

Analysis

Hoyan Mok¹

March 4, 2021

¹E-mail: victoriesmo@hotmail.com

B

Preface

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

Contents

Preface	i
Contents	ii
 I Mathematical Analysis	 1
1 Metric Space and Continuous Mapping	2
§1 Metric Space	2
§2 Topological Space	4
§3 Compact Set	5
§4 Connected Set	8
§5 Complete Metric Spaces	9
§6 Continuous Mapping	11
§7 Contraction	14
 2 Normed Linear Space and Differential Calculus	 16
§8 Normed Linear Space	16
§9 Linear Operators	17
§10 Differentiation	20
§11 Finite-Increment Theorem	25
§12 Higher-Order Derivative	29
§13 Applications of Differentiation	31
13.1 Taylor's Formula	31
13.2 Interior Extrema	33
§14 Implicit Function Theorem	35
 3 Integration	 39
§15 Lebesgue Measure	39
§16 Riemann Integral over n -D cuboids	43
§17 Riemann Integral over Jordan Measurable Sets	47
§18 Properties of Riemann Integrals	49
§19 Fubini's Theorem	52
§20 Change of Variables	54
§21 Improper Integral	58

<i>CONTENTS</i>	iii
4 Surfaces and differential forms in \mathbb{R}^n	63
§22 Surfaces	63
§23 Orientation of a Surface	65
II Real Analysis	66
III Functional Analysis	67
IV Complex Analysis	68
V Appendices	69
A Several Important Inequality	70
§1 Mean-value Inequality	70
§2 Cauchy's Inequality	70
§3 Jensen's Inequality	70
§4 Brunn-Minkowski Inequality	70
§5 Hölder's Inequality and Minkowski's Inequality	70
Bibliography	71
Symbol List	72
Index	73

Part I

Mathematical Analysis

Chapter 1

Metric Space and Continuous Mapping

§1 Metric Space

Definition 1.1 (Metric). A function

$$d: X^2 \rightarrow \mathbb{R}$$

$\forall x, y, z \in X$ 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 equipped 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 $\mathbb{R}_p^n := (\mathbb{R}^n, d_p)$, where

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

while

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

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

$$d_p(f, g) = \left(\int_a^b |f - g|^p dx \right)^{1/p}. \quad (1-3)$$

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 equivalent if they are equal up to a null set.

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

$$\forall a, b, u, v \in X, |d(a, b) - d(u, v)| \leq d(a, u) + d(b, v) \quad (1-4)$$

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. \square

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 **diameter** 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. $B(x, \delta) \subset G$.

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 $\forall \alpha \in A, G_\alpha$ 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 $\mathcal{C}_X(\bigcap_{\alpha \in A} F_\alpha) = \bigcup_{\alpha \in A} \mathcal{C}_X(F_\alpha)$ and a).

d) Similarly, $\mathcal{C}_X(F_1 \cup F_2) = \mathcal{C}_X(F_1) \cap \mathcal{C}_X(F_2)$.

\square

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 $\mathcal{U}(x)$.

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 $\forall U(a)$, $\text{card}(E \cap U(a)) = \infty$, 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. \rightarrow : $\mathcal{C}_X(F)$ is open, hence its elements are all its interior points. Therefore $\overline{F} - F = \overline{F} \cup \mathcal{C}_X(F) = \emptyset$, also we know that $F \subset \overline{F}$, hence $F = \overline{F}$.

\leftarrow : $F = \overline{F}$ means that $\forall x \in \mathcal{C}_X(F)$, 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. \square

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. $U(x) \cap \overline{E} = \emptyset$, which implies that x is an exterior point of E , therefore \overline{E} is closed. \square

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. \square

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 equipped with a **topology** if we assigned a $\mathcal{T} \subset 2^X$, with the following properties:

- a) $\emptyset \in \mathcal{T}$; $X \in \mathcal{T}$.
- b) $(\forall \alpha \in A, G_\alpha \in \mathcal{T}) \rightarrow \bigcup_{\alpha \in A} G_\alpha \in \mathcal{T}$.
- c) $\forall G_1, G_2 \in \mathcal{T}$, $G_1 \cap G_2 \in \mathcal{T}$.

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

These conditions are the intrinsic properties of the open sets we have defined in the metric space¹. 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 \mathcal{T} 's elements are called **open sets**, and their complements are called **closed sets**.

¹See proposition 1

Definition 2.3 (Base). Let (X, \mathcal{T}) be a topological space, and $\mathfrak{B} \subset 2^X$. If $\forall G \in \mathcal{T}, \exists \{B_\alpha\}_{\alpha \in A} \in 2^{\mathfrak{B}}$ s.t. $\bigcup_{\alpha \in A} B_\alpha = G$, we called \mathfrak{B} a (topological or open) **base** of the topology \mathcal{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 $U(x) \in \mathcal{T}$, then $U(x)$ is a **neighbourhood** of x in topological space (X, \mathcal{T}) . All neighbourhoods of a point x is denoted by $\mathcal{U}(x)$.

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

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 $a \in \mathbb{R}$:

$$f \sim_a g \leftrightarrow \exists U(a)(\forall x \in U(a), f(x) = g(x)). \quad (2-1)$$

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 $(C(\mathbb{R}; \mathbb{R}); \mathcal{T})$, where \mathcal{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, \mathcal{T}) a **Hausdorff space**, **separated space** or T_2 **space**, if $\forall x, y \in X, x \neq y \rightarrow (\exists U(x), U(y) \text{ s.t. } U(x) \cap U(y) = \emptyset)$ ².

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

Definition 2.8 (Separable). If there is a *countable* dense set in topological space (X, \mathcal{T}) , then (X, \mathcal{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 \mathcal{T} can be given a **subspace topology** \mathcal{T}_Y whose elements G_Y are intersections of the subset with an open set G in (X, \mathcal{T}) i.e. $\forall G_Y \in \mathcal{T}_Y, \exists G \in \mathcal{T} \text{ s.t. } G_Y = G \cap Y$. Subsets equiped with such topology construct a **topological subspace** (Y, \mathcal{T}_Y) .

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

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

§3 Compact Set

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

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

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

Specially, \emptyset is compact.

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

This theorem tells a truth that whether K is compact or not doesn't depend 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, \mathcal{T}_K) is an intersection of an open set G in (X, \mathcal{T}) and K .

Theorem 3.2 (Compact \rightarrow closed (Hausdorff)). If K is compact in a Hausdorff space (X, \mathcal{T}) ³, then K is a closed set in (X, \mathcal{T}) .

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

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

Assume that $x_0 \notin K$. In a Hausdorff space, $\forall x \in K - \{x_0\}$, $\exists U(x)$ s.t. $U(x) \cap U(x_0) = \emptyset$. 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. $\text{card } \Omega' \in \mathbb{N}$.

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

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. $\forall (K_n)_{n \in \mathbb{N}} \in \mathcal{P}(X)^{\mathbb{N}}$ s.t. $\forall n \in \mathbb{N}_+$, $K_n \supset K_{n+1} \wedge K_n \neq \emptyset \wedge (K_n \text{ is compact})$: $K_n \downarrow K \neq \emptyset$.

Proof. Assume that $K = \emptyset$. Compact subsets of K_1 are all closed, while their complements are all open. An open cover Ω can be constructed as $\{K_1 - K_n | n \in \mathbb{N}_+\}$. 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 $K_{n_0} = \emptyset$ 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, $\Omega = (\cup \Omega_F) \cup \{K - F\}$ constructs an open cover over K . Since K is compact there must be a finite cover $\Omega' \subset \Omega$ which obviously also covers over F . □

The following properties 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) .

³See 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 $\Omega' = \{B(x_i, \varepsilon)\}_{i \in n}$ ($n \in \mathbb{N}_+$). 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 $\{B(x_n; \varepsilon_0)\}_{n \in \mathbb{N}}$ 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 contradictory. \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, $\exists U(a) \in \mathcal{U}(a)$ s.t. $U(a) \cap (\bigcap_{n \in \mathbb{N}} F_n) = \emptyset$, therefore $U(a) \cap (\bigcap_{n \in \mathbb{N}} F_{k_n}) = \emptyset$. But this conflict the fact that $\exists N \in \mathbb{N}$, s.t. $n > N \rightarrow x_{k_n} \in U(a)$ while $x_{k_n} \in F_{k_n}$. \square

Then we get back to the Theorem 3.6.

Proof. \rightarrow : If $\text{card}\{x_n\}_{n \in \mathbb{N}} \in \mathbb{N}$, it is obvious; Now we let $\text{card}\{x_n\}_{n \in \mathbb{N}} \notin \mathbb{N}$. 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 simplification, 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) $\lim_{n \rightarrow \infty} x_{k_n} \in K$.

\leftarrow : Assume that there were an open cover Ω 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 Ω , which means Ω 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 -dimensional cuboids are compact). *Let I be a cuboid in \mathbb{R}^n i.e.*

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

The cuboid I is compact.

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

Denote $S_0 := I$. We divide S_m ($m \in \mathbb{N}$) into 2^n parts by equally dividing every $I_i := \{\mathbf{x} \in \mathbb{R}_n \mid a_i \leq x_i \leq b_i\}$ into two. Choose one that contains infinite points of $(\mathbf{x}_i)_{i \in \mathbb{N}}$ 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 learned that $\exists! \xi := (\xi_i)_{i \in \mathbb{N}}$ s.t. $\{\xi\} := \bigcap S$ from the theory of real numbers.

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

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

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

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

Assume that we find a finite subcover $\Omega' := \{B(\mathbf{x}; n_k) \mid k \in m\}$, 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). \square

§4 Connected Set

Definition 4.1 (Connected space). Topological space (X, \mathcal{T}) is called **connected** if there is no **open-closed set** (i.e. both open and closed) besides \emptyset 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, \mathcal{T}) be a topological space. Subset C is said to be **connected** if subspace (C, \mathcal{T}_C) is connected.

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

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

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

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

Theorem 4.2. *Connected sets have connected closure.*

Proof. \square

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

Proof. \rightarrow : 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 $E \neq \emptyset$ 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, $c_1 = \inf(E \cap [a, b]) \in E$ while $c_2 = \inf((C - E) \cap [a, b]) \in C - E$. However $E \cap (C - E) = \emptyset$, hence $c_1 < c_2$, which means $(c_1, c_2) \cap E = \emptyset$. Here's the contradiction. \square

Definition 4.3 (Locally connected). A topological space (X, \mathcal{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 closed 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, $\lim_{n \rightarrow \infty} x_n = x \in X$. If $x \notin A$, then $\forall U \in \mathcal{U}(x)$, $\text{card}(U \cap A) = \infty$ 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** $f: X_1 \rightarrow X_2$ when (X_1, d_1) and (X_2, d_2) are both metric space, i.e. f is a bijective and $\forall a, b \in X_1$, $d_2(f(a), f(b)) = d_1(a, b)$, 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 metric 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 $f: Y_1 \rightarrow Y_2$ can be defined: if $x_1, x_2 \in X$,

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

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 $\forall y'_1, y''_1 \in Y_1 - X$,

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

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 $C_X := \{(x_n)_{n \in \mathbb{N}} \in X^{\mathbb{N}} \mid \forall \varepsilon \in \mathbb{R}_+, \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \in \mathbb{N} (n > N \wedge m > N \rightarrow d_X(x_n, x_m) < \varepsilon)\}$, 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 $\lim_{n \rightarrow \infty} d(x_n, x'_n) = 0$.

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

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

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

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, $\lim_{n \rightarrow \infty} d(x_n, x'_n)$ always exists. This fact allows us to introduce⁴:

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

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

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

Let $([(x_n^i)_{n \in \mathbb{N}}])_{i \in \mathbb{N}}$ 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 $j > N$, $k > N$, $d_X(x_j^i, x_k^i) < 1/i$. Choose $a^i := x_k^i$ for such $k > N$, so that $d([(a^i)_{n \in \mathbb{N}}], [(x_n^i)_{n \in \mathbb{N}}]) < 1/i$.

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

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

⁴We 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)

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

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

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

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

Note: We have already seen 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 $\mathcal{B} \subset 2^X$ is called a **(filter) base** in X if the following conditions hold:

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

Here is a list of some important filter bases:

- (1) $x \rightarrow a$, where $a \in X$, means $\mathcal{U}(a)$;
- (2) $x \rightarrow \infty$, means $\{V \mid X - V \in \mathcal{U}(a) - \{X\}\}$;
- (3) $E \ni x \rightarrow a$, means $\{\dot{U}(a) \cap E \mid \dot{U}(a) \in \mathcal{U}(a)\}$;
- (4) $E \ni x \rightarrow \infty$, means $\{E \cap V \mid X - V \in \mathcal{U}(a) - \{X\}\}$.

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 $\mathcal{B} \subset 2^{\mathcal{D}(f)}$ of a mapping $f: \mathcal{D}(f) \rightarrow Y$, in which Y is equipped with a topology \mathcal{T} .

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

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

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

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

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

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

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

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

Theorem 6.2 (Cauchy criterion for existence of limit). Let \mathcal{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 \mathcal{B} iff $\forall \varepsilon \in \mathbb{R}_+, \exists B \in \mathcal{B}$ s.t. $\omega(f; B) < \varepsilon$.

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

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

\leftarrow : $\forall n \in \mathbb{N}_+, \exists B_n \in \mathcal{B}$ s.t. $\omega(f; B_n) < 1/n$. Since $B_n \neq \emptyset$ (the definition of filter base), we can choose⁵ $x_n \in B_n$ 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

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

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

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

the right-hand side can be arbitrary small, if n is even greater. \square

Definition 6.4 (Continuity). A mapping $f: X \rightarrow Y$, where X, Y is equipped with topology $\mathcal{T}_X, \mathcal{T}_Y$, 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, \mathcal{T}_X) and (Y, \mathcal{T}_Y) be two topological spaces, $f \in Y^X$. The function f is continuous iff $\forall G_Y \in \mathcal{T}_Y, f^{-1}(G_Y) \in \mathcal{T}_X$.

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

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

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

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

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

Homeomorphic topological spaces are identical with respect to their topological properties 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, $E \in \mathcal{P}(X)$. $f \in C(E, Y)$, $g \in C(f(E), Z)$, then*

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

Theorem 6.5 (Continuous then locally bounded). *Let (X, \mathcal{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 $\exists U(x) \in \mathcal{U}(x)$ 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, \mathcal{T}_X) and (Y, \mathcal{T}_Y) be two topological spaces. Let $K \subset X$ be a compact set. If $f: X \rightarrow Y \in C(X, Y)$, then $f(K)$ is compact.*

Proof. For each open cover $\Omega_Y = \{G_Y \in \mathcal{T}_Y\} \subset \mathcal{T}_Y$ over $f(K)$, $f^{-1}(G_Y) \in \mathcal{T}_X$ (Theorem 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 $f(K)$. \square

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 3.7). If $M \notin f(K)$, then open covers $\{B(M; (M - m)/n) - \bar{B}(M; (M - m)/(n + 1)) \mid n \in \mathbb{N}_+\}$ would not have a finite subcover, which is a contradiction to the compactness of $f(K)$. \square

Theorem 6.9 (Bijective from compact space to Hausdorff space is homeomorphism). *Let (K, \mathcal{T}_K) be a compact space and (Y, \mathcal{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. $G \in \mathcal{T}_K$ 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). \square

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. $\forall E \in \mathcal{P}(X)$,

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

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

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

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

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

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

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

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

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 $\omega_0 + \varepsilon$. \square

Theorem 6.12 (Conservation of connectedness). *Let (X, \mathcal{T}_X) and (Y, \mathcal{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 $(f(E), \mathcal{T}_{f(E)})$ have concurrently open-closed pre-images in (E, \mathcal{T}_E) . \square

Theorem 6.13 (Intermediate-value theorem). *Let (X, \mathcal{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 $\forall C \in [A, B]$, $C \in f(X)$. \square

§7 Contraction

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

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

$$d(f(x_1), f(x_2)) \leq qd(x_1, x_2). \quad (7-1)$$

Lemma 5. *A contraction $f: X \rightarrow X$ is always continuous.*

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

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

\square

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

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

Proof. By the inequality 7-1:

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

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

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

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

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

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

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

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

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

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. $\forall x_1, x_2 \in X$,

$$d_Y(f(x_1), f(x_2)) \leq M d_X(x_1, x_2), \quad (7-4)$$

then f is said to be **Lipschitz continuous**. Inequality 7-4 is called the **Lipschitz 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 $\|\cdot\|: X \rightarrow \mathbb{R}$ assigning to each vector $\mathbf{x} \in X$ a real number $\|\mathbf{x}\|$ is called a **norm** in the linear space X if:

- a) $\|\mathbf{x}\| = 0 \leftrightarrow \mathbf{x} = \mathbf{0}$ (nondegeneracy);
- b) $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$ (homogeneity);
- c) $\|\mathbf{x}_1 + \mathbf{x}_2\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$ (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:

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

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 $\langle \cdot, \cdot \rangle: X^2 \rightarrow \mathbb{C}$ defined, s.t. $\forall \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in X, \forall \lambda \in \mathbb{C}$.

