Contents

IV Measure Theory and Probability Theory 105

Part I

Real Analysis

Chapter 1

Real Numbers

1.1 Axiomatic Formulation of Real Numbers

The building axioms of real numbers is devided into three groups based on the properties they are describing.

1. Field axioms.
2. Order axioms.
3. Completeness axiom.

1.1.1 Field Axioms

A field is a non-empty set $\mathbb{F}$ with two binary operations *addition*, $+$, and *multiplication*, $\cdot$. For all $x, y, z \in \mathbb{F}$:

Axiom 1. Addition and multiplication are commutitive.

$$x + y = y + x, \quad x \cdot y = y \cdot x$$

Axiom 2. Addition and multiplication are associative.

$$x + (y + z) = (x + y) + z, \quad x \cdot (y \cdot z) = (x \cdot y) \cdot z$$

Axiom 3. Multiplication distributes over addition.

$$x \cdot (y + z) = x \cdot y + x \cdot z$$

Axiom 4. There exists a number 0 such that for every number x:

$$x + 0 = 0 + x = x$$

Axiom 5. There exists a number 1 such that for every number x:

$$x \cdot 1 = 1 \cdot x = x$$

Axiom 6. For every number x, there exists a number y such that:

$$x + y = 0$$

y is called the negative of x and is denoted by $-x$.

Axiom 7. For every number $x \neq 0$, there exists a number y such that:

$$x \cdot y = 1$$

y is called the reciprocal of x and is denoted by x^{-1} or $\frac{1}{x}$.

1.1.2 Order Axioms

The order axioms establishes an ordering on the numbers of $\mathbb{F}$ to determine which element is larger or smaller. To achieve an ordering, we define the set of positive real numbers $\mathbb{F}^+ \subset \mathbb{F}$.

Axiom 8. The $\mathbb{F}^+$ is closed under addition and multiplication.

$$\forall x, y \in \mathbb{F}^+, \quad (x + y) \in \mathbb{F}^+ \text{ and } (x \cdot y) \in \mathbb{F}^+$$

Axiom 9. $0 \notin \mathbb{F}^+$.

Axiom 10. For every number $x \neq 0$, either $x \in \mathbb{F}^+$ or $-x \in \mathbb{F}^+$.

We then define the binary operator $>$ such that $x > y$ whenever $(x - y) \in \mathbb{F}^+$.

1.1.3 Completeness Axiom

Given that $(\mathbb{F}, +, \cdot, >)$ is an ordered field, we define the followings:

Definition (Upper bound): A set $S \in \mathbb{F}$ has an upper bound if for some $a \in \mathbb{F}$ is greater or equal to all element of S. That is, $\forall x \in S\ a \geq x$. We say that S is bounded from above.

Definition (The least upper bound): The element $a \in \mathbb{F}$ is the least upper bound of a set $S \subset \mathbb{F}$ if it is smaller than every upper bound of S. We say a is the supremum of S, denoted by $a = \sup S$.

Note that, if the least upper bound exists, it must be unique.

Axiom 11. If S is a non-empty set that bounded from above that it has supremum.

Theorem 1.1. *There exists a unique set that satisifies all the axioms above. It is denoted by $\mathbb{R}$, the set of real numbers.*

Proof. The existence of $\mathbb{R}$ is proved in many ways. One way to construct real numbers is by *Dedekind Cuts*. Let the pair of rational sets $(A, B = \mathbb{Q} \backslash A)$ be a partition of $\mathbb{Q}$ such that:

1. $A \neq \emptyset$ and $A \neq \mathbb{Q}$.
2. $\forall x, y \in \mathbb{Q}$ s.t. $x < y$, $y \in A \implies x \in A$. That is. if $y \in A$, then all rationals less than y are in A.

3. $\nexists\, x \in A$ s.t. $\forall y \in A, x \geq y$. That is, A does not have a maximum element.

For convenience we let A represent the pair (A, B) as A completely determines B. We define $+$, $\cdot$, and $>$ on these cuts as follows:

$$
\begin{aligned}
A + B &= \{a + b \,|\, a \in A, b \in B\} \\
\boldsymbol{0} &= \{a \,|\, a < 0\} \\
-A &= \{a' \,|\, \forall a \in A, a' < -a\}
\end{aligned}
$$

For $\cdot$, we first take two set A and B that have some positive elements.

$$A \cdot B = \{a \cdot b \,|\, a \in A \wedge a \leq 0, b \in B \wedge b \leq 0\} \cup \boldsymbol{0}$$

If A or B did not have any positive elements, we first take the negative of the set, and then multiply the two sets and take the negative of the product. Similarly, we define the reciprocal of A if A has a positive element.

$$
\begin{aligned}
\boldsymbol{1} &= \{a \,|\, a < 1\} \\
A^{-1} &= \left\{a' \,\middle|\, \forall a \in A, a > 0, a' < \frac{1}{a}\right\}
\end{aligned}
$$

Lastly,

$$A > B \text{ if } A \supset B$$

If S is a non-empty set of cuts that is bounded from above, then it has a supremum which is equal to $\bigcup S$. Let us denote the set of all rational cuts by $\mathbb{R}$. It is left to the reader that the $(\mathbb{R}, +, \cdot, >)$ satisifies the axioms above.

The set of real numbers is unique in sense that if $(\mathbb{R}, +, \cdot, >)$ and $(\mathbb{R}', +', \cdot', >')$ both satisify the axioms, then there exists bijective mapping $\alpha : \mathbb{R} \to \mathbb{R}'$ such that:

$$
\begin{aligned}
\alpha(x + y) &= \alpha(x) +' \alpha(y) \\
\alpha(0) &= 0' \\
\alpha(x \cdot y) &= \alpha(x) \cdot' \alpha(y) \\
\alpha(1) &= 1' \\
x < y &\iff \alpha(x) <' \alpha(y)
\end{aligned}
$$

And if S is a non-empty set in $\mathbb{R}$ and $\alpha(S) = \{\alpha(x) \,|\, x \in S\}$, then S is has an upper bound if and only if $\alpha(S)$ has an upper bound, such that, $\alpha(\sup S) = \sup \alpha(S)$. ∎

Exercises

1. The set of natural numbers $\mathbb{N}$ in $\mathbb{R}$ is not bounded from above.

2. Let $x \in \mathbb{R}$ be such that for all $n \in \mathbb{N}$

$$0 \leq x \leq \frac{1}{n}$$

then $x = 0$.

3. (Archimedean Property) For all $a, b > 0$ there exists $n \in \mathbb{N}$:

$$na > b$$

4. Consider $I_n = [a_n, b_n]$ $\forall n \in \mathbb{N}$ such that $I_1 \supset I_2 \supset \ldots$. Show that $\bigcap I_n$ is not empty. Moreover, if for each $e > 0$ there exists n such that $b_n - a_n < e$, then $\bigcap I_n$ is a single point.

5. Show that $\sqrt{2} \in \mathbb{R}$ and for all $p > 0$, there is a positive real number q such that $q^2 = p$

6. Prove that the addition and multiplication identity elements are unique.

7. Show that the Trichotomy law holds for $>$. That is, exactly one of the following three is true.

$$x > y \qquad x = y \qquad y > x$$

8. Show that $1 \in \mathbb{F}^+$.

9. Show that if $x > -1$ and $n \in \mathbb{N}$:

$$(1 + x)^n \geq 1 + nx$$

and equality only holds when $n = 1$.

10. Let $F_p = \{0, 1, \ldots, p - 1\}$ where p is a prime number. Define $+$ and $\cdot$ to be the modular addition and product modulus p, respectively. Investigate whether if F_p can be ordered.

11. Consider the set of all rational polynomials $\mathbb{Q}[x]$:

$$\mathbb{Q}[x] = \left\{ \frac{a_m x^m + \cdots + a_1 x + a_0}{b_n x^n + \cdots + b_1 x + b_0} \;\middle|\; a_i, b_j \in \mathbb{Q}, b_n \neq 0 \right\}$$

Show that $\mathbb{Q}[x]$ under the normal addition and multiplication is a field. Furthermore, show that $\mathbb{Q}^+[x] = \{q \in \mathbb{Q}[x] \,|\, a_m \cdot b_n > 0\}$ constitutes an ordering on $\mathbb{Q}[x]$.

Chapter 2

Topology and Metric Spaces

2.1 Topology

Let X be a set. A **topology** on X is a collection $\mathscr{T}$ of subsets of X called **open set** having the following properties

1. If $U_\alpha \in \mathscr{T}$ where $\alpha \in A$ for any set A then, $\cup_{\alpha \in A} U_\alpha \in \mathscr{T}$.

2. If $U_\alpha \in \mathscr{T}$ where $\alpha \in A$ for any finite set A then, $\cap_{\alpha \in A} U_\alpha \in \mathscr{T}$.

3. $X, \emptyset \in \mathscr{T}$.

A topological space is a pair $(X, \mathscr{T})$, where $\mathscr{T}$ is a topology on X.

Example 2.1. On any set X we can define two topologies. The *trivial topology* on X consists of $\{\emptyset, X\}$ and the *discrete topology* on X is $\mathcal{P}(X)$.

If $\mathscr{T}_1$ and $\mathscr{T}_2$ are topologies on X, we say that $\mathscr{T}_1$ is *weaker/coarser* than $\mathscr{T}_2$, or that $\mathscr{T}_2$ is *stronger/finer* than $\mathscr{T}_1$, if $\mathscr{T}_1 \subset \mathscr{T}_2$.

Definition: Let X be a topological space and $x \in X$, N is called a **neighbourhood** of x if there exists an open set G such that $x \in G \subset N$. We say that x is an interior point of N if N is neighbourhood of x.

Proposition 2.1. *A subset U is open if and only if U is a neighbourhood for all $x \in U$.*

Proof. Let U be a neighbourhood for all of its points. That is, for every $x \in U$ there is an open set G_x such that $x \in G_x \subset U$. Then, $\bigcup_x G_x \subset U$ however, $U \subset \bigcup_x G_x$. Therefore, $U = \bigcup_x G_x$ and by the first axiom U is an open set. If U is an open set, then U is a neighbourhood for each point $x \in U$. ■

Definition: A subset F of X is called **closed** if F^c is open. The **Closure** of a set E is the intersection of all closed sets that include E and it is denoted by $\operatorname{cl} E$ or $\bar{E}$.

Proposition 2.2. *Let X be a topological space and E a subset of X. Then, $\operatorname{cl} E$ is closed.*

Proof. We know that $\operatorname{cl} E = \bigcap_{E \subset F} F$ where F are closed set. Then, $(\operatorname{cl} E)^c = \bigcup_{E \subset F} F^c$ which is an open set and hence $\operatorname{cl} E$ is closed. ■

Proposition 2.3. *If E is subset of a topological space X and $xinX$, then $x \in \operatorname{cl} E$ if and only if $U \cap E = \emptyset$ for every open neighbourhood of U of x.*

Proof. Suppose there exists an open neighbourhood of x, U such that, $U \cap E = \emptyset$. Then, $F = \operatorname{cl} E \cap U^c$ is a closed that constains E hence, $\operatorname{cl} E \subset F$ and $x \notin \operatorname{cl} E$. If $x \notin \operatorname{cl} E$, then $(\operatorname{cl} E)^c$ is an open neighbourhood of x which does not meet E. ■

Definition: A **limit point** of E is a point $x \in X$ such that $E \cap U \backslash \{x\} \neq \emptyset$ for every neighbourhood U of x. The set of all limit points of E is denoted by E' or $\lim E$. If $x \in E$ but $x \notin \operatorname{cl} E$, then x is called an **isolated point** of E.

Definition: A subset E of a topological space X is **perfect** if every point of E is a limit point.

Definition: Let E be subset of a topological space X then, the **interior** of E is the union of all open sets that are contained in E, denoted by E° or $\operatorname{int} E$.

Proposition 2.4. *For any set E in topological space X, interior of E is the set of all interior points of E and $\operatorname{cl} E = (\operatorname{int} E)^c$.*

Definition: The **boundary** of E is defined as $\operatorname{cl} E \backslash \operatorname{int} E$ and it is denoted by $\operatorname{bdry} E$ or ∂E.

Proposition 2.5. *Let E be subset of a topological space X. E is closed if and only if $E = \operatorname{cl} E$ and E is open if and only if $E = \operatorname{int} E$. Furthermore, $\operatorname{bdry} E = \emptyset$ if and only if E is both open and closed.*

Proof. Note that $E \subset \operatorname{cl} E$ and when E is closed, $\operatorname{cl} E \subset E$ therefore, $E = \operatorname{cl} E$. ■

Definition: A set D in a topological space X is called **dense** when, $\operatorname{cl} D = X$. More generally, D is dense in subset E, if $E \subset \operatorname{cl} D$. A topological space in which there exists a countable dense set is called **separable**.

Proposition 2.6. *D is dense in E if and only if for all $x \in E$, $D \cap U \neq \emptyset$ for any open neighbourhood U of x.*

Proof. If there exists $x \in E$ such that an open neighbourhood U of x does not intersect D, then $x \notin \operatorname{cl} D$ and hence $D \not\supset E$. If for each $x \in E$ every neighbourhood U of x intersects D, then $x \in \operatorname{cl} D$ hence $E \subset \operatorname{cl} D$. ■

Let $(X, \mathscr{T})$ be a topological space and $Y \subset X$. Then, $(Y, \mathscr{T}_y)$ is a **topological subspace** where $\mathscr{T}_y = \{U \cap Y \,|\, U \in \mathscr{T}\}$ is the **relative topology** of Y.

Definition: A **base** for a topolgy $\mathscr{T}$ is a collection of open set $\mathscr{B} \subset \mathscr{T}$ such that each $U \in \mathscr{T}$ is a union of open sets in $\mathscr{B}$. That is, $U = \bigcup_{G \in \mathscr{B}} G$.

2.2 Metric spaces

Let X be a non-empty set and $x, y \in X$ then if there exists a non-negative real number $d(x, y)$ with following three properties:

1. (Positive definiteness.) $d(x, y) = 0$ if and only if $x = y$.
2. (Symmetry.) $d(x, y) = d(y, x)$.
3. (Triangle inequality.)$d(x, y) \leq d(x, z) + d(z, y)$.

the combination (X, d) is called a **metric space** and $d(x, y)$ is called the **metric**, or also **distance** function.

Example 2.2. The Euclidean space $\mathbb{R}^n = \{x_1, x_2, \ldots, x_n \mid x_i \in \mathbb{R}\}$ with the metric $d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$ makes a metric space. To prove this we must show the above properties hold.

1. if $d(x, y) = 0$ then:
$$\sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2} = 0$$
Therefore eac terms must be zero:
$$\begin{aligned}(x_i - y_i)^2 &= 0 \quad \forall i \leq n \\ x_i - y_i &= 0 \implies x_i = y_i\end{aligned}$$
Thus $x = y$.
2. It is obvious that $(x_i - y_i)^2 = (y_i - x_i)^2$ and hence $d(x, y) = d(y, x)$
3. The triangle inequality immediately follows from the Cauchy-Schwartz inequality.

We can expand the Euclidean norm by defining Minkowski *p-norm* also called L^p*-norm* for $1 \leq p \leq \infty$ as follows:

$$d_p(x, y) = \left(\sum_i |x_i - y_i|^p\right)^{\frac{1}{p}}$$

and by taking the limit, $p \to \infty$ we find out that:

$$d_\infty(x, y) = \max_i\{|x_i - y_i|\}$$

Example 2.3. We can define **discrete distance** as follows:

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

and it is pretty straightforward to show that the three properties hold.

Definition: The **open ball** $B_r(a)$ with radius r centered at a is the set of all points:

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

and the **closed ball** $\overline{B}_r(a)$ with radius r centered at a is the set of all points:

$$\overline{B}_r(a) = \{x \in X \mid d(x, a) \leq r\}$$

The **sphere** $S_r(a)$ with radius r centered at a is the set of all the points:

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

Definition: Let (X, d) be a metric space. A subset $U \subset X$ is an open set if for all $a \in U$ there exists $\rho > 0$ such that $B_\rho(a) \subset U$.

Given this defintion of open sets, we can define a topolgy on metric space X.

1. Firstly, we need to show that every union of open sets is open itself. Let U_α be some open sets indexed by A and let $x \in \bigcup_\alpha U_\alpha$. Then, there exists a $\alpha \in A$ such that $x \in U_\alpha$. Since, U_α is open, there exists a ball $B_r(x)$ which is contained in U_α. Clearly, $B_r(x) \in \bigcup_\alpha U_\alpha$ and hence $\bigcup_\alpha U_\alpha$ is open.

2. Secondly, we show that intersection of finite collection of open sets in open. Let U_α be open sets indexed by a finite set A and let $x \in \bigcap_\alpha U_\alpha$. For each $\alpha \in A$, there exists a ball $B_{r_\alpha}(x)$ such that $B_{r_\alpha}(x) \subset U_\alpha$. Let $r = \min_\alpha r_\alpha$ and note that $B_r(x) \subset B_{r_\alpha}(x) \subset U_\alpha$ for all $\alpha \in A$. Thus, $B_r(x) \subset \bigcap_\alpha U_\alpha$ hence, $\bigcap_\alpha U_\alpha$ is open.

3. Thirdly, we show that X and $\emptyset$ are open. $\emptyset$ is trivially open as it has no element. And $B_r(x) \subset X$ by defintion for all r hence, X is open as well.

Consider the following re-defintions of concepts introduced in the previous section.

Definition (Internal Point): A point $a \in X$ is called an internal point of U if $\exists \rho > 0$ that the ball $B_\rho(a)$ contained in U.

Definition (Interior): The interior of a set U denoted by U° or $\operatorname{int}(U)$ is the set of all its interior points.

Definition (Adherent Point): A point $a \in X$ is called an adherent point of U if $\forall \rho > 0$ the ball $B_\rho(a)$ contains a point in U.

Definition (Limit Point): A point $a \in X$ is called a limit point of U if $\forall \rho > 0$ the set $B_\rho(a) - \{a\}$ contains a point in U. The set of all limit points is denoted by S' or $\lim S$.

Note: For any limit point $a \in U$ every open ball $B_r(a)$ contains infinitely many points in U.

Definition (Closed Set): A subset $C \subset X$ is closed set if it contains all of its adherent point.

Definition (Closure): The closure of a set U denoted by $\overline{U}$ or $\operatorname{cl} U$ is set of all its adherent points.

Note: The closure of a set is a closed set.

We then, show that these re-definitions are equivalent to the topological defintions.

Theorem 2.7. *Subset $C \subset X$ is closed if and only if $X \backslash C$ is open.*

Proof. Firstly, we prove the necessity condition that is C is closed if $X \backslash C$ is open . We employ proof by contradiction. Let C be a closed subset of X such that its complement is not open. That is, for some $a \in (X \backslash C)$ there is no $\rho > 0$ exists such that $B_\rho(a) \subset (X \backslash C)$. In other words, for all ρ, $\exists p \in B_\rho(a)$ s.t $p_\rho \in C$. Which implies that a is an adherent point of C but since C is closed then $a \in C$ which is a contradiction. Similarly, one can show the sufficiency condition. ∎

Corollary 2.8. *X and $\emptyset$ are both closed and open.*

Remark 1. (Equivalent Definitions)

1. An open set is a union of open balls. Conversely, a union of open balls is an open set.

   Proof. For every $a \in U$ there is a ball $B_\rho(a) \subset U$ thus $\bigcup_{a\in U} B_\rho(a) \subset U$ and since $a \in B_\rho(a)$ we must have $\bigcup_{a\in U} B_\rho(a) \supset U$ hence $U = \bigcup_{a\in U} B_\rho(a)$.

   Now let $U = \bigcup B_\rho(a)$ we need to show that U is open. Let $b \in U$ then b must be a point in at least one of those balls. Let $b \in B_r(c)$ and $\rho = r - d(b, c)$. We will show that $B_\rho(b) \subset B_r(c) \subset U$, for any $x \in B_\rho(b)$ by triangle inequality we have $d(x, c) \leq d(x, b) + d(b, c) < \rho + d(b, c) = r$ which means $x \in B_r(c)$. □

2. A set is open if and only if all of its members are interior points. Therefore, $U = \operatorname{int} U$.

3. Let $I = \{S \subset U \mid S \text{ is open}\}$ then $\operatorname{int} U = \bigcup_{S\in I} S/$

4. Let $I = \{U \subset S \mid S \text{ is closed}\}$ then $\operatorname{cl} U = \bigcap_{S\in I} S$.

Let (X, d) be a metric space and $Y \subset X$ then Y may inherit its metric from X and (Y, d) would also be a metric space and is called a **metric subspace** of X. We will investigate the nature of open and closed sets in subspaces. Let $B^Y_\rho(y) = \{p \in Y \mid d(y, p) < \rho\}$ Then, it is easy to see that:

$$B^Y_\rho(y) = B_\rho(y) \bigcap Y$$

Corollary 2.9. *Let (X, d) be a metric space and $Y \subset X$ is a metric subspace of X then $U \subset Y$ is an open subset of Y if and only if there is a open set $V \subset X$ such that $U = V \cap Y$. Similarly, for any closed set $C \subset Y$ there is a closed set $D \subset X$ such that $C = D \cap Y$.*

Proof. Ofcourse, if $U \subset Y$ is open in Y then by definition it can be represent as a union of open ball $B^Y_r(a)$. Each of these balls is the intersection of a $B^X_r(a) \cap Y$. Therefore

$$U = \bigcup B^Y_r(a) = \bigcup \big(B^X_r(a) \cap Y\big) = \Big(\bigcup B^X_r(a)\Big) \cap Y = V \cap Y$$

Furthermore, if $a \in V \cap Y$ then there exists a ball $B^X_r(a) \subset V$. Therefore

$$B^Y_r(a) = B^X_r(a) \cap Y \subset V \cap Y = U$$

The case for closed subsets can be proved using the complements. ■

Exercises

1. Show that $\operatorname{cl} S = S \cup \lim S$

2.3 Convergence

Let X be a topological space. A **sequence** is a function of form $a : \{k, k+1, k+2, \ldots\} \to X$ where $k \in \mathbb{Z}$. Conventionally, instead of $a(n)$, a_n is used. The sequence $\{a_n\}$ is **convergent** to $a \in X$ if for all neighbourhood U of a there exists N such that:

$$n \geq N \implies a_n \in U$$

and it is denoted by $a_n \to a$ or $a = \lim_{n\to\infty} a_n$. Given a topolgy, convergence is not necessarily well-behaved. For example, in the trivial topolgy, if $a_n \to a$, then $a_n \to b$ for any $b \in X$. To do away with this we consider **Hausdorff spaces**.

Definition: Let X be a topological space. X is a Hausdorff space if for any two $x, y \in X$ where $x \neq y$, there exists disjoint open sets U and V such that $x \in U$ and $y \in V$.

Proposition 2.10. *Let X be a Hausdorff space and x_n is a sequence in X. If $x_n \to x$ and $x_n \to y$ as $n \to \infty$, then $x = y$.*

Proof. If $x \neq y$, then there are disjoint open set U and V with $x \in U$ and $y \in V$. If $x_n \to x$, then there exists N such that for $n \geq N$, $x_n \in U$. But this implies that for $n \geq N$, $x_n \notin V$. Meaning, there exists no N' that for $n \geq N'$, $x_n \in V$ hence, x_n does not converge to y. ■

In the case of a metric space, the set $\{a_k, a_{k+1}, \ldots\}$ is bounded in X, that is, there exist $K > 0$ and a point $b \in X$ such that $\forall n,\ a_n \in B_K(b)$.

Another problem with the defintion of convergence is its dependence on a convergence point. So naturally the following question comes up. Is there a way to show the convergence of sequence based on itself? For that, we need to define **Cauchy sequence**. A sequence $\{a_n\}$ in a metric space X is a Cauchy sequence if:

$$\forall \epsilon > 0,\ \exists N \quad \text{s.t.} \quad n, m \geq N \implies d(a_n, a_m) < \epsilon$$

Lemma 2.11. *In a metric space X, convergence of a_n to a is equivalent to*

$$\forall \epsilon > 0,\ \exists N \quad \text{s.t.} \quad n \geq N \implies d(a_n, a) < \epsilon$$

Proof. If $a_n \to a$, then for $B_\epsilon(a)$ there exists N such that $n \geq N \implies a_n \in B_\epsilon(a)$. On the other hand, for any neighbourhood U we can find $\epsilon > 0$ such that $B_\epsilon(a) \subset U$. Hence, if the metric convergence condition holds, then topological convergence holds, as well. ■

Theorem 2.12. *Every convergent sequence is a Cauchy sequence.*

Proof. For a given $\epsilon > 0$ we know there exist N such that:

$$n \geq N \implies d(a_n, a) < \frac{\epsilon}{2}$$

and equivalently:

$$m \geq N \implies d(a_m, a) < \frac{\epsilon}{2}$$

and since by the triangle inequality we have:

$$d(a_m, a_n) \leq d(a_m, a) + d(a_n, a) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

■

Definition (Subsequence): We call $\{b_n\}$ a **subsequence** of $\{a_n\}$ if there is a sequence of positive integers $n_1 < n_2 < n_3 < \dots$ such that for each k, $b_k = a_{n_k}$.

Exercises

1. Show that if a sequence $\{a_n\}$ is convergent, then its limit is unique. That is, if $a_n \to a$ and $a_n \to b$ as $n \to \infty$ then $a = b$.

2. Prove that every subsequence of a convergent sequence converges and it converges to the same limit.

2.4 Completeness

A metric space (X, d) is **complete** if every Cauchy sequence converges.

Proposition 2.13. *$\mathbb{R}$ with the normal Euclidean norm is a complete metric space.*

To prove it, we need the following lemmas.

Lemma 2.14. *If $\{a_n\}$ is a Cauchy sequence in a metric space (X, d) then the set $S = \{a_k, a_{k+1}, \ldots\}$ is bounded.*

Proof. For a fixed $\epsilon > 0$ we know there exists N such that:

$$m, n \geq N \implies d(a_n, a_m) < \epsilon$$

especially:

$$n \geq N \implies d(a_n, a_N) < \epsilon$$

Since there is only finitely many indices less than N then we can determine the largest $d(a_N, a_m)$for all m less than N lets denote it by A. Finally, let $K = \max\{\epsilon, A\}$ then, $B_K(a_N)$ contains all the elements of sequence. □

Lemma 2.15. *If one of the subsequences of Cauchy sequence is convergent, then the Cauchy sequence is convergent to the same element.*

Proof. Let $a_{n_k} \to a$ when $k \to \infty$ That is, for a given $\epsilon > 0$, $\exists N_1$ such that:

$$k \geq N_1 \implies d(a_{n_k}, a) < \frac{\epsilon}{2}$$

and since $\{a_n\}$ is a Cauchy sequence then we also know that there exists N_2 such that:

$$q, m \geq N_2 \implies d(a_m, a_q) < \frac{\epsilon}{2}$$

Let $N = \max\{N_1, N_2\}$ and $n_q \geq N$ consequently:

$$n_q, m \geq N \implies d\big(a_m, a_{n_q}\big) < \frac{\epsilon}{2}$$

and by the triangle inequality we have:

$$d(a_m, a) \leq d\big(a_m, a_{n_q}\big) + d\big(a_{n_q}, a\big) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which proves the convergence of $a_n \to a$. □

We present a proof for the completeness of $\mathbb{R}$ under the Euclidean norm.

Proof. Let $\{a_n\}$ be a Cauchy sequence. Then by Lemma 2.14, the sequence is bounded and there is a closed interval $I_0 = [a, b]$ in which all a_n lie. Consider the closed intervals $\left[a, \dfrac{a+b}{2}\right]$ and $\left[\dfrac{a+b}{2}, b\right]$. Since the sequence has infinitely many terms then there are infinitely many terms in at least one of the two intervals. Let that interval be I_1 and choose $x_1 \in I_1$ where $x_1 = a_{n_1}$ for some n_1. Repeat the process for I_1 to get I_2 and $x_2 = a_{n_2}$ where $n_2 > n_1$. Since there are infinitely many terms in I_2 we can find such n_2. By continuing this process we have a subsequence $\{x_k\}$ and a sequence of nested closed sets $\{I_k = [a_k, b_k]\}$. Since for all $\epsilon > 0$ there exists K such that $b_K - a_K < \epsilon$ then the intersection of $\{I_k\}$ is a point, say y. We claim that that $x_k \to y$, that is:

$$\forall \epsilon > 0 \, \exists N \in \mathbb{N} \quad \text{s.t.} \quad n \geq N \implies |x_n - y| < \epsilon$$

Since $y = \bigcap I_k$ then $y \in I_k$ for all k, especially $y \in I_n$. Therefore, $|x_n - y|$ is smaller than or equal to the length of I_n which is $\dfrac{b-a}{2^n} \leq \dfrac{b-a}{2^N}$. By setting $N > \log_2 \dfrac{b-a}{\epsilon}$ we have:

$$|x_n - y| \leq \frac{b-a}{2^n} \leq \frac{b-a}{2^N} < \epsilon$$

Therefore $\mathbb{R}$ is a complete metric under Euclidean norm. ■

Let (X, d) and (X', d')be two metric spaces. Define the following norms on the Cartesian product $X \times X'$:

1. $D_1((x, x'), (y, y')) = d(x, y) + d'(x, y)$
2. $D_2((x, x'), (y, y')) = \sqrt{d(x, y)^2 + d'(x', y')^2}$
3. $D_3((x, x'), (y, y')) = \max\{d(x, y), d'(x', y')\}$

Let $p_1 = (x, x')$ and $p_2 = (y, y')$:

$$D_3(p_1, p_2) \leq D_2(p_1, p_2) \leq D_1(p_1, p_2) \leq 2D_3(p_1, p_2)$$

Then, it is easy to see that if a sequence $\{a_n\}$ is convergent under one of these norms, it is convergent to the same value under the other two. The same is true if the sequence is a Cauchy sequence.

By induction we can generalize it to $X_1 \times X_2 \times \ldots \times X_n$. For example, $\mathbb{R}^n$ is complete metric under all the three norms introduced above. That is, every Cauchy sequence in $\mathbb{R}^n$ is convergent. To show this assume the sequence $\{x_i\}$ is a Cauchy sequence under, WLOG, D_1:

$$\forall \epsilon > 0, \, \exists N \quad \text{s.t.} \quad i, j \geq N \implies D_1(x_i, x_j) < \epsilon$$

Then for the k-th coordinate:

$$|x_{i_k} - x_{j_k}| < D_1(x_i, x_j) < \epsilon$$

Therefore, for every coordinate, the image of the sequence on that coordinate is a Cauchy sequence. Since $\mathbb{R}$ is complete then $\{x_{i_k}\}_i$ is convergent to some x_k for all k. We claim that $x_i \to x = (x_1, \ldots, x_n)$ as $i \to \infty$:

$$D_1(x, x_i) = |x_{i_1} - x_1| + |x_{i_2} - x_2| + \cdots + |x_{i_n} - x_n|$$

We have shown that $\{x_{i_k}\}_i$ is convergent to x_k then there must be $N_1, N_2, \ldots N_n$ such that for all k:

$$\forall \epsilon, \quad i \geq N_k \implies |x_{i_k} - x_k| < \frac{\epsilon}{n}$$

Setting $N = \underset{1 \leq k \leq n}{max} N_k$:

$$D_1(x, x_i) < n \cdot \frac{\epsilon}{n} = \epsilon$$

Theorem 2.16. *Let (X, d) be a complete metric space and $Y \subset X$ is a complete metric space if and only Y is a closed subset of X.*

Proof. It is clear that Y being closed is necessary for Y being a complete metric subspace. To show that is also sufficient, we need ot show that if Y is a complete metric subspace then it is closed. Assume the contrary, that is there exists an adherent point of Y, $a \notin Y$. Since a is an adherent point of Y then for all $\rho > 0$ there exists a point $x \in B_\rho(a)$ such that $x \in Y$. For each n let $\rho = \frac{1}{n}$ and choose a point $x_n \in Y$ It is clear that $\{x_n\}$ is convergent to a. From Theorem 2.12 $\{x_n\}$ is a Cauchy sequence. Since Y is complete then a must be in Y which is a contradiction. ∎

Theorem 2.17 (Baire Category Theorem). *Let X be a complete metric space and G_n be a sequence of dense open sets in X. Then, $\cap_{n=1}^{\infty} G_n$ is dense in X.*

Theorem 2.18. *Let (X, d) be a complete metric space, and suppose that $f : X \to X$ has the property that there exists $\alpha < 1$ such that*

$$d(f(x), f(y)) < \alpha d(x, y)$$

for all $x, y \in X$. Then, there exists a unique point $x \in X$ such that $f(x)x$. Moreover, if $x_0 \in X$, and $x_{n+1} = f(x_n)$, then $x = \lim_{n\to\infty} x_n$.

