$, again by Theorem 15.2, the reverse of the inequality holds. □

Theorem 15.4 (Finite additivity of Lebesgue measure)**.** *Let $\langle E_i \rangle_{i\in n} \in \mathscr{F}^n$ be pairwise disjoint. $\forall A \in 2^X$,*

$$\lambda^* \left(A \cap \bigcup_{i\in n} E_i \right) = \sum_{i\in n} \lambda^*(A \cap E_i) .$$

Proof. We might prove this inductively.

As $n = 1$, the proposition is trivial. Assume that for $n \in \mathbb{N}_+$ the proposition holds:

$$\begin{aligned}\lambda^* \left(A \cap \bigcup_{i\in n+1} E_i \right) &= \lambda^* \left(A \cap \bigcup_{i\in n+1} E_i \cap E_n \right) + \lambda^* \left(A \cap \bigcup_{i\in n+1} E_i - E_n \right) \\ &= \lambda^*(A \cap E_n) + \lambda^* \left(A \cap \bigcup_{i\in n} E_i \right) = \sum_{i\in n+1} \lambda^*(A \cap E_i) .\end{aligned}$$

□

Theorem 15.5 (Lebesgue measurable sets are σ-algebra)**.** *$\mathscr{F}$ is a σ-algebra.*

Proof. It is obvious that $\varnothing \in \mathscr{F}$. Since $A \cap E = A - (X - E)$, $A - E = A \cap (X - E)$, the complement of a Lebesgue measurable set is also Lebesgue measurable.

For any (convergent) sequence of sets $\langle E_i \rangle_{i \in \mathbb{N}}$, a pairwise disjoint sequence can be consturcted:

$$F_i = E_i - \bigcup_{j \in i} E_j \, ,$$

so that $\bigcup_{i \in n} F_i = \bigcup_{i \in n} E_i$.

By the monotone of λ^* (Theorem 15.1):

$$\lambda^* \left(A - \bigcup_{i \in n} F_i \right) \geq \lambda^* \left(A - \bigcup_{i \in \mathbb{N}} F_i \right) \, .$$

Since $\mathscr{F}$ is closed under finite unions(Theorem 15.3), $\bigcup_{i \in n} F_i$ is also Lebesgue measurable. Also, λ^* is countably subadditive:

$$\begin{aligned} \lambda^*(A) &= \lambda^* \left(A - \bigcup_{i \in n} F_i \right) + \lambda^* \left(A \cap \bigcup_{i \in n} F_i \right) \\ &\geq \lambda^* \left(A - \bigcup_{i \in \mathbb{N}} F_i \right) + \sum_{i \in n} \lambda^*(A \cap F_i) \, . \end{aligned}$$

Pass n to the infinity, the inequality becomes:

$$\lambda^*(A) \geq \lambda^* \left(A - \bigcup_{i \in \mathbb{N}} F_i \right) + \sum_{i \in \mathbb{N}} \lambda^*(A \cap F_i) \geq \lambda^* \left(A - \bigcup_{i \in \mathbb{N}} F_i \right) + \lambda^* \left(A \cap \bigcup_{i \in \mathbb{N}} F_i \right) \, .$$

The validity of the second '$\leq$' is again by the countable subadditivity of λ^*. □

Theorem 15.6 (Countable additivity of Lebesgue measure)**.** *If $\langle E_i \rangle_{i \in \mathbb{N}} \in \mathscr{F}^{\mathbb{N}}$ is pairwise disjoint, then*

$$\lambda \left(\bigcup_{i \in \mathbb{N}} A_i \right) = \sum_{i \in \mathbb{N}} \lambda(A_i) \, .$$

Proof. By Theorem 15.4, we have:

$$\lambda^* \left(\bigcup_{i \in \mathbb{N}} E_i \right) \geq \lambda^* \left(\bigcup_{i \in n} E_i \right) = \sum_{i \in n} \lambda^*(E_i) \, .$$

Passing $n \to \infty$,

$$\lambda^* \left(\bigcup_{i \in \mathbb{N}} E_i \right) \geq \sum_{i \in \mathbb{N}} \lambda^*(E_i) \, ,$$

while the subadditive of λ^* yields the reverse. □

Definition 15.6 (Measure zero)**.** If a set $E \in 2^X$ is Lebesgue measurable ($E \in \mathscr{F}$), and its measure is 0, we say it is a set of (Lebesgue) ***measure zero***. In another word, E has measure zero, meaning, $\forall \varepsilon \in \mathbb{R}_+$, $\exists \langle I_i \rangle_{i\in\mathbb{N}}$ s.t. $E \subset \bigcup_{i\in\mathbb{N}} I_i$ and $\sum_{i\in\mathbb{N}} \mu(I_i) < \varepsilon$. Here I_i are cuboids.