- a) $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \overline{\langle \mathbf{x}_2, \mathbf{x}_1 \rangle}$;
- b) $\langle \lambda \mathbf{x}_1, \mathbf{x}_2 \rangle = \lambda \langle \mathbf{x}_1, \mathbf{x}_2 \rangle$;
- c) $\langle \mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_3 \rangle = \langle \mathbf{x}_1, \mathbf{x}_3 \rangle + \langle \mathbf{x}_2, \mathbf{x}_3 \rangle$.

A Hermitian form is said to be **positive semi-definite**, if $\forall \mathbf{x} \in X, \langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ ¹. A Hermitian form is said to be **degenerate**, if $\exists \mathbf{x} \in X - \{\mathbf{0}\}$ s.t. $\langle \mathbf{x}, \mathbf{x} \rangle = 0$. 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 form² is said to be an **inner product**. A space equipped with an inner product is said to be a **inner product space**.

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

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 equipped with an inner product \langle, \rangle . $\forall \mathbf{x}, \mathbf{y} \in X$,*

$$|\langle \mathbf{x}, \mathbf{y} \rangle|^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle. \quad (8-2)$$

Proof. The theorem is trivial as $\mathbf{y} = \mathbf{0}$. Let us assume that $\mathbf{y} \neq \mathbf{0}$, therefore $\langle \mathbf{y}, \mathbf{y} \rangle > 0$.
 $\forall \lambda \in \mathbb{C}$,

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

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

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

□

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

$$\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}, \quad (8-3)$$

and a metric

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

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

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

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

□

Definition 8.5 (Hilbert space). If a linear space is equipped with an inner product, 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 \mathcal{A} be a n -multilinear operator space over normed space $(\mathbf{X}_i)_{i \in n}$ to a normed space Y i.e. $\mathcal{A} \in \mathcal{L}(X_0, X_1, \dots, X_{n-1}; Y)$. We define the norm $\|\mathcal{A}\|$ as:

$$\|\mathcal{A}\| := \sup \left\{ \frac{\|\mathcal{A}(\mathbf{x}_i)_{i \in n}\|_Y}{\prod_{i \in n} \|\mathbf{x}_i\|_{X_i}} \mid \forall i \in n, \mathbf{x}_i \in X_i - \{\mathbf{0}\} \right\}, \quad (9-1)$$

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

The following theorem gives an equivalent definition:

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

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

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

$$\|\mathcal{A}(\mathbf{x})_{i \in n}\|_Y \leq \|\mathcal{A}\| \prod_{i \in n} \|\mathbf{x}_i\|_{X_i}. \quad (9-3)$$

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

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

Proof. First assume that \mathcal{A} is bounded.

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

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

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

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

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

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

$$\|\mathcal{A}(\mathbf{e}_i)_{i \in n}\|_Y = \frac{1}{\delta^n} \|\mathcal{A}(\delta \mathbf{e}_i)_{i \in n}\|_Y \leq \frac{\varepsilon}{\delta^n},$$

which implies that the operator \mathcal{A} is bounded. \square

Theorem 9.4 (Continuous at zero then at everywhere). Let $\mathcal{A} \in \mathcal{L}(X_0, X_1, \dots, X_{n-1}; Y)$. 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.

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

$$\begin{aligned} d_Y(\mathcal{A}(\mathbf{x} + \Delta \mathbf{x}), \mathcal{A}(\mathbf{x})) &= \|\mathcal{A}(\mathbf{x} + \Delta \mathbf{x}) - \mathcal{A}(\mathbf{x})\|_Y \\ &= \left\| \mathcal{A}(\Delta \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \mathcal{A}(\mathbf{x}_1, \Delta \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\mathbf{x}_0, \mathbf{x}_1, \dots, \Delta \mathbf{x}_{n-1}) \right. \\ &\quad \left. + \mathcal{A}(\Delta \mathbf{x}_0, \Delta \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\mathbf{x}_0, \dots, \Delta \mathbf{x}_{n-2}, \Delta \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\Delta \mathbf{x}) \right\|_Y \\ &\leq \|\mathcal{A}(\Delta \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1})\|_Y + \dots + \|\mathcal{A}(\mathbf{x}_0, \mathbf{x}_1, \dots, \Delta \mathbf{x}_{n-1})\|_Y \\ &\quad + \dots + \|\mathcal{A}(\Delta \mathbf{x})\|_Y \\ &\leq \|\mathcal{A}\| \sum_{S \in \mathcal{P}(n) - \{\emptyset\}} \prod_{i \in n-S} \|\mathbf{x}_i\|_{X_i} \prod_{j \in S} \|\Delta \mathbf{x}_j\|_{X_j}. \end{aligned}$$

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

³Be reminiscent of the Definition 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, \dots, X_{n-1} to Y by $\mathcal{B}(X_0, \dots, X_{n-1}; Y)$.

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

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

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

$$\|\mathcal{B}\mathcal{A}\| \leq \|\mathcal{B}\| \|\mathcal{A}\|.^4$$

Proof.

$$\begin{aligned} \|\mathcal{B}\mathcal{A}\| &= \sup \{ \|\mathcal{B}\mathcal{A}\mathbf{x}\|_Z / \|\mathbf{x}\|_X \mid \mathbf{x} \in X - \{\mathbf{0}\} \} \\ &\leq \|\mathcal{B}\| \sup \{ \|\mathcal{A}\mathbf{x}\|_Y / \|\mathbf{x}\|_X \mid \mathbf{x} \in X - \{\mathbf{0}\} \} = \|\mathcal{B}\| \|\mathcal{A}\|. \end{aligned}$$

□

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

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

$$\|\mathcal{A}_\ell \mathbf{x} - \mathcal{A}_m \mathbf{x}\|_Y = \|(\mathcal{A}_\ell - \mathcal{A}_m) \mathbf{x}\|_Y \leq \|\mathcal{A}_\ell - \mathcal{A}_m\| \prod_{i \in n} \|\mathbf{x}_i\|_{X_i},$$

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

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

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

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

$$0 \leq \|\mathcal{A} \mathbf{e}\|_Y \leq \|\mathcal{A}_\ell \mathbf{e}\|_Y + \varepsilon \leq \|\mathcal{A}_\ell\| + \varepsilon,$$

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

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

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

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

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

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

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

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

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

□

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

§10 Differentiation

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

$$f(\mathbf{x} + \Delta \mathbf{x}) - f(\mathbf{x}) = \mathcal{L}(\mathbf{x})\Delta \mathbf{x} + \alpha(\mathbf{x}; \Delta \mathbf{x}), \quad (10-1)$$

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

Such $\mathcal{L}|_{\mathbf{x}}$ is called the **differential** of f at \mathbf{x} ⁶, denoted by $df(\mathbf{x})$ or $f'(\mathbf{x})$.

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

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

$$(\mathcal{L}_1(\mathbf{x}) - \mathcal{L}_2(\mathbf{x}))\Delta \mathbf{x} = o(\Delta \mathbf{x}),$$

⁵ \mathbf{x} here is an argument.

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

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

$$\lim_{\|\Delta\mathbf{x}\|_X \rightarrow 0} \left\| (\mathcal{L}_1(\mathbf{x}) - \mathcal{L}_2(\mathbf{x})) \frac{\Delta\mathbf{x}}{\|\Delta\mathbf{x}\|_X} \right\|_Y = 0,$$

This means that whatever the direction of unit vector $\Delta\mathbf{x}/\|\Delta\mathbf{x}\|_X$ is, the norm of $\|(\mathcal{L}_1(\mathbf{x}) - \mathcal{L}_2(\mathbf{x}))\Delta\mathbf{x}/\|\Delta\mathbf{x}\|_X\|_Y$ is always zero, therefore $\|\mathcal{L}_1(\mathbf{x}) - \mathcal{L}_2(\mathbf{x})\| = 0$. By the definition of norms, this means that $\mathcal{L}_1(\mathbf{x}) - \mathcal{L}_2(\mathbf{x}) = \mathcal{O}$, or $\mathcal{L}_1(\mathbf{x}) = \mathcal{L}_2(\mathbf{x})$. \square

Theorem 10.1 gives us the right to define:

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

$$f' : \Delta(f) \rightarrow \mathcal{B}(X, Y); \mathbf{x} \mapsto df(\mathbf{x})$$

is called the *derivative mapping* of f .

Warning: We use $f'(\mathbf{x})$ 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 $\forall \mathcal{A} \in \mathcal{B}(X; Y)$, $\forall \mathbf{x} \in X$, $d\mathcal{A}(\mathbf{x}) = \mathcal{A}$, which is different from the usual notations that writes $f(x) = e^x \rightarrow f'(x) = e^x = f(x)$ and $f(x) = ax \rightarrow f'(x) = a$.

To make it clear, we must remember: $f \in Y^X$, $f' \in \mathcal{B}(X; Y)^X$, $f'(\mathbf{x}) \in \mathcal{B}(X; Y)$, $f'(\mathbf{x})\Delta\mathbf{x} \in Y$. 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 $d\mathbf{x}_i$ as:

$$d\mathbf{x}_i \Delta\mathbf{x} = \Delta\mathbf{x}_i,$$

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

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

$$df(\mathbf{x}) = f'(\mathbf{x}) d\mathbf{x},$$

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 $D \in \mathcal{P}(X)$ is differentiable at \mathbf{x} which is an interior point of D , then f is continuous at \mathbf{x} .

Proof. as $\|\Delta\mathbf{x}\| \rightarrow 0$

$$\|f(\mathbf{x} + \Delta\mathbf{x}) - f(\mathbf{x})\|_Y \leq \|\mathcal{L}(\mathbf{x})\Delta\mathbf{x}\|_Y + \|\alpha(\mathbf{x}; \Delta\mathbf{x})\|_Y \leq \|\mathcal{L}(\mathbf{x})\| \|\Delta\mathbf{x}\|_X + \|\alpha(\mathbf{x}; \Delta\mathbf{x})\|_Y \rightarrow 0.$$

\square

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

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

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

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

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

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

Consider a mappings $f: X \rightarrow Y$, where $Y := \prod_{i \in n} Y_i$, normed with $\|\mathbf{y}\|_Y := \sqrt[p]{\sum_{i \in n} \|\mathbf{y}_i\|_{Y_i}^p}$.

By writing f as $(f_i)_{i \in n}$ such that $f(\mathbf{x}) = (f_i(\mathbf{x}))_{i \in n}$, and

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

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

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

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

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

$$d\mathcal{A}(\mathbf{x}) = \mathcal{A}(d\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\mathbf{x}_0, \dots, \mathbf{x}_{n-2}, d\mathbf{x}_{n-1}).$$

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

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

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

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

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

$$\begin{aligned} \mathcal{A}(\mathbf{x} + \Delta\mathbf{x}) - \mathcal{A}(\mathbf{x}) &= \mathcal{A}(\Delta\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \mathcal{A}(\mathbf{x}_1, \Delta\mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\mathbf{x}_0, \mathbf{x}_1, \dots, \Delta\mathbf{x}_{n-1}) \\ &\quad + \mathcal{A}(\Delta\mathbf{x}_0, \Delta\mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\mathbf{x}_0, \dots, \Delta\mathbf{x}_{n-2}, \Delta\mathbf{x}_{n-1}) + \dots + \mathcal{A}(\Delta\mathbf{x}) \\ &= \mathcal{A}(\Delta\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\mathbf{x}_0, \dots, \mathbf{x}_{n-2}, \Delta\mathbf{x}_{n-1}) + o(\mathbf{x}; \Delta\mathbf{x}), \end{aligned}$$

⁷Remember, we write the composition of two linear operators omitting the “o” in the middle.

where we utilize the fact that

$$\|\mathcal{A}(\Delta \mathbf{x}_0, \Delta \mathbf{x}_1, \dots, \mathbf{x}_{n-1})\|_Y \leq \|\mathcal{A}\| \|\Delta \mathbf{x}_0\|_{X_0} \|\Delta \mathbf{x}_1\|_{X_1} \prod_{i \in n \setminus 2} \|\mathbf{x}_i\|_{X_i} = o(\mathbf{x}; \Delta \mathbf{x}), \dots$$

Therefore

$$d\mathcal{A}(\mathbf{x})\Delta \mathbf{x} = \mathcal{A}(\Delta \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\mathbf{x}_0, \dots, \mathbf{x}_{n-2}, \Delta \mathbf{x}_{n-1})$$

or

$$d\mathcal{A}(\mathbf{x}) = \mathcal{A}(d\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n-1}) + \dots + \mathcal{A}(\mathbf{x}_0, \dots, \mathbf{x}_{n-2}, d\mathbf{x}_{n-1}).$$

□

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

$$\mathcal{A}\mathcal{A}^{-1} = \text{id}_Y; \quad \mathcal{A}^{-1}\mathcal{A} = \text{id}_X.$$

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

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

as $\delta \mathcal{A} \rightarrow \mathcal{O}$.

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

$$\|-\mathcal{A}^{-1}\delta \mathcal{A}\| \leq \|\mathcal{A}^{-1}\| \|\delta \mathcal{A}\| < \|\mathcal{A}^{-1}\| \|\mathcal{A}^{-1}\|^{-1} = 1,$$

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

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

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

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

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

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

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

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

and

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

□

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

$$\varphi_i: X_i \rightarrow X; \mathbf{x}_i \mapsto (\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_{i-1}, \mathbf{x}_i, \mathbf{a}_{i+1}, \dots, \mathbf{a}_{n-1}), \quad (10-3)$$

so that $f \circ \varphi_i$ means the mapping of alone \mathbf{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 $\mathbf{a}_i \in X_i$, we call its derivative at this point the **partial derivative** of f with respect to \mathbf{x}_i at $\mathbf{a} := (\mathbf{a}_i)_{i \in n}$, denoted by $\partial_i f(\mathbf{a})$ or $\frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{a})$.

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

$$df(\mathbf{a}) = \sum_{i \in n} \partial_i f(\mathbf{a}) d\mathbf{x}_i. \quad (10-4)$$

Definition 10.5 (Continuously differentiable). Let $f \in Y^X$ and differentiable at $\mathbf{x} \in X$. If the derivative mapping $f' \in \mathcal{B}(X; Y)^X$ is continuous at \mathbf{x} , we say that f is **continuously differentiable** at point \mathbf{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, \dots, X_{n-1} , Y be normed spaces, $X := \prod_{i \in n} X_i$, $\mathbf{x} \in X$, $f \in Y^X$ is differentiable at \mathbf{x} . f is continuously differentiable at \mathbf{x} iff $\forall i \in n$, $\partial_i f \in \mathcal{B}(X_i; Y)^X$.