Exercises

1. Show that if a sequence $\{a_n\}$ is convergent, then its limit is unique. That is, if $a_n \to a$ and $a_n \to b$ as $n \to \infty$ then $a = b$.

2. Prove that every subsequence of a convergent sequence converges and it converges to the same limit.

2.5 Continuity

Definition (Continuity): Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be two topological spaces and $f : X \to Y$ be a function. We say f is continuous if for every open subset V of Y the pre-image of it is an open set in X.

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

Furthermore, f is continuous at a point $x \in X$ when for all subset W of Y that $f(x)$ is an internal point of, there is an open set U containing x such that $\{f(y) \,|\, y \in U\} \subset W$. In other words x is an internal point of $f^{\text{pre}}(W)$.

Proposition 2.19. *f is continuous if and only if f is continuous at every point $x \in X$.*

Proof. Firstly, if f is continuous we show that f is continuous at every point $x \in X$. Let V be an open set around $f(x)$ then $x \in f^{\text{pre}}(V)$ must be an internal point since $f^{\text{pre}}(V)$ is open. Secondly, if f is continuous at every point $x \in X$ then f is continuous. Let $V = \{f(x) \,|\, x \in U\}$ be an open set in Y. For any $x \in U$, $f(x)$ is an internal point of V and since f is continuous at x, x is an internal point of U which means every point $x \in U$ is an internal point of U and thus $U = f^{\text{pre}}(V)$ is open. ■

Theorem 2.20 ($\epsilon - \delta$ condition). *Let (X, d) and (Y, d') be two metric space. Continuity at a point x is equivalent to the existence a $\delta > 0$ for all $\epsilon > 0$ such that:*

$$d(x, y) < \delta \implies d'(f(x), f(y)) < \epsilon$$

Proof. Let $V = \{f(y) \,|\, d'(f(x), f(y)) < \epsilon\}$ then V is open and hence $f(x)$ is an internal point of V. By continuity at point x, x must be an internal point of $f^{\text{pre}}(V)$. In other words, there exists a $\delta > 0$ such that $U = \{y \,|\, d(x, y) < \delta\} \subset f^{\text{pre}}(V)$. Take an open set $U \subset Y$, then assuming the $\epsilon - \delta$ condition, we will show that $f^{\text{pre}}(U)$ is open. Let $y \in U$ then there is $x \in f^{\text{pre}}(U)$ such that $f(x) = y$. From openness of U, there is a $\epsilon > 0$ such that $B_\epsilon(y) \subset U$, also by continuity condition, there exists a $\delta > 0$ such that:

$$d(x, z) < \delta \implies d'(f(x), f(z)) < \epsilon$$

The openness of $f^{\text{pre}}(U)$ is equivalent to $B_\delta(x) \subset f^{\text{pre}}(U)$, which clearly holds, since for any $z \in B_\delta(x) \implies f(z) \in B_\epsilon(y) \subset U$. ■

Example 2.4. Let (X, d) be a metric space with $d(x, y)$ being the discrete metric, $f : X \to X'$ where (X', d') is an arbitary metric space. Then f is always continuous. Since for every point a the open ball $B_{\frac{1}{2}}(a) = \{a\}$, and union of open sets is an open set itself, then every subset of X is open.

Equivalently, f is continuous at a if for all $\epsilon > 0$, a is an internal point of $f^{\text{pre}}(B_\epsilon(f(a)))$. That is there exists $\delta > 0$ such that, $B_\delta(a) \subset f^{\text{pre}}(B_\epsilon(f(a)))$. More generally, if X has the discrete topolgy or X' has the trivial topolgy, then $f : X \to X'$ is always continous.

Proposition 2.21. *Let $f : (X, \mathscr{T}_1) \to (X, \mathscr{T}_2)$ be the identity function. f is continous if and only if $\mathscr{T}_1$ is stronger that $\mathscr{T}_2$.*

Proof. If f is continous, then every open set $V \in \mathscr{T}_2$ is an open set in $\mathscr{T}_1$. Hence, $\mathscr{T}_2 \subset \mathscr{T}_1$. If $\mathscr{T}_2 \subset \mathscr{T}_1$ and V is an open set in $\mathscr{T}_2$, then $V = f^{\text{pre}}(V) \in \mathscr{T}_1$ is open and thus f is continous. ■

Theorem 2.22. *Let (X, d) and (X', d') be two metric spaces and $f : X \to X'$. f is continuous at $a \in X$ if and only if for every sequence $\{a_n\}$ in X with $a_n \to a$ we have $f(a_n) \to f(a)$.*

Proof. Let f be continuous at a and $a_n \to a$. From continuity of f, for each given ϵ, there is a δ such that:

$$d(x, a) < \delta \implies d'(f(x), f(a)) < \epsilon$$

From the convergence of $\{a_n\}$, for each given δ, there is a N such that:

$$\forall n \geq N \implies d(a_n, a) < \delta$$

By merging these two equations we will get:

$$\forall n \geq N \implies d(a_n, a) < \delta \implies d'(f(a_n), f(a)) < \epsilon$$

which was what was wanted.

If f is not continuous, there must be an $\epsilon > 0$ that for all $\delta > 0$, for some $x \in B_\delta(a)$, $d'(f(x), f(a)) \geq \epsilon$. Especially, for each $n \in \mathbb{N}$, let $\delta = \frac{1}{n}$ and x_n have the described property. Since $x_n \to a$ by our assumption $f(x_n) \to f(a)$, which is a contradiction and thus f is continuous. ■

Definition: Let X and Y be two topological spaces and $f : X \to Y$. f is a **homeomorphism** of X to Y if f is bijective and, f and f^{-1} are continous. Furthermore, X and Y are **homeomorphic** if there exists a homeomorphism between them.

Exercises

1. Let $(X, d), (X', d')$, and (X'', d'') be metric spaces and $f : X \to X'$, $g : X' \to X''$ be two functions. If f is continous at a and g is continous at $f(a)$, then $g \circ f$ is continuous at a.

2. Let $(X_i, d_i),\ i = 1, \ldots k$ be metric spaces. Define D to be any of the three discussed metric over $X = X_1 \times X_2 \times \ldots \times X_k$. Then, the projection function, $\pi_j(x) : X \to X_j$ is continuous for all j.
$$\pi_j(x_1, x_2, \ldots, x_n) = x_j$$

3. Let (X, D) be defined as above and let (X', d') be a matic space, and $f : X' \to X$. f is continous at $a' \in X'$ if and only if $\pi_j \circ f$ is continuous for all $j = 1, \ldots, k$.

4. The four algebraic operations are continuous on their domain.
$$+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad +(x, y) = x + y$$
$$- : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad -(x, y) = x - y$$
$$\times : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad \times(x, y) = x \times y$$
$$\div : \mathbb{R} \times (\mathbb{R} - \{0\}) \to \mathbb{R}, \quad \div(x, y) = x \div y$$
Where the metric of $\mathbb{R}$ on the right hand side is the common Euclidean metric, and on the left hand side is any of the three metric.

2.6 Covering compactness

Definition: Let X be a topological space. A **covering** for s set $E \subset X$ is a collection of U_α of open subsets of X such that $E \subset \bigcup U_\alpha$.

Definition: A **subcovering** of $\mathscr{U} = \{U_\alpha \mid \alpha \in A\}$ is a collection $\mathscr{V} = \{U_\alpha \mid \alpha \in B\}$ where $B \subset A$ and $\mathscr{V}$ covers the same set.

Definition: A subset K of X is called **compact** if every open cover of K has a finite subcover.

Example 2.5. $\mathscr{U} = \{]x-1, x+1[\mid x \in \mathbb{R}\}$ is covering for $\mathbb{R}$. It Obviously has a countable subcovering, however, it does not have finite subcovering.

Theorem 2.23. *Closed and bounded intervals in* $\mathbb{R}$ *are compact.*

Proposition 2.24. *If X is a compact space and $\{K_n\}$ is a sequence of non-empty closed subsets of X, with $K_{n+1} \subset K_n$ for all n, then $\bigcap_{n=1}^{\infty} K_n$ is non-empty.*

Proposition 2.25. *If E is an infinite subset of a compact set, then E has a limit point in K.*

Proposition 2.26. *A subset of a topological space is compact if and only if it is compact in iteself with the relative topology.*

Proposition 2.27. *If X is a Hausdorff space and $K \subset X$ is compact, then K is closed.*

Definition: Let X be a metric space and E a subset of X. The **diameter** of E is defined to be $\operatorname{diam} E = \sup\{d(x,y) \mid x, y \in E\}$. E is said to be bounded if its diameter is finite.

Definition: Let E be subset of a metric space X. E is **totally bounded** if for all $\epsilon > 0$, there exists a finite subset $\{x_1, \ldots, x_n\}$ of X such that $E \subset \bigcup_{k=1}^{n} B_\epsilon(x_k)$. A set F such that $E \subset \bigcup_{x \in F} B_\epsilon(x)$ is called an ϵ-**net** for E.

Proposition 2.28. *If E is totally bounded, then:*

1. *E is bounded.*
2. *$\operatorname{cl} E$ is totally bounded.*
3. *Any subset of E is totally bounded.*

Theorem 2.29. *A metric space X is complete if and only if X is complete and totally bounded.*

Definition: A subset E of topological space is **relatively compact** or **precompact** if $\operatorname{cl} E$ is compact.

Corollary 2.30. *A subset of a complete metric space is relatively compact if and only if it is totally bounded.*

Theorem 2.31. *A set E in $\mathbb{R}^k$ is closed and bounded if and only if it is compact.*

Note: This is generally not true for other metric spaces.

Theorem 2.32 (Weierstrass). *Every bounded infinite subset of $\mathbb{R}^k$ has a limit point in $\mathbb{R}^k$.*

Exercises

1. Show that $\mathbb{Q} \cap [0,1]$ is not covering compact, directly from the definition.

2.7 Sequential compactness

A subset $K \subset X$ is **compact** if it has the Boltzano-Weierstrass property, if for all sequences $\{a_n\}$ in K there exists a subsequence of $\{a_n\}$ that converges to a point $a \in K$.

Theorem 2.33. *Compactness is equivalent to the existence a finite subcovering for every covering.*

Proof. To prove the theorem, let us define:

We will show for metric spaces compactness is equivalent to covering compact. lebegue number ■

Let $\{a_n\}$ be a sequence in $\mathbb{R}$. We define:

$$\limsup a_n = \overline{\lim}\, a_n = \lim_{n\to\infty}\Big(\sup\{a_k : k \geq n\}\Big)$$
$$\liminf a_n = \underline{\lim}\, a_n = \lim_{n\to\infty}\Big(\inf\{a_k : k \geq n\}\Big)$$

Note: The limits, $\limsup a_n$ and $\liminf a_n$, always exists. Albeit they might be infinite.

Let $\{a_n\}$ be a bounded sequence in $\mathbb{R}$, and A^* is the set of all limit points of all subsequence of $\{a_n\}$. We know that A^* is not empty and since $\{a_n\}$ is bounded and then A^* must be bounded as well. Thus, by completeness axiom, A^* has infimum and supremum. Moreover, $\sup A^*, \inf A^* \in A^*$.

Proposition 2.34. *A bounded sequence $\{a_n\}$ is convergent if and only if $\limsup a_n = \liminf a_n$.*

Corollary 2.35. *If K is a compact subset of $\mathbb{R}$ then K has minimum and maximum. That is, there are $M, m \in K$ such that $\forall x \in K,\ m \leq x \leq M$.*

Proof. Since K is bounded then it has supremum and infimum in $\mathbb{R}$. Obviously, there are convergent sequences $\{a_n\}$ and $\{b_n\}$ such that $a_n \to m = \inf K$ and $b_n \to M = \sup K$. By compactness of K, $M, m \in K$. ■

Theorem 2.36. *(X, d) and (X', d') are metric spaces and $K \subset X$ is compact. If $f : X \to X'$ is continuous, then $f(K)$ is a compact subset of X'.*

Proof. Let $\{y_n\} \in f(K)$ and $\{x_n\} \in K$ are such that $f(x_n) = y_n$. Since K is compact there is a convergent subsequence $\{x_{n_k}\}$ and since f is continous $\{y_{n_k} = f(x_{n_k})\}$ is also convergent. Hence $f(K)$ is compact. ■

Corollary 2.37. *Let (X, d) be a metric space and $f : X \to \mathbb{R}$ is continuous. If K is a compact subset of X. Then f attains maximum and minimun in $\mathbb{R}$.*

Note: For a continuous function $f : X \to X'$ it is not necessary that the image of an open/closed set to be open/closed.

Definition (Uniform continuity): Let (X, d) and (X', d') be metric spaces. $f : X \to X'$ is uniformly continuous if:

$$\forall \epsilon > 0\ \exists \delta > 0,\ x, y \in X,\ d(x, y) < \delta \implies d'(f(x), f(y)) < \epsilon$$

Proposition 2.38. *$f : X \to X'$ is uniformly continous if and only if for every pair sequence $\{(x_n, y_n)\}$ in X satisfying $d(x_n, y_n) \to 0$ we have $d'(f(x_n), f(y_n)) \to 0$.*

Proof. Necessity: We have

$$\begin{aligned}
&\forall \epsilon\ \exists \delta \text{ s.t. } \forall x, y \in X, d(x, y) < \delta \implies d'(f(x), f(y)) < \epsilon \\
&\forall \delta\ \exists N \text{ s.t. } n \geq N \implies d(x, y) < \delta
\end{aligned}$$

combining the two brings us at the conclusion.

Sufficiency: Suppose for the sake of contradtion that:

$$\exists \epsilon\ \forall \delta\ \exists x, y \in X \text{ s.t. } d(x, y) < \delta \wedge d'(f(x), f(y)) \geq \epsilon$$

then let $\delta = \frac{1}{n}$ and make the sequence pair $\{(x_n, y_n)\}$. Clearly, $d(x_n, y_n) \to 0$ therefore, $d'(f(x), f(y)) \to 0$. Which is a contradition since $d'(f(x), f(y)) \geq \epsilon$. ■

Proposition 2.39. *(X, d) and (X', d') are matric spaces and X is compact. If $f : X \to X'$ is continuous then it is uniformly continuous.*

Proof. Similarly, for the sake of contradicition suppose

$$\exists \epsilon\ \forall \delta\ \exists x, y \in X \text{ s.t. } d(x, y) < \delta \wedge d'(f(x), f(y)) \geq \epsilon$$

and let $\delta = \frac{1}{n}$ and make the sequence pair (x_n, y_n). By compactness of X, there are two convergent subsequence $\{x_{n_k}\}$ and $\{y_{n_k}\}$. Since $d(x_n, y_n) \to 0$ then if $x_{n_k} \to x$, $y_{n_k} \to x$ as well. By continuity of f, $f(x_{n_k}) \to f(x)$ and $f(y_{n_k}) \to f(x)$ and thus $d'(f(x_{n_k}), f(y_{n_k})) \to 0$. Which is a contradicition as for sufficiently large K, $k \geq K \implies d'(f(x), f(y)) \geq \epsilon$ ■

Exercises

1. Prove that $\sqrt{|\ x\ |} : \mathbb{R} \to \mathbb{R}$ is uniformly continuous.

2.8 Connectedness

Definition: Let X be a topological space. X is **disconnected** if there are non-empty open sets A and B are found such that

$$A \bigcap B = \emptyset, \quad A \bigcup B = X$$

X is said to be **connected** if it is not disconnected. A subset S of X is connected if it is connected when considering its relative topolgy.

Definition:

Example 2.6. The following subsets of $\mathbb{R}$ are disconnected:

1. $S = [-1, 0[\ \cup\ ]0, 1]$.
2. $\mathbb{Q}$.
3. $S = [1, 0] \cup [1, 2]$.

Definition: Two sets A and B in a topological space X are said to be **separable** if $A \cap \operatorname{cl} B = \operatorname{cl} A \cap B = \emptyset$.

Proposition 2.40. *X is connected if and only if it can not be written in form of two non-empty separable sets.*

Definition: $S \subset \mathbb{R}$ is an **intervals** if when $a, c \in S$ and $a < b < c$ then $b \in S$.

Example 2.7. $\mathbb{R}$ and its intervals are connected. In fact the only connected subsets of $\mathbb{R}$ are its intervals.

Theorem 2.41. *Let X and X' be two topological spaces. Let $f : X \to X'$ be continuous and S be a connected subset of X. Then, $f(S)$ is connected in X'.*

Corollary 2.42 (Mean value theorem). *If $f : [a, b] \to \mathbb{R}$ is a continous function and $f(a) = A, f(b) = B$ then for every C between A and B there exists a $c \in [a, b]$ such that $f(c) = C$.*

Proposition 2.43. *If $S \subset X$ is a connected set then every $S \subset T \subset \bar{S}$ is connected.*

Proposition 2.44. *Let $\{E_\alpha\}$ be collection of connected sets with $\cap_\alpha E_\alpha \neq \emptyset$. Then, $\cap_\alpha E_\alpha$ is connected.*

Definition: Let x be a point in a topological space X. The **connected component** of x, C_x, is the union of the all the connected set including x.

Proposition 2.45. *For any $x, y \in X$*

1. *C_x is connected and closed.*
2. *C_x and C_y are either disjoint or equal to each other.*

Definition: A topological space X is said to be **totally disconceted** if $C_x = \{x\}$ for all $x \in X$. A subset S of X is totally disconceted if it is totally disconceted when considering its relative topology.

Definition: The *graph* of a function $f : M \to N$, G_f, given by $G_f = \{(x, f(x)) \,|\, x \in M\}$.

Theorem 2.46. *The graph of a continous function over a connected set is connected.*

Example 2.8. Topological curve is connected and also its closure is connected.

Definition (Path connected): A set S is path connected if for every pair of points $p, q \in S$ there exists a continous function $\gamma : [a, b] \to S$ such that $\gamma(a) = p$ and $\gamma(b) = q$.

Theorem 2.47. *If a set S is path connected, then it is connected but the inverse is not true.*

Example 2.9. Infinite broom is path connected but toplogical sine curve is not.

Proposition 2.48. *If f is continuous function and S is a path connected set, then the image of S under f is path connected.*

Proposition 2.49. *Every open set of $\mathbb{R}$ is the union of countably many disjoint open intervals.*

2.9 Cantor set

Definition: define cantor set

Proposition 2.50. *Cantor set is a perfect set.*

Proposition 2.51. *Cantor set is totally disconnected.*

Theorem 2.52. *Let K be a complete, totally disconnected, and compact metric space. Then, K is homeomorphic to Cantor set.*

Theorem 2.53. *Let P be a non-empty perfect set in $\mathbb{R}^k$. Then, P is uncountable.*

Chapter 3

Differentiation

Definition: Let I be an interval in $\mathbb{R}$. If a is an interior point of I, then we say that $f : I \to \mathbb{R}$ is differentiable at a when the following limit exists:

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

The limit, if exists, is denoted by $f'(a)$. If a is an end point and the length of the interval is greater than zero, then the limit only exists from one direction.

Equivalently, there exists a line l, not parallel to y-axis, in form of $l : A(x) = mx + b$, that is tangent to f at $x = a$. In this case:

$$\lim_{x \to a} \frac{f(x) - A(x)}{x - a} = 0 \qquad\qquad A(a) = f(a)$$

In a general case, two functions f, g are tangent to each other at $x = a$ if:

$$\lim_{x \to a} \frac{f(x) - g(x)}{x - a} = 0 \qquad\qquad f(a) = g(a) \tag{3.1}$$

Corollary 3.1.

1. *f is differentiable at a if it is continuous at a.*
2. *If $f'(a) > 0$, there exists $\delta > 0$ such that for $x \in \ ]a - \delta, a[\ \cap I \implies f(x) < f(a)$ and for $x \in \ ]a, a + \delta[\ \cap I \implies f(x) < f(a)$. And if $f'(a) < 0$ the inequality sign are reversed. Therefore, if f has a local extremum at a, then in case $f'(a)$ exists, $f'(a) = 0$.*

Example 3.1. – a function that its derivate is not continuous (with $\sin \frac{1}{x}$).

Theorem 3.2 (Rolle's theorem). *Let $f : [a, b] \to \mathbb{R}$ be a continuous and differentiable on the interval. If $f(a) = 0, f(b) = 0$, then there exists $c \in \ ]a, b[$ such that:*

$$f'(c) = 0$$

Proof. If $f \equiv 0$ on $[a, b]$ then its derivative $f'(x) \equiv 0$ on $[a, b]$. If $f(x) \neq 0$ for some $x \in]a, b[$ then it must have a non-zero maximum or minimum at some $c \in]a, b[$. Since $[a, b]$ is compact then by continuity of f, $f([a, b])$ is also compact in $\mathbb{R}$ and therefore f attains its maximum or minimum. We know that at least one of its extremities must lie in $]a, b[$, say point c, hence by Item 2 $f'(c) = 0$. ■

Theorem 3.3 (Mean value theorem). *Let $f : [a,b] \to \mathbb{R}$ be a continuous and differentiable on the interval,then there exists $c \in\ ]a,b[$ such that:*

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Proof. Define $g(x) = f(x) - f(a) - \dfrac{f(b) - f(a)}{b - a}(x - a)$. Then it is clear that $g(a) = g(b) = 0$ and g is continous and differentiable on the interval. Then, by Theorem 3.2 there exists $c \in]a,b[$ such that $g'(c) = 0$. Equivalently:

$$\begin{aligned} g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} &= 0 \\ \implies f'(c) &= \frac{f(b) - f(a)}{b - a} \end{aligned}$$

which concludes the proof. ■

Corollary 3.4 (Growth Estimate). *If $|f'(x)| \leq M$ in $]a,b[$ then f satisfies the global lipschitz condition for all $x, y \in [a,b]$ $|f(x) - f(y)| \leq M|x - y|$.*

Corollary 3.5. *Let $f : [a,b] \to \mathbb{R}$ is continuous and $f'(x) < 0$ – or $f'(x) > 0$– for all $x \in]a,b[$ then f is strictly increasing –or decreasing– on $[a,b]$.*

Theorem 3.6. *$f : [a,b] \to \mathbb{R}$ is continuous and differentiable on $]a,b[$ then for $f'(]a,b[)$ the intermediate value theorem holds and thus it is an interval.*

Proof. Let $x_1, x_2 \in]a,b[$. WLOG assume $f'(x_1) < f'(x_2)$, we wish to prove that for all $y^* \in]f'(x_1), f'(x_2)[$ there is a $x^* \in]x_1, x_2[$ such that $f(x^*) = y^*$. Put $g(x) = f(x) - y^* x$. By differentiability of f on $[a,b]$, g is differentiable on $[a,b]$. Then, $g'(x_1) = f'(x_1) - y^* < 0$ and $g'(x_2) = f'(x_2) - y^* > 0$, therefore there are $t_1, t_2 \in]x_1, x_2[$ such that $f(t_1) < f(x_1)$ and $f(t_2) < f(x_2)$. Since g is continuous on $[x_1, x_2]$ then it must attains its minimum at some $x^* \in [x_1, x_2]$. However x^* can't be x_1 or x_2 and hence $x^* \in]x_1, x_2[$. It is then easy to see that $f'(x^*) = y^*$. ■

Definition (Darboux continous): A function f is Darboux continous if it posseses the intermediate value property.

For example f' of differentiable function is Darboux continuous.

Theorem 3.7 (Cauchy's mean value theorem). *$f, g : [a,b] \to \mathbb{R}$ are continuous then there exists a $c \in\ ]a,b[$, such that:*

$$f'(c)(g(b) - g(a)) = g'(c)(f(b) - f(a))$$

Proof. Define $h(x) = (f(x) - f(a))(g(b) - g(a)) - (g(x) - g(a))(f(b) - f(a))$, then clearly $h(a) = 0, h(b) = 0$ and $h(x)$ is continous and differentiable on $[a,b]$. Hence by applying the theorem 3.2 for some $c \in]a,b[$ we have:

$$\begin{aligned} &h'(c) = 0 \\ &\implies f'(c)(g(b) - g(a)) - g'(c)(f(b) - f(a)) = 0 \\ &\implies f'(c)(g(b) - g(a)) = g'(c)(f(b) - f(a)) \end{aligned}$$

■

Theorem 3.8 (L'Hopital's rule). *Suppose that* $\lim_{x\to a^+} f(x) = 0, \lim_{x\to a^+} g(x) = 0$ *where* f, g *are differentiable on a open interval* $I =]a, b[$ *for some* b *such that* $g'(x) \neq 0$ *in* I *except maybe at* $x = a$ *and the limit*

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

exists, then:

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

Proof. For a fixed $\epsilon > 0$ there exists a $\delta > 0$ such that:

$$|x - a| < \delta \implies \left|\frac{f'(x)}{g'(x)} - L\right| < \frac{\epsilon}{2}$$

then since $f(t), g(t) \to 0$ as $t \to a^+$ from right side then there must be a $t \in]a, x[$ such that

$$\left|\frac{f(x) - f(t)}{g(x) - g(t)} - \frac{f(x)}{g(x)}\right| < \frac{\epsilon}{2}$$

then simply:

$$\begin{aligned}
\left|\frac{f(x)}{g(x)} - L\right| &\leq \left|\frac{f(x)}{g(x)} - \frac{f(x) - f(t)}{g(x) - g(t)}\right| + \left|\frac{f(x) - f(t)}{g(x) - g(t)} - L\right| && (3.2)\\
&< \frac{\epsilon}{2} + \left|\frac{f'(\theta)}{g'(\theta)} - L\right| && (3.3)\\
&< \epsilon && (3.4)
\end{aligned}$$

Note that $\theta \in \;]t, x[$ and thus $|\theta - a| < \delta$ ∎

Definition (Higher order derivatives): f is said to be r_{th}-differentiable at x if it is differentiable r times. The r_{th} derivative of f is denoted as $f^{(r)}$. If $f^{(r)}$ exists for all r and x then f is said to be infinitely differentiable or smooth.

Definition (Smoothness classes): The set of all f is continuosly r_{th}-differentiable is called class $\mathcal{C}^r$.

Definition (Taylor polynomial): The r_{th}-order Taylor polynomial of an r_{th}-order differentiable function f at x is

$$P_r(x, h) = f(x) + f'(x)h + \frac{f''(x)}{2}h^2 + \cdots + \frac{f^{(r)}(x)}{r!}h^r = \sum_{n=0}^{r} \frac{f^{(n)}(x)}{n!}h^n$$

Theorem 3.9 (Taylor approximation theorem). *Let* f *be a* r*-differentiable function at* x *then:*

1. $$\frac{f(x + h) - P_r(x, h)}{h^r} \to 0 \text{ as } h \to 0$$

2. *and* P_r *is the only* r_{th} *degree polynomial that has such property.*

3. *Furthermore, if f is r-differentiable on an interval I for every $x, y \in I$, there exists ξ between x, y such that:*

$$f(y) - P_{r-1}(x, y-x) = \frac{f^{(r)}(\xi)}{(r)!}(y-x)^r$$

Proof. 1. For the base case $r = 1$

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - f'(x)h}{h} = f'(x) - f'(x) = 0$$

and by induction we prove the case $r = n \geq 2$

$$\begin{aligned}&\lim_{h \to 0} \frac{f(x+h) - P_n(x,h)}{h^n} = 0 \\ &\iff \forall \epsilon > 0,\ \exists \delta > 0 \text{ such that } \ |h| < \delta \implies |f(x+h) - P_n(x,h)| < \epsilon |h^n|\end{aligned}$$

Let $g(h) = f(x+h) - P_n(x,h)$ then since both $f(x+h)$ and $P_n(x,h)$ are differentiable then we apply Theorem 3.3

$$\begin{aligned} g(h) - g(0) &= g'(c) \\ &= f'(x+c) - \sum_{k=1}^{n} \frac{f^{(k)}(x)}{(k-1)!} c^{k-1} \\ &= f'(x+c) - \sum_{k=0}^{n-1} \frac{f^{(k+1}(x)}{k!} c^{k} \\ &= f'(x+c) - \sum_{k=0}^{n-1} \frac{f'^{(k)}(x)}{k!} c^{k} \end{aligned}$$

for some $c \in]0, h[$. Note that f' is $(n-1)$-differentiable at x. Thus by induction for any $\epsilon > 0$ there exists a δ such that if $c < \delta$ then:

$$\left| f'(x+c) - \sum_{k=0}^{n-1} \frac{f'^{(k)}(x)}{k!} c^k \right| < \epsilon \left| c^{n-1} \right|$$

which means

$$\begin{aligned} |f(x+h) - P_n(x,h)| = |g(h)| &= |h| \left| f'(x+c) - \sum_{k=0}^{n-1} \frac{f'^{(k)}(x)}{k!} c^k \right| \\ &< |h| \epsilon \left| c^{n-1} \right| < \epsilon |h^n| \end{aligned}$$

Therefore, for any ϵ if $h < \delta$ then $c < \delta$ and the result holds.

2. Let $Q_r(x,h)$ be another r_{th} degree polynomial such that

$$\lim_{h \to 0} \frac{f(x+h) - Q_r(x,h)}{h^r} = 0$$

then

$$\lim_{h \to 0} \frac{P_r(x,h) - Q_r(x,h)}{h^r} = 0$$

however this can only happen if $Q_r(x,h) = P_r(x,h)$.

3. Again for the base case $r = 1$

$$f(y) - f(x) = f'(\xi)(y - x)$$

which is the Theorem 3.3. for $r = n$ we have that

$$g(h) = f(x+h) - P_{n-1}(x,h) + Ch^n \implies g(0) = g'(0) = \cdots = g^{(n-1)}(0) = 0$$

Set C such that $g(y-x) = 0$. Then by applying Theorem 3.2 $(n-1)$ times

$$\begin{aligned}
&g(0) = g(y-x) = 0 &&\implies g'(c_1) = 0 && c_1 \in \,]0, y-x[\\
&g'(0) = g'(c_1) = 0 &&\implies g'(c_2) = 0 && c_2 \in \,]0, c_1[\\
& && \vdots && \\
&g^{(n-2)}(0) = g^{(n-2)}(c_{n-2}) = 0 &&\implies g^{(n-1)}(c_{n-1}) = 0 && c_{n-1} \in \,]0, c_{n-2}[\\
&g^{(n-1)}(0) = g^{(n-1)}(c_{n-1}) = 0 &&\implies g^{(n)}(\xi - x) = 0 && \xi - x \in \,]0, c_{n-1}[
\end{aligned}$$

Expanding $g^{(n)}(\xi - x)$ gives the following.

$$\begin{aligned}
&\implies g^{(n)}(\xi - x) = f^{(n)}(\xi) + Cn! \\
&\implies f^{(n)}(\xi) - \frac{n!}{(y-x)^n}(f(y) - P_{n-1}(x, y-x)) = 0 \\
&\implies f(y) - P_{n-1}(x, y-x) = \frac{f^{(n)}(\xi)}{n!}(y-x)^n \qquad \xi \in \,]x, y[
\end{aligned}$$

which completes the proof. ∎

Theorem 3.10 (Inverse function). *Let I be an open set and $f : I \to \mathbb{R}$ is continuous and differentiable such that its derivate is non-zero. Thus, f is either monotonic. Furthermore, it is one to one then it has a differentiable inverse f^{-1}:*

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

Proof. limit algebra ∎

Part II

Multivariate Analysis

Chapter 4

Linear Algebra

4.1 Vector Spaces

Definition (Normed vector space): Let V be a vector space. A **norm** is a real valued function $\|\cdot\| : V \to \mathbb{R}$ which has the following properties

1. $\forall x \in V,\ \|x\| > 0.$
2. $\|x\| = 0 \implies x = 0.$
3. $\forall x \in V\ \forall \alpha \in \mathbb{F},\ \|\alpha x\| = |\alpha| \|x\|.$
4. $\forall x, y \in V\ \|x + y\| \leq \|x\| + \|y\|.$

Each normed vector space induces a metric space (V, d) where $d(x, y) = \|y - x\|$.

Theorem 4.1. *In every normed space $(V, \|\cdot\|)$ we have*

$$|\|v\| - \|w\|| \leq \|v - w\|$$

Hence the norm is Lipschitz continuous.

Definition: Assume V is a vector space and let $\|\cdot\|_1,\ \|\cdot\|_2$ be two norms for V. They are said to be equivalent when

$$\exists c_1, c_2 > 0\ \forall x : \qquad c_1\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|$$

To check if the above definition is indeed an equivalence relation, we must show that following:

Reflexive $\|\cdot\|_1 \sim \|\cdot\|_1$.