We can easily conclude that the sets containing only a point is of measure zero.

Theorem 15.7 (The countable union of sets of measure zero is also of measure zero)**.** *Let E_k, $k \in \mathbb{N}$ be sets of measure zero, then $\bigcup_{k\in\mathbb{N}} E_k = 0$.*

Proof. For each set E_k, find it a cover with cuboids that the sum of the measure of the cuboids are less than $\varepsilon/2^k$. □

Theorem 15.8 (The subset of a set of measure zero is also of measure zero)**.** *Let E be of measure zero, $F \subset E$. F is of measure zero.*

Lemma 7. *E has measure zero iff $\forall \varepsilon \in \mathbb{R}_+$, $\exists \langle I'_i \rangle_{i\in\mathbb{N}}$ s.t. $E \subset \bigcup_{i\in\mathbb{N}} I'_i$ and $\sum_{i\in\mathbb{N}} \mu(I'_i) < \varepsilon$. Here I'_i are open cuboids, defined by products of n open intervals.*

Proof. For any ε we multiply it by λ^n where $\lambda < 1$, the definition yield that we can find a sequence of cuboids, the sum of the measure of which is less than $\lambda^n\varepsilon$. We extend the cuboids by λ^{-1}, we see that the interior point of which contain the previous cuboids. □

Theorem 15.9. *A compact set K has measure zero iff $\forall \varepsilon \in \mathbb{R}_+$, $\exists \langle I_i \rangle_{i\in\mathbb{N}}$ s.t. $E \subset \bigcup_{i\in\mathbb{N}} I_i$ and $\sum_{i\in\mathbb{N}} \mu(I_i) < \varepsilon$.*

Proof. By Lemma 7, we can find an open cover with measure less than ε of E, and therefore there is a finite subcover. The measure of the subcover is of course less than that of the cover. □

§16 Riemann Integral on n-D cuboids

Now we introduce the partition of the cuboid:

Definition 16.1 (Partition of a cuboid)**.** A ***partition*** P of a cuboid $I_{\boldsymbol{a},\boldsymbol{b}}$, is defined as a *finite* collection of cuboids which have no common interior point pairwisely, and the union of which is the cuboid itself $I_{\boldsymbol{a},\boldsymbol{b}}$.

Definition 16.2 (Mesh)**.** The ***mash*** of a partition P is the maximum diametre of the cuboids in P:

$$\lambda(P) := \max\{d(I') \mid I' \in P\}.$$

Definition 16.3 (Distinguished points)**.** The image of a choose function from P to I is the distinguished points of P, denoted by $\boldsymbol{\xi}_j \in I_j$, $I_j \in P$, $j \in \operatorname{card} P$. $\boldsymbol{\xi} := (\boldsymbol{\xi}_j)_{j\in\operatorname{card} P}$.

All partitions of a cuboid I is denoted by $\mathfrak{P}(I)$. Now define a filter base $\lambda(P) \to 0$, the elements of which are $B_\delta := \{(P, \boldsymbol{\xi}) \in \mathfrak{P}(I) \mid \lambda(P) < \delta\}$, $\delta \in \mathbb{R}_+$.

Definition 16.4 (Riemann sum)**.** Let $X := \prod_{i\in n} X_i$, Y be normed spaces, where X_i are 1-D spaces. Let I be a cuboid in X, $f \in Y^I$, $(P, \boldsymbol{\xi}) \in \mathfrak{P}(I)$. $N := \operatorname{card} P$, $P := \{I_j \mid j \in N\}$. The ***Riemann sum*** of f over P with distinguished points $\boldsymbol{\xi}$ is defined as:

$$\sigma(f, P, \boldsymbol{\xi}) := \sum_{j\in N} f(\boldsymbol{\xi}_j)\mu(I_j).$$

Definition 16.5. Riemann integral If the following limit exists, we define:

$$\int_I f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} := \lim_{\lambda(P)\to 0} \sigma(f, P, \boldsymbol{\xi})$$

as the ***Riemann integral*** of f on I.