Proof.

$$\begin{aligned} \|\partial_i f(\mathbf{x} + \Delta\mathbf{x}) - \partial_i f(\mathbf{x})\| &\leq \left\| \sum_{j \in n} (\partial_j f(\mathbf{x} + \Delta\mathbf{x}) - \partial_j f(\mathbf{x})) \right\| = \|df(\mathbf{x} + \Delta\mathbf{x}) - df(\mathbf{x})\| \\ &\leq \sum_{j \in n} \|\partial_j f(\mathbf{x} + \Delta\mathbf{x}) - \partial_j f(\mathbf{x})\| \end{aligned}$$

□

⁸or $C^{(1)}(X)$ 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 , $f \in Y^U$, $\mathbf{x} \in U$. The derivative of f with respect to a vector $\boldsymbol{\ell}$ is defined as:

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

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 , $f \in Y^U$, $\mathbf{x} \in U$. If f is differentiable at \mathbf{x} , then $\forall \boldsymbol{\ell} \in X$, the derivative of f with respect to $\boldsymbol{\ell}$ exists, and

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

Proof.

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

□

§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} , $\mathbf{a}, \mathbf{b} \in X$. The **closed** and **open interval** is defined as:

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

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

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

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

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

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

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

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

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

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. $a_i \leq a_{i+1} < b_{i+1} \leq b_i$, $\forall i \in \mathbb{N}$, over which the inequality

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

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.

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

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

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

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

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

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

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

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

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

□

The equality can be satisfied for some points when f is a real-valued function. cf. Theorem 13.7.

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

$$\|f(\mathbf{x} + \Delta \mathbf{x}) - f(\mathbf{x}) - \mathcal{A} \Delta \mathbf{x}\|_Y \leq \sup\{\|f'(\xi) - \mathcal{A}\| \|\Delta \mathbf{x}\|_X \mid \xi \in [\mathbf{x}_0, \mathbf{x}_0 + \Delta \mathbf{x}]\}.$$

Proof. Define:

$$F: [0, 1] \rightarrow Y; t \mapsto f(\mathbf{x} + t \Delta \mathbf{x}) - \mathcal{A} t \Delta \mathbf{x}.$$

By the finite-increment theorem 11.1,

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

□

Theorem 11.2 (Continuously differentiable then Lipschitz continuous). *Let K be a convex⁹ 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.*

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

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

$$\|f(\mathbf{x}_2) - f(\mathbf{x}_1)\|_Y \leq \sup \{ \|f'(\mathbf{x})\| \mid \mathbf{x} \in [\mathbf{x}_1, \mathbf{x}_2] \} \|\mathbf{x}_2 - \mathbf{x}_1\|_X \leq M \|\mathbf{x}_2 - \mathbf{x}_1\|_X.$$

□

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. $\lim_{x \rightarrow +0} \omega(x) = 0$, and $\forall \mathbf{x} \in X$, if $\Delta \mathbf{x} \in K \cap B(\mathbf{x}; \delta)$, then*

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

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

Proof. By Corollary 4,

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

Let

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

□

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, \dots, X_{n-1}, Y be normed spaces, $X := \prod_{i \in n} X_i$, $G \in \mathcal{T}_X$, $f \in Y^G$.*

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

Proof. \rightarrow : 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

$$\mathcal{L} := \sum_{i \in n} \partial_i f(\mathbf{x}) d\mathbf{x}_i,$$

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

Let us introduce a notation,

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

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

Then

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

By Corollary 4, we have:

$$\begin{aligned} & \|f(\mathbf{x} + \Delta \mathbf{x}) - f(\mathbf{x}) - \mathcal{L} \Delta \mathbf{x}\|_Y \\ & \leq \|\Delta_0 f(\mathbf{x}_0, \mathbf{x}_1 + \Delta \mathbf{x}_1, \dots, \mathbf{x}_{n-1} + \Delta \mathbf{x}_{n-1}) - \partial_0 f(\mathbf{x}) \Delta \mathbf{x}_0\|_Y \\ & \quad + \dots + \|\Delta_{n-1} f(\mathbf{x}) \Delta \mathbf{x}_{n-1} - \partial_{n-1} f(\mathbf{x}) \Delta \mathbf{x}_{n-1}\|_Y \\ & \leq \sup \left\{ \|\partial_0 f(\xi_0, \mathbf{x}_1 + \Delta \mathbf{x}_1, \dots, \mathbf{x}_{n-1} + \Delta \mathbf{x}_{n-1}) \right. \\ & \quad \left. - \partial_0 f(\mathbf{x}_0, \mathbf{x}_1 + \Delta \mathbf{x}_1, \dots, \mathbf{x}_{n-1} + \Delta \mathbf{x}_{n-1})\|_Y \mid \xi_0 \in [\mathbf{x}_0, \mathbf{x}_0 + \Delta \mathbf{x}_0] \right\} \|\Delta \mathbf{x}_0\|_{X_0} \\ & \quad + \dots + \sup \left\{ \|\partial_{n-1} f(\mathbf{x}_0, \dots, \xi_{n-1}) - \partial_{n-1} f(\mathbf{x})\|_Y \mid \xi_{n-1} \in [\mathbf{x}_0, \mathbf{x}_0 + \Delta \mathbf{x}_0] \right\} \|\Delta \mathbf{x}_{n-1}\|_{X_{n-1}}. \end{aligned}$$

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

$$\begin{aligned} & \lim_{\Delta \mathbf{x}_i \rightarrow 0} \sup \left\{ \|\partial_0 f(\mathbf{x}_0, \dots, \xi_i, \dots, \mathbf{x}_{n-1} + \Delta \mathbf{x}_{n-1}) \right. \\ & \quad \left. - \partial_0 f(\mathbf{x}_0, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{n-1} + \Delta \mathbf{x}_{n-1})\|_Y \mid \xi_i \in [\mathbf{x}_0, \mathbf{x}_0 + \Delta \mathbf{x}_0] \right\} \\ &= 0. \end{aligned}$$

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

$$f(\mathbf{x} + \Delta \mathbf{x}) - f(\mathbf{x}) - \mathcal{L} \Delta \mathbf{x} = o(\Delta \mathbf{x}),$$

which means $df(\mathbf{x}) = \mathcal{L}$. □

Then we shall use finite-increment theorem (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, $\mathbf{x}_0 \in X$, $U \in \mathcal{U}(\mathbf{x}_0)$, $f \in Y^U$. If f is differentiable in $\mathring{U} := U - \{\mathbf{x}_0\}$, and*

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

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

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

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

By the definition of differential, we know $f'(\mathbf{x}_0) = \mathcal{L}$. □

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 $\forall \mathbf{x} \in U$, f is differentiable at \mathbf{x} , and $f'(\mathbf{x}) = \mathcal{O}$, then f is a constant function from U i.e. $\exists \mathbf{y}_0 \in Y, \forall \mathbf{x} \in U, f(\mathbf{x}) = \mathbf{y}_0$.*

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

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

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

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 $\forall \mathbf{x} \in U$, f is differentiable at \mathbf{x} , and $f'(\mathbf{x}) = \mathcal{O}$, then f is a constant function from U .*

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

First, E is open. $\forall \mathbf{x} \in E$, $\exists B(\mathbf{x}; \delta) \subset U$. Since $\forall \mathbf{x}' \in B(\mathbf{x}; \delta)$, $[\mathbf{x}, \mathbf{x}'] \subset B(\mathbf{x}; \delta)$, f is constant in $B(\mathbf{x}; \delta)$ and therefore $B(\mathbf{x}; \delta) \subset E$. 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, ($\mathbf{x}_0 \in E$), the only choice for a open-closed set in a connected set U is U itself, i.e. $\forall \mathbf{x} \in U$, $f(\mathbf{x}) = f(\mathbf{x}_0)$. \square

§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 \mathcal{T}_X and \mathcal{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 differentiation. If the n -th differential $f^{(n)} \in Y_n^U$ is differentiable in $U \in \mathcal{T}_X$ ¹⁰, we can define the $(n+1)$ -th differential $f^{(n+1)}(\mathbf{x})$ by:

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

Like $C^{(1)}$, we can define $C^{(p)}$.

Definition 12.2 (Diffeomorphism). $f \in U^V$, where U, V are two open subsets of normed spaces X, Y . If:

- 1) $f \in C^{(p)}(U)$;
- 2) $\exists f^{-1} \in V^U$;
- 3) $f^{-1} \in C^{(p)}(V)$,

then we call f a $C^{(p)}$ -**diffeomorphism** or a **diffeomorphism with smoothness p** .

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

$$((f^{(n)}(\mathbf{x})\ell_0) \cdots \ell_{n-1}) = \frac{\partial}{\partial \ell_0} \cdots \frac{\partial}{\partial \ell_{n-1}} f(\mathbf{x}). \quad (12-1)$$

Proof. See Theorem 10.10. \square

¹⁰ Y_n is also a normed space.

Theorem 12.2 (Symmetry of higher-order differentiation). *Let $\sigma \in S_n$ where S_n is the symmetric group¹¹ on n . Let X and Y be normed spaces, with induced topologies \mathcal{T}_X and \mathcal{T}_Y , $U \in \mathcal{T}_X$, $\mathbf{x} \in U$, $(\ell_i)_{i \in n} \in X^n$. If $f \in Y^U$ has n -th differential $f^{(n)}$ in U , then*

$$\frac{\partial}{\partial \ell_{\sigma(0)}} \cdots \frac{\partial}{\partial \ell_{\sigma(n-1)}} f(\mathbf{x}) = \frac{\partial}{\partial \ell_0} \cdots \frac{\partial}{\partial \ell_{n-1}} f(\mathbf{x}).$$

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

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

Let

$$\begin{aligned} \Delta(t) &:= f(\mathbf{x} + t\ell_0 + t\ell_1) - f(\mathbf{x} + t\ell_0) - f(\mathbf{x} + t\ell_1) + f(\mathbf{x}), \\ D(t, t') &:= f(\mathbf{x} + t\ell_0 + t'\ell_1) - f(\mathbf{x} + t'\ell_1), \end{aligned}$$

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,

$$\begin{aligned} \|\Delta(t) - t^2[f''(\mathbf{x})\ell_0]\ell_1\|_Y &= \|D(t, t) - D(t, 0) - t^2[f''(\mathbf{x})\ell_0]\ell_1\|_Y \\ &\leq t \sup \left\{ \left\| \frac{\partial D}{\partial t'}(t, \theta) - t\theta[f''(\mathbf{x})\ell_0]\ell_1 \right\|_Y \mid \theta \in (0, t) \right\} \\ &\leq t\|\ell_1\|_X \sup \{ \|f'(\mathbf{x} + t\ell_0 + \theta\ell_1) - f'(\mathbf{x} + \theta\ell_1) - t\theta f''(\mathbf{x})\ell_0\| \mid \theta \in (0, t) \} \\ &= t\|\ell_1\|_X \sup \{ \|\theta f''(\mathbf{x})(t\ell_0 + \theta\ell_1 - \theta\ell_1) - t\theta f''(\mathbf{x})\ell_0 + o(t)\| \mid \theta \in (0, t) \} \\ &= o(t^2). \end{aligned}$$

Hence,

$$[f''(\mathbf{x})\ell_0]\ell_1 = \lim_{t \rightarrow 0} \frac{\Delta(t)}{t^2}.$$

Substituting (ℓ_0, ℓ_1) by (ℓ_1, ℓ_0) in the definition of $\Delta(t)$ doesn't change its value, hence we have proved the theorem in the case when $n = 2$. \square

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

$$f^{(n)}(\mathbf{x})(\ell_i)_{i \in n} := ((f^{(n)}(\mathbf{x})\ell_0) \cdots) \ell_{n-1}, \quad (12-2)$$

and

$$f^{(n)}(\mathbf{x})\ell^n := f^{(n)}(\ell, \dots, \ell). \quad (12-3)$$

Theorem 12.3. *Let X_0, \dots, 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 $\forall (i_k)_{k \in n} \in m^n$, $\forall \mathbf{x} \in U$, n -th partial derivative*

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

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

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).

¹¹Or permutation

¹²By Corollary 3, these two spaces are isomorphic

§13 Applications of Differentiation

13.1 Taylor's Formula

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

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

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

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

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

with respect to $\Delta\mathbf{x}$ is¹³:

$$T'_k(\Delta\mathbf{x})\ell = \frac{1}{(k-1)!} f^{(k)}(\mathbf{x}) \Delta\mathbf{x}^{k-1} \ell.$$

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

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

□

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

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

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

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

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

¹³cf. Theorem 10.6

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

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

We have:

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

therefore

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

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 $1-t \leq \delta$, then:

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

or

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

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

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

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

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

and

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

which contradicts to the definition of δ' .

Hence $\delta' = 1$, or:

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

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

$$\|g(0)\|_Y \leq \frac{M}{(n+1)!} \|\Delta\mathbf{x}\|_X^{n+1}. \quad (13-2)$$

Eq. (13-2) is to prove. □

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

Proof.

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

□

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 $\mathbf{x} \in U$, and $\forall k \in n+1$, exists k -linear operators \mathcal{L}_k s.t.

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

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

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

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

hence:

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

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

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

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

13.2 Interior Extrema

Definition 13.1 (Extremum). Let X be a normed space, and $f \in \mathbb{R}^X$. If $\mathbf{x} \in X$ satisfies: $\exists U \in \mathcal{U}(\mathbf{x})$ s.t. $\forall \mathbf{x}' \in U - \{\mathbf{x}\}$, $f(\mathbf{x}) > f(\mathbf{x}')$, then \mathbf{x} is a **locally maximum point** of f . Similarly, we can define **locally minimum point**. Both locally maximum point and minimum point are called **extremum 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 $\mathbf{x} \in U$, where $n \in \mathbb{N}_+$. $\forall k \in n+1$, $f^{(k)}(\mathbf{x}) = \mathcal{O}$, and $f^{(n+1)}(\mathbf{x}) \neq \mathcal{O}$.

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

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

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

hence

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

If the difference remains its sign, then $n+1$ must be an even number. \square

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 $\mathbf{x} \in U$, where $n \in \mathbb{N}_+$. $\forall k \in n+1$, $f^{(k)}(\mathbf{x}) = \mathcal{O}$, and $f^{(n+1)}(\mathbf{x}) \neq \mathcal{O}$.

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

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

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

\square

With the study of extrema, we can rewrite and specify the finite-increment theorem (Theorem 11.1). First let's prove a useful theorem:

Theorem 13.6 (Rolle's theorem). Let $f \in C([x_0, x_0 + \Delta x]; \mathbb{R})$, where $[x_0, x_0 + \Delta x] \subset \mathbb{R}$. If f is differentiable in $(x_0, x_0 + \Delta x)$, and $f(x_0 + \Delta x) - f(x_0) = 0$, then $\exists \xi \in (x_0, x_0 + \Delta x)$ s.t. $f'(\xi) = 0$.

Proof. By the Weierstrass maximum value theorem (Theorem 6.8), f reaches its extrema in $[x_0, x_0 + \Delta x]$.

Either the maximum and minimum point are equal, then the function is constant on $[x_0, x_0 + \Delta x]$ (hence the theorem is trivial), or, one of the maximum and minimum must locates in $(x_0, x_0 + \Delta x)$. By Theorem 13.4, the point when f reach its extremum is the point when $f' = 0$. \square

Theorem 13.7 (Lagrange's finite-increment theorem). Let X be a normed space, $G \in \mathcal{T}_X$, where \mathcal{T}_X is the topology induced by the norm $\|\cdot\|_X$. Let $f \in C(G; \mathbb{R})$. If $\forall \mathbf{x} \in (x_0, x_0 + \Delta \mathbf{x})$, f is differentiable at \mathbf{x} , then $\exists \xi \in (x_0, x_0 + \Delta \mathbf{x})$,

$$|f(\mathbf{x}_0 + \Delta \mathbf{x}) - f(\mathbf{x}_0)| = |f'(\xi)|\|\Delta \mathbf{x}\|_X.$$

Proof. Let:

$$F(t) = f(\mathbf{x}_0 + t\Delta \mathbf{x}) + t[f(\mathbf{x}_0 + \Delta \mathbf{x}) - f(\mathbf{x}_0)],$$

and apply Rolle's theorem (Theorem 13.6) on $t \in [0, 1]$. \square

§14 Implicit Function Theorem

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

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

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

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

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

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

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

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

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

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

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

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

Making use of our estimate of the error to correct \mathbf{y} , we get $\mathbf{y}' = \mathbf{y} - \tilde{\Delta}$. However, since we made a approximation (which is too much!), \mathbf{y}' is also not $f(\mathbf{x})$. 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 $B(\mathbf{y}_0; \beta)$ to Y :

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

Obviously, if $\mathbf{y} = f(\mathbf{x}) \leftrightarrow F(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ then $\Delta_{\mathbf{x}}(f(\mathbf{x})) = f(\mathbf{x})$ i.e. $f(\mathbf{x})$ is a fix-point of the function $\Delta_{\mathbf{x}}(\mathbf{y})$ of \mathbf{y} with fixed \mathbf{x} . Now we need to prove such fix-point exists.