Symmetric $\|\cdot\|_1 \sim \|\cdot\|_2 \implies \|\cdot\|_2 \sim \|\cdot\|_1$.

Transitive $\|\cdot\|_1 \sim \|\cdot\|_2,\ \|\cdot\|_2 \sim \|\cdot\|_3 \implies \|\cdot\|_1 \sim \|\cdot\|_3$.

Remark 2. Equivalent norms induce equivalent metrics, hence they induce the same topology.

Theorem 4.2. *All norms defined on a finite dimensional vector space V are equivalent.*

Proof. Let $\|\cdot\|$ be an arbitrary norm on V and $\{e_1, e_2, \ldots, e_n\}$ be a basis of V. Let $\|\cdot\|_2$ be L_2-norm (Euclidean norm). It will suffice to show $\|\cdot\| \sim \|\cdot\|_2$. Let

$$M = \max(\|e_1\|, \ldots, \|e_n\|)$$

Take $x \in V$, writing $x = \sum_{i=1}^n \xi_i e_i$ we have:

$$\|x\| = \left\| \sum_{i=i}^{n} \xi_i e_i \right\| \leq \sum_{i=1}^{n} |\xi_i| \|e_i\| \leq M\sqrt{n}\|x\|_2$$

Taking $c_2 = M\sqrt{n}$ proves the right inequality. For the left inequality we need the following lemma

Lemma 4.3. *If V is a normed vector space with $\|\cdot\|_2$, as defined above, is viewed as metric space $(V, \|\cdot\|_2)$ then $\|\cdot\| : V \to \mathbb{R}$ is continuous.*

Proof. Let $x_0 \in V$ and M be defined as above. For any $\epsilon > 0$ consider $\delta = \frac{\epsilon}{M\sqrt{n}}$ then if $\|x - x_0\|_2 < \delta$

$$|\|x\| - \|x_0\|| \leq \|x - x_0\| \leq M\sqrt{n}\|x - x_0\| \leq \epsilon$$ □

Now consider the sphere of radius $r = 1$ centered at 0, $S_1(0) = S_1 = \{x \in V : \|x\|_2 = 1\}$. One can show that S is compact (Theorem 4.4). Therefore, $\|x\|$ assumes its minimum on S. Let $a = \|x_0\|$ be the minimum. Since $0 \notin S$ then $a > 0$. By letting $y = x/\|x\|_2$, we have $y \in S$ and thus $a \leq \|y\|$ which is

$$a\|x\|_2 \leq \|x\|$$

Taking $c_1 = a$ proves the theorem. ■

Theorem 4.4. *Let $(V, \|\cdot\|)$ be a normed space over a normed complete field $\mathbb{F}$. The following are equivalent*

1. *V is finite dimensional.*
2. *every bounded closed set in V is compact.*
3. *the closed unit ball in V is compact.*

Proof. Item 1 $\implies$ Item 2: It is similar to proving a closed set $\mathbb{R}^n$ is compact using the fact a closed interval is compact in $\mathbb{R}$.

Item 2 $\implies$ Item 3: Trivial.

Item 3 $\implies$ Item 1: Requires the following lemma:

Lemma 4.5 (Riesz's lemma). *If V is a normed vector space and W is a closed proper space of V and $\alpha \in \mathbb{R}$ with $0 < \alpha < 1$, then there exists an $v \in V$ with $\|v\| = 1$ such that $\|v - w\| \geq \alpha$ for all $w \in W$.*

Now suppose V were to be an infinite dimensional vector space. Then by the Lemma 4.5 there is sequence of unit vectors x_n such that $\forall m, n \in \mathbb{N},\ \|x_n - x_m\| > \alpha$ for some $0 < \alpha < 1$. Which implies that no subsequence of $\{x_n\}$ is convergent and hence the closed unit ball can not be compact. ■

Example 4.1. The closed unit ball in the infinite dimensional vector space $C([0,1],\mathbb{R})$ with $\|f\| = \max f(x)$ is not compact. Take $f_n(x) = x^n$. Obviously $\|f_n\| = 1$, however f_n doesn't uniformly converge and hence f_n doesn't have a limit in $C([0,1],\mathbb{R})$ with the max norm. Consider the following norm

$$\|f\|_I = \int_0^1 |f(x)|\,\mathrm{d}x$$

Note that $\|\cdot\|_I$ and $\|\cdot\|_{\max}$ are not equivalent. Let $g(x) = 0$ for all $x \in [0,1]$. Then

$$\|f_n - g\|_I = \frac{1}{n+1} \to 0 \quad \mathrm{as} n \to \infty.$$

Definition (Banach space): A normed vector space V that is complete is a **Banach space**. A **Hilbert Space** is a Banach space whose norm is induced by an inner product.

Proposition 4.6. *A normed finite dimensional vector space V over a normed complete field $\mathbb{F}$, is Banach space.*

Proof. Let $\{v_i\} \in V$ be a Cauchy sequence, and $\{e_1,\ldots,e_n\}$ be a basis for V with the norm L^1, that is if $v = (\xi^1,\ldots,\xi^n)$ then $\|v\| = \sum_{m=1}^n |\xi^m|$. Then if $v_i = (\xi_i^1,\ldots,\xi_i^n)$

$$\left|\xi_i^m - \xi_j^m\right| \le \sum_{m=1}^n \left|\xi_i^m - \xi_j^m\right| \le \|v_i - v_j\| < \epsilon$$

then $\{\xi_i^m\}_i$ are a Cauchy sequence in $\mathbb{F}$ and hence they converge $\xi_i^m \to \xi^m$. Then, clearly $v_i \to v = (\xi^1,\ldots,\xi^n)$ as each component converges. ∎

Example 4.2. $\mathbb{Q}$ form a vector space itself over itself. It is finite dimensional as $\{\mathbb{1}_{\mathbb{Q}}\}$ is the basis, however the sequence

$$1,\ 1.4,\ 1.41,\ \ldots$$

does not converge even though it is Cauchy.

4.2 Linear Maps

Let V and W be a vector spaces over $\mathbb{F}$. A map $T : V \to W$ is **linear** if

$$T(x + \lambda y) = T(x) + \lambda T(y)$$

for all $x, y \in V$ and $\lambda \in \mathbb{F}$.

Definition: Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed spaces then, a linear transformation $T : V \to W$ is **bounded** if there exists a constant $C > 0$ such that

$$\|Tv\|_W \le C\|v\|_V$$

for all $v \in V$. We denote the set of all linear map from $V \to W$ as $\mathcal{L}(V,W)$ and the set of all bounded linear maps as $\mathcal{B}(V,W)$. If $T \in \mathcal{L}(V,W)$ is bijective such that $T^{-1} \in \mathcal{L}(V,W)$, then T is called an **isomorphism** and V, W are **isomorphic**. An operator $T \in \mathcal{L}(V,W)$ is called **isometric** if $\|Tv\|_W = \|v\|_V$ for all $v \in V$.

Definition: If $(V, \|\cdot\|_V), (W, \|\cdot\|_W)$ are normed spaces then the **operator norm** of a linear transformation $T : V \to W$ is

$$\|T\| = \sup\left\{\frac{\|Tv\|_W}{\|v\|_V}\middle| v \neq 0\right\}$$

Proposition 4.7. *Let $T : U \to V$ and $T' : V \to W$ be two linear transformations.*

$$\|T' \circ T\| \leq \|T\|\|T'\|$$

Proof. for an arbitrary non-zero $x \in U$

$$\|T' \circ T(x)\|_W \leq \|T'\|\|Tx\|_V \leq \|T'\|\|T\|\|x\|_U$$

which implies

$$\|T' \circ T\| \leq \|T\|\|T'\|$$

■

Theorem 4.8. *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed spaces and $T : V \to W$ be a linear transformation. The following are equivalent*

1. *$\|T\|$ is finite.*
2. *T is bounded.*
3. *T is Lipschitz continuous.*
4. *T is continuous at a point.*
5. *$\sup_{\|v\|_V=1}\|Tv\|_W < \infty$.*

Proof. item 1 ⇒ item 2: Obviously

$$\begin{aligned} \frac{\|Tv\|_W}{\|v\|_V} &\leq \|T\| \\ \implies \|Tv\|_W &\leq \|T\|\|v\|_V \end{aligned}$$

note that if $v = 0$ then $Tv = 0$ as well and thus the last inequality holds for all $v \in V$.

item 2 ⇒ item 3:

$$\|Tv - Tu\|_W = \|T(u - v)\|_W \leq C\|u - v\|_V$$

item 3 ⇒ item 4: Trivial.

item 4 ⇒ item 5: Let T be continuous at $u \in V$. Then there is a $\delta > 0$ such that

$$\|v - u\| < \delta \implies \|Tv - Tu\|_W = \|T(v - u)\|_W < 1$$

Now for an arbitrary non-zero v we have

$$\left\|\left(\frac{\delta v}{2\|v\|_V} + u\right) - u\right\|_V < \delta$$

Therefore

$$\left\|T\left(\frac{\delta v}{2\|v\|_V}\right)\right\|_W < 1$$
$$\left\|T\left(\frac{v}{\|v\|_V}\right)\right\|_W < \frac{2}{\delta}$$

item 5 ⇒ item 1: Let $v \in V$ be an arbitrary vector. Then

$$\sup\left\|T\left(\frac{v}{\|v\|_V}\right)\right\|_W < \infty$$
$$\implies \sup\frac{\|Tv\|_W}{\|v\|_W} < \infty$$ ■

Theorem 4.9. *If V is a finite dimensional normed vector space then any linear transformation $T : V \to W$ is continuous.*

Proof. Since V is finite dimensional, according to Theorem 4.2, any two norms are equivalent. Hence, take $\|\cdot\|_2$ to be Euclidean norm over a basis $\{e_1, \ldots, e_n\}$. Let x be such that $\|x\|_2 < \delta$ for some $\delta > 0$. Therefore, $|\xi_i| < \delta^2$

$$\|Tx\|_W = \left\|\sum_{i=1}^{n}\xi_i T(e_i)\right\|_W \leq \sum_{i=1}^{n}|\xi_i|\|T(e_i)\|_W \leq \delta^2 K$$

where $K = \max\|T(e_i)\|_W$. By letting $\delta = \sqrt{\frac{\epsilon}{K}}$ we proved continuity at 0 and hence the continuity by Theorem 4.8. ■

Another proof of Propostion 4.6

Proof. Let $\{e_1, \ldots, e_n\}$ be a basis for V and $\phi : V \to \mathbb{F}^n$ be the representation map for the basis. Since ϕ is a linear map and a bijection then ϕ is homeomorphism. Consider a Cauchy sequence $\{v_k\} \in V$ and let $x_k = \phi(v_k)$ then by continuity of ϕ and ϕ^{-1} we have

$$|x_i - x_j| = |\phi(v_i) - \phi(v_j)| \leq \|\phi\|\|v_i - v_j\| \leq \|\phi\|\left\|\phi^{-1}(x_i) - \phi^{-1}(x_j)\right\| \leq \|\phi\|\left\|\phi^{-1}\right\||x_i - x_j|$$

hence $\{x_k\}$ are Cauchy in $\mathbb{F}^n$ which by completeness of $\mathbb{F}$ implies that they are convergent, $x_k \to x$. Let $v = \phi^{-1}(x)$ then by the right side of the inequality $v_k \to v$. ■

Remark 3. As seen in the last proof, for a bijective linear transformation T

$$1 \leq \|T\|\left\|T^{-1}\right\|$$

Definition (Dual space): Let V be a normed space over the normed field $\mathbb{F}$, then the **topological/continuous dual space** of the normed space V is

$$V^* = \mathcal{L}(V, \mathbb{F})$$

elements of V^* are called **bounded functionals** on V.

Remark 4. Dual space is defined for all vector spaces, however, in analysis we study the topological dual space which only in the finite dimensional case coincide with the algebraic dual space.

Proposition 4.10. *For a finite dimensional normed vector space V, $\dim V^* = \dim V$.*

Proof. Let $\{e_1, \dots, e_n\}$ be a basis for V then, consider the following linear functions

$$e_1^*, \dots, e_n^* \in V^*$$

where

$$e_i^*(e_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

we claim that $\{e_1^*, \dots, e_n^*\}$ is a basis for V^*. It is easy to see that they are as for each j

$$\left[\sum_{i=1}^n c_i e_i^*\right] e_j = c_j$$

and for each $\phi \in \mathcal{L}(V, \mathbb{F})$ we have

$$\phi(e_j) = \alpha_i = \sum_{i=1}^n \alpha_i e_o^*(e_i)$$

hence $\dim V^* = n = \dim V$. ∎

Theorem 4.11. *For two normed vector spaces V, W, $(\mathcal{B}(V, W), \|T\|)$ is a normed vector space. Moreover, it is a Banach space when W is a Banach space.*

Proof. Clearly $\mathcal{B}(V, W)$ is a vector space. For its norm $\|T\|$ we have

1. $\|T\| \geq 0$ by definition.
2. if $\alpha \in \mathbb{F}_W$ then
$$\|\alpha T\| = \sup\left\{\frac{\|(\alpha T)v\|_W}{\|v\|_V}\middle| v \neq 0\right\} = |\alpha| \sup\left\{\frac{\|Tv\|_W}{\|v\|_V}\middle| v \neq 0\right\} = |\alpha|\|T\|$$
3. for the triangle inequality
$$\begin{aligned}\|T_1 + T_2\| &= \sup\left\{\frac{\|(T_1 + T_2)v\|_W}{\|v\|_V}\right\} \\ &\leq \sup\left\{\frac{\|T_1 v\|_W + \|T_2 v\|_W}{\|v\|_V}\right\} \\ &= \sup\left\{\frac{\|T_1 v\|_W}{\|v\|_V}\right\} + \sup\left\{\frac{\|T_2 v\|_W}{\|v\|_V}\right\} \\ &= \|T_1\| + \|T_2\|\end{aligned}$$

Suppose W is a Banach space and $\{T_i\} \in \mathcal{B}(V, W)$ is a Cauchy sequence. Then for all $v \in V$

$$\forall \epsilon \, \exists N \text{ s.t. } m, n > N \implies \|T_m v - T_n v\|_W \leq \|T_m - T_n\| \|v\|_V < \epsilon$$

$\{T_i v\}$ is a Cauchy sequence. Since W is complete then $T_i v \to Tv$ for some function T. We claim that T is a bounded linear map and is the limit of $T_i \to T$.

$$\begin{aligned}T(v + cu) &= \lim_{i\to\infty} T_i(v + cu) = \lim_{i\to\infty} T_i v + cT_i u \\ &= Tv + cTu\end{aligned}$$

Note that $|\|T_m\| - \|T_n\|| \le \|T_m - T_n\|$ and hence $\|T_i\|$ is a Cauchy in sequence in $\mathbb{R}$ that has a limit t. There exists a N such that $|\|T_n\| - t| < 1$ for all $n \ge N$.

$$\frac{\|Tv\|_W}{\|v\|_V} = \lim_{i\to\infty} \frac{\|T_i v\|_W}{\|v\|_V} < t + 1$$

hence T is bounded and $T \in \mathcal{B}(V, W)$. Finally, we show that $T_i \to T$. For an arbitrary $v \neq 0$ and $\epsilon > 0$ there exist N such that

$$n \ge N \implies \|T_i v - Tv\|_W < \epsilon\|v\|_V$$

which means that

$$\|T_i - T\| = \sup \frac{\|T_i v - Tv\|_W}{\|v\|_V} < \epsilon$$

Therefore $T_i \to T$ as desired. ■

Theorem 4.12. *Let $(V, \|\cdot\|)$ be a normed space. Then any linear transformation $T : \mathbb{R}^n \to V$ is continuous. Furthermore, if T is a bijection, it is a homeomorphism.*

Proof. Since $\mathbb{R}^n$ is finite then by Theorem 4.9, T is continuous. Assuming T is bijective, we must show that its inverse T^{-1} is continuous as well. Since T is a bijection then T is a linear isomorphism and $\dim V = \dim \mathbb{R}^n = n$ hence $T^{-1} : V \to \mathbb{R}^n$ is a continuous map. ■

Theorem 4.13. *Let V, W be two finite dimensional normed vector spaces. $T : V \to W$ linear transformation is invertible if and only if there exists a c such that:*

$$c\|v\|_V \le \|Tv\|_W$$

Proof. If T is invertible then $T^{-1} : W \to V$ is bounded and thus

$$\left\|T^{-1}w\right\|_V \le c\|w\|_W$$

and since T is bijective then there exists v such that $w = Tv$ which implies

$$\|y\|_V \le c\|Ty\|_W$$

If there exists such c then $\|Tx\| > 0$ for all non-zero x and hence $\ker T = 0$ which implies that T is a bijection and is invertible. ■

Remark 5. the supremum of such c is $\|T^{-1}\|^{-1}$ which is called the **conorm** of T.

Definition (General linear group): The **general linear group** of a vector space, written $\mathrm{GL}(V)$ is the set of all bijective linear transformation.

Proposition 4.14. *If V is a finite (also works for infinite) vector space then $\mathrm{GL}(V)$ is open in $\mathcal{L}(V, V)$, in fact, if $f \in \mathrm{GL}(V)$ then the open ball centered at f with radius $\|f^{-1}\|^{-1}$ remains in $\mathrm{GL}(V)$. Furthermore, the inverse operator $i : \mathrm{GL}(V) \to \mathrm{GL}(V)$, $i(T) = T^{-1}$ is continuous.*

Proof. First assume $f = \nVdash_V$ then we prove that any linear g that $\|\nVdash_V - g\| < 1$ is invertible which then implies bijectivity (true for linear maps). Let $\|v\| = 1$ then

$$\big|\|v\| - \|gv\|\big| \leq \|v - gv\| \leq \|\nVdash_V - g\|\|v\| < 1$$

Therefore

$$0 < \|gv\| < 2$$

which means $\ker g = \{0\}$ and since V is finite then then g is invertible. For a general f, we have that

$$\left\|1 - f^{-1} \circ g\right\| \leq \left\|f^{-1}\right\|\|f - g\| < 1$$

therefore $f^{-1} \circ g$ is invertible and as a consequence $g = f \circ f^{-1} \circ g$ is invertible. To prove inverse operator is continuous, fix $\epsilon > 0$ then for a $\delta > 0$ if $\|T - S\| < \delta$ then

$$\begin{aligned} &\left\|\nVdash_V - T^{-1} \circ S\right\| = \left\|T^{-1} \circ T - T^{-1} \circ S\right\| \leq \left\|T^{-1}\right\|\|T - S\| < \delta\left\|T^{-1}\right\| \\ \implies &\left\|T^{-1} - S^{-1}\right\| \leq \left\|T^{-1} \circ S - \nVdash_V\right\|\left\|S^{-1}\right\| < \delta\left\|T^{-1}\right\|\left\|S^{-1}\right\| \end{aligned}$$

note that by letting $\delta = \|T^{-1}\|^{-1}/2$ then

$$\|S\| > -\frac{\|T^{-1}\|^{-1}}{2} + \|T\| > \frac{\|T^{-1}\|^{-1}}{2}$$

also if for any invertible linear map R

$$\|R\| > a \implies \|Rx\| > a\|x\| \implies \frac{\|y\|}{a} = \frac{\|R \circ R^{-1}(y)\|}{a} > \left\|R^{-1}y\right\|$$

which means that $\|S^{-1}\| < 2\|T^{-1}\|$, hence by letting

$$\delta = \min\left\{\frac{\epsilon\|T^{-1}\|^{2}}{2}, \frac{\|T^{-1}\|^{-1}}{2}\right\}$$

we proved the continuity. ■

Definition: Let $V_1, V_2, \dots, V_n$ be normed vector spaces. Then $\phi : V_1 \times \ldots \times V_n \to W$ is **n-linear** if by fixing any $n - 1$ component, ϕ is linear relative to the remaining component.

Proposition 4.15. *If $V_1, V_2, \dots, V_n$ are normed vector spaces and $\phi : V_1 \times \ldots \times V_n \to W$ is a n-linear then the followings are equivalent*

1. *ϕ is continuous.*
2. *ϕ is continuous at 0.*
3. *ϕ is bounded, that is there exists a constant $C > 0$ such that*

$$\|\phi(v_1, \dots, v_n)\|_W \leq C\|v_1\|_{V_1} \cdots \|v_n\|_{V_n}$$

Remark 6. As oppose to linear transformation, n-linear function's continuity does not imply uniform continuity.

Proof. Item 1 $\implies$ Item 2: Trivial.

Item 2 $\implies$ Item 3: For the sake of contradiction, suppose Item 3 is false. That is, for every $k \in \mathbb{N}$ there exists a point $v_k = (v_k^1, \ldots, v_k^n)$ such that

$$\left\|\phi\left(v_k^1, \ldots, v_k^n\right)\right\|_W > n^n \left\|v_k^1\right\|_{V_1} \cdots \|v_k^n\|_{V_k}$$

Note that v_k^m can not be zero for any k and m, otherwise $\phi(v_k) = 0$. Define

$$w_k^m = \frac{v_k^m}{n\|v_k^m\|_{V_k}} \to 0$$

which from the continuity at 0 implies that $w_k = (w_k^1, \ldots, w_k^n) \to 0$. However,

$$\|\phi(w_k) - \phi(0)\|_W > n^n \frac{1}{n} \cdots \frac{1}{n} = 1$$

which is a contradiction.

Item 3 $\implies$ Item 1. Let $v_n \to v$ and define the points

$$\bar{v}_k^m = (v^1, \ldots, v^m, v_k^{m+1}, \ldots, v_k^n), \qquad \bar{v}_k^0 = v_k$$

and $\bar{v}_k^n = v$. Note that v_k^m are bounded for sufficiently large $k \geq N_1$, therefore there exists M such that $\forall m,\ \|v_k^m\|_{V_m} \leq M$. Also, pick M such that $\forall m,\ \|v^m\|_{V_m} \leq M$ as well. Then

$$\begin{aligned}
\|\phi(v_k) - \phi(v)\|_W &\leq \sum_{i=1}^n \left\|\phi\left(\bar{v}_k^{i-1}\right) - \phi\left(\bar{v}_k^i\right)\right\|_W \\
&= \sum_{i=1}^n \left\|\phi\left(\bar{v}_k^{i-1} - \bar{v}_k^i\right)\right\|_W \\
&\leq \sum_{i=1}^n C\left\|v^1\right\|_{V_1} \cdots \left\|v^{i-1}\right\|_{V_{i-1}} \left\|v_k^i - v^i\right\|_{V_i} \left\|v_k^{i+1}\right\|_{V_{i+1}} \cdots \|v_k^n\|_{V_n} \\
&\leq CM^{n-1} \sum_{i=1}^n \left\|v_k^i - v^i\right\|_{V_i}
\end{aligned}$$

pick N_2 such that for all $k \geq N_2$, for each i, $\|v_k^i - v^i\|_{V_i} < \frac{\epsilon}{nCM^{n-1}}$ then

$$\|\phi(v_k) - \phi(v)\|_W < CM^{n-1} \sum_{i=1}^n \frac{\epsilon}{nCM^{n-1}} = \epsilon$$ ■

We denote the set of all n-linear functions from $V_1 \times \ldots \times V_n \to W$ by $\mathcal{L}^n(V_1 \times \ldots \times V_n, W)$.

Proposition 4.16. *Let $V_1, \ldots, V_n, W$ be normed vector spaces. Then $\mathcal{L}^n(V_1 \times \ldots \times V_n, W)$ and $\mathcal{L}(V_1, \mathcal{L}(V_2, \ldots, \mathcal{L}(V_n, W)))$ are isomorphic.*

Proof. We want to prove

$$\mathcal{L}^n(V_1 \times \ldots \times V_n, W) \cong \mathcal{L}(V_1, \mathcal{L}(V_2, \ldots, \mathcal{L}(V_n, W)))$$

consider the mapping $T : \mathcal{L}(V_1, \mathcal{L}(V_2, \ldots, \mathcal{L}(V_n, W))) \to \mathcal{L}^n(V_1 \times \ldots \times V_n, W)$, such that for any $v_1 \in V_1,\ \ldots,\ v_n \in V_n$

$$\alpha((v_1)(v_2)\ldots(v_n)) = T(\alpha)(v_1, v_2, \ldots, v_n)$$

First note that T is linear. Then if $T(\alpha) = 0$ implies $\alpha = 0$, thus T is injective and hence bijective. ■

Definition (Positive definite): Matrix $A \in M_n(\mathbb{R})$ is **positive definite** whenever A is symmetric and

$$\forall x \in \mathbb{R}^n \backslash \{0\},\ x^T A x > 0$$

Theorem 4.17. *Every positive definite matrix A is diagonizable. In face, there exists an orthognal matrix P such that*

$$PAP^T = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$$

where $\lambda_i > 0$ for each i.

Exercises

1. Show that for a linear transformation T, $\|T\| = \sup_{\|v\|_V \leq 1} \|Tv\|_W$.
2. Prove or disprove that if $x^T A x = x^T A^T x$ for all $x \in \mathbb{R}^n \backslash \{0\}$ then $A = A^T$.

Chapter 5

Differentiation

Let V, W be finite dimensional vector spaces and $f : U \subset V \to W$ where U is open. Then f is differentiable at x_0 when a linear transformation $T : V \to W$ such that

$$\lim_{\|h\| \to 0} \frac{\|f(x_0 + h) - f(x_0) - T(h)\|}{\|h\|} = 0$$

Or equivalently there exists a sublinear function $R(h)$ such that

$$f(x_0 + h) - f(x_0) - Th = R(h) \qquad \frac{R(h)}{\|h\|} \to 0$$

T if it exists is unique, represented by $f'(x_0)$, $\mathrm{D}f$, or $\mathrm{d}f(x)$ and called the **total derivative** or **Fréchet derivative**.

Example 5.1. Any linear function $f : V \to W$ with $f(v) = Tv + b$ where $b \in W$ is differentiable and $\mathrm{D}f(v) = T$. Since

$$\|h\|_V < \delta \implies \|f(v+h) - f(v) - (\mathrm{D}f(v))(h)\|_W = \|T(v+h) - Tv - Th\|_W = 0 < \epsilon\|h\|_V$$

Hence, the derivative of any linear function is constant. Consider $S : V \times V \to V$ with $S(v, u) = v + u$. S is differentiable because S is linear (why?). We claim that $\mathrm{D}S = S$ as

$$\|S((v+h), (u+k)) - S(v, u) - S(h, k)\| = 0$$

Example 5.2. Let $\mu : \mathbb{R} \times V \to V$ with $\mu(r, x) = rx$. Then μ is differentiable and $(\mathrm{D}\mu(r, x))(t, h) = rh + tx$ as

$$\begin{aligned} \|\mu((r+t), (x+h)) - \mu(r, x) - (\mathrm{D}\mu(r, x))(t, h)\| &= \|rx + rh + tx + th - rx - rh - tx\| \\ &= |t|\|h\| \leq \epsilon\|(t, h)\| \end{aligned}$$

by letting $\|(t, h)\| = \sqrt{t^2 + \|h\|^2}$ and $\delta = \epsilon$.

Proposition 5.1. *Differentiability of f at x implies continuity at x.*

Proof.

$$\|f(x+h) - f(x)\| = \|(\mathrm{D}f(x))(h) + R(h)\| \leq \|\mathrm{D}f(x)\|\|v\| + \|R(v)\| \to 0$$

as $v \to 0$. ∎

Proposition 5.2. *Assume $f : U \subset V \to W$ is differentiable at x_0 and let $u \in V$ be a non-zero vector then*
$$f'(x_0)(u) = \lim_{t \to 0} \frac{f(x_0 + tu) - f(x_0)}{t}$$

Proof. Let $h = tu$ then

$$\begin{aligned} R(tu) &= f(x_0 + tu) - f(x_0) - T(tu) \\ &= f(x_0 + tu) - f(x_0) - tT(u) \\ \implies \frac{R(tu)}{t} &= \frac{f(x_0 + tu) - f(x_0)}{t} - T(u) \\ \implies \lim_{t \to 0} \frac{f(x_0 + tu) - f(x_0)}{t} &= T(u) \end{aligned}$$
■

Definition (Directional derivative): If we let $\|u\| = 1$ then the limit in Propostion 5.2 becomes the **directional derivative** of f in the direction of u and is denoted by $\mathrm{D}_u f$.

Remark 7. The existence of all directional derivatives of f doesnt imply its differentiability or even its continuity.

Remark 8. If $\mathrm{D}f : U \to \mathcal{L}(V, W)$ is continuous then each $\frac{\partial f_i}{\partial x_j}$ is continuous. Since

$$\mathrm{D}f(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$$

and the reverse is true as well.

Theorem 5.3. *$f : V \to W$ has all of its partial derivative in a neighbourhood of $u \in U$ and they're continuous at u then f is differentiable at u. Especially, if $\frac{\partial f_i}{\partial x_j}$ exist and are continuous at every point of U then $f \in \mathcal{C}^1$.*

Proof. We prove that each f_i is differentiable. Let $\{e_1, \dots, e_n\}$ be a basis for V and take $\|x\| = \sum|\xi_j|$. Consider a convex neighbourhood E of a. Then, for a given $\epsilon > 0$ we will show there exists a $\delta > 0$ such that

$$\|h\| < \delta \implies \left\| f_i(a + h) - f_i(a) - \sum_{j=1}^{n} \big(\mathrm{D}_{e_j} f_i(a)\big)(h_j) \right\| \leq \epsilon \|h\|$$

Cosider the point sequence $a^k = \sum_{j<k} a_j e_j + \sum_{j \geq k} (a_j + h_j) e_j$ where $a^1 = a + h$ and $a^{n+1} = a$ then

$$\left\| f_i(a + h) - f_i(a) - \sum_{j=1}^{n} \big(\mathrm{D}_{e_j} f_i(a)\big)(h_j) \right\| \leq \sum_{k=1}^{n} \left\| f_i\big(a^k\big) - f_i\big(a^{k+1}\big) - (\mathrm{D}_{e_k} f_i(a))(h_k) \right\|$$

hence we are done if

$$\left\| f_i\big(a^k\big) - f_i\big(a^{k+1}\big) - (\mathrm{D}_{e_k} f_i(a))(h_k) \right\| \leq \epsilon |h_k|$$

for $k = n$

$$\|f_i(a^n) - f_i(a) - (\mathrm{D}_{e_n} f_i(a))(h_n)\|$$

which equivalent to the existence n_{th} partial derivative of a. and for $k < n$

$$\begin{aligned}&\left\|f_i\left(a^k\right) - f_i\left(a^{k+1}\right) - (\mathrm{D}_{e_k} f_i(a))(h_k)\right\| \\ &\le \left\|f_i\left(a^k\right) - f_i\left(a^{k+1}\right) - \left(\mathrm{D}_{e_k} f_i\left(a^k\right)\right)(h_k)\right\| + \left\|\left(\mathrm{D}_{e_k} f_i\left(a^k\right)\right)(h_k) - (\mathrm{D}_{e_k} f_i(a))(h_k)\right\|\end{aligned}$$

which uses the existence of partial derivatives in neighbourhood and its continuity. ■

Proposition 5.4. *Let $f, g : V \to W$ be differentiable at x and $h : W \to U$ be differentiable at $y = f(x)$. Furthermore, let c be an scalar then*

1. $\mathrm{D}(f + cg) = \mathrm{D}f + c\,\mathrm{D}g$.
2. *$h \circ f$ is differentiable at x and*

$$\mathrm{D}(h \circ f) = ((\mathrm{D}h) \circ f) \circ \mathrm{D}f$$

Proof.