Definition 16.6 (Riemann integrable)**.** If the integral of f in I exists, we call f Riemann integrable. The Riemann integrable functions on I is denoted by $\mathfrak{R}(I)$.

Theorem 16.1 (Riemann integrable then bounded)**.** $f \in \mathfrak{R}(I) \ \to\ \exists M \in \mathbb{R}_+ \forall \boldsymbol{x} \in I\,(\|f(\boldsymbol{x})\|_Y < M)$.

Proof. If f is not bounded, $\forall M \in \mathbb{R}_+$ there always exists a $\boldsymbol{x}_j \in I_j \in P$, $\forall \delta \in \mathbb{R}_+$, even if $\lambda(P) < \delta$, $\sigma(f, P, \boldsymbol{\xi}) \geq \|f(\boldsymbol{x}_j)\|\mu(I_j) > M$. □

We say a proposition $p(\boldsymbol{x})$ holds ***almost everywhere*** or ***a.e.*** on X, meaning $\exists E \subset X$, s.t. E is of measure zero and $\forall \boldsymbol{x} \in (X - E)\big(p(\boldsymbol{x})\big)$.

Theorem 16.2 (Lebesgue's criterion)**.** *Let $f \in \mathbb{R}^I$, where I is a cuboid in a n-D space. $f \in \mathfrak{R}(I)$ $\leftrightarrow$ f is bounded in I and f is almost everywhere continuous on I.*

Proof. $\to$: $f \in \mathfrak{R}(I) \to f$ is bounded on I (Theorem 16.1). Denote the discontinuous points of f on I by E. In another word, $E = \{\boldsymbol{x} \in I \mid \omega(f; \boldsymbol{x}) > 0\}$.

Now consider a sequence of sets $E_k := \{x \in I \mid \omega(f; \boldsymbol{x}) \geq 1/k\}$, which is monotone, and limits at E: $E = \bigcup_{k\in\mathbb{N}_+} E_k$.

If E is not of measure zero, since it is a union of a countable sequence, $\exists k_0 \in \mathbb{N}_+$, E_{k_0} is not of measure zero.

Assume that there were a partition $P = \{I_j \mid j \in N\}$ of I. Let:

$$A = \{I_j \in P \mid I_j \cap E_{k_0} \neq \varnothing \wedge \omega(f; I_i) \geq 1/2k_0\},$$

and $B = P - A$.

Now we prove: $E_{k_0} \subset \cup A$. If a point $\boldsymbol{x}$ of E_{k_0} locates as an interior point in I_j, then there exists a neighbourhood of $\boldsymbol{x}$, where the oscillation of f is larger than $1/k_0 - 1/2k_0 = 1/2k_0$ [1].

Else, if $\boldsymbol{x}$ locates as a boundary point of cuboids in P, we denote these cuboids by $C(\boldsymbol{x}) := \{I_j \in P \mid \boldsymbol{x} \in I_j\}$. If (assuming) $\forall I_j \in C(\boldsymbol{x})$, $\omega(f; I_j) < 1/2k_0$ (that is, $C(\boldsymbol{x}) \cap A \neq \varnothing$). $\forall \varepsilon \in \mathbb{R}_+$, $\exists \delta \in \mathbb{R}_+$ s.t. $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in B(\boldsymbol{x}; \delta) \subset \cup C(\boldsymbol{x})$,

$$d(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2)) \leq d(f(\boldsymbol{x}), f(\boldsymbol{x}_1)) + d(f(\boldsymbol{x}), f(\boldsymbol{x}_2)) < \frac{1}{k_0} - \varepsilon.$$

Passing $\delta \to 0$, we have: $\forall \varepsilon$, $\omega(f; \boldsymbol{x}) \leq 1/k_0 - \varepsilon$, or $\omega(f; \boldsymbol{x}) < 1/k_0$, which contradicts with the fact that $\boldsymbol{x} \in E_{k_0}$. Hence: there must be a $I_j \in C(\boldsymbol{x})$, $\omega(f; I_j) \geq 1/2k_0$, therefore such $I_j \in A$.

In conclusion, we have proved that A covers E_{k_0}.