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

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

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

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

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

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

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

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

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

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

$$\|\Delta_x(\mathbf{y}) - \mathbf{y}_0\|_Y \leq \|\Delta_x(\mathbf{y}) - \Delta_x(\mathbf{y}_0)\|_Y + \|\Delta_x(\mathbf{y}_0) - \mathbf{y}_0\|_Y < \varepsilon \|\mathbf{y} - \mathbf{y}_0\|_Y + (\varepsilon - 1)\delta < \varepsilon\varepsilon + (\varepsilon - 1)\delta = \delta.$$

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

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

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

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

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

- a) $F(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$;
 - b) $F \in C(W; Z)$;
 - c) There exists the partial derivative of $F(\mathbf{x}, \mathbf{y})$ with respect to $\mathbf{y} \in Y$: $\partial_{\mathbf{y}}F(\mathbf{x}, \mathbf{y})$ in W , and $\partial_{\mathbf{y}}F$ is continuous at point $(\mathbf{x}_0, \mathbf{y}_0)$;
 - d) $\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0) \in \mathcal{B}(Y; Z)$ is reversible,
- then, $\exists U \in \mathcal{U}(\mathbf{x}_0), \exists V \in \mathcal{U}(\mathbf{y}_0), \exists f \in C(U; Y)$ s.t. $U \times V \subset W$ and $\forall \mathbf{x} \in U, \forall \mathbf{y} \in V$,

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

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

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

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

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

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

- a) $F(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$;
- b) F is continuous at $\mathbf{x}_0, \mathbf{y}_0$;
- c) There exist the partial derivatives of $F(\mathbf{x}, \mathbf{y})$ with respect to $\mathbf{y} \in Y$: $\partial_{\mathbf{y}}F(\mathbf{x}, \mathbf{y})$ and with respect to \mathbf{x} : $\partial_{\mathbf{x}}F(\mathbf{x}, \mathbf{y})$, in W , and $\partial_{\mathbf{y}}F, \partial_{\mathbf{x}}F$ are continuous at point $(\mathbf{x}_0, \mathbf{y}_0)$;
- d) $\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0) \in \mathcal{B}(Y; Z)$ is reversible,

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

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

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

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

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

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

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

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

However,

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

hence we have:

$$\begin{aligned} & \|f(\mathbf{x}_0 + \Delta \mathbf{x}) - \mathbf{y}_0 + [\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0)]^{-1} \partial_{\mathbf{x}}F(\mathbf{x}_0, \mathbf{y}_0) \Delta \mathbf{x}\|_Y \\ &\leq \|[\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0)]^{-1}\| \left[(1 + \|[\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0)]^{-1} \partial_{\mathbf{x}}F(\mathbf{x}_0, \mathbf{y}_0)\|) \|\Delta \mathbf{x}\|_X \right. \\ &\quad \left. + \|f(\mathbf{x}_0 + \Delta \mathbf{x}) - f(\mathbf{x}_0) + [\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0)]^{-1} \partial_{\mathbf{x}}F(\mathbf{x}_0, \mathbf{y}_0) \Delta \mathbf{x}\|_Y \right] o(1), \end{aligned}$$

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

or,

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

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

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

□

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

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

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

- a) $F(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$;
- b) $F \in C^{(1)}(W; Z)$;
- c) $\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0) \in \mathcal{B}(Y; Z)$ is reversible i.e. $\exists[\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0)]^{-1} \in \mathcal{B}(Z; Y)$ s.t.

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

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

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

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

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. $\mathbf{x}_0 \in X, \mathbf{y}_0 \in Y$. Denote*

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

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

- a) $F(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$;
 - b) $F \in C^{(k)}(W; Z)$;
 - c) $\partial_{\mathbf{y}}F(\mathbf{x}_0, \mathbf{y}_0) \in \mathcal{B}(Y; Z)$ is reversible,
- then, $\exists U \in \mathcal{U}(\mathbf{x}_0), \exists V \in \mathcal{U}(\mathbf{y}_0), \exists f \in C^{(k)}(U; Y)$ s.t. $U \times V \subset W$ and $\forall \mathbf{x} \in U, \forall \mathbf{y} \in V$,

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

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** $I_{\mathbf{a}, \mathbf{b}}$ in $X := \prod_{i \in n} X_i$, where $\mathbf{a}, \mathbf{b} \in X$, are defined as:

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

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:

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

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 pairwise, 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:

$$\mu(I) \leq \sum_{i \in n} \mu(I_i).$$

We shall expect the measure 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 $\mathcal{F} \in 2^X$ be a collection of subsets of a set X . If \mathcal{F} satisfies:

- 1) $\emptyset \in \mathcal{F}$;
 - 2) $\forall A \in \mathcal{F}, X - A \in \mathcal{F}$ (closed under complementation);
 - 3) $\forall \langle A_i \rangle_{i \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}, \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$ (closed under countable unions),
- then \mathcal{F} is said to be a **σ -algebra**.

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

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

- 1) $\mu(\emptyset) = 0$;
- 2) (Countable additivity) If $\langle A_i \rangle_{i \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ 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, \mathcal{F}, μ) is called a **measurable space**, and the sets in \mathcal{F} are called **measurable sets**. The image of a set in \mathcal{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:

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

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 $\lambda^*(A) \leq \lambda^*(B)$.*

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

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

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

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

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

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 \mathcal{F} .

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

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 $\langle E_i \rangle_{i \in n} \in \mathcal{F}^n$ 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). \mathcal{F} is a σ -algebra.

Proof. It is obvious that $\emptyset \in \mathcal{F}$. 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 constructed:

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

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

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

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

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

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

Pass n to the infinity, the inequality becomes:

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

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

Theorem 15.6 (Countable additivity of Lebesgue measure). If $\langle E_i \rangle_{i \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ 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:

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

Passing $n \rightarrow \infty$,

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

while the subadditive of λ^* yields the reverse. \square

Definition 15.6 (Measure zero). If a set $E \in 2^X$ is Lebesgue measurable ($E \in \mathcal{F}$), 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 $\varepsilon/2^k$. \square

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. \square

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. \square

§16 Riemann Integral over n -D cuboids

Now we introduce the partition of the cuboid:

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

Definition 16.2 (Mesh). The **mesh** 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 $\xi_j \in I_j$, $I_j \in P$, $j \in \text{card } P$. $\xi := (\xi_j)_{j \in \text{card } P}$.

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

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 , $f \in Y^I$, $(P, \xi) \in \mathfrak{P}(I)$. $N := \text{card } P$, $P := \{I_j \mid j \in N\}$. The **Riemann sum** of f over P with distinguished points ξ is defined as:

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

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

$$\int_I f(\mathbf{x}) d\mathbf{x} := \lim_{\lambda(P) \rightarrow 0} \sigma(f, P, \xi)$$

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). $f \in \mathfrak{R}(I) \rightarrow \exists M \in \mathbb{R}_+ \forall \mathbf{x} \in I (\|f(\mathbf{x})\|_Y < M)$.

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

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

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

Now consider a sequence of sets $E_k := \{\mathbf{x} \in I \mid \omega(f; \mathbf{x}) \geq 1/k\}$, 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:

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

and $B = P - A$.

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

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

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

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

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

¹We 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 ξ and ξ' , when they belong to $I_j \in A$, we let $d(f(\xi_j), f(\xi'_j)) > 1/3k_0$ ², and when $I_j \in B$, $\xi_j = \xi'_j$.

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

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

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

Now we prove that E_ε is closed. If $\mathbf{x} \notin E_\varepsilon$ i.e. $\omega(f; \mathbf{x}) \leq \varepsilon'$ where $\varepsilon' < \varepsilon$. By the definition of the oscillation at a point, $\forall \varepsilon'' \in \mathbb{R}_+$, there exists a ball $B(\mathbf{x}; \delta)$ on which $\omega(f, B(\mathbf{x}; \delta)) < \varepsilon' + \varepsilon''$. Let $\varepsilon'' = \varepsilon - \varepsilon'$, and notice that $\omega(f; \mathbf{x}') \leq \omega(f, B(\mathbf{x}, \delta))$ where $\mathbf{x}' \in B(\mathbf{x}, \delta)$. Therefore, $B(\mathbf{x}; \delta) \subset I - E_\varepsilon$. 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 $\delta = d(\cup C_2, \partial(\cup C_3))$. 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 $E_\varepsilon \subset I - K$. $\forall \mathbf{x} \in K$, since $\mathbf{x} \notin E_\varepsilon$, $\omega(f; \mathbf{x}) < \varepsilon$.

By Theorem 6.11, $\exists \delta' \in \mathbb{R}_+$, if $\mathbf{x}', \mathbf{x}'' \in K$ satisfies that $d(\mathbf{x}', \mathbf{x}'') < \delta'$, $d(f(\mathbf{x}'), f(\mathbf{x}'')) < 2\varepsilon$. 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 $P'' := \{I''_{jj'} := I_j \cap I'_{j'} \mid I_j \in P \wedge I'_{j'} \in P'\}$.

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

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. $I_j \cap (I - \cup C_3) \neq \emptyset$, since $\lambda(I_j) < \delta'' \leq \delta$, there is no way that $I_j \cap \cup C_2 \neq \emptyset$.

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

Therefore:

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

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

³Lemma 4

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

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

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:

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

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

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

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

□

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:

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

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 $\lambda(P) \rightarrow 0$):

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

Proof. We will only prove the lower Darboux theorem.

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

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

$$\sum_{\substack{I_j \in P; \\ I_j \cap \Gamma_\varepsilon \neq \emptyset}} \mu(I_j) < \varepsilon.$$

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 Γ_ε , which contradicts with the fact that $\lambda(\Gamma_\varepsilon) = 0$.

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

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

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

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

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

□

Theorem 16.4 (Darboux'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 on I , and $\underline{\mathfrak{J}} = \overline{\mathfrak{J}}$.*

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

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

§17 Riemann Integral over Jordan Measurable Sets

Definition 17.1 (Jordan Measurable). A set E in an n -D normed space X is said to be **Jordan measurable** if it is bounded, and its boundary ∂E is of measure zero.

In fact, Jordan measurable set is not a σ -algebra. For example, sets of a single points in \mathbb{R}^n is Jordan measurable, but their countable union $\mathbb{Q}^n \cap [0, 1]^n$ is not Jordan measurable.

Lemma 9 (Jordan measurable sets is closed under finite union, finite intersection and difference). *If A, B are two Jordan measurable sets, $A \cap B, A \cup B, A - B$ are also Jordan measurable.*

Proof. Only to notice that $\partial(A \cup B) \subset \partial A \cup \partial B$, $\partial(A \cap B) \subset \partial A \cup \partial B$, $\partial(A - B) \subset \partial A \cup \partial B$. □

Definition 17.2 (Characteristic function). Let E be a set in an n -D normed space X , $\chi_E \in X \rightarrow \mathbb{R}$ is defined as:

$$\chi_E(\mathbf{x}) := \begin{cases} 1 & \mathbf{x} \in E, \\ 0 & \mathbf{x} \notin E. \end{cases}$$

If a function f is defined on E , and $E \subset I$ where I is a cuboid. By default, we assign any values to $f(\mathbf{x})$ when $\mathbf{x} \in I - E$, so that $\chi_E(\mathbf{x})f(\mathbf{x})$ is considered equal to $f(\mathbf{x})$ when $\mathbf{x} \in E$, $\chi_E(\mathbf{x})f(\mathbf{x})$ is zero when $\mathbf{x} \notin E$. We might denote the function $\mathbf{x} \mapsto \chi_E(\mathbf{x})f(\mathbf{x})$ by $\chi_E f$.

Lemma 10. Let $f \in Y^E$ where E is a set in n -D normed space X and Y is a normed space. $E \subset I' \cap I''$ where I' and I'' are cuboids in X . If $\chi_E \cdot f|_I \in \mathfrak{R}(I)$, then $\chi_E \cdot f|_{I'} \in \mathfrak{R}(I')$, and

$$\int_I \chi_E(\mathbf{x})f(\mathbf{x}) \, d\mathbf{x} = \int_{I'} \chi_E(\mathbf{x})f(\mathbf{x}) \, d\mathbf{x}.$$

Proof. Let $I = I' \cap I''$.

Since all discontinuous points of $\chi_E f$ are contained in $E \cup \partial E = \overline{E} \subset I^4$, therefore, by Lebesgue's criterion 16.2 if $\chi_E f$ is Riemann integrable in either all of or none of I , I' and I'' .

If the integrals exist, we choose partitions such that $P \in \mathfrak{P}(I)$ is a subset of $P' \in \mathfrak{P}(I')$. Passing $\lambda(P') \rightarrow 0$, we can prove the equality. \square

Definition 17.3 (Riemann integrals over a set). Let $f \in Y^E$ where E is a bounded set in n -D normed space X and Y is a normed space. The Riemann integral of f over E is defined as:

$$\int_E f(\mathbf{x}) \, d\mathbf{x} := \int_I \chi_E(\mathbf{x})f(\mathbf{x}) \, d\mathbf{x},$$

where I is a arbitrary cuboid that contains E .

Theorem 17.1 (Lebesgue's criterion over a set). Let $f \in \mathbb{R}^E$, where E is a Jordan measurable set in a n -D space. $f \in \mathfrak{R}(E) \Leftrightarrow f$ is bounded in E and f is almost everywhere continuous on E .

Proof. The Lebesgue's criterion over a cuboid 16.2 and the definition of Jordan measurable set. \square

The definition of Darboux integrals and the Darboux's criterion can be generalised to Riemann integrals over a bounded set. We might denote the Darboux lower and upper integrals by:

$$\overline{\int}_E f(\mathbf{x}) \, d\mathbf{x}, \quad \underline{\int}_E f(\mathbf{x}) \, d\mathbf{x}.$$

Definition 17.4 (Jordan content). Let E be a Jordan measurable set. The **Jordan content** or the **Jordan measure**⁵ is defined as:

$$\mu(E) := \int_E 1 \, d\mathbf{x}.$$

We might call the Jordan content of a set the content, the area or the volume of it.

Definition 17.5 (Zero content). A set E is said to be of **zero content**, if it is Jordan measurable, and $\mu(E) = 0$.

A set of zero content must be of zero measure.

Theorem 17.2. A set E is of zero content, iff $\forall \varepsilon \in \mathbb{R}_+, \exists \langle I_j \rangle_{j \in N}$ s.t.

$$E \subset \bigcup_{j \in N} I_j, \quad \sum_{j \in N} \mu(I_j) < \varepsilon.$$

⁴The closure of E is the smallest closed set that contain E .

⁵Though Jordan content is not a measure.