1. we have

$$\begin{aligned}&\|(f + cg)(x + k) - (f + cg)(x) - (\mathrm{D}f(x) + c\,\mathrm{D}g(x))(k)\| \\ &\le \|f(x + k) - f(x) - (\mathrm{D}f(x))(h)\| + |c|\|g(x + k) - g(x) - (\mathrm{D}g(x))(h)\|\end{aligned}$$

2. we know that

$$\begin{cases} f(x + k) - f(x) - (\mathrm{D}f(x))(k) = R(k) \\ h(y + l) - h(y) - (\mathrm{D}h(y))(l) = S(l) \end{cases}$$

and we wish to prove that

$$h \circ f(x + k) - h \circ f(x) - (\mathrm{D}h(f(x)) \circ \mathrm{D}f(x))(k) = T(k)$$

where $\|T(k)\| \le \epsilon\|k\|$ whenever $\|k\| < \delta$. Let $l = f(x + k) - f(x)$ and substituting into the second equation

$$\begin{aligned} h(f(x + k)) - h(f(x)) - (\mathrm{D}h(y))(f(x + k) - f(x))& \\ = h(f(x + k)) - h(f(x)) - (\mathrm{D}h(y))((\mathrm{D}f(x))(k) + R(k))& \\ = h(f(x + k)) - h(f(x)) - (\mathrm{D}h(y) \circ \mathrm{D}f(x))(k) - (\mathrm{D}h(y))(R(k))& \\ = T(k) - (\mathrm{D}h(y))(R(k)) = S(l)& \\ \implies T(k) = S(l) + (\mathrm{D}h(y))(R(k))& \end{aligned}$$

■

Proposition 5.5. *$f : U \subset V \to W_1 \times \ldots \times W_n$ is differentiable at x_0 if and only if all its component is differentiable at x_0. Furthermore, $\mathrm{D}f = (\mathrm{D}f_1, \ldots, \mathrm{D}f_n)$.*

Proof. Define the following norm on $W_1 \times \ldots \times W_n$

$$\|(w_1, \ldots w_n)\| = \sum_{i=1}^{n} \|w_i\|_{W_i} \tag{5.1}$$

then

$$\|f(x_0 + h) - f(x_0) - (\mathrm{D}f(a))(h)\| = \sum_{i=1}^{n} \|f_i(x_0 + h) - f_i(x_0) - (\mathrm{D}f_i(a))(h)\|$$

and since every other norm is equivalent to the norm defined above, we are done. ■

Theorem 5.6 (Leibnitz rule). *Let $V_1, V_2, \dots, V_n$ be finite dimensional vector spaces and $f : V_1 \times \dots \times V_n \to W$ is a n-linear function. f is differentiable at $a = (a_1, \dots, a_n)$ and*

$$(\mathrm{D}f(a))(h_1, \dots h_n) = f(h_1, a_2, \dots, a_n) + f(a_1, h_2, \dots, a_n) + \dots + f(a_1, a_2, \dots, h_n)$$

Proof. we have that

$$f(a+h) = \sum_{\xi_i \in \{a_i, h_i\}} f(\xi_1, \dots, \xi_n)$$

therefore

$$f(a+h) - f(a) - \sum_{i=1}^{n} f(a_1, \dots, a_{i-1}, h_i, a_{i+1}, \dots, a_n) = \sum_{\substack{\xi_i \in \{a_i, h_i\} \\ \text{at least two } h_i}} f(\xi_1, \dots, \xi_n)$$

Let $\delta = 1$ then $\|h\| = \sum \|h_i\| < 1$ also $i, j,\ \|h_i\| \|h_j\| \le \|h\|^2$. Hence if we define

$$A = \max\{\prod\}_{i \in I} \|a_i\| I \subset \mathbb{N}_n$$

then

$$\sum_{\substack{\xi_i \in \{a_i, h_i\} \\ \text{at least two } h_i}} f(\xi_1, \dots, \xi_n) \le (2^n - n - 1) A \|h\|^2$$

and letting $\delta = \min\left\{1, \dfrac{\epsilon}{(2^n - n - 1)(A+1)}\right\}$ we arrive at the conclusion. ■

Example 5.3. Let $Z : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ with $Z(u, v) = u \times v$ be a bilinear function, $f, g : \mathbb{R} \to \mathbb{R}^3$ and $h(t) = f(t) \times g(t)$. $h = Z \circ \phi$ where $\phi(t) = (f(t), g(t))$. Then we have:

$$\begin{aligned} \mathrm{D}h(t) &= (\mathrm{D}Z)(\phi(t)) \circ \mathrm{D}\phi(t) \\ &= (\mathrm{D}Z)(\phi(t)) \circ (\mathrm{D}f(t), \mathrm{D}g(t)) \\ &= Z(\mathrm{D}f(t), g(t)) + Z(f(t), \mathrm{D}g(t)) \\ &= \mathrm{D}f(t) \times g(t) + f(t) \times \mathrm{D}g(t) \end{aligned}$$

Example 5.4. Consider $A = [f_{ij}(x_1, \dots, x_n)]$ where each f_{ij} is differentiable. Then

$$\mathrm{D}\det(A)$$

can be calculated using the Leibnitz rule, since determinant is n-linear function.

5.1 Mean value theorem

Mean value theroem of 1-dimensional does not generalize very well. For example, the continuous function $f(t) : [0, 1] \to \mathbb{R}^2$ with

$$t \mapsto (t^2, t^3)$$

is differentiable on $]0, 1[$, however

$$\begin{aligned} f(1) - f(0) = (1, 1) &= \mathrm{D}f(c)(1 - 0) \\ &= (2c, 3c^2) \end{aligned}$$

which has no solution for $c \in\ ]0, 1[$. Although it must be said that for $f : U \to \mathbb{R}$ where $U \subset V$ is convex, the mean value theorem holds.

Theorem 5.7. *Let V, W be normed finite dimensional vector spaces and $f : U \to W$ is differentiable and $A, B \in U$ are such that the line connecting in completely contained in U and for each p on that line*

$$\|\mathrm{D}f(p)\| \leq M$$

then

$$\|f(B) - f(A)\|_W \leq M\|B - A\|_V$$

First consider the following lemma:

Lemma 5.8. *If $\phi : [0,1] \to W$ is continuous, differentiable on $]0,1[$ and $\|\phi'(t)\| \leq M$ for all $t \in]0,1[$ then*

$$\|\phi(1) - \phi(0)\|_W \leq M$$

Proof. We provide three proofs for the lemma

1. Assuming the norm on W is induced by an inner product. Then, let $e = \frac{\phi(1)-\phi(0)}{\|\phi(1)-\phi(0)\|}$ be a unit vector in W then $\psi : [0,1] \to \mathbb{R}$, $\psi(t) = e \cdot \phi(t)$ is continuous and differentiable on $]0,1[$. By the mean the value theorem

$$\begin{aligned} |\psi(1) - \psi(0)| &= |\psi'(t_0)| \\ |e \cdot (\phi(1) - \phi(0))| &= |e \cdot \phi'(t)| \\ \|\phi(1) - \phi(0)\| &\leq M \end{aligned}$$

2. Using the Hahn-Banach theorem, that is for a finite dimensional vector space V and $e \in V$ with $\|e\| = 1$ there exists a linear function $\theta : V \to \mathbb{R}$ such that $\|\theta\| = 1$ and $\theta(e) = 1$. Now let $\psi(t) = \theta(\phi(t))$ and take e as defined above then

$$\begin{aligned} |\psi(1) - \psi(0)| &= |\psi'(t_0)| \\ |\theta(\phi(1) - \phi(0))| &= (\mathrm{D}\theta(\phi(t_0)))(\phi'(t_0)) \\ \|\phi(1) - \phi(0)\| &= \theta(\phi'(t_0)) \leq \|\theta\|\|\phi'(t_0)\| \leq M \end{aligned}$$

3. From Hoimander. For any ϵ consider the set T_ϵ.

$$T_\epsilon = \{t\} \in [0,1] \forall s,\ 0 \leq s \leq t,\ \|\phi(s) - \phi(0)\| \leq (M + \epsilon)s + \epsilon$$

first note that $T_\epsilon = [0, c]$ for some $c > 0$ because for $s = 0$ the inequality is strict and both sides are continuous with respect to s. We claim that $c = 1$ because otherwise $c < 1$ and by differentiability of ϕ, there exists a $\delta < 1 - c$ such that if

$$\begin{aligned} \|h\| < \delta \implies \|\phi(c+h) - \phi(c) - (\mathrm{D}\phi(c))(h)\| &\leq \epsilon\|h\| \\ \implies \|\phi(c+h) - \phi(c)\| &\leq \|h\| \left(\epsilon + \|\phi'(c)\|\right) \\ &\leq \|h\|(\epsilon + M) \end{aligned}$$

also since $c \in T_\epsilon$

$$\begin{aligned} \|\phi(c) - \phi(0)\| &< (M + \epsilon)c + \epsilon \\ \implies \|\phi(c+h) - \phi(0)\| &< (M + \epsilon)(c + h) + \epsilon \qquad 0 < h < \delta \end{aligned}$$

hence $c + h \in T_\epsilon$ which is a contradiction and thus $c = 1$. $\square$

Proof. Let $\sigma : [0,1] \to U$ be a parameterization of the line connecting the point A to point B, $\sigma(t) = (1-t)A + tB$. Let $\phi = f \circ \sigma$, then clearly ϕ is continuous on $[0,1]$ and differentiable on $]0,1[$ and we have

$$\begin{aligned} \phi'(t) &= (\mathrm{D}f(\sigma(t)))(\sigma'(t)) \\ &= (\mathrm{D}f(\sigma(t)))(B-A) \\ \implies \|\phi'(t)\| &\leq \|\mathrm{D}f(\sigma(t))\| \|B-A\|_V \leq M\|B-A\|_V \end{aligned}$$

therefore by the Lemma 5.8

$$\|f(B) - f(A)\|_W = \|\phi(1) - \phi(0)\|_W \leq M\|B-A\|_V$$

which concludes the proof. ■

Corollary 5.9. *Let $U \subset V$ is connected and open and $f : U \to W$ is differentiable and $\mathrm{D}f(u) = 0$ for all $u \in U$ then f is constant.*

Proof. Let $p \in U$ and $S = \{q\} \in Uf(q) = f(p)$. S is closed because f is continuous and hence the pre-image closed set $\{f(p)\}$ is closed. For each $q \in S$ there exists $r > 0$ such that $B_r(q) \subset U$ and since $B_r(q)$ is convex then for each $l \in B_r(q)$ we apply the Theorem 5.7

$$\|f(l) - f(q)\| \leq \sup\|\mathrm{D}f(t)\|\|l-q\| = 0$$

which implies that $f(l) = f(q) = f(p)$ hence S is open in U which by the connectedness of U means $S = U$. Therefore, f is constant on U. ■

Corollary 5.10. *Let V_1, V_2, W be finite dimensional normed vector space and $U \subset V_1 \times V_2$ is open such that for every $y \in V_2$ the intersection $(V_1 \times \{y\}) \cap U$ is connected. Assumne $f : U \to W$ is differentiable and $\mathrm{D}_{V_1}f(x,y) = 0$ for all $(x,y) \in U$ then for any two point $(x_1,y),(x_2,y) \in U$,$f(x_1,y) = f(x_2,y)$.*

Proof. Fix $y \in V_2$ and define the function $g : V_1 \to W$

$$g(x) = f(x,y)$$

therefore

$$\mathrm{D}g(x) = \mathrm{D}_{V_1}f(x,y) = 0$$

and since $(V_1 \times \{y\}) \cap U$, the domain of g is connected. Hence by applying the Corollary 5.9 we get that

$$g(x) = c \implies f(x_1,y) = f(x_2,y)$$

for all $y \in V_2$. ■

5.2 Fundamental theorem of calculus

Theorem 5.11. *Let U be an open set of V such that for every $A, B \in U$ the line segment connecting A and B remains in U and let $\sigma : [0,1] \to U$ be that line, $\sigma(t) = (1-t)A + tB$, and lastly let $f : U \to W$ is continuously differentiable. Then*

$$f(B) - f(A) = T(B-A)$$

where T is

$$T = \int_0^1 \mathrm{D}f(\sigma(t))\,\mathrm{d}t$$

Proof. Let $g_i : [0,1] \to \mathbb{R}$ be

$$g_i(t) = \pi_i \circ f(\sigma(t))$$

is continuously differentiable then by the fundamental theorem of calculus for the real-valued functions we have

$$\begin{aligned}
g(1) - g(0) &= \int_0^1 g'(t)\,\mathrm{d}t \\
&= \int_0^1 \pi_i \circ \mathrm{D}f(\sigma(t))\,\mathrm{d}t \\
&= \pi_i \circ \int_0^1 \mathrm{D}f(\sigma(t))\,\mathrm{D}\sigma(t)\,\mathrm{d}t \\
&= \pi_i \circ \int_0^1 \mathrm{D}f(\sigma(t))(B-A)\,\mathrm{d}t \\
\implies \pi_i \circ (f(B) - f(A)) &= \pi_i \circ T(B-A) \\
\implies f(B) - f(A) &= T(B-A)
\end{aligned}$$

which was what was wanted. ■

Theorem 5.12. *Consider the continuous function $T : U \times U \to \mathcal{L}(V, W)$ which is such that*

$$f(B) - f(A) = (T(A,B))(B-A)$$

then $f \in \mathcal{C}^1$ and $\mathrm{D}f(A) = T(A,A)$

Proof. We have

$$f(A+h) - f(A) = (T(A+h, A))(h)$$

hence

$$\begin{aligned}
\|f(A+h) - f(A) - (T(A,A))(h)\| &= \|(T(A+h,A))(h) - (T(A,A))(h)\| \\
&\le \|T(A+h,A) - T(A,A)\|\|h\|
\end{aligned}$$

now by continuity of T, there exists a $\delta > 0$ such that

$$\|(h,k)\| < \delta \implies \|T(A+h, A+k) - T(A,A)\| < \epsilon$$

By letting $k = 0$ we get $\mathrm{D}f(A) = T(A,A)$. Since T is continuous then $f \in \mathcal{C}^1$ as well. ■

Corollary 5.13. *Let V be a normed finite dimensional vector space and U is open subset of V. If*

$$f : [a,b] \times U \to \mathbb{R}$$

is continuous then

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

is continuous. Furthermore, if $\frac{\partial f}{\partial y_i}$ exists and is continuous then $\frac{\partial F}{\partial y_i}$ exists and is continuous as well.

$$\frac{\partial F}{\partial y_i} = \int_a^b \frac{\partial f}{\partial y_i}(x,y)\,\mathrm{d}x$$

Proof. Firstly, we want to show that there exists a $\delta > 0$ such that for each $y \in U$

$$\|h\| < \delta \implies \|F(y+h) - F(y)\| < \epsilon$$

we have that

$$\begin{aligned}\|F(y+h) - F(y)\| &= \left\| \int_a^b f(x, y+h) f(x, y) \,\mathrm{d}x \right\| \\ &\leq (b-a) \sup_{x \in [a,b]} \{f(x, y+h) f(x, y)\}\end{aligned}$$

note that from the continuity of f for each $x \in [a, b]$ and $y \in U$ there are open balls $I_{x,y}$ around x and $J_{x,y}$ around y such that

$$x' \in I_{x,y},\ y' \in J_{x,y} \implies \|f(x', y') - f(x, y)\| < \frac{\epsilon}{b-a}$$

Fix y_0, then $\cup I_{x,y_0} \supset [a, b]$ which by the compactness of the interval implies that there is a finite family of there open set the covers $[a, b]$. Setting δ to the minimum radius of J_{x,y_0} yields the result. Secondly, we show that there exists a $\delta > 0$ such that

$$|h| < \delta \implies \left\| \frac{F(y+he_i) - F(y)}{h} \right\| < \epsilon$$

and we have that

$$\begin{aligned}\frac{F(y+he_i) - F(y)}{h} &= \frac{1}{h} \int_a^b f(x, y+he_i) - f(x, y) \,\mathrm{d}x \\ &= \frac{1}{h} \int_a^b \frac{\partial f}{\partial y_i}(x, y+the_i) h \,\mathrm{d}x \\ &= \int_a^b \frac{\partial f}{\partial y_i}(x, y+the_i) \,\mathrm{d}x\end{aligned}$$

from the previous part we know that we can make

$$\left\| \frac{\partial f}{\partial y_i}(x, y') - \frac{\partial f}{\partial y_i}(x, y) \right\|$$

as small as we want by making $\|y - y'\| < \delta$ small independently of x. Therefore, there exist a $\delta > 0$ such that if $|th| < |h| < \delta$ then

$$\left\| \frac{\partial f}{\partial y_i}(x, y') - \frac{\partial f}{\partial y_i}(x, y) \right\| < \frac{\epsilon}{b-a}$$

hence

$$\left\| \frac{F(x, y+he_i) - F(x, y)}{h} - \int_a^b \frac{\partial f}{\partial y_i}(x, y) \,\mathrm{d}x \right\| = \left\| \int_a^b \frac{\partial f}{\partial y_i}(x, y+the_i) - \frac{\partial f}{\partial y_i}(x, y) \,\mathrm{d}x \right\| < \frac{\epsilon}{b-a}$$

and the continuity of $\frac{\partial F}{\partial y_i}$ comes as a result of applying the first part to $\frac{\partial f}{\partial y_i}$. ∎

5.3 Higher derivative

Let V, W be finite dimensional normed vector spaces with $(e_1, \dots, e_n)$ is an ordered basis for V. Consider $U \subset V$ is an open set and $f : U \to W$. If f is differentiable then its partial derivatives

$$\mathrm{D}_i f : U \to E \quad \text{with} \quad (\mathrm{D}_i f)(x) = (\mathrm{D} f(x))(e_i)$$

Then, clearly if $\mathrm{D}_i f$ is differentiable one can define its partial derivatives $(\mathrm{D}_j)(\mathrm{D}_i f)$ also denoted by

$$(\mathrm{D}_j)(\mathrm{D}_i f) = \frac{\partial^2 f}{\partial x_j \partial x_i} = \mathrm{D}_{ji} f$$

For Fréchet derivative, if $\mathrm{D} f : U \to \mathcal{L}(V, W)$ is differentiable at x, then f is twice differentiable and

$$\mathrm{D}^2 f(x) = (\mathrm{D}(\mathrm{D} f))(x) : U \xrightarrow{\text{linear map}} \mathcal{L}(V, W)$$

is a linear map. Therefore,

$$\mathrm{D}^2 f : U \to \mathcal{L}(V, \mathcal{L}(V, W))$$

which by the Propostion 4.16 is equivalent to $\mathcal{L}^2(V \times V, W)$ and one can define

$$\mathrm{d}^2 f : U \to \mathcal{L}^2(V \times V, W)$$

where $\mathrm{d}^2 = T(\mathrm{D}^2)$ as defined in Propostion 4.16. With this definition, for the higher order derivatives $n \geq 2$

$$\mathrm{d}^n : U \to \mathcal{L}^n(V^n, W)$$

Example 5.5. Let $A : V \to W$ be a affine function $A(x) = Lx + b$ where L is linear. Then, $\mathrm{D}A(x) = L$ and hence $\mathrm{D}^2 A = 0$.

Example 5.6. Let $\beta : V \times V \to W$ be a bilinear function. By the Leibnitz rule

$$(\mathrm{D}\beta(x_1, x_2))(h_1, h_2) = \beta(x_1, h_2) + \beta(h_1, x_2)$$

therefore $\mathrm{D}\beta : V \times V \to \mathcal{L}(V \times V, W)$ is a linear a function itself, since

$$(\mathrm{D}\beta(x_1 + x_1', x_2 + x_2'))(h_1, h_2) = \beta(x_1, h_2) + \beta(x_1', h_2) + \beta(x_1, h_2) + \beta(b_1, x_2')$$

which means $(\mathrm{D}(\mathrm{D}\beta))(x) = \mathrm{D}\beta$ independent of x.

Theorem 5.14. *If f is twice differentiable at p then its second partial derivatives exist at p. Conversely, if its second partial derivatives exist at a neighbourhood of p and they are continuous, then f is differentiable.*

Proof. Assume that $\mathrm{D}^2 f(p)$ exists. Then

$$\begin{aligned} \mathrm{D}_j\left(\mathrm{D}_i f(p)\right) &= \lim_{h \to 0} \frac{(\mathrm{D} f(p + h e_j))(e_i) - (\mathrm{D} f(p))(e_i)}{h} \\ &= \left(\lim_{h \to 0} \frac{\mathrm{D} f(p + h e_j) - \mathrm{D} f(p)}{h}\right)(e_i) \end{aligned}$$

which exists sincce $\mathrm{D} f$ is differentiable at p. Conversely, assume that the second partials exist and are continuous at p. Then

$$(\mathrm{D}(\mathrm{D} f))(p) = \begin{bmatrix} \frac{\partial \mathrm{D} f}{\partial x_1}(p) & \dots & \frac{\partial \mathrm{D} f}{\partial x_n}(p) \end{bmatrix}$$

note that each $\frac{\partial \mathrm{D}f}{\partial x_i}(p)$ is in $\mathcal{L}(V, W)$. In fact, since

$$\mathrm{D}f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

then

$$\frac{\partial \mathrm{D}f}{\partial x_i} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_i \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_i \partial x_n} \end{bmatrix}$$

which is continuous at p and hence $\frac{\partial \mathrm{D}f}{\partial x_i}(p)$ is continuous and by Theorem 5.3, $\mathrm{D}f$ is differentiable at p. ■

Remark 9. In general, one can show that $f \in \mathcal{C}^r$ is equivalent to its partial being in $\mathcal{C}^r$.

Let $f : U \to \mathbb{F}$ then $\mathrm{D}f : U \to \mathcal{L}(V, \mathbb{F})$ which is the topological dual space V^* therefore

$$\mathrm{D}f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} e_i^*$$

then for the second derivative of f, $\mathrm{D}^2 f(x) : U \to V^*$

$$\begin{aligned}
\mathrm{D}^2 f(x) &= \sum_{i=1}^{n} \frac{\partial \mathrm{D}f}{\partial x_i}(x) e_i^* \\
&= \sum_{i=1}^{n} \frac{\partial \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}}{\partial x_i}(x) e_j^* e_i^* \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} e_j^* e_i^* \\
\implies \big(\big(\mathrm{D}^2 f(x)\big)(e_i)\big)(e_j) &= \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \\
\implies \mathrm{d}^2 f(x)(e_i, e_j) &= \frac{\partial^2 f}{\partial x_i \partial x_j}(x)
\end{aligned}$$

Definition (Hessian matrix): If for a function $f : U \to \mathbb{F}$ all of its second partial derivatives exist then **hessian matrix** is

$$\begin{bmatrix}
\dfrac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\
\vdots & \ddots & \vdots \\
\dfrac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \dfrac{\partial^2 f}{\partial x_n^2}(x)
\end{bmatrix}$$

Theorem 5.15. *If f is twice differentiable at x, $\mathrm{d}^2 f(x)$ is symmetric. That is,*

$$\mathrm{d}^2 f(x)(h, k) = \mathrm{d}^2 f(x)(k, h)$$

Proof. Let $\|h\|$ and $\|k\|$ be sufficiently small such that $a + th, a + tk, a + th + tk$ and the lines connecting them stays in U for some $t \in \mathbb{R}$. Consider

$$\Delta(t, h, k) = f(a + th + tk) - f(a + th) - f(a + tk) + f(a)$$

Assuming f is a real-valued twice differentiable function then if we prove

$$\big(\mathrm{d}^2 f(x)\big)(h,k) = \lim_{t\to 0} \frac{\Delta(t,h,k)}{t^2}$$

we are done since, Δ is symmetric with respect to h and k. Now consider

$$g(s) = f(a + th + tsk) - f(a + tsk)$$

then by the the Mean value theorem

$$\begin{aligned}\Delta(t,h,k) = g(1) - g(0) &= g'(\xi)\\ &= (\mathrm{D}f(a + th + t\xi k))(tk) - (\mathrm{D}f(a + t\xi k))(tk)\end{aligned}$$

and since $\mathrm{D}f$ is differentiable then by definition

$$\implies \mathrm{D}f(a + x) = \mathrm{D}f(a) + \big(\mathrm{D}^2 f(a)\big)(x) - R(x)$$

therefore

$$\begin{aligned}\Delta(t,h,k) = t\big(\mathrm{D}f(a) + \big(\mathrm{D}^2 f(a)\big)(th + t\xi k) - R(th + t\xi k)\big)(k)&\\ - t\big(\mathrm{D}f(a) + \big(\mathrm{D}^2 f(a)\big)(t\xi k) - R(t\xi k)\big)(k)&\end{aligned}$$

then

$$\begin{aligned}\Delta(t,h,k) &= t\big(\big(\mathrm{D}^2 f(a)\big)(th + t\xi k) - \big(\mathrm{D}^2 f(a)\big)(t\xi k)\big)(k) - t(R(t\xi k) - R(th + t\xi k))(k)\\ &= t^2\big(\big(\mathrm{D}^2 f(a)\big)(h)\big)(k) - t(R(t\xi k) - R(th + t\xi k))(k)\\ \implies \frac{\Delta(t,h,k)}{t^2} &= \big(\big(\mathrm{D}^2 f(a)\big)(h)\big)(k) - \frac{(R(t\xi k) - R(th + t\xi k))(k)}{t} \to \big(\big(\mathrm{D}^2 f(a)\big)(h)\big)(k)\end{aligned}$$

which is what we wanted. ∎

Theorem 5.16. *The k_{th} derivative of a k-times differentiable function is a symmetric k-linear function.*

Proof. it's generalization of above. ∎

Proposition 5.17. *If $f, g \in \mathcal{C}^r$ are two functions then $f \circ g \in \mathcal{C}^r$.*

Proof. Let $f : V' \to V''$ and $g : V \to V'$ be two $\mathcal{C}^r$ functions and $\beta : \mathcal{L}(V', V'') \times \mathcal{L}(V, V') \to \mathcal{L}(V, V'')$ is a bilinear function such that

$$\beta(\phi, \psi) = \phi \circ \psi$$

Now note that

$$\begin{aligned}(\mathrm{D}(f \circ g))(a) &= (\mathrm{D}f \circ g)(a) \circ \mathrm{D}g(a)\\ &= \beta((\mathrm{D}f \circ g)(a), \mathrm{D}g(a))\end{aligned}$$

Consider the following functions

$$a \xmapsto[\mathcal{C}^\infty]{\Delta} (a,a) \xmapsto[\mathcal{C}^{r-1}]{(\mathrm{D}f\circ g, \mathrm{D}g)} ((\mathrm{D}f \circ g)(a), \mathrm{D}g(a)) \xmapsto[\mathcal{C}^\infty]{\beta} (\mathrm{D}(f \circ g))(a)$$

therefore $\mathrm{D}(f \circ g) \in \mathcal{C}^{r-1}$ and hence $f \circ g \in \mathcal{C}^r$. ∎

Example 5.7. The inverse operator $i : \mathrm{GL}(V) \to \mathcal{L}(V,V)$ is in C^∞. Remember that

$$((\mathrm{D}i)(A))(M) = -A^{-1}MA^{-1}$$

Let $\gamma : \mathcal{L}(V,V) \times \mathcal{L}(V,V) \to \mathcal{L}(\mathcal{L}(V,V), \mathcal{L}(V,V))$ with

$$(\gamma(A,B))(M) = -AMB$$

is a bilinear function. Therefore

$$((\mathrm{D}i)(A))(M) = \big(\gamma\big(A^{-1}, A^{-1}\big)\big)(M)$$

now

$$A \overset{i}{\mapsto} A^{-1} \xmapsto[\mathcal{C}^\infty]{\Delta} (A^{-1}, A^{-1}) \xmapsto[\mathcal{C}^\infty]{\gamma} (\mathrm{D}i)(A)$$

Since we have proved that i is differentiable then $\mathrm{D}i$ is differentiable which means i is twice differentiable and so on. Hence $i \in C^\infty$.

As a matter of notation if $\phi : V_1 \times \ldots \times V_n$ be an n-linear then

$$\phi \cdot h_1 \ldots h_n := \phi(h_1, \ldots, h_n)$$

particularly if $V_1 = \cdots = V_n$

$$\phi \cdot h^n := \phi(h, \ldots, h)$$

Now one can describe a homogeneous polynomial of degree k with a symmetric k-linear function

$$p(x) = \phi \cdot x^k$$

Then, $p(x)$ is differentiable since

$$x \xmapsto[\mathcal{C}^\infty]{\Delta} (x, \ldots, x) \xmapsto[\mathcal{C}^\infty]{\phi} p$$

and

$$\begin{aligned}
(\mathrm{D}p(x))(h) &= (\mathrm{D}\phi(\Delta(x)) \circ \mathrm{D}\Delta(x))(h) \\
&= \mathrm{D}\phi \cdot x^n \circ \Delta(h) \\
&= k\phi \cdot x^{k-1} h \\
\implies \mathrm{D}p(x) &= k\phi \cdot x^{k-1}
\end{aligned}$$

Theorem 5.18 (Taylor approximation). *Let $f : U \to W$ be k-times differentiable at a, then*

$$p_k(x) = f(a) + \mathrm{d}f(a) \cdot (x-a) + \frac{1}{2!}\mathrm{d}^2 f(a) \cdot (x-a)^2 + \cdots + \frac{1}{k!}\mathrm{d}^k f(a)(x-a)^k$$

*is k_{th} degree **Taylor** polynomial. Then the followings hold*

1.

$$\lim_{x \to a} \frac{f(x) - p_k(x)}{\|x-a\|^k} = 0$$

2. *$p_k(x)$ is the only k_{th} degree polynomial with such property.*

3. *Additionally, if f is $(k+1)$-times differentiable in a neighbourhood of a then the remainder*
$$R(x) = f(x) - p_k(x)$$
can be estimated with
$$\|R(b)\| \leq \frac{1}{(k+1)!} \sup\left\{\left\|\mathrm{D}^{k+1} f(\xi)\right\|\right\} \|b-a\|^{k+1}$$
where ξ is on line connecting a to b.

Proof.

1. for $k=1$ it is equivalent to differentiability of f. By induction, assume it is true for $k=n-1$ and let $g(x) = f(x) - p_k(x)$ then [1]
$$\begin{aligned}
\mathrm{D}g(x) &= \mathrm{D}f(x) - \mathrm{D}p_k(x) \\
&= \mathrm{d}f(x) - \mathrm{D}\left[f(a) + \mathrm{d}f(a)\cdot(x-a) + \cdots + \frac{1}{n!}\,\mathrm{d}^n f(a)(x-a)^n\right] \\
&= \mathrm{d}f(x) - \left[\mathrm{d}f(a) + \frac{1}{1!}\,\mathrm{d}^2 f(a)\cdot(x-a) + \cdots + \frac{1}{(n-1)!}\,\mathrm{d}^n f(a)(x-a)^{n-1}\right]
\end{aligned}$$
which is equivalent to the proposition at $n-1$ for $\mathrm{d}f(a)$ and hence there exists a $\delta > 0$ such that if $\|x-a\| < \delta$
$$\|\mathrm{D}g(x)\| \leq \epsilon \|x-a\|^{n-1}$$
by the Theorem 5.7 we have
$$\begin{aligned}
\|g(x)\| = \|g(x) - g(a)\| &\leq \|x-a\| \sup \|\mathrm{D}g(\xi)\| \\
&\leq \epsilon \|x-a\| \|\xi - a\|^{k-1} \\
&\leq \|x-a\|^k
\end{aligned}$$

2. If there were two such polynomial p_1, p_2 then for $q = p_1 - p_2$ we have that
$$\lim_{x \to a} \frac{q(x)}{\|x-a\|^k} = 0$$
then one can show that $q(x) \equiv 0$.

3. Define $g : [0,1] \to W$ as such
$$g(t) = f(a + t(b-a))$$
therefore
$$g^{(n)}(t) = \mathrm{d}^k f(a + t(b-a)) \cdot (b-a)^k$$

[1] Differentiablity of order k implies differentiability of order $k-1$ in a neighbourhood.