[1] We take the $\varepsilon = 1/2k_0$ in the definition of oscillation at a point (as a limit)

Since E_{k_0}, by our assumption, is not of measure zero, then $\exists \varepsilon_0 \in \mathbb{R}_+$, $\sum_{I_j \in A} \mu(I_j) > \varepsilon_0$. Take two sets of distinguished points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$, when they belong to $I_j \in A$, we let $d(f(\boldsymbol{\xi}_j), f(\boldsymbol{\xi}'_j)) > 1/3k_0$ [2], and when $I_j \in B$, $\boldsymbol{\xi}_j = \boldsymbol{\xi}'_j$.

$$d\big(\sigma(f, P, \boldsymbol{\xi}), \sigma(f, P, \boldsymbol{\xi}')\big) = \sum_{I_j \in A} \mu(I_j) d(f(\boldsymbol{\xi}_j), f(\boldsymbol{\xi}'_j)) > \frac{\varepsilon_0}{3k_0}.$$

By Cauchy's criterion, $\sigma(f, P, \boldsymbol{\xi})$ would have no limit.

$\leftarrow$: Let $\varepsilon \in \mathbb{R}_+$ and $E_\varepsilon = \{\boldsymbol{x} \in I \mid \omega(f; \boldsymbol{x}) \geq \varepsilon\}$. Since f is a.e. continuous on I, $\mu(E_\varepsilon) = 0$.

Now we prove that E_ε is closed. If $\boldsymbol{x} \notin E_\varepsilon$ i.e. $\omega(f; \boldsymbol{x}) \leq \varepsilon'$ where $\varepsilon' < \varepsilon$. By the definition of the oscillation at a point, $\forall \varepsilon'' \in \mathbb{R}_+$, there exists a ball $B(\boldsymbol{x}; \delta)$ on which $\omega(f, B(\boldsymbol{x}; \delta)) < \varepsilon' + \varepsilon''$. Let $\varepsilon'' = \varepsilon - \varepsilon'$, and notice that $\omega(f; \boldsymbol{x}') \leq \omega(f, B(\boldsymbol{x}, \delta))$ where $\boldsymbol{x}' \in B(\boldsymbol{x}, \delta)$. Therefore, $B(\boldsymbol{x}; \delta) \subset I - E_\varepsilon$. Hence: E_ε is closed.

Since E_ε is closed in a compact set I [3], we know by Theorem 3.4 that E_ε is also compact. By Theorem 15.9 we can find a finite cover $C_1 = \{I_j \mid j \in k\}$ of E_ε with $\sum_{j \in k} \mu(I_j) < \varepsilon$. Now we extend these cuboids by $\alpha > 1$, $\beta > \alpha$ to get $C_2 = \{\alpha I_j \mid j \in k\}$ and $C_3 = \{\beta I_j \mid j \in k\}$.

Let $\delta = d\big(\cup C_2, \partial(\cup C_3)\big)$. Since any point in $\cup C_2$ is an interior point of one of the βI_j, we claim: $\delta > 0$.

Let $K = I - (\cup C_2 - \partial(\cup C_2))$. Obviously K is also compact, and $E_\varepsilon \subset I - K$. $\forall \boldsymbol{x} \in K$, since $\boldsymbol{x} \notin E_\varepsilon$, $\omega(f; \boldsymbol{x}) < \varepsilon$.

By Theorem 6.11, $\exists \delta' \in \mathbb{R}_+$, if $\boldsymbol{x}', \boldsymbol{x}'' \in K$ satisfies that $d(\boldsymbol{x}', \boldsymbol{x}'') < \delta'$, $d(f(\boldsymbol{x}'), f(\boldsymbol{x}'')) < 2\varepsilon$. Let $\delta'' = \min\{\delta, \delta'\}$.

Assume that there were two partitions $P, P' \in \mathfrak{P}(I)$ s.t. $\lambda(P) < \delta''$, $\lambda(P') < \delta''$. Let $P'' := \{I''_{jj'} := I_j \cap I'_{j'} \mid I_j \in P \wedge I'_{j'} \in P'\}$.

$$\begin{aligned} d\big(\sigma(f, P, \boldsymbol{\xi}), \sigma(f, P'', \boldsymbol{\xi}'')\big) &= d\left(\sum_{j \in N} \sum_{j' \in N'} f(\boldsymbol{\xi}_j)\mu(I''_{jj'}), \sum_{j' \in N'} \sum_{j \in N} f(\boldsymbol{\xi}''_{jj'})\mu(I''_{jj'})\right) \\ &\leq \sum_{j \in N} \sum_{j' \in N'} d\big(f(\boldsymbol{\xi}_j), f(\boldsymbol{\xi}''_{jj'})\big)\mu(I''_{jj'}). \end{aligned}$$