§18 Properties of Riemann Integrals

Theorem 18.1 (Integrals are linear operators). *Let E be a bounded set in an n -D normed space X , $\mathfrak{R}(E)$ is a linear space, and $\int_E d\mathbf{x}: \mathfrak{R}(E) \rightarrow Y$ is a linear operator.*

Theorem 18.2. *If a Riemann integrable function $f \in \mathfrak{R}(E)$ is a.e. zero over E , $\int_E f(\mathbf{x}) d\mathbf{x} = 0$.*

Proof. Choosing distinguished point ξ such that $f(\xi_j) = 0$, the Riemann sum $\sigma(f, P, \xi)$ must be zero, therefore limits to zero as $\lambda(P) \rightarrow 0$. \square

We can define a equivalence relation \sim on $\mathfrak{R}(E)$, so that $f \sim g$ as long as $f(\mathbf{x}) = g(\mathbf{x})$ a.e. in E , which induces a equivalence class $\tilde{\mathfrak{R}}(E)$. $\tilde{\mathfrak{R}}(E)$ is also a linear space, and $\int_E d\mathbf{x}$ is also a linear operators from $\tilde{\mathfrak{R}}(E)$.

Theorem 18.3 (Additivity of integrals). *Let E_1 and E_2 be two Jordan measurable sets in an n -D normed space X , $f \in Y^{E_1 \cup E_2}$. 1. $\mathfrak{R}(E_1 \cup E_2) = \mathfrak{R}(E_1) \cap \mathfrak{R}(E_2)$; 2. If $\mu(E_1 \cap E_2) = 0$, then:*

$$\int_{E_1 \cup E_2} f(\mathbf{x}) d\mathbf{x} = \int_{E_1} f(\mathbf{x}) d\mathbf{x} + \int_{E_2} f(\mathbf{x}) d\mathbf{x}.$$

Proof. 1. By Lebesgue's criterion 16.2.

2. By the linearity of integrals 18.1, we have:

$$\int_{E_1} f(\mathbf{x}) d\mathbf{x} + \int_{E_2} f(\mathbf{x}) d\mathbf{x} = \int_{E_1 \cup E_2} (\chi_{E_1}(\mathbf{x}) + \chi_{E_2}(\mathbf{x})) f(\mathbf{x}) d\mathbf{x}.$$

The value of $(\chi_{E_1}(\mathbf{x}) + \chi_{E_2}(\mathbf{x})) f(\mathbf{x})$ is the same as $f(\mathbf{x})$ except for $\mathbf{x} \in E_1 \cap E_2$, where its value is $2f(\mathbf{x})$. By Theorem 18.2, the integral of $(\chi_{E_1}(\mathbf{x}) + \chi_{E_2}(\mathbf{x})) f(\mathbf{x})$ is the same as $\int_{E_1 \cup E_2} f(\mathbf{x}) d\mathbf{x}$. \square

Theorem 18.4. *$f \in Y^E$, where E is a set in an n -D normed space and Y is a complete normed space. If $f \in \mathfrak{R}(E)$, then*

$$\left\| \int_E f(\mathbf{x}) d\mathbf{x} \right\|_Y \leq \int_E \|f(\mathbf{x})\|_X d\mathbf{x}.$$

Proof. First we can see, the discontinuous points of $|f|$ is the same as f 's. Use the Riemann sum to give the inequality and pass the limits $\lambda(P) \rightarrow 0$. \square

Theorem 18.5. *$f \in \mathbb{R}^E$, where E is a set in an n -D normed space. If $f(\mathbf{x}) \geq 0$ a.e. $\mathbf{x} \in E$, and $f \in \mathfrak{R}(E)$, then*

$$\int_E f(\mathbf{x}) d\mathbf{x} \geq 0.$$

Proof. We can always find a function $f'(\mathbf{x})$ that is always non-negative (hence a.e. equal to $f(\mathbf{x})$), hence:

$$\int_E f(\mathbf{x}) d\mathbf{x} \geq \int_{\underline{E}} f(\mathbf{x}) d\mathbf{x} \geq 0.$$

\square

Corollary 5. $f, g \in \mathbb{R}^E$, where E is a set in an n -D normed space. If $f(\mathbf{x}) \geq g(\mathbf{x})$ a.e. $\mathbf{x} \in E$, and $f, g \in \mathfrak{R}(E)$, then

$$\int_E f(\mathbf{x}) \, d\mathbf{x} \geq \int_E g(\mathbf{x}) \, d\mathbf{x}.$$

Corollary 6. $f \in \mathbb{R}^E$, where E is a Lebesgue measurable set in an n -D normed space. If $f \in \mathfrak{R}(E)$, and $m \leq f(\mathbf{x}) \leq M$ a.e. in E , we have:

$$\lambda(E)m \leq \int_E f(\mathbf{x}) \, d\mathbf{x} \leq \lambda(E)M.$$

Corollary 7. $f \in \mathbb{R}^E$, where E is a bounded, Lebesgue measurable set in an n -D normed space. Let $m := \inf\{f(\mathbf{x}) \mid \mathbf{x} \in E\}$, $M := \sup\{f(\mathbf{x}) \mid \mathbf{x} \in E\}$. If $f \in \mathfrak{R}(E)$, then $\exists \theta \in [m, M]$ s.t.

$$\int_E f(\mathbf{x}) \, d\mathbf{x} = \theta \mu(E).$$

Corollary 8. E is a Jordan measurable set in \mathbb{R}^n , $f: E \rightarrow \mathbb{R}$. If f is continuous at $\mathbf{x}_0 \in E$ and $\forall \delta \in \mathbb{R}_+$, $\mu(B(\mathbf{x}_0; \delta) \cap E) > 0$, then:

$$\lim_{\delta \rightarrow +0} \frac{1}{\mu(B(\mathbf{x}_0; \delta) \cap E)} \int_{\mu(B(\mathbf{x}_0; \delta) \cap E)} f(\mathbf{x}) \, d\mathbf{x} = f(\mathbf{x}_0).$$

Theorem 18.6 (First mean-value theorem for the integral). $f \in \mathbb{R}^E$, where E is a connected, bounded, Lebesgue measurable set in an n -D normed space. Let $m := \inf\{f(\mathbf{x}) \mid \mathbf{x} \in E\}$, $M := \sup\{f(\mathbf{x}) \mid \mathbf{x} \in E\}$. If $f \in C(E)$ and f is bounded on E , then $\exists \boldsymbol{\xi} \in E$ s.t.

$$\int_E f(\mathbf{x}) \, d\mathbf{x} = f(\boldsymbol{\xi}) \mu(E).$$

Proof. By Theorem 6.12, the image of E under continuous function f must be also connected, which, in \mathbb{R} , must be intervals (or a single point).

If $\lambda(E) = 0$, the theorem is trivial. Since:

$$m \leq \frac{1}{\lambda(E)} \int_E f(\mathbf{x}) \, d\mathbf{x} \leq M,$$

there must be a point in E where f takes value $\int_E f(\mathbf{x}) \, d\mathbf{x} / \lambda(E)$. □

Theorem 18.7 (Mean-value theorem for the integral). $f, g \in \mathbb{R}^E$, where E is a connected, bounded, Lebesgue measurable set in an n -D normed space. $f, g \in \mathfrak{R}(E)$. If a.e. $\mathbf{x} \in E$, $m \leq f(\mathbf{x}) \leq M$ and $g(\mathbf{x}) \geq 0$, we have s.t.

$$m \int_E g(\mathbf{x}) \, d\mathbf{x} \leq \int_E f(\mathbf{x}) g(\mathbf{x}) \, d\mathbf{x} \leq M \int_E g(\mathbf{x}) \, d\mathbf{x}.$$

Proof. The discontinuous points of $fg: \mathbf{x} \mapsto f(\mathbf{x})g(\mathbf{x})$ must be either the discontinuous points of f or g , which are both of measure zero. Therefore, $fg \in \mathfrak{R}(E)$.

$$mg(\mathbf{x}) \leq f(\mathbf{x})g(\mathbf{x}) \leq Mg(\mathbf{x}).$$

□

Theorem 18.8 (Limit of p -norm). *E is Jordan measurable, $\mu(E) > 0$. If $f \in C(E; \mathbb{R})$, $\forall \mathbf{x} \in E$, $f(\mathbf{x}) \geq 0$ and $\sup\{f(\mathbf{x}) \mid \mathbf{x} \in E\}$, then:*

$$\lim_{n \rightarrow \infty} \sqrt[n]{\int_E [f(\mathbf{x})]^n d\mathbf{x}} = M.$$

Proof. $\forall \varepsilon \in (0, M)$, since f is continuous, $\exists \delta \in \mathbb{R}^+$ and $\exists \mathbf{x}_0 \in E$ s.t. $B(\mathbf{x}_0; \delta) \subset E_\varepsilon = \{\mathbf{x} \mid f(\mathbf{x}) \geq M - \varepsilon\}$:

$$\sqrt[n]{\int_E [f(\mathbf{x})]^n d\mathbf{x}} \geq \sqrt[n]{\int_{B(\mathbf{x}_0; \delta)} (M - \varepsilon)^n d\mathbf{x}} = (M - \varepsilon) \sqrt[n]{\mu(B(\mathbf{x}_0; \delta))},$$

while

$$\sqrt[n]{\int_E [f(\mathbf{x})]^n d\mathbf{x}} \leq M \sqrt[n]{\mu(E)}.$$

Pass $n \rightarrow \infty$ and use the fact that ε is arbitrary. □

Theorem 18.9. *$f: E \rightarrow \mathbb{R}$, where E is a bounded set in \mathbb{R}^n , $f_+ := (|f| + f)/2$. If $\int_E f(\mathbf{x}) d\mathbf{x} > A$, then there exists a subset $F \subset E$ s.t. F consists of finite cuboids in \mathbb{R}^n , and:*

$$\int_F f(\mathbf{x}) d\mathbf{x} > A.$$

Proof. Let I be a cuboid that contains E . $\exists d \in \mathbb{R}_+$, $\forall P \in \mathfrak{R}(I) (\lambda(P) < d)$, Darboux lower sum $s(f_+, P) > A$.

$$\begin{aligned} s(f_+, P) &= \sum_{I_i \in P} \frac{\mu(I_i)}{2} \inf\{|f(\mathbf{x})| + f(\mathbf{x}) \mid \mathbf{x} \in I_i\} \\ &= \sum_{I'_i \in P'} \mu(I'_i) \inf\{f(\mathbf{x}) \mid \mathbf{x} \in I'_i\} = s(f|_{\cup P'}, P') > A, \end{aligned}$$

where P' can be considered as a partition of $\cup P'$, where f is non-negative. Let $F = \cup P'$

Hence:

$$\int_F f(\mathbf{x}) d\mathbf{x} \geq \int_{\underline{F}} f(\mathbf{x}) d\mathbf{x} > A.$$

□

§19 Fubini's Theorem

Let X be a cuboid in \mathbb{R}^n and Y be a cuboid in \mathbb{R}^m , $f \in \mathbb{R}^{X \times Y}$.

We use the following notation to denote one of the *iterated integrals*:

$$\int_X dx \int_Y f(x, y) dy,$$

which should be understood as the integral of $F(x)$ over X , where $F(x)$ is the integral of $f(x, y)$ over Y with fixed x . What if at some point $x_0 \in X$, $y \mapsto f(x_0, y) \notin \mathfrak{R}(Y)$? If the function f is bounded, we can assign a value to $F(x_0)$ between $\overline{\int}_Y f(x_0, y) dy$ and $\underline{\int}_Y f(x_0, y) dy$. The way of assignment can be arbitrary if $f \in \mathfrak{R}(X \times Y)$, which we might see in the following famous theorem:

Theorem 19.1 (Fubini's theorem). *Let X be a cuboid in \mathbb{R}^n and Y be a cuboid in \mathbb{R}^m , $f \in \mathbb{R}^{X \times Y}$. If $f \in \mathfrak{R}(X \times Y)$, then all three of the following integrals⁶:*

$$\int_{X \times Y} f(x, y) dx dy, \quad \int_X dx \int_Y f(x, y) dy, \quad \int_Y dy \int_X f(x, y) dx$$

exist and equal.

Proof. $P_X \in \mathfrak{P}(X)$, $P_Y \in \mathfrak{P}(Y)$. Let $N_X = \text{card } P_X$, $N_Y = \text{card } P_Y$. Let $P := \{I_X \times I_Y \mid I_X \in P_X \wedge I_Y \in P_Y\}$, which is a partition of $X \times Y$.

The Darboux lower sum:

$$\begin{aligned} s(f, P) &= \sum_{\substack{I_X \in P_X, \\ I_Y \in P_Y}} \mu_{\mathbb{R}^n}(I_X) \mu_{\mathbb{R}^m}(I_Y) \inf\{f(x, y) \mid x \in I_X \wedge y \in I_Y\} \\ &\leq \sum_{\substack{I_X \in P_X, \\ I_Y \in P_Y}} \mu_{\mathbb{R}^n}(I_X) \mu_{\mathbb{R}^m}(I_Y) \inf\left\{\inf\{f(x, y) \mid y \in I_Y\} \mid x \in I_X\right\} \\ &= \sum_{I_X \in P_X} \mu_{\mathbb{R}^n}(I_X) \inf\left\{\sum_{I_Y \in P_Y} \mu_{\mathbb{R}^m}(I_Y) \inf\{f(x, y) \mid y \in I_Y\} \mid x \in I_X\right\} \\ &= \sum_{I_X \in P_X} \mu_{\mathbb{R}^n}(I_X) \inf\left\{\int_{\underline{Y}} f(x, y) dy \mid x \in I_X\right\} \leq \sum_{I_X \in P_X} \mu_{\mathbb{R}^n}(I_X) \inf\{F(x) \mid x \in I_X\} \\ &\leq \sum_{I_X \in P_X} \mu_{\mathbb{R}^n}(I_X) \sup\{F(x) \mid x \in I_X\} \\ &\leq \sum_{\substack{I_X \in P_X, \\ I_Y \in P_Y}} \mu_{\mathbb{R}^n}(I_X) \mu_{\mathbb{R}^m}(I_Y) \sup\{f(x, y) \mid x \in I_X \wedge y \in I_Y\} = S(f, P). \end{aligned}$$

As $\lambda(P) \rightarrow 0$, $s(f, P)$ and $S(f, P)$ must converge to the multiple integral, hence the upper and lower Darboux integrals of F are equal, hence $F \in \mathfrak{R}(X)$. \square

⁶The first one is known as the *multiple integral*.

Corollary 9. Let X be a cuboid in \mathbb{R}^n and Y be a cuboid in \mathbb{R}^m , $f \in \mathbb{R}^{X \times Y}$. If $f \in \mathfrak{R}(X \times Y)$, then $\mathbf{x} \mapsto f(\mathbf{x}, \mathbf{y}) \in \mathfrak{R}(X)$ a.e. $\mathbf{y} \in Y$, $\mathbf{y} \mapsto f(\mathbf{x}, \mathbf{y}) \in \mathfrak{R}(Y)$ a.e. $\mathbf{x} \in X$.

Proof. By the Fubini's theorem 19.1, whatever value we assign to $F(\mathbf{x})$ when $\mathbf{y} \mapsto f(\mathbf{x}, \mathbf{y}) \notin \mathfrak{R}(Y)$, the multiple integrals are the same.

Hence:

$$\int_X \left(\int_{\underline{Y}} f(\mathbf{x}, \mathbf{y}) d\mathbf{y} \right) d\mathbf{x} = \int_X \left(\overline{\int_Y f(\mathbf{x}, \mathbf{y}) d\mathbf{y}} \right) d\mathbf{x}, \Rightarrow \int_X \left(\overline{\int_Y f(\mathbf{x}, \mathbf{y}) d\mathbf{y}} - \int_{\underline{Y}} f(\mathbf{x}, \mathbf{y}) d\mathbf{y} \right) d\mathbf{x}.$$

Hence, the upper and lower integrals of $\mathbf{y} \mapsto f(\mathbf{x}, \mathbf{y})$ are equal a.e. $\mathbf{x} \in X$. \square

Theorem 19.2. Let $\varphi_1, \varphi_2 \in \mathbb{R}^X$ be bounded functions on X , where X is a bounded set in \mathbb{R}^n . Let $E := \{(\mathbf{x}, y) \mid \mathbf{x} \in X \wedge y \in [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]\}$. $f \in \mathbb{R}^E$.