For each component of g we apply the single variable Taylor's approximation

$$g_i(1) - \sum_{n=0}^{k} \frac{g_i^{(n)}(0)}{n!} = \frac{g_i^{(k+1)}(\xi_i)}{(k+1)!}$$

or equivalently

$$\begin{aligned}\|R(b)\| &= \left\| f(b) - \sum_{n=0}^{k} \frac{\mathrm{d}^n f(a) \cdot (b-a)^n}{n!} \right\| \\ &= \frac{1}{(k+1)!} \left\| \begin{bmatrix} \mathrm{d}^{k+1} f_1(a + \xi_1(b-a)) \cdot (b-a)^k & \dots & \mathrm{d}^{k+1} f(a + \xi_m(b-a)) \cdot (b-a)^k \end{bmatrix} \right\|\end{aligned}$$

which was what was wanted. ∎

Theorem 5.19. *Let $f : U \to \mathbb{R}$ and p is an extremum of the function then*

$$\forall h, \ (\mathrm{D}f(p))(h) = 0$$

Proof. For all h define $g_h :]-\epsilon, \epsilon[\to \mathbb{R}$

$$g_h(t) = f(p + th)$$

then $g_h'(0) = 0$. ∎

Theorem 5.20. *Let $f : U \to \mathbb{R}$ be of $\mathcal{C}^2$, p be a critical point of f, and $\mathrm{D}^2 f(p)$ be positive definite. Then, p is a local minimum of f. (If $\mathrm{D}^2 f(p)$ is negative definite then p is local maxima.)*

Assuming the following lemma

Lemma 5.21. *If $\mathrm{D}^2 f$ is continuous and positive definite at point p then it is positive definite in a neighbourhood of p.*

Proof. We wish to prove that there exists a $\delta > 0$ for all unit vectors in V, e, $0 < t < \delta$

$$f(p) \le f(p + te)$$

To do so, define $g_e :]0, \delta[\to \mathbb{R}$

$$g_e(t) = f(p + te)$$

then by the Taylor's theorem

$$g_e(t) = g(0) + g'(0)t + \frac{g''(\xi)}{2!} t^2$$

where $\xi \in]0, t[$. Equivalently

$$\begin{aligned} f(p + te) &= f(p) + (\mathrm{D}f(p))(e) + \frac{\mathrm{d}^2 f(p + t\xi) \cdot e^2}{2} t^2 \\ &= f(p) + \frac{\mathrm{d}^2 f(p + t\xi) \cdot e^2}{2} t^2 \end{aligned}$$

Using the Lemma 5.21 there exists a neighbourhood of p such that

$$\mathrm{d}^2 f(p + t\xi) \cdot h^2 > 0$$

for all h in the neighbourhood. Therefore,

$$f(p + te) > f(p)$$

which is what we wanted. ∎

5.4 Smoothness Classes

Let $f \in \mathcal{C}^r$ then one can define the norm

$$\|f\|_r = \max\left\{\sup_{x\in U}\|f(x)\|, \ldots, \sup_{x\in U}\|\mathrm{D}^r f(x)\|\right\}$$

and let the set of all such f with $\|f\|_r < \infty$ be denoted as $\mathcal{C}^r(U, W)$.

Theorem 5.22. *Uniform convergence in $\mathcal{C}^r$ is equivalent to Cauchy.*

Proof.

Theorem 5.23. *$\mathcal{C}^r(U, W)$ under $\|\cdot\|_r$ is a Banach space.*

Definition (Local convergence): A functional sequence f_n is **locally convergent** if for each $x \in U$ there exists a open set $x \in V \subset U$ such that $f_n|_V$ is uniformly convergent.

Theorem 5.24. *Let V, W be normed finite dimensional spaces, $U \subset V$ is open and connected, $x_0 \in U$ and $f_n : U \to W$ is a sequence of differentiable function that*

1. *$f_n(x_0)$ is convergent.*
2. *$\mathrm{D}f_n : U \to \mathcal{L}(V, W)$ is locally convergent to some function $g : U \to \mathcal{L}(V, W)$*

then the sequence f_n is locally convergent to $f : U \to W$ and $\mathrm{D}f = g$. Furthermore, because of connectedness of U for each $x \in U$, $f_n(x)$ is convergent.

Proof. take open ball W around x_0 such that $\mathrm{D}f_n|_W$ is uniformly convergent. then prove the first statement.

$$\|f_m(x) - f_n(x)\| \leq \|(f_m - f_n)(x) - (f_m - f_n)(x_0)\| + \|f_m(x_0) - f_n(x_0)\|$$

apply MVT here and make the bounds smaller using (2). Then prove the differentiability with e/3. To prove (3) use open/close argument. ■

5.5 Inverse function theorem

Consider a function f, we wish to find all the solutions to the equation

$$f(x) = y_0$$

To do so, we can define another function F_{y_0} such that

$$F_{y_0}(x) = x - f(x) + y_0$$

then if x is a solution to the equation, it is a fixed point of F_{y_0}.

Theorem 5.25 (Banach fixed point). *Let (X, d) be a complete metric space and $f : X \to X$ is such that for some $0 \leq \lambda < 1$*

$$\forall x, y \in X,\ d(f(x), f(y)) \leq \lambda d(x, y)$$

Then for each $x \in X$ the sequence $\{f^n(x)\}$ is convergent to $p \in X$ such that $f(p) = p$.

Proof. Let $x_n = f^n(x)$ for $n \geq 0$ then

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

thereofore

$$d(x_n, x_m) \leq \sum_{i=n}^{m-1} \lambda^i d(x_0, x_1) \leq d(x_0, x_1) \frac{\lambda^n}{1-\lambda}$$

hence $\{f^n(x)\}$ is Cauchy and it is convergent to a point p. Lastly,

$$\begin{aligned} d(f(p), p) &\leq d(f(p), f(x_n)) + d(f(x_n), p) \\ &\leq \lambda d(p, x_n) + d(x_{n+1}, p) < \epsilon \end{aligned}$$

■

Theorem 5.26 (Inverse function theorem). *Let V, W be finite dimensional normed vector space such that $\dim V = \dim W$ and $U \subset V$ is open. If $f : U \to W$ is continuously differentiable and for some $a \in U$, $\mathrm{D}f(a)$ is invertible. Then, there are open set $S \subset V$ and $T \subset W$ such that $a \in S \subset U$ and $f(a) \in T$ such that $f|_S$ is bijective and $(f|_S)^{-1} = g$ where $g \in \mathcal{C}^1$ and*

$$\mathrm{D}g(f(x)) = (\mathrm{D}f(x))^{-1}$$

Proof. Let S be an open convex set around a such that for all $x \in S$

$$\|\mathrm{D}f(x) - \mathrm{D}f(a)\| < \frac{1}{2}\left\|\mathrm{D}f^{-1}(a)\right\|^{-1}$$

hence $\mathrm{D}f(x)$ is invertible. Let $T = f(S)$ then we shall prove the following

1. $f|_S$ is bijective.

   Let $\psi : S \to V$ with

$$\begin{aligned} \psi_y(x) &= x - (\mathrm{D}f(a))^{-1}(f(x) - y) \\ \implies \mathrm{D}\psi_y(x) &= \nVdash_V - (\mathrm{D}f(a))^{-1}\mathrm{D}f(x) \\ &= (\mathrm{D}f(a))^{-1} \circ (\mathrm{D}f(a) - \mathrm{D}f(x)) \\ \implies \|\mathrm{D}\psi_y(x)\| &\leq \left\|(\mathrm{D}f(a))^{-1}\right\| \|\mathrm{D}f(a) - \mathrm{D}f(x)\| \\ &< \frac{1}{2}[(\mathrm{D}f(a))^{-1}][(\mathrm{D}f(a))^{-1}]^{-1} = \frac{1}{2} \end{aligned}$$

   therefore by mean value theorem

$$\|\psi_y(x_1) - \psi_y(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|$$

   which follows that ψ_y has at most one fixed point because

$$\|\psi_y(x_1) - \psi_y(x_2)\| = \|x_1 - x_2\| \leq \frac{1}{2}\|x_1 - x_2\|$$

   is a contradiction, and for that fixed point

$$\psi_y(x) = x - (\mathrm{D}f(a))^{-1}(f(x) - y) = x \implies y = f(x)$$

   which means f is injective. By the definition of T, f is surjective as well.

2. T is open.

   We wish to prove that for each $f(x_0) = y_0 \in T$ we wish to prove there exist a $\sigma > 0$ such that $B_\sigma(y_0)$ is contained in T. In other words, $\forall y \in B_\sigma(y_0)$

   $$\exists x \in S,\ f(x) = y \iff \psi_y(x) = x$$

   To apply the contraction fixed point we must find complete metric space X such that $\psi_y(X) = X$. Choose ρ as small as needed that $\overline{B_\rho(x_0)} \subset S$, which makes a complete metric space. Let $\sigma = \dfrac{r\rho}{2}$ where $r = \left\|(\mathrm{D}f(a))^{-1}\right\|^{-1}$. Lastly, we show that for each $y \in \overline{B_\sigma(y_0)}$, $\psi_y\left(\overline{B_\rho(x_0)}\right) = \overline{B_\rho(x_0)}$. That is, $x \in \overline{B_\rho(x_0)}$ implies that $\psi_y(x) \in \overline{B_\rho(x_0)}$.

   $$\begin{aligned}\|\psi_y(x) - x_0\| &\leq \|\psi_y(x) - \psi_y(x_0)\| + \|\psi_y(x_0) - x_0\| \\ &\leq \frac{1}{2}\|x - x_0\| + \left\|(\mathrm{D}f(a))^{-1}(y - y_0)\right\| \leq \frac{\rho}{2} + \frac{\rho}{2} = \rho\end{aligned}$$

3. $g = (f|_S)^{-1} : T \to S$ is continuously differentiable. Writting the differentiability criteria

   $$\|g(y+h) - g(y) - (\mathrm{D}g(y))(h)\| \leq \epsilon\|h\|$$

   Let $y = f(x)$ and $y + h = f(x + k)$ then $h = f(x+k) - f(x)$ and note

   $$\|\psi_y(x+k) - \psi_y(x)\| = \left\|k - \left((\mathrm{D}f(a))^{-1}\right)(h)\right\| \leq \frac{1}{2}\|k\|$$

   which implies

   $$\frac{1}{2}\|k\| \leq \left\|\left((\mathrm{D}f(a))^{-1}\right)(h)\right\| \leq \frac{3}{2}\|k\|$$

   $$\begin{aligned}\left\|k - \left((\mathrm{D}f(x))^{-1}\right)(f(x+k) - f(x))\right\| &= \left\|\left((\mathrm{D}f(x))^{-1}\right)((\mathrm{D}f(x))(k) - f(x+k) - f(x))\right\| \\ &\leq \left\|(\mathrm{D}f(x))^{-1}\right\|\|f(x+k) - f(x) - (\mathrm{D}f(x))(k)\| \\ &\leq \left\|(\mathrm{D}f(x))^{-1}\right\|\epsilon\|k\| \\ &\leq 2\left\|(\mathrm{D}f(x))^{-1}\right\|\left\|(\mathrm{D}f(a))^{-1}\right\|\epsilon\end{aligned}$$

   which proves the differentiability of g as $\left\|(\mathrm{D}f(x))^{-1}\right\|$ is bounded in S. As shown, the inverse operator is i is continuous and therefore if $\mathrm{D}f$ is continuous, then $i(\mathrm{D}f)$ is continuous. In fact, if $f \in C^k$ then $g \in C^k$ as well. ■

Corollary 5.27. *If f is continuously differentiable and $\mathrm{D}f(x)$ is invertible for every $x \in U$, then for any open set S, $f(S)$ is an open set as well.*

Proof. By the inverse function theorem for each $x \in S$ there is an open set U_x in S and V_x in W such that $f(U_x) = V_x$, therefore

$$f(S) = f\left(\bigcup U_x\right) = \bigcup f(U_x) = \bigcup V_x$$

which is an open set. ■

5.6 Implicit function

Theorem 5.28. *Let V, W be finite dimensional normed vector spaces and $U \subset V \times W$ is open. If $f : U \to W$, $f \in \mathcal{C}^1$ where $f(x_0, y_0) = z_0$ and $\big(\mathrm{D}f|_{\{x_0\}\times W}\big)(x_0, y_0)$ is invertible then there exist open set S around x_0 and T around y_0 that $S \times T \subset U$, such that for each $x \in S$ there exists a unique $y \in T$ with*

$$f(x, y) = z_0$$

hence there is a continuously differentiable function $\phi : S \to T$ such that $\phi(x) = y$ where $f(x, y) = z_0$ and

$$\mathrm{D}\phi = -\left(\mathrm{D}_y f\right)^{-1} \mathrm{D}_x f$$

Proof. To apply the inverse function theorem, we need a function whose domain and range have the same dimension. So define, $F : U \to V \times W$

$$F(x, y) = (x, f(x, y))$$

Then

$$\mathrm{D}F(x_0, y_0) = \left[\begin{array}{c|c} I_n & \nvdash \\ \hline \big(\mathrm{D}f|_{\{y_0\}\times U}\big)(x_0, y_0) & \big(\mathrm{D}f|_{\{x_0\}\times W}\big)(x_0, y_0) \end{array}\right]$$

Since I_n and $\big(\mathrm{D}f|_{\{x_0\}\times W}\big)(x_0, y_0)$ are both invertible then $\mathrm{D}F(x_0, y_0)$ is invertible as well. By inverse function theorem there are open set Ω_1 around (x_0, y_0) and Ω_2 around (x_0, z_0) such that $F|_{\Omega_1}$ is $\mathcal{C}^1$ diffeomorphism from Ω_1 to Ω_2. Let $S = \{x\}(x, z_0) \in \Omega_2$ and the set $T = \{y\}(x, y) \in \Omega_1$, then we shall prove that for each $x \in S$ there exists exactly one $y \in T$ [2]. If $x \in S$ then there exists $(x, y) \in \Omega_1$ such that $F(x, y) = (x, z_0)$, suppose that there are two such y, y_1 and y_2 for x

$$F(x, y_1) = (x, f(x, y_1)) = (x, z_0) = (x, f(x, y_2)) = F(x, y_2)$$

which since $F|_{\Omega_1}$ is injective then $y_1 = y_2$.

Let $G : \Omega_2 \to \Omega_1$ be the local inverse of F and $\phi : S \to T$

$$\phi(x) = (\pi_2 \circ G)(x, z_0)$$

which implies that ϕ is continuously differentiable. Lastly,

$$\begin{aligned} \mathrm{D}f(x, \phi(x)) &= \mathrm{D}f(x, \phi(x)) \circ (I, \mathrm{D}\phi(x)) \\ &= \mathrm{D}_x f(x, \phi(x)) + \mathrm{D}_y f(x, \phi(x))\, \mathrm{D}\phi(x) = 0 \\ \implies \mathrm{D}\phi &= -\left(\mathrm{D}_y f\right)^{-1} \mathrm{D}_x f \end{aligned}$$

■

5.7 Rank theorem

A generalization of $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$

[2]show that they are open

Theorem 5.29. *Let $f : U \to W$ be of class $\mathcal{C}^1$ and*

$$\forall x \in U, \ \operatorname{rank} \mathrm{D}f(x) = k$$

then for each $p \in U$ there exist open subsets $p \in U_0$ and $f(p) \in W_0$ and diffeomorphisms

$$\begin{aligned} \alpha &: U_0 \to U_0' \\ \beta &: V_0 \to V_0' \end{aligned}$$

such that

$$\beta \circ f \circ a^{-1}(x_1, \dots, x_n) = (x_1, \dots, x_k, 0, \dots, 0)$$

Proof. Without loss of generality, assume $V = \mathbb{R}^n$, $W = \mathbb{R}^m$, with a transition $p = 0 \in \mathbb{R}^n$, $f(p) = 0 \in \mathbb{R}^m$, and with a change of basis,

$$\mathrm{D}f(p) = \begin{bmatrix} I_k & \nvdash \\ \nvdash & \nvdash \end{bmatrix}$$

Suppose

$$f(x, y) = (f_1(x, y), f_2(x, y))$$

where $x \in \mathbb{R}^k$, $y \in \mathbb{R}^{n-k}$, $f_1(x, y) \in \mathbb{R}^k$, and $f_2(x, y) \in \mathbb{R}^{m-k}$. Then

$$\mathrm{D}f(0, 0) = \begin{bmatrix} I_k & \nvdash \\ \nvdash & \nvdash \end{bmatrix}$$

Let $H(x, y, u) := u - f_1(x, y)$, where $u \in \mathbb{R}^k$. Note that $H \in \mathcal{C}^1$ and $\mathrm{D}_x H = -\mathrm{D}_x f_1$ is invertible at $(0, 0, 0)$. Therefore, there are open set Ω_1 around $(0, 0) \in \mathbb{R}^k \times \mathbb{R}^{n-k}$ and Ω_2 around $0 \in \mathbb{R}^k$ such that $H|_{\Omega_1}$ is bijective and there exists a function $\phi : \Omega_1 \to \Omega_2$ such that

$$x = \phi(u, y)$$

whenever $H(x, y, u) = 0$. Then let $\alpha^{-1} : \Omega_1 \to \Omega_2 \times \mathbb{R}^{n-k}$ be

$$\alpha^{-1}(u, y) = (\phi(u, y), y)$$

which is a diffeomorphism around $(0, 0) \in \mathbb{R}^k \times \mathbb{R}^{n-k}$ and

$$f \circ \alpha^{-1}(u, y) = (f_1(\phi(u, y), y), f_2(\phi(u, y), y)) = (u, \eta(u, y))$$

$\mathrm{D}_y \eta \equiv 0$ that is $\eta(u, y) = \eta(u, 0)$ since, $\operatorname{rank}(\mathrm{D}f \circ \alpha^{-1}(u, y)) = \operatorname{rank}(\mathrm{D}f(u, y))$ and thus

$$\mathrm{D}f \circ \alpha^{-1}(u, y) = \begin{bmatrix} I_k & \nvdash \\ \mathrm{D}_u \eta & \mathrm{D}_y \eta \end{bmatrix}$$

Let $\beta : \mathbb{R}^k \times \mathbb{R}^{m-k} \to \mathbb{R}^k \times \mathbb{R}^{m-k}$

$$\beta(u, z) = (u, z - \eta(u, z))$$

which is clear diffeomorphism around $(0, 0) \in \mathbb{R}^k \times \mathbb{R}^{m-k}$ which means that

$$\beta \circ f \circ \alpha^{-1}(x, y) = (x, 0)$$

■

5.8 Lagrange Multiplier

Theorem 5.30. *Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable and $g : U \to \mathbb{R}$ is continuously differentiable, $S = g^{-1}(0) \subset U$ and $\nabla g(s) \neq 0$, $\forall s \in S$. Assume $x_0 \in S$*

$$\max_{x \in S} f(x) = f(x_0)$$

then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f|_{x_0} = \lambda \nabla g|_{x_0}$$

Proof. Since $\operatorname{rank} \mathrm{D}g = 1$ everywhere, then there are diffeomorphism α, β such that

$$\beta \circ g \circ \alpha^{-1}(t_1, \ldots, t_n) = t_n$$ ■

Exercises

1. Using l'Hopital's rule show that

$$\lim_{t \to 0} \frac{\Delta(t,h,k)}{t^2} = \frac{(\mathrm{d}f(a))(h,k) + (\mathrm{d}f(a))(k,h)}{2}$$

Part III

Complex Analysis

Chapter 6

Topology

6.1 Topology

A set $N \subset S$ is called a neighbourhood of $x \in S$ if it contains a ball $B_r(x)$. A set is open if it is a neighbourhood of all its points. A point $x \in X$ is an isolated point if it has a neighbourhood whose intersection with X reduces to x. An accumulation point is a point that is not isolated.

Proposition 6.1. *A non-empty open set in plane is connected if and only if any two points can be joined by a polygon which lies in the set.*

Definition: A non-empty connected open set is called a region. A component or a maximal connected set of a set S is a connected subset which is not contained in any larger connected subset.

Proposition 6.2. *Every set has a unique decomposition into component.*

Proposition 6.3. *In $\mathbb{R}^n$ the components of any open set are open.*

Definition: A set A is dense in X if $\operatorname{cl} E = X$. A metric space is separable if there exists a countable subset union of disjoint regions. A topological space is locally connected if for each neighbourhood of its points there is a connected open subset for that neighbourhood. A metric space that open balls $B_r(x)$ are connected is locally connected.

Proposition 6.4. *In a locally connected separable space every open set is union of disjoint regions.*

Definition: A set S is totally bounded if for every $\epsilon > 0$, S can be covered by finitely many balls of radius ϵ.

6.2 Compact sets

Definition (Point of accumulation): point v is a **point of accumulation** for the sequence $\{z_n\}$ if for given $\epsilon > 0$ there exists infinitely many n such that

$$|z_n - v| < \epsilon$$

Similarly, a point of accumulation of an infinite set S is a point v that for each open set U containing v there are infinitely many elements of S.

Theorem 6.5 (Weierstrass-Bolzano theorem). *If S is an infinite bounded set of real numbers, then S has a point of accumulation.*

Definition: A **compact** set S is a set that every sequence of its elements has a point of accumulation in S. The following defintions are equivalent

1. Every infinite subset of S has a point of accumulation in S.
2. Every sequence of elements of S has a convergent subsequent whose limit is in S.

Theorem 6.6. *A complex set is compact if and only if it is closed and bounded.*

Theorem 6.7. *Let $S_1 \supset S_2 \supset \dots$ be a sequence of non-empty closed subsets of a compact set S. Then, the interestion of all S_n is not empty.*

Theorem 6.8. *Let S be a compact set and f be continuous function on S. Then the image of f is compact.*

Theorem 6.9. *Let S be a compact set and f be continuous function on S. Then f is uniformly continuous.*

Definition: Let A, B be two sets of complex numbers. The **distance** between them is

$$d(A,B) = \min_{\substack{\alpha\in A\\ \beta\in B}} |\alpha - \beta|$$

Theorem 6.10. *Let S be a closed set and let v be a complex number. There exists a point $w \in S$ such that*

$$d(S, \{v\}) = |w - v|$$

Theorem 6.11. *Let K be a compact set and let be S a closed set. Then, there are elements $\alpha \in K$ and $\beta \in S$ such that*

$$d(K, S) = |\alpha - \beta|$$

Theorem 6.12. *A set is compact if and only if it is complete and totally bounded.*

Corollary 6.13. *Let K be compact. Let r be a real number greater than zero. There exists a finite number of discs of radius r whose union contains K.*

We say that a family of open set $\{U_i\}$ covers a set S when for every $z \in S$, $z \in U_i$ for some i as well. A subcovering of S is a covering of S with a subfamily of $\{U_i\}$. If that subfamily is finite we say that it is a finite subcovering of S. S is **covering compact** if every open convering can reduce to a finite subcovering.

Theorem 6.14. *Let S be a set then S is sequentially comapct if and only if covering compact.*

6.3 Connectedness

Proposition 6.15. *A path connected set is connected but the converse is true when the set is open.*

Chapter 7

Complex Numbers

7.1 Algebra of Complex Numbers

We define complex numbers to be all the pairs of real numbers (x, y) with following addition and multiplication:

$$(x_1, y_1) +' (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$

$$(x_1, y_1) \cdot' (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1)$$

where $+$ and $\cdot$ are real addition and multiplication, respectively. We denote the set of complex numbers with $\mathbb{C}$. It is easy to check that addition and multiplication defined above have the following:

1. Addition and multiplication are commutitive:

$$z + w = w + z \quad z, w \in \mathbb{C}$$

$$z \cdot w = w \cdot z \quad z, w \in \mathbb{C}$$

2. Addition and multiplication are associative:

$$(z + w) + u = z + (w + u) \quad z, w, u \in \mathbb{C}$$

$$(z \cdot w) \cdot u = z \cdot (w \cdot u) \quad z, w, u \in \mathbb{C}$$

3. Addition and multiplication are distributive:

$$(z + w) \cdot u = z \cdot u + w \cdot u \quad z, w, u \in \mathbb{C}$$

4. Addition and multiplication have unique identity elements $0 = (0, 0)$ and $1 = (1, 0)$, respectively.

5. Every complex number z has a unique addition inverse. Denoted by $-z$.

6. Every non-zero complex number z has a unique multiplication inverse. Denoted by z^{-1} or $\dfrac{1}{z}$.

Which means $\mathbb{C}$ is a field.

We can represent in many other forms. Two of the most commonly for $z = (x, y)$ used are :

$$z = x + iy \quad \text{where } i = (0,1)$$
$$z = re^{i\theta} \quad r \geq 0\,, \theta \in \mathbb{R}$$
$$z = \begin{bmatrix} x & -y \\ y & x \end{bmatrix}$$

It is easy to see that $i^2 = -1$. We also define the functions $\Re(z) = x$ and $\Im(z) = y$. In the second representation r is the distance from origin and θ is the angle between the positive real axis and the ray passing through z. One can also view complex number as an extension of real numbers in which every polynomial has a root.

Theorem 7.1. *If K is field such that every odd degree polynomial has at a least a root and for all $a \in K$, either $x^2 = a$ or $x^2 = -a$ has a root then it is sufficient to add the root of $x^2 = -1$ to K so that every polynomial has a root.*

For every complex number $z = x+iy$ there exists the mapping $\bar{z} : \mathbb{C} \to \mathbb{C}$ where $\bar{z} = x - iy$ and is called the *conjugate* of z.

Proposition 7.2. *The following properties are satisfied.*

1. $\overline{z+w} = \bar{z} + \bar{w}$.
2. $\overline{zw} = \bar{z}\bar{w}$.
3. $z = \bar{\bar{z}}$.
4. $z = \bar{z}$ *if and only if* $z \in \mathbb{R}$.
5. *The following relations hold:*

$$\Re(z) = \frac{z+\bar{z}}{2} \quad , \quad \Im(z) = \frac{z-\bar{z}}{2i}$$

Another mapping is the *norm* or *modulus* function, $|z| : \mathbb{C} \to \mathbb{R}$ where $|z| = \sqrt{x^2+y^2}$. Geometrically speaking the norm of z gives the distance of z form origin.

Proposition 7.3. *The following properties are satisfied.*

1. $|z| \geq 0 \quad \forall z \in \mathbb{C}$ *and especially* $|z| = 0$ *if and only if* $z = 0$
2. $|zw| = |z||w|$.
3. $|z|^2 = z\bar{z}$.
4. $|z| = |-z| = |\bar{z}|$
5. *The following inequalities hold:*

$$-|z| \leq \Re(z)\,,\ \Im(z) \leq |z|$$

6. *Triangle inequality:*

$$|z+w| \leq |z| + |w|$$
$$||z| - |w|| \leq |z - w|$$

The point at infinity, ∞, is informally a point that is unboundedly far from the origin. The extended complex plane is defined as $\mathbb{C} \cup \{\infty\}$. Every line passes through ∞ but no half plane contains it.

7.2 Riemann sphere

Riemann sphere is a unit sphere centered at the origin complex plane. We can map every point (except $(0,0,1)$) on the sphere to a point in complex plane by the following bijective transformation.

$$\phi(z) = (x_1, x_2, x_3) = \left(\frac{z+\bar{z}}{1+|z|^2}, \frac{z-\bar{z}}{i(1+|z|^2)}, \frac{|z|^2-1}{|z|^2+1}\right)$$

Geometrically, this transformation maps every point z to the interestion of the line connecting it to the $(0,0,1)$ and the sphere. Furthermore, we can define $\phi\infty(=)(0,0,1)$.

7.3 Limits and continuity

Let $f : U \to \mathbb{C}$, α be an adherent point of U, and w be a complex number. Then

$$w = \lim_{z\to\alpha} f(z)$$

when

$$\forall\epsilon > 0,\ \exists\delta > 0 \text{ s.t. } z \in U, |z-\alpha| < \delta \implies |f(z) - w| < \epsilon$$

By defining the $\epsilon-$neighbourhood of ∞ to the

$$\left\{ z \,\middle|\, \frac{1}{z} < \epsilon \right\}$$

we can extend the defintion of limit to when α or w are ∞. We say that f is continuous at α if

$$\lim_{z\to\alpha} f(z) = f(\alpha)$$

Definition: A function f is said to be uniformly continuous if

$$\forall\epsilon > 0,\ \exists\delta > 0 \text{ s.t. } z, w \in S, |z-w| < \delta \implies |f(z) - f(w)|$$

7.4 Sequences and convergence

A Sequence $\{z_n\}_{n\in\mathbb{N}}$ is said to be convergent to z if:

$$\forall\epsilon > 0,\ \exists N \text{ s.t. } n \geq N \implies |z - z_n| < \epsilon$$

and it is denoted as

$$z = \lim_{n\to\infty} z_n$$

A sequence $\{z_n\}$ is a **Cauchy sequence** if

$$\forall\epsilon,\ \exists N \text{ s.t. } n, m \geq N \implies |z_m - z_n| < \epsilon$$

If we write $z_n = x_n + iy_n$, since $|z_n - z_m| = \sqrt{(x_n - x_m)^2 + (y_n - y_m)^2}$ and $|x_n - x_m| \leq |z_n - z_m|$, $|y_n - y_m| \leq |z_n - z_m|$, we can conclude that $\{z_n\}$ is Cauchy if and only if $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences. Thus, since for real sequences a Cauchy sequence is convergent then, a complex Cauchy sequence is convergent as well.

Due to similarities between the definition given above and their real counterpart, most results dealing with limits can be easily extended to complex numbers.

A series $\sum a_n$ is absolutely convergent if $\sum|a_n|$ is convergent.

Proposition 7.4. *If the series $\sum a_n$ is absolutely convergent then it can be summed in any order.*

Proposition 7.5. *If double sum $\sum_m \sum_n a_{mn}$ is absolutely convergent then the summation order can be interchanged.*

$$\sum_m \sum_n a_{mn} = \sum_n \sum_m a_{mn}$$

The resulting series obtained is absolutely convergent and converges to the same value.

Definition: Let $\{a_n\}$ and $\{b_n\}$ be two sequences of positive real numbers. Then

$$a_n \equiv b_n \qquad \text{for} \quad n \to \infty$$

if for each n, there exists a $u_n \in \mathbb{R}^+$ such that $\lim_{n\to\infty} u_n^{\frac{1}{n}} = 1$ and $a_n = b_n u_n$.

7.5 Function spaces and power series

A family of functions $\{f_n(x) : X \to Y\}$ converges uniformly to $f : X \to Y$, denoted $f_n \rightrightarrows f$

$$\forall \epsilon,\ \exists N \text{ s.t. } n \geq N \implies d_Y(f_n(x) - f(x)) < \epsilon,\ \forall x$$

if Y is a complete metric space then Cauchy convergence becomes equivalent to uniform convergence

$$\forall \epsilon,\ \exists N \text{ s.t. } n, m \geq N \implies d_Y(f_n(x) - f_m(x)) < \epsilon,\ \forall x$$

$\sum f_n(x)$ converges uniformly if the partial sums $s_m = \sum^m f_n$ converges uniformly.

Theorem 7.6 (Weierstrass M-test). *Suppose that $\{f_n(x)\}$ is a sequence of real/complex-valued functions defined on a set X, and there is a sequence of non-negative numbers M_n satisfying the conditions*

- *$|f_n(x)| \leq M_n$ for all $n \geq 1$ and all $x \in X$*
- *$\sum_{n=1}^{\infty} M_n$ converges*

then the series $\sum_{n=1}^{\infty} f_n(x)$ converges absolutely and uniformly on X.

Theorem 7.7. *Let $\{f_n(x)\}$ be a sequence of continuous functions that $f_n \rightrightarrows f$. Then f is continuous.*

Chapter 8

Analytic Functions

8.1 Differentiability

Let $f : U \overset{open}{\subset} \mathbb{C} \to \mathbb{C}$ and z be a point in U. f is said to be complex differentiable at z if the following limit exists.

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

Equivalently, there exists a $c \in \mathbb{C}$ such that

$$f(z_0 + h) - f(z_0) - hc = o(h) \tag{8.1}$$

Differentiation rules are similar to real differentiation and are proved the same way.

f is **holomorphic** if it is differentiable at every point of U. Taking the limit along the real and imaginary axes and equating them gives the **Cauchy-Riemann equations**

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

Assuming that u and v are twice differentiable, we can see that

$$\begin{aligned} \Delta u &= \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \\ &= \frac{\partial^2 v}{\partial x \partial y} - \frac{\partial^2 v}{\partial y \partial x} = 0 \end{aligned}$$

Which means that u, v satisfy the *Laplace's equation* and thus are harmonic functions.

Theorem 8.1. *If u, v and have continuous first order partial derivatives and satisfy the Cauchy-Riemann equations then $f(z) = u(x, y) + v(x, y)$ is analytic with continuous derivative.*

8.2 Linear transformation

A linear fractional transformation, also called *Möbius transformation*, is a function in form of

$$w = S(z) = \frac{az + b}{cz + d}$$

with $ad - bc \neq 0$. If $c \neq 0$ then $S(\infty) = \frac{a}{c}$ and $S\left(-\frac{d}{c}\right) = \infty$. The inverse of a linear fractional transformation is linear fractional itself and is equal to

$$S^{-1}(w) = \frac{dw - b}{-cw + a}$$

If $ad - bc = 1$ we say that S is normalized and since multiplying a, b, c, d by a non-zero factor does not change the function we can assume that all Möbius transformations are normalized.