Now we divide P'' into two parts: $A := \{I''_{jj'} \in P'' \mid I_j \subset \cup C_3\}$, $B = P'' - A$. We shall see $\cup B \subset K$: if there were a cuboid I_j in P s.t. $I_j \cap (I - \cup C_3) \neq \varnothing$, since $\lambda(I_j) < \delta'' \leq \delta$, there is no way that $I_j \cap \cup C_2 \neq \varnothing$.

We assume the function f to be bounded, let $2M \geq \sup\{d(f(\boldsymbol{x}), f(\boldsymbol{x}')) \mid \boldsymbol{x}, \boldsymbol{x}' \in I\}$.

Therefore:

$$\begin{aligned} d\big(\sigma(f, P, \boldsymbol{\xi}), \sigma(f, P'', \boldsymbol{\xi}'')\big) &\leq \sum_{I''_{jj'} \in A} d\big(f(\boldsymbol{\xi}_j), f(\boldsymbol{\xi}''_{jj'})\big)\mu(I''_{jj'}) + \sum_{I''_{jj'} \in B} d\big(f(\boldsymbol{\xi}_j), f(\boldsymbol{\xi}''_{jj'})\big)\mu(I''_{jj'}) \\ &< 2M \sum_{I''_{jj'} \in A} \mu(I''_{jj'}) + \varepsilon \sum_{I''_{jj'} \in B} \mu(I''_{jj'}) \\ &\leq 2M \cdot \beta^n \varepsilon + \varepsilon\mu(I) = (2M \cdot \beta^n + \mu(I))\varepsilon. \end{aligned}$$

[2] which is possible, because $\omega(f; I_j) \geq 1/2k_0$.

[3] Lemma 4

Similarly we have $d\big(\sigma(f,P',\boldsymbol{\xi}'),\sigma(f,P'',\boldsymbol{\xi}'')\big) < (2M\cdot\beta^n+\mu(I))\varepsilon$, then by triangle inequality:

$$d\big(\sigma(f,P,\boldsymbol{\xi}),\sigma(f,P',\boldsymbol{\xi}')\big) < 2(2M\cdot\beta^n+\mu(I))\varepsilon.$$

Therefore, $f\in\mathfrak{R}(I)$. □

Definition 16.7 (Darboux sum)**.** Let $f\in\mathbb{R}^I$, where I is a cuboid in a n-D space. $P=\{I_j\mid j\in N\}\in\mathfrak{P}(I)$, the ***Darboux lower sum*** and the ***Darboux upper sum*** is defined as:

$$s(f,P)=\sum_{I_j\in P}\mu(I_j)\inf\{f(\boldsymbol{x})\mid \boldsymbol{x}\in I_j\},\quad S(f,P)=\sum_{I_j\in P}\mu(I_j)\sup\{f(\boldsymbol{x})\mid \boldsymbol{x}\in I_j\}.$$

Lemma 8. *$\forall P,P'\in\mathfrak{P}(I)$, $s(f,P)\le S(f,P')$.*

Proof. Let $P'':=\{I_j\cap I'_j\mid I_{j'}\in P\wedge I'_{j'}\in P'\}$, we have:

$$s(f,P)\le s(f,P'')\le S(f,P'')\le S(f,P').$$

□

Definition 16.8 (Darboux integrals)**.** Let $f\in\mathbb{R}^I$, where I is a cuboid in a n-D space. The ***lower Darboux integral*** and the ***upper Darboux integral*** are defined as:

$$\underline{\mathfrak{I}}:=\sup\{s(f,P)\mid P\in\mathfrak{P}\},\quad \overline{\mathfrak{I}}:=\inf\{S(f,P)\mid P\in\mathfrak{P}\}.$$

Theorem 16.3 (Darboux theorem)**.** *Let $f\in\mathbb{R}^I$, where I is a cuboid in a n-D space. If f is bounded on I, then the limits of Darboux sums exist (as $\lambda(P)\to 0$):*

$$\underline{\mathfrak{I}}=\lim_{\lambda(P)\to 0}s(f,P),\quad \overline{\mathfrak{I}}=\lim_{\lambda(P)\to 0}S(f,P).$$

Proof. We will only prove the lower Darboux theorem.