If $f \in \mathfrak{R}(E)$, then:

$$\int_E f(\mathbf{x}, y) d\mathbf{x} dy := \int_X d\mathbf{x} \int_{\varphi_1(\mathbf{x})}^{\varphi_2(\mathbf{x})} f(\mathbf{x}, y) dy.$$

Proof. Find a cuboid $I := I_X \times I_Y$ that is large enough to contain E , we shall see $X \subset I_X, \forall \mathbf{x} \in X, [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})] \subset I_Y$.

$$\begin{aligned} \int_E f(\mathbf{x}, y) d\mathbf{x} dy &= \int_I \chi_E(\mathbf{x}, y) f(\mathbf{x}, y) d\mathbf{x} dy = \int_{I_X} d\mathbf{x} \int_{I_Y} \chi_X(\mathbf{x}) \chi_{[\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]}(y) f(\mathbf{x}, y) dy \\ &= \int_{I_X} \chi_X(\mathbf{x}) d\mathbf{x} \int_{I_Y} \chi_{[\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]}(y) f(\mathbf{x}, y) dy = \int_X d\mathbf{x} \int_{\varphi_1(\mathbf{x})}^{\varphi_2(\mathbf{x})} f(\mathbf{x}, y) dy. \end{aligned}$$

\square

Theorem 19.3. Let $\varphi_1, \varphi_2 \in C(X)$ be bounded and continuous real-valued functions on X , where X is a bounded set in \mathbb{R}^n . Let $E := \{(\mathbf{x}, y) \mid \mathbf{x} \in X \wedge y \in [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x})]\}$.

If X is Jordan measurable, E is also Jordan measurable, and its volume can be formulated as:

$$\mu(E) = \int_X [\varphi_2(\mathbf{x}) - \varphi_1(\mathbf{x})] d\mathbf{x}.$$

Proof. First we examine the component of ∂E , which is a compact set in \mathbb{R}^n .

$$\partial E = \bigcup_{i \in \{1, 2\}} \{(\mathbf{x}, \varphi_i(\mathbf{x})) \mid \mathbf{x} \in X \wedge \varphi_1(\mathbf{x}) \leq \varphi_2(\mathbf{x})\} \cup (\partial X \times [\varphi_1(\partial X), \varphi_2(\partial X)]).$$

$\forall i \in \{1, 2\}$: since φ_i is continuous and bounded on \overline{X} , which is a closed and bounded (therefore compact) set, by Theorem 6.10, φ_i is also uniformly continuous on it. Then, $\forall \varepsilon \in \mathbb{R}_+$, we can find a δ , so that $\{(\mathbf{x}, \varphi_i(\mathbf{x})) \mid \mathbf{x} \in X\}$ is covered by a finite collection of products of a cuboid with side less than δ and an interval with length ε , where the cuboids have no common interior points

pairwisely. Let I be a cuboid that large enough to cover X , the total volume must less than $\mu(I)\varepsilon$. Hence: $\{(\mathbf{x}, \varphi_i(\mathbf{x})) \mid \mathbf{x} \in X\}$ is of zero measure.

Since φ_i , $i = 1, 2$ are bounded and continuous, they must be also bounded on the boundary of X , hence there exists a $M \in \mathbb{R}_+$ s.t. $\varphi_2(\mathbf{x}) - \varphi_1(\mathbf{x}) < M$ when $\mathbf{x} \in \partial X$ and $\varphi_2(\mathbf{x}) \geq \varphi_1(\mathbf{x})$. Hence, $\forall \varepsilon \in \mathbb{R}_+$, there exists a finite cover made of cuboids, with total volume less than ε/M . Therefore: $\partial X \times [\varphi_1(\partial X), \varphi_2(\partial X)]$ is of zero measure. \square

§20 Change of Variables

In this section the following formula will be establish:

$$\int_{\varphi(D_t)} f(\mathbf{x}) d\mathbf{x} = \int_{D_t} f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| d\mathbf{t}. \quad (20-1)$$

To get rid of some boundary problem, we want the formula to be applied on some compact set, however a linear approximation is only valid in an open set, hence we introduce:

Definition 20.1 (Support). The **support** of a function $f \in Y^X$ is defined as:

$$\text{supp } f := \overline{\{\mathbf{x} \in X \mid f(\mathbf{x}) \neq \mathbf{0}\}}.$$

Lemma 11. Let X be an n -D normed space, U be a bounded open subset of X . There exists a collection of cuboids P s.t. the cuboids in P have no common interior points pairwisely, and

$$U = \bigcup_{I \in P} I.$$

Proof. We divide the n -D normed space into grids with gap $1/2^k$, collect those cubes that fully contained in $U - \cup P_{k-1}$, and add it to P_{k-1} to get P_k . $P = \bigcup_{k \in \mathbb{N}} P_k$ is what we want. \square

Lemma 12 (Conservation of zero-measure under diffeomorphism). Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between open sets in X , where X is an n -D normed space. If $E_t \subset D_t$ is of measure zero, its image $\varphi(E_t)$ is also of measure zero.

Proof. Let P be a collection of cuboids that their union is D_t and they have no common interior points pairwisely (Lemma 11). Now consider $\forall I \in P$, let $E_t(I) := E_t \cap I$. As a subset of E_t , which is of measure zero, $E_t(I)$ is also of measure zero.

$\varphi \in C^{(1)}(I)$, where I is a compact set, hence $\exists M \in \mathbb{R}_+$ s.t. $\forall \mathbf{t} \in I$, $\|\varphi'(\mathbf{t})\| < M$ (Theorem 6.8). By finite-increment theorem (Theorem 11.1), $\forall \mathbf{t}_1, \mathbf{t}_2 \in I$, $\|\varphi(\mathbf{t}_1) - \varphi(\mathbf{t}_2)\|_X \leq M \|\mathbf{t}_1 - \mathbf{t}_2\|_X$.

From any cover Ω of $E_t(I)$, a cover $\Omega(I) := \{O \cap I \mid O \in \Omega\}$ can be constructed. Hence, we can assume that $\Omega = \{I_i \mid i \in \mathbb{N}\} \subset 2^I$, where $E_t(I) \subset \cup \Omega$ and $\sum_{i \in \mathbb{N}} \mu(I_i) < \varepsilon$. Since $\varphi(I_i)$ can be contained in a cuboids that is centered at $\varphi(\mathbf{t}_i)$ where \mathbf{t}_i is the centre of I_i , and with sides of M times, we claim: $\varphi(E_t(I))$ is also of measure zero.

As the countable union of $\varphi(E_t(I))$ ($I \in P$), $\varphi(E_t)$ is also of measure zero. \square

Lemma 13 (Conservation of zero-content under diffeomorphism). Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between open sets in X , where X is an n -D normed space. If $\bar{E}_t \subset D_t$, and E_t is of zero content, its image $\varphi(E_t)$ is also of zero content.

Proof. As the boundary of a set with zero content, ∂E_t must also be of zero content. Hence, compact set $\overline{E_t}$ is also of zero content, therefore of zero measure. By Lemma 12, $\varphi(\overline{E_t}) = \overline{\varphi(E_t)}$ is a compact set of zero measure, therefore of zero content.

As a subset of $\varphi(\overline{E_t})$, $\varphi(E_t)$ is also of zero content. \square

Lemma 14 (Conservation of Jordan-measurability under diffeomorphism). *Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between open sets in X , where X is an n -D normed space. If $\overline{E_t} \subset D_t$, and E_t is Jordan measurable, its image $\varphi(E_t)$ is also Jordan measurable.*

Proof. One must have $\varphi(\partial E_t) = \partial \varphi(E_t)$ since φ is a diffeomorphism (therefore a homeomorphism). \square

Now we are in the position to discuss the existence of integrals at the right hand side of Eq. 20-1.

Lemma 15. *Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between bounded open sets in X , where X is an n -D normed space. If $f \in \mathfrak{R}(D_x)$, and $\text{supp } f$ is a compact set in D_x , then $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \in \mathfrak{R}(D_t)$.*

Proof. Let $\psi(\mathbf{t}) := f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})|$.

Since φ is a diffeomorphism, φ^{-1} must also be differentiable, hence $\varphi'(\mathbf{t})$ must be reversible, therefore with non-zero determinant. Hence: $\text{supp } \psi = \text{supp } f \circ \varphi = \varphi^{-1}(\text{supp } f)$. $\varphi^{-1}(\text{supp } f)$ is a closed set in D_t , hence $\text{supp } \psi$ is a compact set.

Note that the discontinuous points must be contained in the support of the function (the complement of the support is an open set where the value of the function is always a constant: zero). The discontinuous points of $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \chi_{D_t}(\mathbf{t})$ must be all preimage of the discontinuous points of f under φ . Since $f \in \mathfrak{R}(D_x)$, by Lemma 12, the set of these points are also of zero measure. \square

First, we need to prove Eq. 20-1 in 1-D space.

Theorem 20.1 (Change of variables in 1-D space). *Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between bounded open sets in \mathbb{R} . If $f \in \mathfrak{R}(D_x)$, and $\text{supp } f$ is a compact set in D_x , then*

$$\int_{D_x} f(x) dx = \int_{D_t} f \circ \varphi(t) |\varphi'(t)| dt. \quad (20-2)$$

Proof. Since D_x is open, $\forall x \in \text{supp } f$, $\exists \delta \in \mathbb{R}_+$, $B(x; \delta) \subset D_x$. The collection of $B(x; \delta)$ is therefore an open cover of the compact set $\text{supp } f$, hence there exists a finite subcover. Note that $B(x; \delta)$ is an interval in \mathbb{R} , and any intersection of two intervals is either empty or another interval. Hence, we can find a finite collection of closed intervals (as the closure of open intervals), the intervals of which have no common interior points pairwise, and the union of the collection contains $\text{supp } f$.

Assume one of the intervals is $I_x = [a, b]$. In 1-D space, diffeomorphism maps intervals into intervals, hence any partition P_x of I_x would induce a partition P_t of $I_t = \varphi^{-1}(I_x)$.

Let $P_x := \{[x_i, x_{i+1}] \mid i \in N\}$ where $x_0 = a$, $x_N = b$, and $x_i < x_{i+1}$. Let $P_t = \{[t_i, t_{i+1}] := \varphi^{-1}([x_i, x_{i+1}]) \mid i \in N\}$ ⁷. By Lagrange's finite-increment theorem (Theorem 13.7), $\forall i \in N$, $\exists \tau_i \in [t_i, t_{i+1}]$ s.t.

$$|x_{i+1} - x_i| = |\varphi(t_{i+1}) - \varphi(t_i)| = |\varphi'(\tau_i)| |t_{i+1} - t_i|.$$

⁷Be careful! φ could be decreasing.

Let $\xi_i = \varphi(\tau_i)$, and denote $\psi(t) := f \circ \varphi(t)|\varphi'(t)|$. The Riemann sum:

$$\sigma(f, P_x, \xi) = \sum_{i \in N} f(\xi_i)|x_{i+1} - x_i| = \sum_{i \in N} f \circ \varphi(\tau_i)|\varphi'(\tau_i)||t_{i+1} - t_i| = \sigma(\psi, P_t, \tau).$$

Since $\psi \in \mathfrak{R}(I_x)$, the limit of $\sigma(\psi, P_t, \tau)$ is the integral, and by the fact that $\lambda(P_x) \rightarrow 0$ iff $\lambda(P_t) \rightarrow 0$, we have the equation:

$$\int_{I_x} f(x) dx = \int_{I_t} f \circ \varphi(t)|f'(t)| dt.$$

Summing all I_x (finite), we have Eq. 20-2. □

In a finite-dimensional space V over \mathbb{R} , vectors can be represented as coordinates, in another word, V and $\mathbb{R}^{\dim V}$ are isomorphic.

Definition 20.2 (Elementary diffeomorphism). An *elementary diffeomorphism* $\varphi: D_t \mapsto D_x$ is defined as a diffeomorphism, being a change of one of the coordinates in a given basis, i.e.

$$x_i = \varphi(\mathbf{t}) = \begin{cases} t_i, & i \neq k; \\ \varphi_k(\mathbf{t}), & i = k. \end{cases}$$

Theorem 20.2 (Change of variables for elementary diffeomorphism). Let $\varphi: D_t \rightarrow D_x$ be an elementary diffeomorphism between bounded open sets in \mathbb{R}^n . If $f \in \mathfrak{R}(D_x)$, and $\text{supp } f$ is a compact set in D_x , then

$$\int_{D_x} f(x) dx = \int_{D_t} f \circ \varphi(t)|\det \varphi'(t)| dt.$$

Proof. Use Theorem 20.1 and Fubini's theorem (Theorem 19.1). □

Theorem 20.3 (Change of variables for composite diffeomorphism). Let $\varphi: D_t \rightarrow D_x$ and $\psi: D_\tau \rightarrow D_t$ be two elementary diffeomorphisms between bounded open sets in \mathbb{R}^n . If Eq. 20-1 holds for both φ and ψ (for any $f \in \mathfrak{R}(D_x)$ or $f \in \mathfrak{R}(D_t)$), then $\forall f \in \mathfrak{R}(D_\tau)$:

$$\int_{D_x} f(\mathbf{x}) d\mathbf{x} = \int_{D_\tau} f \circ \varphi \circ \psi(\boldsymbol{\tau})|\det(\varphi \circ \psi)'(\boldsymbol{\tau})| d\boldsymbol{\tau}$$

Lemma 16 (Decomposition diffeomorphism). Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between bounded open sets in \mathbb{R}^n . $\forall \mathbf{t} \in D_t$, $\exists U \in \mathcal{U}(\mathbf{t})$, where φ can be considered as a composition of elementary diffeomorphisms.

Proof. The proof would be taken assuming $n = 2$. For $n > 2$, the proof can be given by induction.

First we prove the existence of a neighbourhood U , where the change of the first coordinates $\psi: (t_0, t_1) \mapsto (\varphi_0(t_0, t_1), t_1)$ is a diffeomorphism.

Consider $F(\mathbf{t}, \mathbf{x}) := \psi(\mathbf{t}) - \mathbf{x}$. Notice that $F(\mathbf{t}, \mathbf{x})$ must be continuously differentiable. Hence $F(\mathbf{t}, \mathbf{x}) = \mathbf{0}$ or $\psi(\mathbf{t}) = \mathbf{x}$ yields a function $\mathbf{t} = \psi^{-1}(\mathbf{x})$ in some neighbourhood, which is also continuously differentiable.

Then, consider $\varphi \circ \psi^{-1}$, which is a composition of two diffeomorphism, hence must also be a diffeomorphism, and we can see it is elementary. Therefore we have: $\varphi = (\varphi \circ \psi^{-1}) \circ \psi$. □

Theorem 20.4 (Change of variables for diffeomorphism). *Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between bounded open sets in \mathbb{R}^n . If $f \in \mathfrak{R}(D_x)$, and $\text{supp } f$ is a compact set in D_x , then*

$$\int_{D_x} f(\mathbf{x}) \, d\mathbf{x} = \int_{D_t} f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \, d\mathbf{t}.$$

Proof. $\forall \mathbf{x} \in K_t := \varphi^{-1}(\text{supp } f)$, $\exists \delta(\mathbf{t}) \in \mathbb{R}_+$ s.t. φ decomposes into elementary diffeomorphisms in $B(\mathbf{t}, \delta(\mathbf{t}))$. It is clear that the collection of $B(\mathbf{t}; \delta(\mathbf{t})/2)$ must be a open cover of K_t , hence there exists a finite subcover $\Omega := \{B(\mathbf{t}_i; \delta(\mathbf{t}_i)/2) \mid i \in m\}$. Let $\delta := \min\{\delta(\mathbf{t}_i)/2 \mid i \in m\}$. Now if a set with diameter less than δ intersects with K_t , it would be contained in a $B(\mathbf{t}_i; \delta(\mathbf{t}_i))$ where $\delta(\mathbf{t}_i) \geq 2\delta$.

Let $d := d(K_t, \partial D_t)$ ⁸.