Theorem 8.2. *The set of all Möbius transformations $\mathcal{M}$ is a group under composition.*

$$\mathcal{M} = \left\{ \frac{az + b}{cz + d} \,\middle|\, ad - bc = 1, a, b, c, d \in \mathbb{C} \right\}$$

Proposition 8.3. *Give 3 different points z_1, z_2, z_3 there exists a linear transformation that takes to $1, 0, \infty$ in that order. If non of the point is ∞ then*

$$S(z) = \frac{z - z_2}{z - z_3} \frac{z_1 - z_3}{z_1 - z_3}$$

if one of them is ∞ then

$$S(z) = \frac{z - z_2}{z - z_3}, \quad S(z) = \frac{z_1 - z_3}{z - z_3}, \quad S(z) = \frac{z - z_2}{z_1 - z_2}$$

for $z_1 = \infty, z_2 = \infty, z_3 = \infty$ respectively.

Proof. It is clear that S as defined above is such a transformation. To prove S is unique, suppose there is another transformation T that satisfies the condition. Suppose that neither of the points is ∞ and

$$T(z) = \frac{az + b}{cz + d}$$

for some $a, b, c, d \in \mathbb{C}$ with $ad - bc \neq 0$. Consider TS^{-1}

$$\begin{aligned} TS^{-1}(z) &= \frac{aS^{-1}(z) + b}{cS^{-1}(z) + d} \\ &= \frac{a(z - z_2)(z_1 - z_3) + b(z - z_3)(z_1 - z_2)}{c(z - z_2)(z_1 - z_3) + d(z - z_3)(z_1 - z_2)} \\ &= \frac{(az_1 + bz_1 - az_3 - bz_2)z + az_2z_3 - az_2z_1 + bz_3z_2 - bz_3z_1}{(cz_1 + dz_1 - cz_3 - dz_2)z + cz_2z_3 - cz_2z_1 + dz_3z_2 - dz_3z_1} \end{aligned}$$

We know that

$$TS^{-1}(1) = 1, \quad TS^{-1}(1) = 1, \quad TS^{-1}(\infty) = \infty,$$

which implies that

$$\begin{cases} az_2z_3 - az_2z_1 + bz_3z_2 - bz_3z_1 = 0 \\ cz_1 + dz_1 - cz_3 - dz_2 = 0 \\ az_1 + bz_1 - az_3 - bz_2 = cz_2z_3 - cz_2z_1 + dz_3z_2 - dz_3z_1 \end{cases}$$

hence

$$TS^{-1}(z) = \frac{(az_1 + bz_1 - az_3 - bz_2)z}{az_1 + bz_1 - az_3 - bz_2} = z$$

and thus TS^{-1} is the identity and we must have $T = S$. ∎

In general we have the following

Proposition 8.4. *For any two sets of distinct complex numbers z_1, z_2, z_3 and w_1, w_2, w_3 there exists a linear fractional transformation that takes z_i to w_i for $i = 1, 2, 3$.*

Proof. We first show that following lemma

Lemma 8.5. *Any linear fractional transformation has at most two fixed points unless it is the identity.*

Proof. It is clear that

$$\frac{az+b}{cz+d} = z \implies cz^2 + (d-a)z + b = 0$$

has at most two solutions unless $c = 0$ and $d - a = 0$. □

Then consider the following linear fractional transformation (interpreted appropriately if any of them is ∞)

$$S = w_1 \frac{(z-z_2)(z-z_3)}{(z_1-z_2)(z_1-z_3)} + w_2 \frac{(z-z_1)(z-z_3)}{(z_2-z_1)(z_2-z_3)} + w_3 \frac{(z-z_1)(z-z_2)}{(z_3-z_1)(z_3-z_1)}$$

Clearly S the required property. Suppose there exists another linear fractional transformation T with such property. We have

$$TS^{-1}w_i = w_i, \quad i = 1, 2, 3$$

which means TS^{-1} is the identity and we must have $T = S$. ■

Definition (Cross ratio): The cross ratio of $(z_1, z_2; z_3, z_4)$ is the image of z_1 under the transformation S that takes z_1, z_2, z_3 to $1, 0, \infty$. That is

$$(z_1, z_2; z_3, z_3) = \frac{z_1 - z_3}{z_1 - z_4} \frac{z_2 - z_4}{z_2 - z_3}$$

Proposition 8.6. *Let $z_1, z_2, z_3, z_4 \in \mathbb{C}$ and T is a linear fractional transformation. Then $(Tz_1, Tz_2; Tz_3, Tz_4) = (z_1, z_2; z_3, z_3)$*

Proof. Suppose $S = (z, z_2; z_3, z_4)$ and $S' = (z, Tz_2; Tz_3, Tz_4)$. Note that

$$ST(z_2) = 1, \quad ST(z_3) = 0, \quad ST(z_4) = \infty$$

therefore $S = S'T$ and for all $z_1 \in \mathbb{C}$, $Sz = S'Tz$. ■

Proposition 8.7. *Let $r, c \in \mathbb{R}$ and $k \in \mathbb{C}$, then the equation*

$$r|z|^2 + k\bar{z} + \bar{k}z + c = 0$$

represents a line if $r = 0$, represents a circle if $r \neq 0$ and $|k|^2 \geq rc$.

The locus of all the points of $r|z|^2 + k\bar{z} + \bar{k}z + c = 0$, if non-empty, is called a circline.

Proposition 8.8. *circlines correspond to circles on Riemann sphere.*

Proposition 8.9. *A linear transformation carries circlines to circlines.*

Proposition 8.10. *The cross ratio is real if and only if the four points lie in a circle or a straight line.*

8.3 Polynomials and rational functions

Theorem 8.11. *If all zero's of a polynomial P lie in a half plane (convex polygon), the all zero's of the derivative P' lie in the same half plane (convex polygon).*

A **rational function** $R(z)$ is the quotient of two polynomials

$$R(z) = \frac{P(z)}{Q(z)} = \frac{a_m x^m + \cdots + a_0}{b_n x^n + \cdots + b_0}$$

We can assume that P and Q have no factor in common. In that case, the zeros of P are the zeros of R and the zeros of Q are called the *poles* of R. We can further define $R(\infty)$ to be $S_R(0)$ where $S_R(z) = R\big(\frac{1}{z}\big)$.

$$S_R(z) = z^{n-m} \frac{a_m + \cdots + a_0 x^m}{b_n + \cdots + b_0 x^n}$$

Then if $n > m$ we say that R has roots at ∞ with multiplicity of $n - m$, if $n < m$ then R has poles at ∞ with multiplicity of $m - n$. If $n = m$ then $R(\infty) = \frac{a_m}{b_n}$. It is now easy to see that the number of roots and poles of R - including ∞ - is the greater of m, n and it is called the *degree* of R.

We can write R in the following form

$$R(z) = G(z) + H(z)$$

where G is a polynomial with no constant term, by carrying out the division. Now let $\beta_1, \ldots, \beta_k$ be distinct poles of R. Then $S_j(\xi) = R\Big(\beta_j + \frac{1}{\xi}\Big)$ is a rational function of ξ and it can be written as

$$S_j(\xi) = G_j(\xi) + H_j(\xi) \implies R(z) = G_j\left(\frac{1}{z - \beta_j}\right) + H_j\left(\frac{1}{z - \beta_j}\right)$$

note that $G_j\Big(\frac{1}{z-\beta_j}\Big)$ has pole(s) only at β_j. Now consider

$$S(z) = R(z) - G(z) - \sum_{j=1}^{k} G_j\left(\frac{1}{z - \beta_j}\right)$$

It can have poles only at $\beta_1, \ldots, \beta_k$ and ∞. R and G_j have poles at $z = \beta_j$, however their difference if finite. R and G have poles at ∞ but their difference is finite as well. Hence S has no poles and thus must be a constant. Incorporating the constant into G allows us to write R as

$$R(z) = G(z) + \sum_{j=1}^{k} G_j\left(\frac{1}{z - \beta_j}\right)$$

which is its *partial fraction decomposition.*

8.4 Conformal maps

A **conformal map** is a transformation that conserves angle and direction. For example every linear $\mathbb{R}^2$ transformation of the form

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \sqrt{a^2+b^2} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

is a conformal mapping since it is a scaling and a rotation, both of which conserve the angle and direction. Let U be an open set in $\mathbb{C}$ and $\gamma, \eta : [a,b] \to U$ be two complex valued curves. Suppose that $z_1 = \gamma(t_1) = \eta(t_2)$ then the angle between γ and η at z_1 is defined as the angle between $\gamma'(t_1)$ and $\eta'(t_2)$, provided that they are nonvanishing. Let $f : U \to \mathbb{C}$ be a holomorphic function then by chain rule

$$\frac{\mathrm{d}f(\gamma(t))}{\mathrm{d}t} = f'(\gamma(t))\gamma'(t)$$

Viewing $w = f'(\gamma(t))$ as a $\mathbb{R}^2$ linear transformation, we can deduce that f is conformal.

$$\frac{\mathrm{d}f(\gamma(t_1))}{\mathrm{d}t} = f'(z_1)\gamma'(t_1)$$
$$\frac{\mathrm{d}f(\eta(t_2))}{\mathrm{d}t} = f'(z_1)\eta'(t_2)$$

Thus given that $f'(z_1) \neq 0$ the angle between γ and η at z_1 is the same as $f \circ \gamma$ and $f \circ \eta$.

Chapter 9

Formal Power Series

9.1 Algbraic definitions

Monoid

A set S equipped with a binary operation $\cdot$ is a **monoid** if it satisfies the following properties:

- For all $a, b, c \in S$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- There exists $e \in S$ such that for each $a \in S$, $a \cdot e = e \cdot a = a$.

Group

A set S equipped with a binary operation $\cdots$ is a **group** if it satisfies the following properties:

- For all $a, b, c \in S$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- There exists $e \in S$ such that for each $a \in S$, $a \cdot e = e \cdot a = a$.
- For each element $a \in S$, there exists $b \in S$ such that $a \cdot b = b \cdot a = e$.

If $\cdot$ is commutative then the group is an *Abelian group*.

Ring

A set S equipped with binary operations $+$ and $\cdots$ is a **ring** if it satisfies the folloing properties:

1. S is an Abelian group under $+$.
2. S is a monoid under $\cdot$.
3. $\cdot$ distributes over $+$.

If $\cdot$ is commutative then the ring is called a *commutative ring*.

Integral domain

It is commutative ring with property that

$$ab = 0 \implies a = 0 \vee b = 0$$

Field

A commutative ring that each non-zero $a \in S$ has a multiplicative inverse.

Example 9.1. A field is an integral domain because if $ab = 0$ and a is non-zero then

$$a^{-1}ab = b = 0$$

9.2 Introduction

Let K be a commutative ring. Then the set of formal power series over K is

$$K[[X]] = \{a\}_i \in K a_0 + a_1 x + a_2 x^2 + \dots$$

The addition and multiplication can be generalized from $K[X]$, the set of polynomials in K.

Definition (Order): Let $S \in K[[X]]$ then $\omega(S)$ is the least $n \geq 0$ such that $a_n \neq 0$. Furthermore, $\omega(0) = \infty$.

Definition (Summability): Let I be a set of indices then the family $\{S_i\}_{i\in I}$ is **summable** if for each $k \geq 0$, $w(S_i) \geq k$ for all S_i except only a finite number of them. Then we can define S to be the sum of this family

$$S = \sum_{i\in I} S_i = \sum_{i\in I} \sum_j a_j^{(i)} x^j$$

which is sensible since only a finite number of $a_j^{(i)}$ are non-zero.

Proposition 9.1. *Let K be an integral domain, then for any two formal series*

$$\omega(ST) = \omega(S) + \omega(T)$$

Asume that $\infty + k = \infty$ for any finite number k and $\infty + \infty = \infty$.

Proof. If either of S or T is zero then the proposition is true. Otherwise, both $n = \omega(S)$ and $m = \omega(T)$ are finite. Equivalently $a_n \neq 0$ and $b_m \neq 0$ while $a_k = 0, b_l = 0$ for $k < n, l < m$. Therefore,

$$ST = \sum_i \sum_{j=0}^{i} a_j b_{i-j} x^i = \sum_i c_i x^i$$

for $i < m + n$, $c_i = 0$ and $c_{m+n} = a_n b_m$ which is non-zero as K is an integral domain which was what was wanted. ■

Corollary 9.2. *For an integral domain K, the ring $K[[X]]$ is an integral domain.*

For simplicity from now on assume K is either an integral domain or a field.

Consider two formal series

$$S = \sum_n a_n x^n, \qquad T = \sum_m b_m x^m$$

with $b_0 = 0$. Then we can define the composition $S \circ T$ as follow

$$S \circ T = \sum_n a_n T^n$$

To show that it is well defined we need to show that only a finite number of T^n have degree less than k for all $k \geq 0$. Which is clear implied by the fact that $\omega(T) \geq 1$ and by Propostion 9.1, $\omega(T^n) \geq n$. It can easily be shown that

$$\begin{cases} (S_1 + S_2) \circ T & = S_1 \circ T + S_2 \circ T \\ (S_1 \cdot S_2) \circ T & = (S_1 \circ T)(S_2 \circ T) \end{cases}$$

Moreover, if $\{S_i\}$ is a summable family of formal series then

$$(\sum_i S_i) \circ T = \sum_i (S_i \circ T)$$

Proposition 9.3. *For three formal series $S, T,$ and U with $\omega(T), \omega(U) \geq 1$*

$$(S \circ T) \circ U = S \circ (T \circ U)$$

Proposition 9.4. *A formal series $S = \sum_i a_i x^i$ has a multiplicative if and only if $a_0 \neq 0$.*

Proposition 9.5. *We have that if $\omega(T) \geq 1$*

$$\omega(S \circ T) = \omega(S)\omega(T)$$

For two formal series f, g we write $f \equiv g \pmod{T^N}$ if $a_n = b_n$ for all $n = 0, 1, \ldots, N-1$. Cleary $f \equiv g \pmod{T^N}$ for all N is equivalent to $f = g$.

Proposition 9.6. *Suppose $f_1 \equiv f_2, g_1 \equiv g_2, h_1 \equiv h_2 \pmod{T^N}$. Then*

$$f_1 + g_1 \equiv f_2 + g_2 \pmod{T^N}, \quad f_1 g_1 \equiv f_2 g_2 \pmod{T^N}, \quad f_1 \circ h_1 \equiv f_2 \circ h_2 \pmod{T^N}$$

9.3 Formal Derivative

The derivative of a formal series is defined as

$$\frac{\mathrm{d}S}{\mathrm{d}x} = a_1 + 2a_2 x + 3a_3 x + \ldots$$

Proposition 9.7. *The derivative has the following properties*

1. $(S_1 + S_2)' = S_1' + S_2'$.
2. *The Leibniz rule holds*
$$(S_1 \cdot S_2)' = S_1' S_2 + S_1 S_2'$$
3. *if S has multiplicative inverse then*
$$\left(\frac{1}{S}\right)' = -\frac{S'}{S^2}$$

4. *The chain rule holds, if $\omega(T) \geq 1$ then*

$$(S \circ T)' = (T' \circ S) \circ S'$$

5. *The higher order derivatives can be defined recursively*

$$S^{(n)} = (S^{(n-1)})', \qquad S^{(0)} = S$$

Then, we can write the formal Taylor series of S as

$$S = \sum_n \frac{S^{(n)}(0)}{n!} x^n$$

9.4 Compositional inverse

Given a formal series S we wish to find a formal series T such that

1. $\omega(T) \geq 1$.
2. $S \circ T = X$.

Theorem 9.8. *S has compositional inverse T, if and only if $\omega(S) = 1$. In this case, T is unique and $T \circ S = X$ as well.*

9.5 Power series

A power series is a complex function of the form

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

Theorem 9.9. *For every power series there exists a number $0 \leq R \leq \infty$, called the radius of convergence, with the following properties*

- *The series converges absolutely for every z with $|z| < R$ and it is an analytic function. The derivative can be obtained by termwise differentiation and it has the same radius of convergence. The convergence is uniform in $|z| \leq \rho$ for $0 \geq \rho \geq R$.*
- *If $|z| > R$ the series diverges.*

Proposition 9.10. *Let f and g be two power series that converge absolutely in $B_r(0)$. Then $f + g$, gf, and αf , $\alpha \in \mathbb{C}$ are absolutely convergent in $B_r(0)$.*

Example 9.2. Let $\alpha \in \mathbb{C}$

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}, \qquad \binom{\alpha}{0} = 1$$

then the binomial formal series is defined as

$$(1+T)^{\alpha} = \sum_{n=0}^{\infty} \binom{\alpha}{n} T^n$$

For non-negative integer α the radius of convergence is infinite but for other values of α the radius is 1.

Theorem 9.11 (Abel's limit theorem). *If $\sum_{n=0}^{\infty}$ converges, then $f(z) = \sum_{n=0}^{\infty} a_n z^n$ tends to $f(1)$ as z approaches 1 in such a way that $\frac{|1-z|}{1-|z|}$ remains bounded.*

Theorem 9.12.

1. *Let f be a non-constant power series with non-zero radius of convergence. If $f(0) = 0$ then there exists $s > 0$ such that $f(z) \neq 0$ for all non-zero z with $|z| \leq s$.*

2. *Let f and g to be two convergent power series. Suppse $f(z) = f(z)$ for all points x in an infinite set having 0 as a point of accumulation. Then $f = g$.*

Let $f = \sum a_n z^n$ and $\phi = \sum c_n z^n$ with $c_n \in \mathbb{R}_+$ be two formal power series. Then f is dominated by ϕ

$$f \prec \phi$$

if $|a_n| \leq c_n$ for all n.

Proposition 9.13. *Suppose $f \prec \phi$ and $g \prec \psi$ then*

$$f + g \prec \phi + \psi, \quad fg \prec \phi\psi$$

Theorem 9.14. *Let f be a power series with non-zero constant term and non-zero radius of convergence and let g be the inverse of f that is, $fg = 1$. Then g has a non-zero radius of convergence.*

Proposition 9.15. *Let*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \quad h(z) = \sum_{n=1}^{\infty} b_n z^n$$

be two power series and f is absolutely convergent for $|z| \leq r$ for some $r > 0$ and let $s > 0$ be such that

$$\sum_{n=1}^{\infty} |b_n| s^n \leq r$$

Then $f \circ h$ converges absolutely for $|z| \leq s$.

9.6 Analytic functions

Let f be defined in a neighbourhood of z_0. f is said to be analytic if there exists some power series

$$\sum a_n (z - z_0)^n$$

which converges absolutely for $|z - z_0| < r$ and for such z

$$f(z) = \sum a_n (z - z_0)^n$$

Proposition 9.16. *If f and g are analytic on U then $f + g$, fg, f/g for $g(z) \neq 0$ are also analytic. If $g : U \to V$ and $f : V \to \mathbb{C}$ then $f \circ g$ is analytic.*

Theorem 9.17. *Let $f(z) = \sum a_n z^n$ with radius of convergence of r. Then f is analytic on $B_r(0)$.*

Theorem 9.18. *Let $f(z) = \sum a_n z^n$ with radius of convergence of r then*

1. *The series $\sum n a_n z^{n-1}$ has the same radius of convergence.*

2. *f is holomorphic on $B_r(0)$ and its derivative is equal to $\sum n a_n z^{n-1}$.*

We shall see later that every holomorphic function admits a power series.

9.7 Inverse and open mapping

f is analytic isomorphism if $f : U \to V$ is analytic and V is open and there exists analytic function $g : V \to U$ such that $f \circ g = \mathrm{id}_V$ and $g \circ f = \mathrm{id}_U$. f is locally analytic isomorphism at z_0 or locally invertible if there exists an open neighbourhood U containing z_0 such that f is an analytic isomorphism on U.

Theorem 9.19. *Let $f(z) = a_1 z + \dots$ be a formal power series with $a_1 \neq 0$. We know there exists a unique $g(z)$ such that $f(g(z)) = z$ and $g(f(z)) = z$. If f is convergent a power series, then g is a convergent power series as well.*

Theorem 9.20. *Let f be an analytic function on open set U containing z_0. Suppose that $f'(z_0) \neq 0$ then, f is a local analytic isomorphism at z_0.*

Definition: Let U be an open set then $f : U \to V$ is open mapping if for each open subset $U' \subset U$, $f(U')$ is open.

Theorem 9.21. *Let f be an analytic function on open set U such that for each point of U, f is not constant on a given neighbourhood of that point. Then f is an open mapping.*

Theorem 9.22. *Let f be an analytic function on open set U and f is injective. Let $V = f(U)$ then, $f : U \to V$ is an analytic isomorphism and $f'(z) \neq 0$ for all $z \in U$.*

9.8 The local maximum modulus principle

Definition: f is locally constant at a point z_0 if there exists an open set D containing z_0 such that f is constant on D.

Theorem 9.23. *Let f be an analytic function on open set U. Let $z_0 \in U$ be a maximum for $|f|$, that is*

$$|f(z_0)| \geq |f(z)| \quad \forall z \in U$$

Then f is locally constant at z_0.

Corollary 9.24. *Above theorem also holds for $\Re$ instead of $|\cdot|$.*

Theorem 9.25. *Let f be a non-constant polynomial*

$$f(z) = a_0 + a_1 z + \dots + a_d z^d$$

with $a_d \neq 0$. Then f has some complex roots.

Chapter 10

Complex Integrals

10.1 Integrals over paths

Path is a sequence of regular curves

$$\gamma = \{\gamma_1, \ldots, \gamma_n\}$$

such that

$$\gamma_j(b_j) = \gamma_{j+1}(a_{j+1})$$

Theorem 10.1. *Suppose U is a connected open set and f is holomorphic on U. If $f' = 0$ then f is constant on U.*

Definition: If f is a function on an open set Ω and g is a holomorphic function on Ω, such that $g' = f$, then g is called a *primitive* if f on Ω.

Corollary 10.2. *On an connected open set U, the primitives of are determined up to a constant.*

Theorem 10.3. *Suppose U is a connected open set then*

1. *If f is analytic on U and not constant, the set of zeroes of f on U is discrete.*
2. *f, g are analytic on U and $S = \{z \mid f(z) = g(z)\}$ is not discrete then $f = g$ on U.*

Let $f : [a, b] \to \mathbb{C}$ be a continuous function with

$$f(t) = u(t) + iv(t)$$

then

$$\int f(t)\,\mathrm{d}t = \int u(t)\,\mathrm{d}t + i \int v(t)\,\mathrm{d}t$$

and by the Fundamental theorem of calculus

$$G(t) = \int_a^t F(\tau)\,\mathrm{d}\tau$$

is differentiable with $G'(t) = F(t)$. We can then expand the notion of complex integrability to continuous complex function $f : U \to \mathbb{C}$, where U is an open subset of $\mathbb{C}$ using regular curves such as $\gamma : [a, b] \to U$. The integral of f along γ is

$$\int_\gamma f(t) = \int_a^b f(\gamma(t))\gamma'(t)\,\mathrm{d}t$$

This definition of integral is invariant to substitution. To see this, let $g : [a, b] \to [c, d]$ be a $\mathcal{C}^1$ function and $\gamma = \psi \circ g$

$$\begin{aligned}
\int_\gamma f &= \int_a^b f(\gamma(t))\gamma'(t)\,\mathrm{d}t \\
&= \int_a^b f(\psi(g(t)))\psi'(g(t))g'(t)\,\mathrm{d}t \\
&= \int_c^d f(\psi(\tau))\psi'(\tau)\,\mathrm{d}\tau \\
&= \int_\psi f
\end{aligned}$$

We can further expand this notion to paths $\gamma = \{\gamma_1, \dots, \gamma_n\}$ as follow

$$\int_\gamma f = \sum_{i=1}^n \int_{\gamma_i} f$$

Theorem 10.4. *Let f be a continuous function on open set U with primitive g. Let $\gamma : [a, b] \to U$ be a path with $\alpha = \gamma(a)$ and $\beta = \gamma(b)$ then*

$$\int_\gamma f = g(\beta) - g(\alpha)$$

Corollary 10.5. *If γ is a closed path in U then*

$$\int_\gamma f = 0$$

Theorem 10.6. *Let U be a connected open set and f a continuous function U. If for any closed path γ on U, the integral of f along γ is zero , $\int_\gamma f = 0$, then f has a primitive g on U.*

Proof. Pick $z_0 \in U$ and define

$$g(z) = \int_{z_0}^z f$$

where the integral is taken over any path from z_0 to z. Let γ, η be two curves from z_0 to z and let η^- be the reverse of η. Since

$$\int_\gamma f + \int_{\eta^-} f = 0 \implies \int_\gamma f = \int_\eta f$$

and hence g is well-defined. Consider the difference quotient

$$\frac{g(z+h) - g(z)}{h} = \frac{1}{h}\int_z^{z+h} f$$

where the integral can be taken along any segments. Furthermore, we can write

$$f(\xi) = f(z) + \phi(\xi)$$

where $\lim_{\xi \to z} \phi(\xi) = 0$. Then,

$$\begin{aligned}
\frac{1}{h}\int_z^{z+h} f(\xi)\,\mathrm{d}\xi &= \frac{1}{h}\int_z^{z+h} f(z)\,\mathrm{d}\xi + \frac{1}{h}\int_z^{z+h} \phi(\xi)\,\mathrm{d}\xi \\
&= f(z)\frac{1}{h}\int_z^{z+h} \phi(\xi)\,\mathrm{d}\xi \\
&\leq f(z) + \max_{\xi}|\phi(x)| \xrightarrow{h\to 0} f(z) \\
\implies g' = f
\end{aligned}$$

■

Let $\gamma : [a,b] \to \mathbb{C}$ be a regular curve. The *speed* is of γ is defined as $|\gamma'|^2$ and the length of γ is

$$l(\gamma) = \int_a^b |\gamma'(t)|\,\mathrm{d}t$$

and for a path $\gamma = \{\gamma_1, \ldots, \gamma_n\}$

$$l(\gamma) = \sum_{i=1}^{n} l(\gamma_i)$$

and sup norm of f on a set S

$$\|f\| = \sup_{z\in S}|f|$$

and over a curve γ the sup-norm of f is

$$\|f\|_\gamma = \sup_{t\in[a,b]} |f(\gamma(t))|$$

Proposition 10.7. *ML inequality Let f be a continuous function on U and γ be a path in U them*

$$\left|\int_\gamma f\right| \leq \|f\|_\gamma l(\gamma)$$

Theorem 10.8. *Let $\{f_n\}$ be a sequence of continuous functions on U with $f_n \rightrightarrows f$ then*

$$\lim_{n\to\infty}\int_\gamma f_n = \int_\gamma f$$

and if $\sum f_n$ is a series of continuous function which converges uniformly on U

$$\int_\gamma \sum f_n = \sum \int_\gamma f_n$$

10.2 Local primitive for a holomorphic function

Theorem 10.9 (Goursat). *Let R be a rectangle, and let f be a holomorphic function on R then*

$$\int_{\partial R} f = 0$$

Theorem 10.10. *Let U be a disc centered at a point z_0 and f a continuous function on U. Assume that for any rectangle $R \subset U$*

$$\int_{\partial R} f = 0$$

For each point z in disk define

$$g(z) = \int_{z_0}^{z} f$$

where the integral is taken over a rectangle R whose opposite vertices are z_0 and z. Then, g is holomorphic on U and is a primitive for f.

Theorem 10.11. *Let U be a disc and f be a holomorphic function on U. Then, f has a primitive on U and the integral of f along any closed path in U is zero.*

10.3 Path integrals for continuous curves

Knowing that local primitive exists for holomorphic functions allows us to describe their integral along a path in a way which makes no use of differentiability of the path and applys to continuous paths.z

Lemma 10.12. *Let $\gamma : [a, b] \to U$ be a continuous curve in an open set U. Then, there exists some positive number $r > 0$ such that every point on the curve lies at distance $\geq r$ from the complement of U.*

Proof. Consider

$$\phi(t) = \min_{w \in U^c} |\gamma(t) - w|$$

$|\gamma(t) - w|$ is continuous in terms of w by considering a sufficiently large disc that contains U we can bound the domain of w to the intersection of this disc and U^c. Thus, due to compactness of the intersection, $|\gamma(t) - w|$ achieves its minimum. Moreover, since $\phi(t)$ is continuous and its domain is compact then, it has a minimum and the minimum is non-zero since U is open. ∎

Definition: Let $P = \{a_0, \ldots, a_n\}$ be a partition of $[a, b]$ and $\{D_0, \ldots, D_n\}$ be a sequence of discs. This sequence is *connected* by curve along the partition P if D_i contains $\gamma([a_i, a_{i+1}])$.

Let $\epsilon < \frac{r}{2}$ where r is the same as preceding lemma. Since γ is uniformly continuous, there exists a δ such that

$$\forall t, s \in [0, 1], \; |t - s| < \delta \implies |\gamma(t) - \gamma(s)| < \epsilon$$

Consider a partition P on $[a, b]$ such that $\|P\| < \delta$. Then, the $\gamma([a_i, a_{i+1}])$ lies in the disc $D_i = B_\epsilon(a_i)$. Let f be a holomorphic function on U.

$$\int_\gamma f = \sum_{i=0}^{n-1} \int_{\gamma_i} f \qquad \gamma_i = \gamma([a_i, a_{i+1}])$$

Let $z_i = \gamma(a_i)$ and g_i be a primitive of f on disc D_i. If each $\gamma_i \in \mathcal{C}^1$ then we know that

$$\int_\gamma f = \sum_{i=0}^{n-1} g_i(z_{i+1}) - g_i(z_i)$$

By considering the following lemma, we can relax the regularity condition for γ and reduce it to continuity.

Lemma 10.13. *Let U be an open set and $\gamma : [a,b] \to U$ be a continuous curver. Let*

$$a = a_0 \leq a_1 \leq \ldots \leq a_n = b$$

be a partition on $[a,b]$ such that $\gamma([a_i, a_{i+1}])$ is contained in a disc D_i which itself is contained in U. Let f be a holomorphic function on U, g_i be a primitive of f on D_i, and $z_i = \gamma(a_i)$, then

$$\sum_{i=0}^{n-1} g_i(z_{i+1}) - g_i(z_i)$$

is independent of the choices of partition, discs D_i, and primitives g_i on D_i subjected to stated condition. Therefore, Lemma 10.13 depends only on γ and hence the integral on continuous path is well-defined.

Proof.

Definition: Let γ, η be two paths defined on $[a,b]$. We say that they are *close* together if there exists a partition

$$a = a_0 \leq a_1 \leq \ldots \leq a_n = b$$

and for each $i = 0, \ldots, n-1$ there exists a disc D_i contained in U such that

$$\gamma([a_i, a_{i+1}]), \eta([a_i, a_{i+1}]) \subset D_i$$

Lemma 10.14. *Let γ, η be continuous paths on open set U, that are close together and have the same endpoints. Let f be holomorphic on U*

$$\int_\gamma f = \int_\eta f$$

Proof.

10.4 Homotopy

Let $\gamma, \eta : [a,b] \to U$ be two paths. γ is *homotopic* to η, if there eists a continuous function

$$\Psi : [a,b] \times [c,d] \to U$$

such that

$$\Psi(t,c) = \gamma(t) \qquad \Psi(t,c) = \eta(t)$$

for all $t \in [a,b]$. Intuitively, Ψ is a continuous deformation of γ to η. Ψ *leave the end point fixed* if we have

$$\Psi(a,s) = \gamma(a) \qquad \Psi(b,s) = \gamma(b)$$

for all $s \in [c,d]$. Similarly, we assume that a homotopy of two closed paths is such that each path $\Psi(\,\cdot\,, s)$ is a closed path.

Theorem 10.15. *Let γ, η be two homotopic continuous paths on open set U with the same endpoints and f be a holomorphic function on U.*

$$\int_\gamma f = \int_\eta f$$

In particular, if γ, η are closed paths and homotopic to a point in U then

$$\int_\gamma f = \int_\eta f = 0$$

Proof. Let $\Psi : [a, b] \times [c, d] \to U$ be the homotopy of the closed paths γ, η. Since Ψ is a continuous function on a compact domain, the image of Ψ is compact and hence it has a positive distance r from U^c. Similarly, by considering the uniform continuity we can have partitions $P = \{a_1, \ldots, a_n\}$ on $[a, b]$ and $Q = \{c_1, \ldots, c_m\}$ on $[c, d]$ and

$$S_{ij} = [a_i, a_{i+1}] \times [c_j, c_{j+1}]$$

such that $\Psi(S_{ij})$ is contained in D_{ij} which itself is contained in U. Let Ψ_j

$$\Psi_j(t) = \Psi(t, c_j)$$

Then Ψ_j and Ψ_{j+1} are closed together and by applying the preceding lemma we get

$$\int_\gamma f = \int_{\Psi_0} f = \int_{\Psi_1} f = \cdots = \int_{\Psi_m} f = \int_\eta f$$ ■

A set S of complex numbers is conex, if for any $z, w \in S$ the segment $[z, w, \subset]S$. For examples, discs and rectangles are convex.