$\forall\varepsilon\in\mathbb{R}_+$, $\exists P_\varepsilon\in\mathfrak{P}(I)$ s.t. $s(f,P_\varepsilon)>\underline{\mathfrak{I}}-\varepsilon$. Let $\mathit{\Gamma}_\varepsilon:=\bigcup_{I_j\in P_\varepsilon}\partial I_j$. Obviously, $\lambda(\mathit{\Gamma}_\varepsilon)=0$.

We claim that: $\exists\delta\in\mathbb{R}_+$ s.t. $\forall P\in\mathfrak{P}$, if $\lambda(P)<\delta$, then

$$\sum_{\substack{I_j\in P;\\ I_j\cap\mathit{\Gamma}_\varepsilon\ne\varnothing}}\mu(I_j)<\varepsilon.$$

This can be proved by assuming the opposite, then there exists a lower bound (that is non-zero) for the sum of the measure of the cuboids that covers $\mathit{\Gamma}_\varepsilon$, which contradicts with the fact that $\lambda(\mathit{\Gamma}_\varepsilon)=0$.

Now let $P':=\{I_j\cap J_{j'}\mid I_j\in P_\varepsilon\wedge J_{j'}\in P\}$, we can see:

$$\underline{\mathfrak{I}}-\varepsilon<s(f,P_\varepsilon)\le s(f,P')\le\underline{\mathfrak{I}}.$$

$$
\begin{aligned}
&|s(f,P') - s(f,P)| \\
&= \left| \sum_{\substack{J_{j'} \in P, \\ J_{j'} \cap \Gamma_\varepsilon \neq \varnothing}} \left(\sum_{I_j \in P_\varepsilon} \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in J_{j'} \cap I_j\} \mu(J_{j'} \cap I_j) - \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in J_{j'}\} \mu(J_{j'}) \right) \right| \\
&\leq M \left| \sum_{\substack{J_{j'} \in P, \\ J_{j'} \cap \Gamma_\varepsilon \neq \varnothing}} \left(\sum_{I_j \in P_\varepsilon} \mu(J_{j'} \cap I_j) + \mu(J_{j'}) \right) \right| = 2M \sum_{\substack{J_{j'} \in P, \\ J_{j'} \cap \Gamma_\varepsilon \neq \varnothing}} \mu(J_{j'}) < 2M\varepsilon.
\end{aligned}
$$

Hence: $s(f,P') > s(f,P) - 2M\varepsilon > \underline{\mathfrak{I}} - \varepsilon$, or, $\underline{\mathfrak{I}} \geq s(f,P) > (2M+1)\underline{\mathfrak{I}} - \varepsilon$. Therefore:

$$\lim_{\lambda(P) \to 0} s(f,P) = \underline{\mathfrak{I}}.$$

□

Theorem 16.4 (Darboux criterion)**.** *Let $f \in \mathbb{R}^I$, where I is a cuboid in a n-D space. $f \in \mathfrak{R}(I)$ $\leftrightarrow$ f is bounded on I, and $\underline{\mathfrak{I}} = \overline{\mathfrak{I}}$.*

Proof. $\rightarrow$: If $f \in \mathfrak{R}(I)$, f is bounded (Theorem 16.1), then both upper integral and lower integral exists. As $\lambda(P) \to 0$, the infimum and supremum of Riemann sums must converge to the Riemann integral itself.

$\leftarrow$: We only need to notice that $s(f,P) \leq \sigma(f,P,\boldsymbol{\xi}) \leq S(f,P)$. □

§17 Riemann Integral on Jordan Measurable sets

Part II

Real Analysis

Part III

Functional Analysis

Part IV

Complex Analysis

Bibliography

[1] R. Coleman. *Calculus on Normed Vector Spaces*. Universitext. Springer New York, 2012. ISBN: 9781461438946. URL: https://books.google.co.jp/books?id=TCzpZuCpc%5C_QC.

[2] Vladimir A Zorich. *Mathematical analysis II; 2nd ed.* Universitext. Berlin: Springer, 2016. DOI: 10.1007/978-3-662-48993-2. URL: https://cds.cern.ch/record/2137923.

Symbol List

Here listed the important symbols used in this notes.

Index