Find a cuboid I that contains D_t . $P \in \mathfrak{P}(I)$ and $\lambda(P) < \min\{\delta, d\}$. Let $P' := \{I_i \mid I_i \cap K_t \neq \emptyset \wedge I_i \in P, i \in N\}$.

$$\int_{D_t} f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \, d\mathbf{t} = \sum_{i \in N} \int_{I_i} f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \, d\mathbf{t}.$$

Now we know that φ decomposes into elementary diffeomorphisms in I_i , which is contained in some $B(\mathbf{t}_i; \delta(\mathbf{t}_i))$. Use Theorem 20.1, Theorem 19.1. \square

The corollary below might be more partially convenient:

Corollary 10. *Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between bounded open sets in \mathbb{R}^n . $\overline{E_x} \subset D_x$, $\varphi^{-1}(E_x) = E_t$. If $f \in \mathfrak{R}(E_x)$, then*

$$\int_{E_x} f(\mathbf{x}) \, d\mathbf{x} = \int_{E_t} f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \, d\mathbf{t}.$$

Theorem 20.5 (Change of variables). *Let D_t, D_x be two Jordan measurable sets in \mathbb{R}^n , $\varphi \in D_x^{D_t}$. $S_t \subset D_t$ and $S_x \subset D_x$ are of measure zero. If $D_t \setminus S_t$ and $D_x \setminus S_x$ are open, $\varphi|_{D_t \setminus S_t}$ is a diffeomorphism from $D_t \setminus S_t$ to $D_x \setminus S_x$, and $\exists M \in \mathbb{R}_+$ s.t. $\forall \mathbf{t} \in D_t \setminus S_t$, $|\det(\varphi|_{D_t \setminus S_t})'(\mathbf{t})| < M$; then, $\forall f \in \mathfrak{R}(D_x)$, $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \in \mathfrak{R}(D_t \setminus S_t)$, and,*

$$\int_{D_x} f(\mathbf{x}) \, d\mathbf{x} = \int_{D_t \setminus S_t} f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \, d\mathbf{t}.$$

If also $|\det \varphi'|$ exists and bounded on D_t , we have:

$$\int_{D_x} f(\mathbf{x}) \, d\mathbf{x} = \int_{D_t} f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \, d\mathbf{t}.$$

Proof. By Lemma 12, the discontinuous points of $f \circ \varphi$ is of zero measure. Now we prove that $D_t \setminus S_t$ is also Jordan measurable.

⁸ d must be positive since K_t is a closed set contained in the open set D_t .

Since $D_x \setminus S_x$ is open, every point of $D_x \setminus S_x$ has a neighbourhood of it contained in $D_x \setminus S_x$, therefore disjoint from S_x , hence by the definition of boundary: $(D_x \setminus S_x) \cap \partial S_x = \emptyset$. Recall that $S_x \subset D_x$, hence $\partial S_x \subset \overline{D_x}$. Therefore,

$$\partial S_x \subset \overline{D_x} \cap \mathbb{G}_{\mathbb{R}^n}(D_x \setminus S_x) = \overline{D_x} \cap (\mathbb{G}_{\mathbb{R}^n} D_x \cup S_x) \subset \partial D_x \cup S_x.$$

Hence, $\partial D_x \cup S_x = \partial D_x \cup \overline{S_x}$, in another word, $\partial D_x \cup S_x$ is a compact set. As a union of sets of zero measure, $\partial D_x \cup S_x$ is of zero measure too, hence it is of zero content (Theorem 15.9). Since $\partial(D_x \setminus S_x) \subset \partial D_x \cup S_x$, $D_x \setminus S_x$ is Jordan measurable. Similarly we can prove that $D_t \setminus S_t$ is also Jordan measurable.

By Lebesgue's criterion 16.2, $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \in \mathfrak{R}(D_t \setminus S_t)$.

Next we need to find a subset of $D_x \setminus S_x$ so that we can use Corollary 10.

$\forall \varepsilon \in \mathbb{R}_+$, we can find a finite collection $\{I_i \mid i \in N\}$ of cuboids that covers $\partial D_x \cup S_x$, and $\forall \mathbf{x} \in \partial(D_x \setminus S_x)$, \mathbf{x} locates interiorly in at least one cuboid in the collection, and $\sum_{i \in N} \mu(I_i) < \varepsilon$. Let $U_x := \bigcup_{i \in N} I_i$ and $V_x := D_x \setminus U_x$.

Any Jordan measurable set E_x that contains $\overline{V_x}$ and is contained in D_x must have:

$$\left| \int_{D_x} f(\mathbf{x}) \, d\mathbf{x} - \int_{E_x} f(\mathbf{x}) \, d\mathbf{x} \right| = \left| \int_{D_x \setminus E_x} f(\mathbf{x}) \, d\mathbf{x} \right| \leq M \mu(D_x \setminus E_x) < M\varepsilon,$$

where $M = \sup\{f(\mathbf{x}) \mid \mathbf{x} \in D_x\}$.

Let $E_t = \varphi^{-1}(E_x)$. By Corollary 10 and the arbitrariness of ε , we have Eq. 20-1.

If $|\det \varphi'|$ exists and bounded on D_t : The discontinuous points of $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})|$ on D_t are made up by the discontinuous points in $D_t \setminus E_t$, which we have proved is of zero measure, and a subset of $S_t \cup \partial D_t$, which is also of zero measure. Hence: $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t}) |\det \varphi'(\mathbf{t})| \in \mathfrak{R}(D_t)$. \square

§21 Improper Integral

Definition 21.1 (Exhaustion). Let $\langle E_n \rangle_{n \in \mathbb{N}}$ be an increasing sequence of sets i.e. $\forall n, m \in \mathbb{N}$, $n \leq m \rightarrow E_n \subseteq E_m$. We denote $E_n \uparrow E := \bigcup_{n \in \mathbb{N}} E_n$, and say that $\langle E_n \rangle_{n \in \mathbb{N}}$ is an *exhaustion* of the set E .

Theorem 21.1. Let $\langle E_n \rangle_{n \in \mathbb{N}}$ be an Jordan measurable exhaustion of a Jordan measurable set E .

1. $\lim_{n \rightarrow \infty} \mu(E_n) = \mu(E)$;
2. $\forall f \in \mathfrak{R}(E)$, $\forall n \in \mathbb{N}$, $f|_{E_n} \in \mathfrak{R}(E_n)$, and

$$\lim_{n \rightarrow \infty} \int_{E_n} f(\mathbf{x}) \, d\mathbf{x} = \int_E f(\mathbf{x}) \, d\mathbf{x}.$$

Proof. By $\mu(E_n) \leq \mu(E)$ we have $\lim_{n \rightarrow \infty} \mu(E_n) \leq \mu(E)$. $\forall \varepsilon \in \mathbb{R}_+$, find ∂E a cover Ω by open cuboids, where $\sum_{I \in \Omega} \mu(I) < \varepsilon$. For $\varepsilon/2^n$, repeat this for ∂E_n to get Ω_n . Let $\tilde{E} := E \cup (\cup \Omega)$ and $\tilde{E}_n := E_n \cup (\cup \Omega)$.

$\{\cup \Omega\} \cup \{\tilde{E}_n \mid n \in \mathbb{N}\}$ is an open cover of the compact set \bar{E} , hence we can find it a finite subcover⁹: $\{\cup \Omega\} \cup \{E_i \mid i \in N\}$. Or: $\{\cup \Omega, E_{N-1}\} \cup \{\cup \Omega_i \mid i \in N\}$ is also a finite open cover of \bar{E} . Hence:

$$\mu(E) \leq \mu(\bar{E}) \leq \mu(E_{N-1}) + \mu(\cup \Omega) + \sum_{i \in N} \mu(\cup \Omega_i) < \mu(E_{N-1}) + \varepsilon + \sum_{i \in N} \varepsilon/2^i < \mu(E_{N-1}) + 3\varepsilon.$$

In conclusion: $\forall \varepsilon \in \mathbb{R}_+, \exists N \in \mathbb{N}, \mu(E) < \mu(E_N) + 3\varepsilon$. Hence: $\mu(E) = \sup\{\mu(E_n) \mid n \in \mathbb{N}\}$, or:

$$\lim_{n \rightarrow \infty} \mu(E_n) = \mu(E).$$

Since f is Riemann integrable, $\exists M \in \mathbb{R}_+$ s.t. $\forall \mathbf{x} \in E, |f(\mathbf{x})| < M$.

$$\left| \int_E f(\mathbf{x}) \, d\mathbf{x} - \int_{E_n} f(\mathbf{x}) \, d\mathbf{x} \right| = \left| \int_{E \setminus E_n} f(\mathbf{x}) \, d\mathbf{x} \right| \leq M \mu(E \setminus E_n) = M(\mu(E) - \mu(E_n)).$$

Passing $n \rightarrow \infty$, we have:

$$\left| \int_E f(\mathbf{x}) \, d\mathbf{x} - \lim_{n \rightarrow \infty} \int_{E_n} f(\mathbf{x}) \, d\mathbf{x} \right| = 0.$$

□

Definition 21.2 (Improper integral). $f: E \mapsto Y$, where Y is a normed space. If $\exists I \in Y$ s.t. \forall Jordan measurable exhaustion $\langle E_n \rangle_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ that $\forall n \in \mathbb{N}, f|_{E_n} \in \mathfrak{R}(E_n)$,

$$I = \lim_{n \rightarrow \infty} \int_{E_n} f(\mathbf{x}) \, d\mathbf{x},$$

we call I the *improper integral* of f over E , denoted by:

$$\int_E f(\mathbf{x}) \, d\mathbf{x} := I.$$

The following statement explains the reason why we can use the same notation as for proper integral (notice that we don't assume that E is Jordan measurable, as in Theorem 21.1):

Theorem 21.2 (Agreement of improper integrals with (proper) integrals). $f: E \mapsto Y$, where Y is a normed space. If $f \in \mathfrak{R}(E)$, then the improper integral of f over E converges, and equals to the integral of f over E .

Proof. Find a cuboid I that contains E , and let $\langle E_n \rangle_{n \in \mathbb{N}}$ be a Joran measurable exhaustion of E . Let D_E be the discontinuous points of $\mathbf{x} \mapsto \chi_E(\mathbf{x})f(\mathbf{x})$. Since $f \in \mathfrak{R}(E)$, D_E must be of measure zero.

⁹Notice that E_n is monotone.

Let D_{E_n} be the discontinuous points of $\mathbf{x} \mapsto \chi_{E_n}(\mathbf{x})f(\mathbf{x})$. $D_{E_n} \subset D_E \cup \partial E_n$, while ∂E_n is of measure zero since E_n is Jordan measurable. Hence $f \in \mathfrak{R}(E_n)$.

Notice:

$$\int_E f(\mathbf{x}) \, d\mathbf{x} - \int_{E_n} f(\mathbf{x}) \, d\mathbf{x} = \int_I (\chi_E(\mathbf{x}) - \chi_{E_n}(\mathbf{x}))f(\mathbf{x}) \, d\mathbf{x} = \int_{E \setminus E_n} f(\mathbf{x}) \, d\mathbf{x}.$$

The remaining part has been done in the proof of Theorem 21.1 □

Theorem 21.3 (Convergence of improper integral for non-negative function). *If $f: E \mapsto \mathbb{R}$ is non-negative, and \exists an exhaustion $\langle E_n \rangle_{n \in \mathbb{N}} \in E^{\mathbb{N}}$ s.t. $\forall n \in \mathbb{N}$, $f|_{E_n} \in \mathfrak{R}(E_n)$, and*

$$I = \lim_{n \rightarrow \infty} \int_{E_n} f(\mathbf{x}) \, d\mathbf{x}$$

exists, then the improper integral of f over E exists and equals to I .

Proof. If there were another exhaustion $\langle E'_n \rangle_{n \in \mathbb{N}}$ where f is integrable over each E'_n : $\langle E'_n \cap E_k \rangle_{k \in \mathbb{N}}$ is an exhaustion of E'_n , hence:

$$\int_{E'_n} f(\mathbf{x}) \, d\mathbf{x} = \lim_{k \rightarrow \infty} \int_{E'_n \cap E_k} f(\mathbf{x}) \, d\mathbf{x} \leq \lim_{k \rightarrow \infty} \int_{E_k} f(\mathbf{x}) \, d\mathbf{x} = I.$$

A monotonely increasing¹⁰ bounded sequence must be convergent:

$$I' = \lim_{n \rightarrow \infty} \int_{E'_n} f(\mathbf{x}) \, d\mathbf{x} \leq I.$$

$$I = \lim_{k \rightarrow \infty} \int_{E_k} f(\mathbf{x}) \, d\mathbf{x} = \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \int_{E'_n \cap E_k} f(\mathbf{x}) \, d\mathbf{x} \leq I',$$

hence $I = I'$. □

Theorem 21.4 (Comparison criterion). *Let E be a set over which the improper integral of $g \in \mathbb{R}^E$ converges. $f \in \mathbb{R}^E$. If for each Jordan measurable exhaustion $\langle E_n \rangle_{n \in \mathbb{N}}$ of E where g is integrable over E_n ($\forall n \in \mathbb{N}$), f is also integrable over E_n for any $n \in \mathbb{N}$, and $\forall \mathbf{x} \in E$, $|f(\mathbf{x})| \leq g(\mathbf{x})$, then:*

$$\int_E |f(\mathbf{x})| \, d\mathbf{x}, \quad \int_E f(\mathbf{x}) \, d\mathbf{x}$$

converge (as improper integrals).

Proof. A function is integrable over two Jordan measurable sets, then it is integrable over the difference of them.

¹⁰ f is non-negative.

$\forall \varepsilon \in \mathbb{R}_+, \exists N \in \mathbb{N}$ s.t. if $n > m > N$, then:

$$\int_{E_n} |f(\mathbf{x})| d\mathbf{x} - \int_{E_m} |f(\mathbf{x})| d\mathbf{x} = \int_{E_n \setminus E_m} |f(\mathbf{x})| d\mathbf{x} \leq \int_{E_n \setminus E_m} |g(\mathbf{x})| d\mathbf{x} < \varepsilon.$$

Hence, by Cauchy's criterion, we know the convergence of the improper integral of $\mathbf{x} \mapsto |f(\mathbf{x})|$ over E .

Let $f_+ := (|f| + f)/2$, $f_- := (|f| - f)/2$. We can verify that both f_+ and f_- are non-negative, and since they are $\leq |f|$, we conclude that the improper integral of f_+ and f_- converges. By the fact that $f = f_+ - f_-$, we know the convergence of the improper integral of f over E . \square

The following theorem tells us, the convergence of f and $|f|$ is in fact equivalent, which is different from the improper integrals defined as the limit of integrals over cuboids as the boundary moves towards infinity:

Theorem 21.5 (Equivalence of convergence and absolute convergence). *Let D be an open set in \mathbb{R}^n , and f is integrable over any compact subset of D . The improper integral of f over D converges \leftrightarrow the improper integral of $|f|$ over D converges.*

Proof. \leftarrow : we have proved in the proof of Theorem 21.4.