Lemma 10.16. *Let S be a convex set and γ, η continuous closed curves in S. Then, γ, η are homotopic in S.*

Proof.

$$\Psi(t, s) = s\gamma(t) + (1 - s)\eta(t)$$ ■

Open set U is simply connected if it is connected and every closed path is homotopic to a point in U. By the preceding lemma, every convec set is simply connected, as path connectednes implies connectedness.

10.5 Existence of global primitives

Theorem 10.17. *Let f be a holomorphic function on a simply connected open set U and let $z_0 \in U$. For an $z \in U$*

$$g(z) = \int_z^{z_0} f(\xi)\,\mathrm{d}\xi$$

is independent of the path in U and g is a primitive for f.

Example 10.1. Let U be a simply connected open set not containing 0. Pick $z_0 \in U$ and w_0 such that $e^{w_0} = z_0$, define

$$\log z = w_0 + \int_{z_0}^z \frac{1}{\xi}\,\mathrm{d}\xi$$

Then, log is a primitive for $\frac{1}{z}$. If $L(z)$ is another primitive for $\frac{1}{z}$ on U such that $e^{L(z)} = z$, then there exists an integer k such that

$$L(z) = \log z + 2\pi i k$$

10.6 Local Cauchy formula

Theorem 10.18 (Local Cauchy formula). *Let $\bar{D}$ be a closed disc of positive radius and f holomorphic on $\bar{D}$ (open disc U containing $\bar{D}$). Let γ be the circle with is the boundary of $\bar{D}$. Then, for every $z_0 \in D$*

$$f(z_0) = \frac{1}{2\pi i} \int_\gamma \frac{f(\xi)}{\xi - z_0} \, \mathrm{d}\xi$$

Theorem 10.19. *Let f be holomorphic on open set U then f is analytic on U.*

Theorem 10.20. *Let f be holomorphic on closed disc $\bar{B}_R(z_0)$, $R > 0$. Let C_R be the circle bounding the disc. Then, f has a power series expansion*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

whose coefficients are given by the formula

$$a_n = \frac{1}{n!} f^{(n)}(z_0) = \frac{1}{2\pi i} \int_{C_R} \frac{f(\xi)}{(\xi - z_0)^{n+1}} \, \mathrm{d}\xi$$

Furthermore, if $\|f\|_{C_R}$ denotes the sup-norm of f on C_R then

$$|a_n| = \frac{\|f\|_{C_R}}{R^n}$$

in particular, the radus of convergence of the series is greater than R.

The preceding theorems imply that a function is analytic if and only if it is holomorphic.

Definition: A function f is entire if it is holomorphic on all $\mathbb{C}$. Hence, entire functions have convergence radius of ∞.

Corollary 10.21. *Let f be entire and $\|f\|_R$ be its sup-norm on the circle of radius R. Suppose that there exists a constant C and a positive integer k such that*

$$\|f\|_R = CR^k$$

for arbitrarily large R. Then, f is a polynomial of degree $\leq k$.

Theorem 10.22 (Liouville's theorem). *A bounded entire function os constant.*

Corollary 10.23. *A polynomial over complex number which does not have root in $\mathbb{C}$ is constant.*

Theorem 10.24. *Let γ be a path in an oopen set U and g be a continuous function on γ. If z is not on γ define*

$$f(z) = \int_\gamma \frac{g(\xi)}{\xi - z} \, \mathrm{d}\xi$$

Then f is analytic on the complement of γ in U and its derivatives are given by

$$f^{(n)}(z) = n! \int_\gamma \frac{g(\xi)}{(\xi - z)^{n+1}} \, \mathrm{d}\xi$$

Corollary 10.25. *Let f be an analytic function on closed disc $\bar{B}_R(z_0)$ with $R > 0$. Let $0 < R_1 < R$ and denote the sup-norm of f on the circle of radius R by $\|f\|_R$. For $z \in \bar{B}_R(z_0)$*

$$\left|f^{(n)}(z)\right| \leq \frac{n!\,R}{(R-R_1)^{n+1}}\|f\|_R$$

Theorem 10.26 (Morera's theorem). *Let U be an open set in $\mathbb{C}$ and let be continuous on U. Assume that the integral of f along the boundary of every rectangle in U is zero. Then, f is analytic.*

Chapter 11

Winding Numbers

11.1 Winding number

Definition: Winding number of a closed path γ with respect to a point α is

$$W(\gamma, \alpha) = \frac{1}{2\pi i} \int_\gamma \frac{1}{z - \alpha} \, \mathrm{d}z = \frac{1}{2\pi i} \int_a^b \frac{\gamma'(t)}{\gamma(t) - \alpha} \, \mathrm{d}t$$

provided that γ does not pass through α.

Lemma 11.1. *If γ is a closed path, then $W(\gamma, \alpha)$ is an integer.*

Proof. Consider a the integral and compare it to $\gamma(t) - \alpha$. ■

Lemma 11.2. *Let γ be a path. Then, the function of α defined by*

$$\alpha \mapsto \int_\gamma \frac{1}{z - \alpha} \, \mathrm{d}z$$

for α no on the path, is a continuous function of α.

Lemma 11.3. *Let γ be a closed path and S a connected set not intersecting γ. Then,*

$$\alpha \mapsto \int_\gamma \frac{1}{z - \alpha} \, \mathrm{d}z$$

is a constant for α in S. If S is not bounded, then this constant is zero.

Definition: A closed path γ in U is homologous to 0 in U if

$$\int_\gamma \frac{1}{z - \alpha} \, \mathrm{d}{=}\, 0$$

for every point α not in U. Equivalently,

$$W(\gamma, \alpha) = 0 \quad \forall \alpha \in U^c$$

Furthermore, two closed paths γ, η are homologous in U if

$$W(\gamma, \alpha) = W(\eta, \alpha) \quad \forall \alpha \in U^c$$

Theorem 11.4. *1. If γ, η are closed paths in U and homotopic then they are homologous. Converse is true if their inside is contained in U.*

2. If γ, η are closed paths and close together in U then they are homologous.

Definition: Let $\gamma_1, \dots, \gamma_n$ be a sequence of curves and let $m_1, \dots, m_n$ be integers. The formal sum

$$\gamma = m_1\gamma_1 + \dots + m_n\gamma_n = \sum_{i=1}^{n} m_i\gamma_i$$

is called a chain. We define

$$\int_\gamma f = \sum_{i=1}^{n} m_i \int_{\gamma_i} f$$

and thus

$$W(\gamma, \alpha) = \frac{1}{2\pi i}\int_\gamma \frac{1}{z-\alpha}\,\mathrm{d}z = \sum_{i=1}^{n} m_i W(\gamma_i, \alpha)$$

Proposition 11.5. *If γ, η are closed chains in U then*

$$W(\gamma + \eta, \alpha) = W(\gamma, \alpha) + W(\eta, \alpha)$$

γ and η are homologous in U if

$$funcW\gamma, \alpha = W(\eta, \alpha) \quad \forall \alpha \in U^c$$

11.2 Cauchy's formula

Theorem 11.6 (Cauchy's theorem). *Let γ be a closed chain an open set U such that γ is homologous to zero in U. Let f be holomorphic in U. Then*

$$\int_\gamma f = 0$$

Corollary 11.7. *If γ, η are closed chains in U then*

$$\int_\gamma f = \int_\eta f$$

Theorem 11.8. *Let U be an open set and γ be a closed chain in U homologous to zero. Let $z_1, \dots, z_n$ be finite number of distinct points of U. Let γ_i be the boundary of closed disc $\bar{D}_i$ contained in U, containing z_i, and oriented counter-clockwise. We assume that $\bar{D}_i$ does not intersect $\bar{D}_j$ if $i \neq j$. Let*

$$m_i = W(\gamma, z_i)$$

Let U^* be the set U without $\{z_1, \dots, z_n\}$. Then γ is homologous to $\sum m_i\gamma_i$ in U^*.

Theorem 11.9 (Cauchy's formula). *Let γ be a closed chain in open set U homologous to zero in U. Let f be an analytic function on U and $z_0 \in U$ not on γ*

$$\frac{1}{2\pi i}\int_\gamma \frac{f(z)}{z - z_0}\,\mathrm{d}z = W(\gamma, z_0) f(z_0)$$

Proof. use analytic ■

Proof. Dixon, define $g : U \times U$

$$(w, z) \mapsto \begin{cases} \dfrac{f(w) - f(z)}{w - z} & w \neq z \\ f'(z) & w = z \end{cases}$$

for each w, $z \mapsto g(w, z)$ is analytic. g is continuous, define a bounded entire function. ■

11.2.1 Artin's proof

Lemma 11.10. *Let γ be a path in open U. Then, there exists a rectangular path η with the same endpoints such that γ and η are close together. In particular, γ and η are homologous in U and for any holomorphic function f on U we have*

$$\int_\gamma f = \int_\eta f$$

This lemma reduces Cauchy's formula to the case when γ is rectangular.

Definition: Let $\gamma_i = \gamma([a_i, a_{i+1}])$ then the chain

$$\gamma_1 + \cdots + \gamma_n$$

is a subdivision of γ. If η_i are obtained by a reparameterization of γ_i the

$$\eta_1 + \cdots + \eta_n$$

are also subdivision of γ.

Theorem 11.11. *Let γ be a rectangular closed chain in U and homologous to zero. Then there exists rectangles $R_1, \ldots, R_N$ contained in U such that if ∂R_i is the boundary of R_i oriented in counter-clockwise then*

$$\sum_{i=1}^{N} m_i \partial R_i$$

for some integers m_i is a subdivision of γ.

Proof. $m_i = W(\gamma, \alpha_i)$, $\alpha_i \in R_i$. If $m_i \neq 0$ for R_i then $R_i \subset U$. $\sum m_i \partial R_i$ is a subdivision. ■

Chapter 12

Application of Cauchy's formula

12.1 Uniform limits of analytical functions

Theorem 12.1. *Let $\{f_n\}$ be a sequence of holomorphic functions on a open set U. Assume that for each compact subset K of U, the sequence converges uniformly on K and let the limit function be f. Then f is holomorphic.*

Theorem 12.2. *Let $\{f_n\}$ be a sequence of holomorphic functions on a open set U. Assume that for each compact subset K of U, the sequence converges uniformly on K and let the limit function be f. Then the sequence of derivatives converges uniformly on every compact subset K, $f' = \lim f'_n$.*

12.2 Laurant seris

The series

$$f(z) = \sum_{n=-\infty}^{\infty} a_n z^n$$

is the Laurant series expansion. Let A be a set of complex number. Laurant series converges absolutely (uniformly) on A if the two series

$$f^+(z) = \sum_{n\geq 0} a_n z^n \qquad f^-(z) = \sum_{n<0} a_n z^n$$

converge absolutely (uniformly). In that case

$$f = f^- + f^-$$

Theorem 12.3. *Let A be the annulus $A = \{z \,|\, r \leq |z| \leq R\}$ for some $r \leq R$. Let f be holomorphic on A and $r < s < S < R$. Then f has Laurant expansion*

$$f(z) = \sum_{n=-\infty}^{\infty} a_n z^n$$

where

$$a_n = \begin{cases} \frac{1}{2\pi i} \int_{C_R} \frac{f(\xi)}{\xi^{n+1}} & n \geq 0 \\ \frac{1}{2\pi i} \int_{C_r} \frac{f(\xi)}{\xi^{n+1}} & n < 0 \end{cases}$$

converges absolutely on $s \leq |z| \leq S$.

12.3 Singularity

Let D be an open disc cenetered at z_0 and $U = D/\{z_0\}$. Let f be analytic on U. Then, f is said to have isolated singularity at z_0.

12.3.1 Removable singularity

Theorem 12.4. *If f is bounded in some neighbourhood of z_0, then one can define $f(z_0)$ in a uninque way such that f is analytic on z_0*

Proof. Laurant expansion ■

This kind of singularity is called Removable singularity.

12.3.2 Poles

If f has finite negative terms in its Laurant expansion

$$f(z) = \frac{a_{-m}}{(z - z_0)^m} + \cdots + a_0 + a_1(z - z_0) + \ldots$$

Then f is said to have a pole of order m. However, the order of f at z_0 is $-m$.

$$\text{ord}_{z_0} fg = \text{ord}_{z_0} f + \text{ord}_{z_0} g$$

Proposition 12.5. *f has a pole of order m at z_0 if and only if $f(z)(z - z_0)^m$ is holomorphic at z_0 and has no zero at z_0.*

Definition: f i defined on U/S, where S is a discrete set of points which are the poles of f, is mermorphic on U. Thus, it is the quotient of two holomorphic functions in the neighbourhood of a point.

12.3.3 Essential singularity

When the Laurant expansion has inifinite negative terms.

Theorem 12.6 (Casorati-Weirestrass). *Let 0 be an essential singularity of the function f. Let D be a disc cenetered at 0 on which f is holomorphic except at 0. Let U be the complement of $\{0\}$ in D. Then, $f(U)$ is dense in the complex numbers. In other words, the values of f on U come arbitrarily close to any complex number. In fact f takes on every complex value except possible one.*

Theorem 12.7. *The only analytic automorphism of $\mathbb{C}$ are functions of the form $f(z) = az + b$, where a, b are constants and $a \neq 0$.*

Chapter 13

Calculus of Residues

13.1 Residue

Let Laurent expansion of f at z_0 be

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$$

Then, the residue of f at z_0 denoted by $\operatorname{Res}(f, z_0)$ is a_{-1}.

Theorem 13.1. *Let z_0 be an isolated singularity of f, and let C be a small circle oriented counter-clockwise cenetered at z_0 such that f is holomorphic on C and its interior except possibly at z_0. Then*

$$\int_C f(\xi)\,\mathrm{d}\xi = 2\pi i \operatorname{Res}(f, z_0)$$

Theorem 13.2 (Residue formula). *Let U be an open set, and γ a closed chain in U such that γ is homologous to 0 in U. Let f be analytic on U except at a finite number of points $z_1, \ldots, z_n$. Then*

$$\int_\gamma f = 2\pi i \sum_{i=1}^{n} W(\gamma, z_i) \operatorname{Res}(f, z_i)$$

Lemma 13.3. *If f has a simple pole at z_0 and g is holomorphic at z_0*

$$\operatorname{Res}(fg, z_0) = g(z_0) \operatorname{Res}(f, z_0)$$

Proof. Suppose f has the following Laurent expansion

$$f(z) = \sum_{n=-1}^{\infty} a_n (z - z_0)^n$$

and g is the following power series expansion

$$g(z) = \sum_{n=0}^{\infty} b_n (z - z_0)^n$$

Then

$$
\begin{aligned}
fg(z) &= \left(\sum_{n=-1}^{\infty} a_n(z-z_0)^n\right)\left(\sum_{n=0}^{\infty} b_n(z-z_0)^n\right)\\
&= \frac{a_{-1}b_0}{z-z_0} + (a_{-1}b_1 + a_0b_0) + \dots
\end{aligned}
$$

therefore

$$\operatorname{Res}(fg, z_0) = a_{-1}b_0 = g(z_0)\operatorname{Res}(f, z_0) \qquad \blacksquare$$

Lemma 13.4. *Suppose $f(z_0) = 0$ but $f'(z_0) \neq 0$. Then, $\frac{1}{f}$ has a pole of order 1 at z_0 with*

$$\operatorname{Res}\left(\frac{1}{f}, z_0\right) = \frac{1}{f'(z_0)}$$

Proof. Suppose f has the following power series expansion

$$f(z) = \sum_{n=1}^{\infty} a_n(z-z_0)^n$$

with $a_1 \neq 0$. Then,

$$
\begin{aligned}
\frac{1}{f}(z) &= \frac{1}{(z-z_0)(a_1 + a_2(z-z_0) + \dots)}\\
&= \frac{1}{z-z_0}\left(a_1^{-1} + O(z-z_0)\right)
\end{aligned}
$$

has residue of $a_1^{-1} = (f'(z_0))^{-1}$ at z_0. $\blacksquare$

Lemma 13.5. *f has a pole order m and g is holomorphic at z_0*

$$\operatorname{Res}(fg, z_0) = \sum_{i=0}^{m-1} g^{(i)}(z_0)\operatorname{Res}\left((z-z_0)^i f(z), z_0\right)$$

Lemma 13.6. *Let f be meromorphic at z_0 then*

$$\operatorname{Res}\left(\frac{f'}{f}, z_0\right) = \operatorname{ord}_{z_0}(f)$$

Theorem 13.7. *Let U be an open set, γ closed chain in U, f meromorphic on U with only finite number of negative terms and poles at $z_0, \dots, z_n$, none of which lie on γ.*

$$\int_\gamma f'(f) = 2\pi i \sum_{i=1}^{n} W(\gamma, z_i)\operatorname{ord}_{z_0}(f)$$

Definition: Let γ be a closed path. It has interior and exterior if for all$\alpha \in \mathbb{C}/\gamma$ the winding the number $W(\gamma, \alpha) = 0, 1$. Then, the interior is defined as all the points with winding number of 1 and exterior is all the points with winding number of 0.

Theorem 13.8 (Rouche's). *Let γ be a closed path homologous to 0 in U and assume γ has in interior. Let f, g be analytic function on U and*

$$|f(z) - g(z)| < |f(z)|$$

for all z on γ. Then f and g have the same number of zeros in the interior of γ.

Theorem 13.9. *Let f be analytic in neighbourhood of a point z_0 and assumer $f'(z_0) \neq 0$. Then, f is a local analytic isomorphism at z_0.*

13.2 Evaluation of definite integrals

add the integrals

Part IV

Measure Theory and Probability Theory

Chapter 14

Measure Theory

14.1 Bernoulli sequences

Let $\mathcal{B}$ be the set of all Bernoulli sequences. $\mathcal{B}$ is uncountable.

Proposition 14.1. *If we delete a countable subset of $\mathcal{B}$, we can index what is left by the points on the real interval $I =]0,1]$. That is, there exists an injective function $f : I \to \mathcal{B}$.*

Proof. Each $\omega \in I$ can be written as

$$\begin{aligned} \omega &= \sum_{i=1}^{\infty} \frac{a_i}{2^i} \qquad a_i = 0, 1 \\ &= 0.a_1 a_2 \ldots \end{aligned}$$

Since ω may not have a unique binary representation, we will only consider non-terminating expansion for ω. Then, by mapping 1 to H and 0 to T we get an injective function from I to $\mathcal{B}$. To show this, suppose $\mathcal{B}_{\text{deg}}$ is the set of all Bernoulli sequences that after a certain point degenerates to all tails.

Lemma 14.2. $\mathcal{B}_{\text{deg}}$ *is countable.*

Proof. Let $\mathcal{B}^k_{\text{deg}}$ be all degenerate Bernoulli sequences where we have only tails after k_{th} toss. Then, $\mathcal{B}^k_{\text{deg}}$ is finite and

$$\mathcal{B}_{\text{deg}} = \bigcup_{k=1}^{\infty} \mathcal{B}^k_{\text{deg}}$$

is the countable union of finite set and hence $\mathcal{B}_{\text{deg}}$ is countable. □

This concludes the proof. ■

Definition (Borel Principle): Suppose E is a probabilistic event occuring in certain sequences. Let $\mathcal{B}_E$ denote the subset of $\mathcal{B}$ for which that event occurs. Let I_E be the corresponding subset of I, then

$$\mathbb{P}(E) = \mu_L(I_E)$$

where μ_L is Lebesgue measure.

Example 14.1. Start with X dollars and at each toss you win 1 dollars if heads shows up and ypu lose 1 dollars if tail shows up. What is the probability you lose all your original stake?

For $\omega \in I$ define the k_{th} *Radamcher* function, R_k, by

$$\begin{aligned} R_k(\omega) &= 2a_k - 1 \\ &= \begin{cases} +1 & a_k = 1 \\ -1 & a_k = 0 \end{cases} \end{aligned}$$

Then, let $S_k(\omega)$ be the total amount won or loss at k_{th} toss.

$$S_k(\omega) = \sum_{l=1}^{k} R_l(\omega)$$

Thus, the event that we lose our stake at k_{th} is

$$I_{E_k} = \left\{ \omega \in I \,\middle|\, S_l(\omega) > -X \text{ for } l < k, S_k(\omega) = -X \right\}$$

hence

$$I = \bigcup_{l=1}^{\infty} I_{E_k}$$

is the event that we loss all our money eventually. We will however postpone calculating $\mu_L(I_E)$ as it is not finite union of intervals.

14.2 Weak law of large numbers

For some fixed ϵ let

$$I_N = \left\{ \omega \in I \,\middle|\, \left| \frac{s_N(\omega)}{N} - \frac{1}{2} \right| > \epsilon \right\}$$

where $s_N(\omega) = a_1 + a_2 + \cdots + a_N$. Then, I_N represents the event that the number of heads and tails <u>are not</u> roughly equal.

Theorem 14.3 (Weak law of large numbers). *WLLN states*

$$\mu_L(I_N) \to 0 \text{ as } N \to \infty$$

Proof. Equivalently, for

$$A_N = \left\{ \omega \in I \,\middle|\, |S_N(\omega)| > 2N\epsilon \right\}$$

then

$$\mu_L(A_N) \to 0 \text{ as } N \to \infty$$

Lemma 14.4 (Chebyshev's inequality). *Let f be a non-negative, piecewise constant function on $]0,1]$. Let $\alpha > 0$ be given. Then,*

$$\mu_L\Big(\Big\{ \omega \in I \,\Big|\, f(\omega) > \alpha \Big\} \Big) < \frac{1}{\alpha} \int_0^1 f \, \mathrm{d}x$$

where $\int$ is the Riemann integral.

Proof. Since f is piecewise constant then there is a parition $0 = x_1 < \cdots < x_k = 1$ such that $f = c_i$ on $]x_i, X_{i+1}]$ for $i = 1, \ldots k-1$. Then,

$$\begin{aligned}\int_0^1 f \, \mathrm{d}x &= \sum_{i=1}^{k-1} c_i(x_{i+1} - x_i) \\ &\geq \sum_{c_i > \alpha} c_i(x_{i+1} - x_i) \\ &> \alpha \sum_{c_i > \alpha} x_{i+1} - x_i = \alpha \mu_L\Big(\Big\{\omega \in I \,\Big|\, f(\omega) > \alpha\Big\}\Big)\end{aligned}$$

which proves the Chebyshev's inequality. □

Then, we have

$$A_N = \Big\{\omega \in I \,\Big|\, |S_N(\omega)| > 2N\epsilon\Big\} = \Big\{\omega \in I \,\Big|\, (S_N(\omega))^2 > 4N^2\epsilon^2\Big\}$$

and hence

$$\begin{aligned}\mu_L(A_N) &< \frac{1}{4N^2\epsilon^2}\int_0^1 (S_N(\omega))^2 \, \mathrm{d}\omega \\ &= \frac{1}{4N^2\epsilon^2}\left[\sum_i \int_0^1 (R_i(\omega))^2 \, \mathrm{d}\omega + \sum_{i\neq j}\int_0^1 R_i(\omega)R_j(\omega) \, \mathrm{d}\omega\right] \\ &= \frac{1}{4N^2\epsilon^2} N = \frac{1}{4N\epsilon^2}\end{aligned}$$

Therefore, as N approaches infinity, $\mu(A_N)$ approaches zero. ■

Now we want to show that for a "typical" Bernoulli sequence

$$\frac{1}{2} - \frac{s_N(\omega)}{N} \to 0 \text{ as } N \to \infty \tag{14.1}$$

and by "typical" we mean Equation (14.1) fails on a set of zero proability or the equivalent event has measure zero.

Definition: A set $A \subset \mathbb{R}$ has Lebesgue measure zero if for every $\epsilon > 0$, there exists a countable covering $\{A_i\}$ of A by intervals such that

$$\sum_{i=1}^{\infty} \mu_L(A_i) < \epsilon$$

- Subset of a measure zero are measure zero.
- A signle point has a measure zero.
- countable union of measure zeros is a measure zero.

Let $N = \Big\{\omega \in I \,\Big|\, \frac{s_N(\omega)}{N} \to \frac{1}{2}\Big\}$. N is called the set of **normal numbers**. Let N^c be the complement of N.

Theorem 14.5 (Strong law of large numbers). *SLLN states that N^c has measure zero.*

Proof. Let $A_n = \{\omega \in I \,|\, |S_n(\omega)| > \epsilon n\}$ then

$$A_n = \left\{\omega \in I \,\middle|\, (S_n(\omega))^4 > \epsilon^4 n^4\right\}$$

By Chebyshev's inequality

$$\begin{aligned}\mu_L(A_n) &< \frac{1}{n^4\epsilon^4}\int_0^1 (S_n(\omega))^4 \,\mathrm{d}\omega \\ &= \frac{1}{n^4\epsilon^4}(n + 3n(n-1)) \le \frac{3}{\epsilon^4 n^2}\end{aligned}$$

Lemma 14.6. *Given $\delta > 0$ there exists a sequence $\epsilon_1, \epsilon_2, \ldots$ such that $\epsilon \to 0$ and*

$$\sum_{n=1}^{\infty} \frac{3}{n^2\epsilon_n^4} < \delta$$

Proof. Choose $\epsilon_n^4 = cn^{-1/2}$ for some constant c.

$$\sum_{n=1}^{\infty} \frac{3}{n^2\epsilon_n^4} = \frac{3}{c}\sum_{n=1}^{\infty} n^{-\frac{3}{2}} = \frac{3}{c}L < \delta$$

if $c > \dfrac{3L}{\delta}$. □

Finally, Let $B_n = \{\omega \in I \,|\, |S_n(\omega)| > \epsilon_n n\}$. Then, by our first computation

$$\mu_L(B_n) < \frac{3}{\epsilon_n^4 n^2}$$

and by the lemma

$$\sum \mu_L(B_n) < \frac{3}{\epsilon_n^4 n^2} < \delta$$

It remains to show that $N^c \subset \cup_{n=1}^{\infty} B_n$ which is obvious. As every $\omega \in N^c$ is some B_n. Therefore, N^c has measure zero. ■

14.3 Measure theory

14.3.1 Measure

Definition: A **ring** of sets in X is a non-empty collection $\mathscr{R}$ of subsets of X satisfying following two properties

1. $A, B \in \mathscr{R} \implies A \cup B \in \mathscr{R}$.
2. $A, B \in \mathscr{R} \implies A - B \in \mathscr{R}$.

Lemma 14.7. *If $A, B \in \mathscr{R}$, then $A \cap B \in \mathscr{R}$. Moreover, R is a ring if and only if it is closed under intersection and symmetric difference.*

Proof.

$$A \cap B = [(A \cup B) - (A - B)] - (B - A)$$

■

Example 14.2. $\mathcal{P}(X)$ is a ring.

Definition: A **semi-ring** is a non-empty collection of $\mathscr{S}$ such that

1. $\emptyset \in \mathscr{S}$.
2. $A, B \in \mathscr{S} \implies A \cap B \in \mathscr{S}$.
3. $A, B \in \mathscr{S} \implies A - B = \cup_{i=1}^{k} C_i$ where $C_i \in \mathscr{S}$ are disjoint.

The ring generated by a semi-ring $\mathscr{S}$ is denoted by $\mathcal{R}(\mathscr{S})$.

Theorem 14.8. *Suppose $\mathscr{S}$ is a semi-ring. Then, $A \in \mathcal{R}(\mathscr{S})$ if and only if $A = \cup_{i=1}^{k} C_i$ for disjoint $C_i \in \mathscr{S}$.*

Example 14.3. Let $\mathscr{R}_{\text{Leb}} = \{A \mid A = \bigcup_{i=1}^{n} A_i\}$ where A_i are disjoint k-cells in $\mathbb{R}^k$. $\mathscr{R}_{\text{Leb}}$ is a ring. Why?

Definition: Let $\mu : \mathscr{R} \to \overline{\mathbb{R}}_0^+$ where $\mathscr{R}$ is a ring of some set X, then μ is **additive** if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \mathscr{R}$ are disjoint.

Proposition 14.9. *Let $\mathscr{R}$ be a ring of X and $\mu : \mathscr{R} \to \overline{\mathbb{R}}_0^+$ is additive, then*

1. $\mu(\emptyset) = 0$.
2. *(Monotonicity) If $A, B \in \mathscr{R}$ with $A \subset B \implies \mu(A) \geq \mu(B)$.*
3. *(Finite Addtivity) For disjoint $A_1, \ldots, A_n \in \mathscr{R}$, $\mu(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$.*
4. *(Lattice property) $A, B \in \mathscr{R} \implies \mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.*
5. *(Finite subaddtivity) If $A_1, \ldots, A_n \in \mathscr{R}$, then*

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

Definition: μ is **countably additive** on $\mathscr{R}$ if given any countable collection $\{A_i\} \subset \mathscr{R}$ with A_i mutually disjoint and such that $A = \cup A_i$ is also in $\mathscr{R}$

$$\mu(A) = \sum_{i=1}^{\infty} \mu(A_i)$$

A countably additive, non-negative set function μ on ring $\mathscr{R}$ in X is called a **measure**.

Theorem 14.10. *If $X = \mathbb{R}^n$, $\mathscr{R} = \mathscr{R}_{\text{Leb}}$ and μ for n-cells is defined as*

$$\mu(A) = \prod_{i=1}^{n} (b_i - a_i)$$

where $A = \{x \in \mathbb{R}^n \mid x_i \in (a_i, b_i)\}$ – (a, b) denotes any of four possibilities, $]a, b[,]a, b], [a, b[$ and $[a, b]$. Then, μ is a measure.

To prove this theorem, we consider the following lemma

Lemma 14.11. *Let $A \in \mathscr{R}_{\text{Leb}}$ and let $\epsilon > 0$. There exists $F, G \in \mathscr{R}_{\text{Leb}}$ such that F is closed and G is open, $F \subset A \subset G$ and*

$$\mu(F) \geq \mu(A) - \epsilon$$
$$\mu(G) \leq \mu(A) + \epsilon$$

Proof.

Proof.

Definition: Let $\{A_n\}$ be a sequence of sets in X. Then,

$$\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \qquad\qquad \liminf A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$$

A_n is said to converge to A if $\limsup A_n = \liminf A_n = A$. A is said to be **limit set** of $\{A_n\}$. The sequence $\{A_n\}$ is increasing if $A_n \subset A_{n+1}$ for all n and it is decreasing if $A_n \supset A_{n+1}$ for all n.

It can be readily seen that if $\{A_n\}$ is increasing/decreasing, then it is convergent to $\cup A_n$/ $\cap A_n$.

Definition: Let $\mathscr{R}$ be a ring of subsets in X and $\mu : \mathscr{R} \to \overline{\mathbb{R}}_0^+$ is a set function. For any $E \in \mathscr{R}$, μ is said to be **continuous from below** if for all increasing sequences E_n, $\mu(E_n) \to \mu(E)$. Similarly, μ is said to be **continuous from above** if for all decreasing sequences E_n such that $\mu(E_i) < \infty$ for at least one i, $\mu(E_n) \to \mu(E)$. μ is continuous at E if it is both continuous from below and above.

Theorem 14.12. *A measure μ is continuous at every $E \in \mathscr{R}$.*

Proof.

Proposition 14.13. *Suppose $\mathscr{R}$ is ring of subsets of X and μ is a finite additive function.*

- *If $\mu : \mathscr{R} \to \overline{\mathbb{R}}_0^+$ is continuous from below at every $E \in \mathscr{R}$, then μ is a measure.*
- *If $\mu : \mathscr{R} \to \mathbb{R}_0^+$ is continuous from above at $\emptyset$, then μ is a measure.*

14.3.2 Caratheodory extension

Definition: Let A be a subset of X. A number $l \geq 0$ is called an **approximate outer measure** of A if there exists a covering of A by countable collection of sets $\{A_i\}$ with each $A_i \in \mathscr{R}$ such that

$$\sum_{i=1}^{\infty} \mu(A_i) \geq l$$

Remark 10. l is allowed to be $+\infty$.

Definition: Let A be a subset of X. The **outer measure** of A is the greates lower bound of the set $\{l \,|\, l \text{ isanapproximateoutermeasure}\}$.

$$\mu^*(A) = \inf\left\{ l \;\middle|\; A \subset \bigcup_{i=1}^{\infty} A_i, \sum_{i=1}^{\infty} \mu(A_i) \le l \right\}$$

If the set is empty, then $\mu^*(A) = +\infty$.

Remark 11. $\mu^* : \mathcal{P}(X) \to \overline{\mathbb{R}}_0^+$ is not a measure. However, μ^* is a measure on larger ring of subsets of X.