\rightarrow : Assume that the improper integral of $|f|$ over D diverges while f doesn't. By Theorem 21.3, if the improper integral of $|f|$ diverges, then for any compact and Jordan measurable exhaustion $\langle E_n \rangle_{n \in \mathbb{N}}$ over which $|f|$ is integrable,

$$\lim_{n \rightarrow \infty} \int_{E_n} |f(\mathbf{x})| d\mathbf{x} = \infty.$$

Let $f_+ := (|f| + f)/2$, $f_- := (|f| - f)/2$. It can be verified that the improper integrals of f_+ and f_- must also diverge to ∞ . $\forall n \in \mathbb{N}, \exists N \in \mathbb{N}$ s.t.

$$\int_{E_N} f_+(\mathbf{x}) d\mathbf{x} > \int_{E_n} f_+(\mathbf{x}) d\mathbf{x} + \int_{E_n} |f(\mathbf{x})| d\mathbf{x} + n.$$

Now rechoose $\langle E_n \rangle_{n \in \mathbb{N}}$ so that $E_{n+1} := E_N$, repeatedly do it so that we have a Jordan measurable exhaustion of E that satisfies $\forall n \in \mathbb{N}$:

$$\int_{E_{n+1} \setminus E_n} f_+(\mathbf{x}) d\mathbf{x} > \int_{E_n} |f(\mathbf{x})| d\mathbf{x} + n.$$

Now by Theorem 18.9, we have $F_n \subset E_{n+1} \setminus E_n$ s.t. F_n consists of finite cuboids and

$$\int_{F_n} f(\mathbf{x}) d\mathbf{x} > \int_{E_n} |f(\mathbf{x})| d\mathbf{x} + n.$$

Let $G_n = F_n \cup E_n$. $\bigcup_{n \in \mathbb{N}} G_n = E$, $G_n \subset G_{n+1}$, $\forall n \in \mathbb{N}$. Hence:

$$\int_{G_n} f(\mathbf{x}) d\mathbf{x} > \int_{E_n} (|f(\mathbf{x})| + f(\mathbf{x})) d\mathbf{x} + n \geq n,$$

therefore: the improper integral of f over E diverges, which contradicts to our previous assumption. \square

Theorem 21.6 (Change of variables for diffeomorphism (improper integrals)). *Let $\varphi: D_t \rightarrow D_x$ be a diffeomorphism between open sets in \mathbb{R}^n . If f is integrable over any Jordan measurable compact subset of D_x , the improper integral of f over D_x converges, then, the improper integral of $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t})|\det \varphi'(\mathbf{t})|$ over D_t also converges, and*

$$\int_{D_x} f(\mathbf{x}) d\mathbf{x} = \int_{D_t} \varphi(\mathbf{t})|\det \varphi'(\mathbf{t})| d\mathbf{t}.$$

Proof. By Lemma 11, we can find an exhaustion $\langle E_{x,n} \rangle_{n \in \mathbb{N}}$ of D_x , where $E_{x,n}$ is the union of finite amount of cuboids (hence Jordan measurable and compact).

Hence we have:

$$\int_{E_{x,n}} f(\mathbf{x}) d\mathbf{x} = \int_{E_{t,n}} \varphi(\mathbf{t})|\det \varphi'(\mathbf{t})| d\mathbf{t},$$

where $E_{t,n} = \varphi^{-1}(E_{x,n})$.

Passing $n \rightarrow \infty$ and by the condition, the left-hand side converges. By Theorem 21.5 and Theorem 21.3, this implies the convergence of the improper integral of $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t})|\det \varphi'(\mathbf{t})|$ over D_t and its equality with f 's. \square

With same technique, we have:

Theorem 21.7 (Change of variables (improper integrals)). *Let D_t, D_x be two open sets in \mathbb{R}^n , $\varphi \in D_x^{D_t}$. $S_t \subset D_t$ and $S_x \subset D_x$ are of measure zero. If $D_t \setminus S_t$ and $D_x \setminus S_x$ are open, $\varphi|_{D_t \setminus S_t}$ is a diffeomorphism from $D_t \setminus S_t$ to $D_x \setminus S_x$, and $\exists M \in \mathbb{R}_+$ s.t. $\forall \mathbf{t} \in D_t \setminus S_t$, $|\det(\varphi|_{D_t \setminus S_t})'(\mathbf{t})| < M$; then, if the improper integral of a function f over D_x converges, so does the improper integral of $\mathbf{t} \mapsto f \circ \varphi(\mathbf{t})|\det \varphi'(\mathbf{t})|$ over $D_t \setminus$, and,*

$$\int_{D_x} f(\mathbf{x}) d\mathbf{x} = \int_{D_t \setminus S_t} f \circ \varphi(\mathbf{t})|\det \varphi'(\mathbf{t})| d\mathbf{t}.$$

If also $|\det \varphi'|$ exists and bounded on D_t , we have:

$$\int_{D_x} f(\mathbf{x}) d\mathbf{x} = \int_{D_t} f \circ \varphi(\mathbf{t})|\det \varphi'(\mathbf{t})| d\mathbf{t}.$$

Chapter 4

Surfaces and differential forms in \mathbb{R}^n

§22 Surfaces

Definition 22.1 (k -D surface). A set $S \subset \mathbb{R}^n$ is called a **k -D surface** if $\forall x \in S, \exists U \in \mathcal{U}(x)$ s.t. $U \cap S$ is homeomorphic to \mathbb{R}^k .

A k -D surface is a k -D manifold embedded in \mathbb{R}^n , which we will introduce later.

A 1-D surface is always called a curve.

Definition 22.2 (Chart). Let S be a k -D surface, and $U \subset S$ be a neighbourhood of a point $x \in S$ in S that is homeomorphic to \mathbb{R}^k . Given a homeomorphism $\varphi: \mathbb{R}^k \rightarrow U$ (or from any parameterised topological space that is homeomorphic to \mathbb{R}^k , e.g. a ball or a (non-degenerate) cuboid in \mathbb{R}^k), the pair (U, φ) , is called a **chart** or **local chart** of S . \mathbb{R}^k is called the **parametre domain** of the chart and U is called the **range of action** or **domain of action** of the chart on S .

Definition 22.3 (Atlas). An **Atlas** of a surface S is a countable collection of charts $\{(U_i, \varphi_i) \mid i \in \mathbb{N}\}$, whose domains of action cover the surface S i.e. $\bigcup_{i \in \mathbb{N}} U_i = S$.

Sometimes there is an atlas with only one chart $\{(S, \varphi)\}$ (when S itself is homeomorphic to \mathbb{R}^k), such surface is called an elementary surface.

Theorem 22.1. Let S be a k -D surface embedded in \mathbb{R}^n , and there is a chart (φ, U) with range of action $U \subset S$, where $\varphi: I_k \rightarrow U$ and

$$I_k = \{\mathbf{t} \in \mathbb{R}^n \mid |t_i| < 1, i \in k\}.$$

If $\varphi \in C^{(1)}(I_k, \mathbb{R}^n)$ and $\forall \mathbf{t} \in I_k, \text{rank } \varphi'(\mathbf{t}) = k$, then $\exists \varphi \in \mathbb{R}_+$, $\exists \varphi_\varepsilon: I_n \rightarrow \mathbb{R}^n$ s.t. $\varphi_\varepsilon|_{I_k \cap I_n} = \varphi|_{I_k \cap I_n}$, where φ_ε is a diffeomorphism from

$$I_n := \{\mathbf{t} \in \mathbb{R}^n \mid |t_i| < \varphi, i \in n\}$$

to \mathbb{R}^n .

Proof. Without loss of generality, we assume that the principal minor

$$\det \left(\frac{\partial \varphi_i}{\partial t_j}(\mathbf{0}) \right)_{i,j \in k} \neq 0.$$

By the implicit function theorem (Theorem 14.1), we can find functions $f_i \in C^{(1)}(I_n; \mathbb{R})$ ($i \in n$):

$$\begin{cases} t_0 = f_0(x_0, \dots, x_{k-1}), \\ \dots\dots\dots \\ t_{k-1} = f_{k-1}(x_0, \dots, x_{k-1}), \\ x_k = f_k(x_0, \dots, x_{k-1}), \\ \dots\dots\dots \\ x_{n-1} = f_{n-1}(x_0, \dots, x_{k-1}), \end{cases}$$

where $\mathbf{x} = \varphi(\mathbf{t})$.

Hence

$$\begin{cases} t_0 = f_0(x_0, \dots, x_{k-1}), \\ \dots\dots\dots \\ t_{k-1} = f_{k-1}(x_0, \dots, x_{k-1}), \\ t_k = x_k - f_k(x_0, \dots, x_{k-1}), \\ \dots\dots\dots \\ t_{n-1} = x_{n-1} - f_{n-1}(x_0, \dots, x_{k-1}) \end{cases}$$

is the diffeomorphism that we are looking for. □

Definition 22.4 (Smooth surface). Let S be a k -D surface in \mathbb{R}^n . If there is an atlas $\{(U_i, \varphi_i) \mid i \in \mathbb{N}\}$ of S s.t. $\varphi_i \in C^{(m)}(U_i)$ ($m \in \mathbb{N}_+$) and $\forall \mathbf{x} \in U_i$, $\text{rank } \varphi'_i(\mathbf{x}) = k$, we shall call S a **smooth surface** (of class $C^{(m)}$).

Surface can be given by either an atlas or a group of equations:

Theorem 22.2 (Surface given by equations). $\forall i \in n - k$, $F_i \in C^{(m)}(\mathbb{R}^n; \mathbb{R})$ ($m \in \mathbb{N}_+$). If S is the solutions of equation:

$$F_i(\mathbf{x}) = 0, \quad i \in n - k, \tag{22-1}$$

and $\forall \mathbf{x} \in S$,

$$\text{rank} \left(\frac{\partial F_i}{\partial x_j}(\mathbf{x}) \right)_{i \in n-k, j \in n} = k,$$

then S is a k -D smooth surface of class $C^{(m)}$.

Proof. Let $\mathbf{x}_0 \in S$. By the condition of the theorem, there are k coordinates, without loss of generality we assume the indices of which are $k, k+1, \dots, n-1$, s.t.

$$\det \left(\frac{\partial F_i}{\partial x_j}(\mathbf{x}_0) \right)_{\substack{i \in n-k \\ j \in n \setminus k}} \neq 0.$$

By implicit function theorem (Theorem 14.1), we have a cuboid I centered at \mathbf{x}_j :

$$x_j = f_j(\mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^k \cap I, \quad j \in n \setminus k,$$

where $f_j \in C^{(m)}(U)$.

Now we can define a homeomorphism $\varphi: \mathbb{R}^k \cap I \rightarrow I \cap S$:

$$\varphi_i(\mathbf{t}) = \begin{cases} t_i, & i \in k; \\ f_i(\mathbf{t}), & i \in n \setminus k. \end{cases}$$

□

§23 Orientation of a Surface

Algebra has taught us that we can find a basis $\langle \hat{\mathbf{e}}_i \rangle_{i \in n}$ for \mathbb{R}^n and the transformation from one basis to another is taken by a non-degenerate matrix \mathbf{P} .

Definition 23.1 (Orientation classes of frames). Two bases are said to belong one **orientation class of frames** if the transform matrix between them has a positive determinant.

It is clear that there are two orientation classes of frames.

Part II

Real Analysis

Part III

Functional Analysis

Part IV

Complex Analysis

Part V

Appendices

Appendix A

Several Important Inequality

§1 Mean-value Inequality

§2 Cauchy's Inequality

§3 Jensen's Inequality

§4 Brunn-Minkowski Inequality

§5 Hölder's Inequality and Minkowski's Inequality

Theorem 5.1 (Young's Inequality). *Let $a \geq 0$, $b \geq 0$ be two non-negative numbers. If $p > 1$, $q > 1$ and*

$$\frac{1}{p} + \frac{1}{q} = 1,$$

then

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

The equality iff $a^p = b^q$.

Proof.

□

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](https://cds.cern.ch/record/2137923). URL: <https://cds.cern.ch/record/2137923>.

Symbol List

Here listed the important symbols used in this notes.

$B(a; \delta)$, 3	$\omega(f; x)$, 12
$\mathcal{B}(X_0, \dots, X_{n-1}; Y)$, 19	\overline{E} , 4
$C^{(p)}$, 29	$\overline{\mathcal{I}}$, 46
$C^{(1)}(X)$, 24	∂E , 4
$C^{(1)}(X, Y)$, 24	$\partial_i f$, 24
$C_\infty[a, b]$, 2	$\frac{\partial f}{\partial \mathbf{x}_i}(\mathbf{a})$, 24
$C^{(n)}(U; Y)$, 30	\mathbb{R}_p^n , 2
$C^{(n)}(U)$, 30	$\Re(I)$, 44
$C_p[a, b]$, 2	$S(f, P)$, 46
d_∞ , 2	$s(f, P)$, 46
d_p , 2	$\sigma(f, P, \xi)$, 43
$d\mathbf{x}$, 21	$\text{supp } f$, 54
$\Delta(f)$, 21	$\tilde{B}(X, \delta)$, 3
$df(\mathbf{x})$, 20	$U(x)$, 3, 5
$E_n \uparrow E$, 58	$\mathring{U}(x)$, 5
$f^{(n)}(\mathbf{x})$, 29	$\mathcal{U}(x)$, 3, 5
$f'(\mathbf{x})$, 20	$\underline{\mathcal{I}}$, 46
$\lambda(P)$, 43	$\ \mathcal{A}\ $, 17
\langle, \rangle , 16	(X, d) , 2
$\mu(E)$, 48	(X, \mathcal{T}) , 4
$\mu(I_{\mathbf{a}, \mathbf{b}})$, 39	(\mathbf{x}, \mathbf{y}) , 25
$\omega(f; E)$, 12	$[\mathbf{x}, \mathbf{y}]$, 25

Index

$C^{(p)}$ -diffeomorphism, 29

T_2 space, 5

ε -net, 6

k -D surface, 63

a.e., 44

almost everywhere, 44

Atlas, 63

ball, 3

Banach space, 16

base, 5, 11

Borel sets, 40

boundary, 4

boundary point, 4

bounded, 18

Carathéodory criterion, 41

Cauchy sequence, 9

Cauchy-Bunyakovskii's inequality, 17

chart, 63

Chebyshev metric, 2

closed ball, 3

closed interval, 25

closed set, 3, 4

closure, 4

compact set, 6

complete, 9

complete normed space, 16

completion, 9

connected, 8

connected set, 8

connected space, 8

continuous, 12

continuously differentiable, 24

contraction, 14

contraction mapping principle, 14

convex set, 27

criterion for continuity, 12

cuboid, 39

Darboux lower sum, 46

Darboux upper sum, 46

deleted neighbourhood, 5

dense set, 5

derivative, 20

derivative mapping, 21

derivative with respect to a vector, 25

diameter, 3

diffeomorphism with smoothness p , 29

differentiable, 20

differential, 20

direct product, 5

distance, 2

domain of action, 63

elementary diffeomorphism, 56

exhaustion, 58

exterior point, 4

extremum point, 33

filter base, 11

fixed point, 14

fundamental sequence, 9

germ, 5

Hausdorff axiom, 5

Hausdorff space, 5

Hermitian space, 16

- Hilbert space, 17
- homeomorphic, 13
- homeomorphism, 12
- implicit function theorem, 35
- improper integral, 59
- inner product, 16
- inner product space, 16
- interior point, 4
- isometric, 9
- isometry, 9
- isomorphic, 19
- isomorphism, 19
- iterated integral, 52
- Jordan content, 48
- Jordan measurable, 47
- Jordan measure, 48
- Lebesgue measurable, 41
- Lebesgue outer measure, 40
- Lipschitz condition, 15
- limit, 11
- limit point, 4
- Lipschitz continuous, 15
- local chart, 63
- locally connected, 9
- locally maximum point, 33
- locally minimum point, 33
- lower Darboux integral, 46
- measurable set, 40
- measurable space, 40
- measure function, 40
- measure zero, 43
- mesh, 43
- metric, 2
- metric space, 2
- multiple integral, 52
- n -th differentiation, 29
- neighbourhood, 3, 5
- nested sequence, 6
- norm, 16
- normed, 16
- open base, 5
- open cover, 5
- open interval, 25
- open set, 3, 4
- open-closed set, 8
- orientation class of frames, 65
- oscillation, 12
- parametre domain, 63
- partial derivative, 24
- partition, 43
- Picard-Banach fixed-point principle, 14
- pre-Hilbert space, 17
- range of action, 63
- reversible operators, 23
- Riemann integral, 44
- Riemann sum, 43
- separable, 5
- separated space, 5
- separation axiom, 5
- sequentially compact, 7
- σ -algebra, 40
- smooth surface, 64
- standard topology, 4
- stronger, 5
- subcover, 5
- subspace, 4, 5
- subspace topology, 5
- support, 54
- tangent mapping, 20
- Taylor's formula, 31
- topological base, 5
- topological space, 4
- topology, 4
- uniformly continuous, 13
- upper Darboux integral, 46
- weight, 5
- zero content, 48