Proposition 14.14.

1. *If $A \in \mathscr{R}$, then $\mu^*(A) = \mu(A)$.*
2. *If $A \subset B$, then $\mu^*(A) \le \mu^*(B)$.*
3. *μ^* is countably subadditive.*

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

Proof.

Definition: A set $A \subset X$ is **measurable** with respect to μ if for all $E \subset X$,

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^c)$$

The set of all measurable sets with respect to μ is denoted by $\mathcal{M}(\mu)$.

Proposition 14.15. *Let μ be a measure defined on the ring $\mathscr{R}$. Then, $\mathcal{M}(\mu^*)$ is a ring and $\mathscr{R} \subset \mathcal{M}(\mu^*)$.*

Proof. $\mathcal{M}(\mu^*)$ is closed under complementing and intersection. ∎

Definition: Let $\mathscr{S}$ be a collection of subsets of a set X. $\mathscr{S}$ is called a **σ-ring** if

1. $\mathscr{S}$ is a ring.
2. $\mathscr{S}$ is closed under countable union. That is, given $\{A_i\} \subset \mathscr{S}$, $\bigcup A_i \in \mathscr{S}$.

Theorem 14.16. *$\mathcal{M}(\mu^*)$ is a σ-ring and the restriction μ^* on $\mathcal{M}(\mu^*)$ is a measure.*

Remark 12. The extension of μ to μ^* is not necessarily unique. However, by placing certain requirements on X we can deduce uniqueness.

Definition: Let $\mathscr{R}$ be a ring of subsets in X and $\mu : \mathscr{R} \to \overline{\mathbb{R}}_0^+$ be a measure. μ is **finite** if for each $A \in \mathscr{R}$, $\mu(A) < \infty$. μ is **σ-finite** if for each $E \in \sigma(R)$, there exists a sequence of subsets $\{E_n\} \subset \mathscr{R}$ such that $E \subset \cup E_n$ and $\mu(E_n) < \infty$ for all n.

Definition: A collection of subsets $\mathscr{C}$ is a **monotone class** if the limit set of every increasing and decreasing sequence of $\mathscr{C}$ is in $\mathscr{C}$. The smallest monotone class of a collection $\mathscr{D}$ is denoted by $\mathcal{C}(\mathscr{D})$.

Theorem 14.17. *If $\mathscr{R}$ is a ring and $\mathscr{C}$ is a monotone class containing $\mathscr{R}$, then $\sigma(\mathscr{R}) \subset \mathscr{C}$. In fact $\sigma(\mathscr{R}) = \mathcal{C}(\mathscr{R})$.*

Corollary 14.18. *Let $\mathscr{R}$ be a ring of subsets of X and $\mu, \nu : \sigma(\mathscr{R}) \to \overline{\mathbb{R}}_0^+$ are two measures. If μ and ν are finite and equal on $\mathscr{R}$, then they are equal on $\sigma(\mathscr{R})$.*

Definition: Let $\mathscr{D}$ be a collection of subset of X and $A \subset X$. Then

$$\mathscr{D}|_A = \{A \cap E \mid E \in \mathscr{D}\}$$

Proposition 14.19. *Let $\mathscr{R}$ be a ring. Then, $\sigma(\mathscr{R}|_A) = \sigma(\mathscr{R})|_A$.*

Theorem 14.20. *Suppose $\mathscr{R}$ is a ring of subsets of X and $\mu : \mathscr{R} \to \overline{\mathbb{R}}_0^+$ is a σ-finite measure. The restriction of μ^* to $\sigma(\mathscr{R})$ is the only extension of μ to $\mathscr{R}$.*

14.3.3 Metric extension

Let $\mu : \mathscr{R} \to \mathbb{R}_0^+$ be a measure on the ring $\mathscr{R}$. For $A, B \in \mathcal{P}(X)$, we define a *psuedo distance* function on $\mathcal{P}(X)$

$$d(A, B) = \mu^*(A \triangle B)$$

Note that, $d(A, B)$ may be $+\infty$ and $d(A, B) = 0$ does not imply that $A = B$. To go around this constraint, we consider the equivalence relation $\sim$ with $A \sim B$ when $d(A, B) = 0$. Then, d is metric on the equivalence classes $\mathcal{P}(X)/\sim$.

Proposition 14.21. *Suppose $A, B, C \in \mathcal{P}(X)$, then*

1. $d(A, B) = d(B, A)$.
2. $d(A, A) = 0$.
3. $d(A, B) \le d(A, C) + d(C, A)$

Proof.

Although, d is not quite a metric, we can still definte the notion of convergence; $A_i \to A$ if $d(A, A_i) \to 0$.

Proposition 14.22. *The Boolean operation in $\mathcal{P}(X)$ are continuous with respect to d. That is, if $A_n \to A$ and $B_n \to B$*

$$A_n \cup B_n \to A \cup B$$
$$A_n \cap B_n \to A \cap B$$
$$A_n^c \to A^c$$

Proposition 14.23. *μ^* is continuous in the following sense that for $A, B \in \mathcal{P}(X)$ where $\mu^*(A)$ or $\mu^*(B)$ is finite*

$$|\mu^*(A) - \mu^*(B)| \le d(A, B)$$

Definition: Let $\mathscr{M}_F$ be the closure of $\mathscr{R}$ in $\mathcal{P}(X)$. That is, $A \in \mathscr{M}_F$ whenever there exists a sequence $\{A_i\} \subset \mathscr{R}$ such that $d(A_i, A) \to 0$ as $i \to \infty$.

Theorem 14.24.

1. *$\mathscr{M}_F$ is a ring.*
2. *For $A \in \mathscr{M}_F$, $\mu^*(A) < +\infty$.*
3. *μ^* is a measure on $\mathscr{M}_F$.*

Definition: A is **measurable set**, $A \in \mathscr{M}$, if there exists a sequence $\{A_i\} \subset \mathscr{M}_F$ such that $A = \bigcup A_i$.

Theorem 14.25. *If $A \in \mathscr{M}$, then $A \in \mathscr{M}_F \iff \mu^*(A) < +\infty$.*

Theorem 14.26. *$\mathscr{M}$ is a σ-ring.*

Theorem 14.27. *If $A_1, A_2, \ldots$ is a countable collection of disjoint sets in $\mathscr{M}$ then*

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

That is, μ^* is a measure on $\mathscr{M}$.

We now investigate the relation between the measurablity in Caratheodory sense and metric sense. Note that, for measurablity in metric sese we assumed that μ is a finite measure. Therefore, we assume that μ^* is σ-finite with respect to $\mathscr{M}_F$.

Theorem 14.28. *Let $\mathscr{R}$ be a ring of subset of X and $\mu : \mathscr{R} \to \mathbb{R}_0^+$ be a measure. If $A \in \mathscr{M}_F$, then for every $E \subset X$*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

Theorem 14.29. *Let $\mathscr{R}$ be a ring of subset of X and $\mu : \mathscr{R} \to \mathbb{R}_0^+$ be a measure. If $\mu^*(A) < \infty$ and for every $E \subset X$*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

then, $A \in \mathscr{M}_F$.

Therefore, from the last two theorems we conclude that if the measure space of Caratheodory extension $(X, \mathscr{M}(\mu^*), \mu^*)$ is σ-finite, then both methods of extension result in the same extension.

14.3.4 Completion of measure spaces

Lemma 14.30. *Suppose $\mathscr{R}$ is a ring of subsets of X and $\mu : \mathscr{R} \to \overline{\mathbb{R}}_0^+$ is measure. Furthermore, let μ^* be the outer measure of μ. If $\mu^*(A) = 0$, then $A \in \mathcal{M}(\mu^*)$. ESpecially, for every subset $B \subset A$, $\mu^*(B) = 0$ and $B \in \mathcal{M}(\mu^*)$.*

Definition: A measure space $(X, \mathscr{F}, \mu)$ is **complete** if every subset of a null set, is in $\mathscr{F}$ and is measure zero.

Theorem 14.31. *Every measure space $(X, \mathscr{F}, \mu)$ can be uniquely extended to a complete measure space.*

Let $(X, \overline{\mathscr{F}}, \overline{\mu})$ be the extended complete measure described above. We shall investigate how $(X, \overline{\mathscr{F}}, \overline{\mu})$ is related to Caratheodory extension. Firstly, consider the following covering lemma.

Lemma 14.32. *Suppose $(X, \mathscr{F}, \mu)$ is a measure space. For every $E \subset X$, there exists a $A \in \mathscr{F}$ such that $E \subset C$ and $\mu^*(E) = \mu(C)$.*

Theorem 14.33. *Suppose $(X, \mathscr{F}, \mu)$ is a σ-finite measure space. If $(X, \overline{\mathscr{F}}, \overline{\mu})$ is the completion and $(X, \mathcal{M}(\mu^*), \mu^*)$ is the Caratheodory extension of $(X, \mathscr{F}, \mu)$, then $\overline{\mu} = \mu^*$ and $\overline{\mathscr{F}} = \mathcal{M}(\mu^*)$.*

Let μ be a measure on the ring $\mathscr{R}$ and $\mathscr{F} = \sigma(\mathscr{R})$ and let ν be the restriction of μ^* to $\mathscr{F}$.

Theorem 14.34. *For every $A \subset X$, $\mu^*(A) = \nu^*(A)$.*

Therefore, if $(X, \mathscr{F}, \nu)$ is σ-finite, then its completion is the same as $(X, \mathcal{M}(\mu^*), \mu^*)$

14.3.5 Lebesgue measure

Example 14.4. In the case of Lebesgue measure μ_L on $\mathscr{R}_{\text{Leb}}$, since it is a σ-finite measure, then its metric and Caratheodory extensions are equal. The restriction of μ_L^* to $\mathcal{M}(\mu_L^*)$, is called the **Lebesgue measure** and it is denoted by $\lambda_1 = \lambda$. The σ-field $\mathcal{M}(\mu_L^*)$ is called the **Lebesgue measurable sets** and it is denoted by $\Lambda^1 = \Lambda$.

Proposition 14.35. *Every open and closed subset of $\mathbb{R}^n$ is in $\mathscr{M}$.*

Corollary 14.36. *All countable unions and intersection of closed and open sets are measurable.*

Definition: The Borel sets, $\mathscr{B}$, is the σ-field generated by $\mathscr{R}_{\text{Leb}}$.

Proposition 14.37. *$\mathscr{B}$ contains all intervals and open sets. Moreover, it is the smallest σ-ring containing the open sets.*

Theorem 14.38. *If $A \in \Lambda$, there exists a Borel set $B \subset A$ such that $\lambda(A - B) = 0$. That is, A can be written as $A = (A - B) \cup B$ where B is Borel set and $\lambda(A - B) = 0$.*

Theorem 14.39. *For each $A \subset \mathbb{R}$ we have*

$$\lambda^*(A) = \inf\{\lambda(U) \mid A \subset U, U \text{ is open}\}$$

Corollary 14.40. *If $A \in \Lambda$ and if $\epsilon > 0$ is given, then there exists a Borel set such that $G \supset A$ and $\lambda(G - A) < \epsilon$.*

Corollary 14.41. *If $A \in \Lambda$, there exsists a Borel set $F \subset A$ with $\lambda(A - F) < \epsilon$.*

Corollary 14.42. *If μ is a measure on Λ that for each Borel set B, $\mu(B) = \lambda(B)$, then $\mu = \lambda$.*

Theorem 14.43. *If E is a Lebesgue measurable set, then*

$$\lambda(E) = \sup\{\lambda(K) \mid K \subset E, K \text{ is compact}\}$$

Theorem 14.44. *For each subset $A \subset \mathbb{R}$ and $c \in \mathbb{R}$, $\lambda^*(A + c) = \lambda^*(A)$ and $\lambda^*(cA) = |c|\lambda^*(A)$. Moreover, if A is Lebesgue measurable, then $A + c$ and cA are Lebesgue measurable as well.*

Theorem 14.45. *There exists a non-Lebesgue measurable set in $\mathbb{R}$.*

14.3.6 Finite signed measures

Definition: Suppose $(X, \mathscr{F})$ is measurable space. The set function $\nu : \mathscr{F} \to \mathbb{R}$ is a **finite signed measure** if it is countably additive. That is, for every sequence of disjoint subsets $\{A_n\}$, $\sum \nu(A_n)$ is convergent and

$$\nu\left(\bigcup_{i=1}^{\infty} A_n\right) = \sum_{i=1}^{n} \nu(A_n)$$

Since the order of right hand side summation does not matter, then the series is absolutely convergent.

Proposition 14.46.

1. *$\nu(\emptyset) = 0$.*
2. *ν is a finitely additive.*
3. *If $A, B \in \mathscr{F}$ and $A \subset B$, then $\nu(B - A) = \nu(B) - \nu(A)$.*
4. *ν is continuous from below at every $E \in \mathscr{F}$.*

Definition: Let ν be a finite signed measure on $\mathscr{F}$. For each $A \in \mathscr{F}$, the signed finite measure ν_A is defined as

$$\nu_A(E) = \nu(A \cap E)$$

Proposition 14.47. *ν_A is a measure if and only if for every subset $F \subset A$ that $F \in \mathscr{F}$, $\nu(F) \geq 0$.*

Proposition 14.48.

1. *$\nu_\emptyset(E) = 0$ for all $E \in \mathscr{F}$.*
2. *If $A, B \in \mathscr{F}$ and $A \cap B = \emptyset$, then $\nu_{A \cup B} = \nu_A + \nu_B$.*
3. *If $A, B \in \mathscr{F}$ and $A \subset B$, $\nu_{B-A} = \nu_B - \nu_A$.*
4. *If $A, B \in \mathscr{F}$, $\nu_{A \cap B} = (\nu_A)_B$.*
5. *If $A, B \in \mathscr{F}$, $\nu_{A \cup B} + \nu_{A \cap B} = \nu_A + \nu_B$.*

Theorem 14.49 (Hann-Jordan decomposition). *If ν is a signed finite measure on a σ-field $\mathscr{F}$, then there exists $A \in \mathscr{F}$ such that $\nu_A \geq 0$ and $\nu_{A^c} \leq 0$ hence*

$$\nu = \nu_A - (-\nu_{A^c})$$

That is, ν is the difference of two finite measures.

Proof. Let $\mathscr{N} = \{B \in \mathscr{F} \mid \nu_B \leq 0\}$. Then, $\mathscr{N}$ is closed under finite and countable union. Moreover if $B \in \mathscr{N}$ and $E \in \mathscr{F}$, then $B \cap E \in \mathscr{N}$. Consider the following lemma

Lemma 14.50. *The set $\{\nu(B) \mid B \in \mathscr{N}\}$ has an smallest element.*

then do some more work. ∎

14.4 Measure theoretic modeling

Definition: Let X be a set and $\mathscr{F}$ a ring of subsets of X.

1. $\mathscr{F}$ is a field if $X \in \mathscr{F}$.
2. $\mathscr{F}$ is a σ-field if $X \in \mathscr{F}$ and $\mathscr{F}$ is a σ-ring.

Definition: Let X be a set and $\mathscr{F}$ be a field of subsets of X. Suppose μ is a measure defined on $\mathscr{F}$. Then, μ is a probability measure if $\mu(X) = 1$. In this case, the triplet $(X, \mathscr{F}, \mu)$ is a called probability space.

Let X be a sample of space of a probabilistic process. A measure theoretic model of the proccess is a σ-field $\mathscr{F}$ of subsets of X and probability measure μ defined on $\mathscr{F}$. So that, for any "plausible" event E in X, we have $B_E \in \mathscr{F}$ and $\mathbb{P}(E) = \mu(B_E)$ where B_E is the set of points in X for which in E occurs.

Definition: Given set $B_1, B_2, \ldots$ in $\mathscr{F}$, then

$$\{B_i;\ \text{i.o.}\} = \limsup B_n = \bigcap_{k=1}^{\infty} \bigcup_{n \geq k} B_n$$

Theorem 14.51 (First Borel-Cantelli lemma). *Given a sequence $B_1, B_2, \ldots$ in $\mathscr{F}$ define $B = \limsup B_n$. Then, $\sum_{i=1}^{\infty} \mu(B_i) < \infty$ implies $\mu(B) = 0$.*

Definition: Let X be a sample space with σ-field $\mathscr{F}$ and probability measure μ. Two sets $A_1, A_2 \in \mathscr{F}$ are **independent** if

$$\mu(A_1 \cap A_2) = \mu(A_1)\mu(A_2)$$

More generally, $A_1, \ldots, A_n$ are independent, if for any subset I of $\mathbb{N}_n$

$$\mu\left(\bigcap_{i \in I} A_i\right) = \mathbb{P}(i \in I)\mu(A_i)$$

Furthermore, a countable collection of sets is independent if every finite subcollection is independent.

Theorem 14.52 (Second Borel-Cantelli lemma). *Assume $(X, \mathscr{F}, \mu)$ is a probability space and let $A_1, A_2, \ldots$ be an independent collection of sets from $\mathscr{F}$. Suppose that $\sum_{i=1}^{\infty} \mu(A_i)$ is not finite, then $\mu(\limsup A_n) = 1$.*

Lemma 14.53. *Let $A_1, A_2, \ldots$ be an independent collection of sets in $\mathscr{F}$. Then, $A_1^c, A_2^c \ldots$ is an independent collection of set in $\mathscr{F}$.*

Chapter 15

Integeration

Definition: A **measure space** is a triplet $(X, \mathscr{F}, \mu)$ where $\mathscr{F}$ is a σ-field of subsets of X and μ is a measure defined on $\mathscr{F}$. A **measurable space** is a pair $(X, \mathscr{F})$.

15.1 Measurable functions

Let $(X, \mathscr{F})$ and $(Y, \mathscr{S})$ be two measurable spaces. The function $f : X \to Y$ is measurable if for all $B \in \mathscr{S}$, $f^{-1}(B) \in \mathscr{F}$. That is, the σ-algebra generated by f, $\sigma(f) = \{f^{-1}(B) \mid B \in \mathscr{S}\}$ is a subset of $\mathscr{F}$. Moreover, if $\mathscr{F} = \mathscr{B}(X)$ and $\mathscr{S} = \mathscr{B}(Y)$ are the Borel set of X and Y, respectively, f is called a **Borel measurable** function.

Let $\overline{\mathbb{R}}$ denote the set of the *extended real numbers*, $\mathbb{R} \cup \{\pm\infty\} = [-\infty, +\infty]$. We may define addition and multiplication as follows

1. $\forall a \in R, -\infty < a < \infty.$
2. $\forall a \in R, a + (\pm\infty) = \pm\infty.$
3. $\forall a \in \mathbb{R}^+, a(\pm\infty) = \pm\infty.$
4. $(-1)(\pm\infty) = \mp\infty.$

The extended Borels sets, $\mathscr{B}(\overline{\mathbb{R}})$ are collection of subsets having the following form

$$A, \quad A \cup \{\pm\infty\}, \quad A \cup \{-\infty, +\infty\}$$

where A is a Borel set. The extended Borel set make a σ-field.

For the rest of this text, we may assume a measurable function $f : X \to \overline{\mathbb{R}}$ where $\overline{\mathbb{R}}$ is equipped with $\mathscr{B}(\overline{\mathbb{R}})$.

Lemma 15.1. *Suppose $f : X \to \overline{\mathbb{R}}$ is a function. The followings are equivalent*

1. *f is measurable.*
2. *For all $a \in \mathbb{R}$, $\{x \in X \mid f(x) > a\} \in \mathscr{F}$.*
3. *For all $a \in \mathbb{R}$, $\{x \in X \mid f(x) \geq a\} \in \mathscr{F}$.*
4. *For all $a \in \mathbb{R}$, $\{x \in X \mid f(x) < a\} \in \mathscr{F}$.*
5. *For all $a \in \mathbb{R}$, $\{x \in X \mid f(x) \leq a\} \in \mathscr{F}$.*

Example 15.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathscr{F} = \mathscr{M}$ the Lebesgue measurable sets. If f is continuous, then f is measurable.

Random variables are measurable functions from a measure space.

Theorem 15.2. *If f and g are measurable functions, then* $\max(f, g)$ *and* $\min(f, g)$ *are also measurable.*

Corollary 15.3. *Suppose f is a measurable function, then*

$$f^+(x) = \begin{cases} f(x) & f(x) \geq 0 \\ 0 & f(x) < 0 \end{cases} \qquad f^-(x) = \begin{cases} -f(x) & f(x) \leq 0 \\ 0 & f(x) > 0 \end{cases}$$

are measurable function. Since $f = f^+ - f^-$, then every function is the difference of two non-negative measurable functions.

Definition: Let f_i be functions of X to $\overline{\mathbb{R}}$. Then

$$\inf f_i(x) = \inf\{f_i(x)\} \qquad \sup f_i(x) = \sup\{f_i(x)\}$$

Theorem 15.4. *If $\{f_i\}$ are measurable functions, then* $\sup f_i$ *and* $\inf f_i$ *are measurable functions.*

Furthermore, we may define lim sup and lim inf as follows

$$\begin{aligned} \limsup f_i &= \lim_{n\to\infty} \sup_{i\geq n} f_i \\ &= \inf_n \sup_{i \geq n} f_i \end{aligned} \qquad \begin{aligned} \liminf f_i &= \lim_{n\to\infty} \inf_{i\geq n} \\ &\sup_n \inf_{i\geq n} f_i \end{aligned}$$

Corollary 15.5. *If $\{f_i\}$ is a collection of measurable functions, then* $\limsup f_i$ *and* $\liminf f_i$ *are measurable.*

Corollary 15.6. *Suppose $\{f_i\}$ are measurable and converge pointwise to f. Then, f is measurable.*

Remark 13. The restriction of a measurable function $f : (X, \mathscr{F}) \to (Y, \mathscr{S})$ to a measurable set $A \in \mathscr{F}$, is measurable as well. Since

$$f|_A^{-1}(B) = A \cap f^{-1}(B) \in \mathscr{F}$$

Hence, if a sequence of measurable function f_n converge pointwise to f on a measurable set A, then $f : A \to Y$ is measurable.

Measurablity of sum/multiplication/inverse of two measurable functions is requires some care. Particulary, to avoid situation like $+\infty + (-\infty)$.

Theorem 15.7. *Let $f_i : X \to \mathbb{R}$ be some measurable functions. Then, for a continuous function $G : \mathbb{R}^n \to \mathbb{R}$, $G(f_1, \ldots, f_n)$ is measurable.*

Definition: Suppose X and Y are two spaces equipped with their respective Borel sets, $\mathscr{B}(X)$ and $\mathscr{B}(Y)$. A function $f : X \to Y$ is Borel measurable if for all $S \in \mathscr{B}(Y)$, $f^{-1}(S) \in \mathscr{B}(B)$.

Theorem 15.8. *Every continuous function is Borel measurable.*

15.2 The Lebesgue Integral

The measurable function $s : X \to \mathbb{R}$ is a **step** function if it takes on only finite number of values. If the distinct values are $c_1, \ldots, c_n$ and $E_i = s^{-1}(c_i)$, then

$$s = \sum_{i=1}^{n} c_i \chi_{E_i}$$

Theorem 15.9. *s is a measurable if and only if $E_i \in \mathscr{F}$ for $i = 1, \ldots, n$.*

Example 15.2. Let $E \in \mathscr{F}$ then

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

is a step function. Furthermore, let s be a simple function that takes on values $c_1, c_2, \ldots, c_n$ and let $E_i = s^{-1}(c_i)$ for $i = 1, \ldots, n$. Then,

$$s = \sum_{i=1}^{n} c_i \chi_{E_i}$$

Definition: Let $s : X \to \mathbb{R}$ be a (non-negative) simple function and let $c_1, \ldots, c_n$ be distinct non-zero values of s with $E_i = s^{-1}(c_i)$. Let $E \in \mathscr{F}$ and define the *integral of s over E with respect to μ* as the sum

$$\int_E s = \sum_{i=1}^{n} c_i \mu(E \cap E_i)$$

Note that, the integral might be $+\infty$ since $\mu(E \cap E_i)$ might be $+\infty$.

Proposition 15.10. *Let s and r be simple non-negative functions and $E \in \mathscr{F}$.*

1. $\int_E s + r = \int_E s + \int_E r$.
2. $\int_E cs = c \int_E s$ *for* $c \geq 0$.
3. *If* $s \leq r$, *then* $\int_E s \leq \int_E r$.

Definition: Let $f : X \to \overline{\mathbb{R}}$ be a non-negative measurable function and $E \in \mathscr{F}$. The *integral of f over E with respect to μ* is defined as

$$\int_E f \, \mathrm{d}\mu = \sup \left\{ \int_E s \,\middle|\, s \leq f, s \text{ is simple} \right\}$$

Let s be a simple function. We need to check that the new definiton of integral is equivalent to the old one in the case of simple functions. That is

$$\int_E s \, \mathrm{d}\mu = \sum_{i=1}^{n} c_i \mu(E \cap E_i)$$

To justify why simple functions are used to approximate consider the following.

Theorem 15.11. *Let f be a non-negative measurable function. Then, there exists a sequence of non-negative simple functions*

$$0 \leq s_1 \leq s_2 \leq \ldots \leq f$$

such that $s_i \to f$ pointwise. Moreover, if f is bounded, $s_i \rightrightarrows f$.

Proof. Fix n and divide the interval $[0, n[$ to $n2^n$ subinterval of length 2^{-n}.

$$I_{n,i} = \left\{ \frac{i-1}{n2^n} \leq x < \frac{i}{n2^n} \right\} \qquad i = 1, \ldots n2^n$$

Then let $E_{n,i} = f^{-1}(I_{n,i})$ and $F_n = f^{-1}([n, +\infty])$. Note that $E_{n,i}$ and F are mutually disjoint and cover X.

$$s_n(x) = \sum_{i=1}^{n2^n} \left(\frac{i-1}{2^n} \right) \chi_{E_{n,i}} + n\chi_{F_n}$$

then $s_n \leq f$ and $s_n \leq s_{n+1}$ for all n. ■

In contrast to Riemann integral we approximate by dividing the range of the function. Removing the conditions on x-axis. Gives good approximation without f having to be continuous.

Proposition 15.12. *Let f and g be non-negative measurable functions and $E, F \in \mathscr{F}$. Then*

1. *If $f \leq g$, then $\int_E f \, \mathrm{d}\mu \leq \int_E g \, \mathrm{d}\mu$.*
2. *If $E \subset F$, then $\int_E f \, \mathrm{d}\mu \leq \int_F f \, \mathrm{d}\mu$.*
3. *If $\mu(E) = 0$, then $\int_E f \, \mathrm{d}\mu = 0$.*

Theorem 15.13 (Chebyshev). *Let f be a non-negative measurable function and let $E \in \mathscr{F}$ and $c > 0$. Define $E_c = \{x \in E \mid f(x) \geq c\}$, then*

$$\mu(E_c) \leq \frac{1}{c} \int_E f \, \mathrm{d}\mu$$

Corollary 15.14. *Let f be a non-negative measurable function with $\int_E f \, \mathrm{d}\mu < \infty$, then*

$$\mu(\{x \in E \mid f(x) = +\infty\}) = 0$$

Definition: If a property holds on a set $E \in \mathscr{F}$ except for a subset of zero measure, we say that the property holds **almost everywhere** on E.

Corollary 15.15. *Let f be a non-negative function and $E \in \mathscr{F}$.*

$$\int_E f \, \mathrm{d}\mu = 0 \implies f \equiv 0 \text{ a.e. on } E$$

Theorem 15.16. *Let f be a non-negative function and $A_1, A_2, \ldots$ pairwise disjoint from $\mathscr{F}$.*

$$\int_{\bigcup_{i=1}^{\infty} A_i} f \, \mathrm{d}\mu = \sum_{i=1}^{\infty} \int_{A_i} f \, \mathrm{d}\mu$$

We can use integrals to define measurse. *Gaussian measure* μ_G is defined on measurable subsets of $\mathbb{R}$

$$\mu_G(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} \, \mathrm{d}\mu_L$$

Moreover, μ_G is a probability measure.

Corollary 15.17. *Let f and g be a non-negative functions and $E \in \mathscr{F}$. If $f = g$ a.e.on E*

$$\int_E f \, \mathrm{d}\mu = \int_E g \, \mathrm{d}\mu$$

15.3 Further properties of integrals

Let $\{f_i\}$ be a sequence of measurable functions with

$$0 \leq f_1 \leq f_2 \leq \dots$$

Then, $f = \lim_{n\to\infty} f_n$ exists and is measurable.

Lemma 15.18. *Let f be a non-negative measurable function on X and let $E_1, E_2, \dots$ be a sequence of sets in $\mathscr{F}$ with $E_1 \subset E_2 \subset \dots$ and $E = \cup_i E_i$. Then*

$$\int_E f \, \mathrm{d}\mu = \lim_{i\to\infty} \int_{E_i} f \, \mathrm{d}\mu$$

Theorem 15.19 (Monotone convergence). *Let f and $\{f_i\}$ be described as above. Then for $E \in \mathscr{F}$*

$$\int_E f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_E f_n \, \mathrm{d}\mu$$

Remark 14. Let f be a non-negative measurable function and s_n be the step functions from the construction. By 15.19

$$\int_E s_n \, \mathrm{d}\mu \to \int_E f \, \mathrm{d}\mu$$

Theorem 15.20. *Suppose f and g are two non-negative measurable function, $c > 0$, and $E \in \mathscr{F}$*

1. $\int_E f + g \, \mathrm{d}\mu = \int_E f \, \mathrm{d}\mu + \int_E g \, \mathrm{d}\mu$.
2. $\int_E cf \, \mathrm{d}\mu = c \int_E f \, \mathrm{d}\mu$.

Corollary 15.21. *Let $\{f_i\}$ be non-negative measurable functions. Then, $\sum f_i$ is a non-negative measurable function and*

$$\int_E \sum_{i=1}^{\infty} f_n \, \mathrm{d}\mu = \sum_{i=1}^{n} \int_E f_n \, \mathrm{d}\mu$$

Lemma 15.22. *The following two conditions are equivalent*

1. $\int_E |f| \, \mathrm{d}\mu < +\infty$.
2. $\int_E f^+ \, \mathrm{d}\mu < +\infty$ *and* $\int_E f^- \, \mathrm{d}\mu < +\infty$.

Proof. $|f| = f^+ + f^-$. ■

Definition: A measurable function f is **integrable** over E if either of the conditions hold. In this case, $f \in \mathscr{L}(\mu, E)$. If $E = X$, then $f \in \mathscr{L}(\mu)$. For $f \in \mathscr{L}(\mu, E)$

$$\int_E f \, \mathrm{d}\mu = \int_E f^+ \, \mathrm{d}\mu - \int_E f^- \, \mathrm{d}\mu$$

Theorem 15.23. *Suppose $f, g \in \mathscr{L}(\mu, E)$ and $c \in \mathbb{R}$*

1. *$cf \in \mathscr{L}(\mu, E)$ and $\int_E cd \, \mathrm{d}\mu = c \int_E f \, \mathrm{d}\mu$.*
2. *$f + g \in \mathscr{L}(\mu, E)$ and $\int_E f + g \, \mathrm{d}\mu = \int_E f \, \mathrm{d}\mu + \int_E g \, \mathrm{d}\mu$.*
3. *If $f \leq g$, then $\int_E f \, \mathrm{d}\mu = \int_E g \, \mathrm{d}\mu$.*

Corollary 15.24. *Let $f \in \mathscr{L}(\mu, E)$, then*

$$\left| \int_E f \, \mathrm{d}\mu \right| \leq \int_E |f| \, \mathrm{d}\mu$$

Lemma 15.25 (Fatou's lemma). *Assume $f_1, f_2, \ldots$ are non-negative measurable function and let $f = \liminf f_n$*

$$\int_E f \, \mathrm{d}\mu \leq \liminf \int_E f_n \, \mathrm{d}\mu$$

Theorem 15.26 (Lebesgue dominated convergence). *Let $f_1, f_2, \ldots$ be a sequence of measurable functions and let $E \in \mathscr{F}$. Suppose the following assumptions hold*

1. *$\lim_{n\to\infty} f_n(x)$ exists for all $x \in E$.*
2. *There is a non-negative measurabe function $g \in \mathscr{L}(\mu, E)$ with $g \geq |f_n|$ on E for all n.*

Then, $f(x) = \lim_{n\to\infty} f_n(x)$ is integrable and

$$\int_E \lim_{n\to\infty} f_n \, \mathrm{d}\mu = \lim_{n\to\infty} \int_E f_n \, \mathrm{d}\mu$$

Corollary 15.27. *Let $\{f_i\}$ be a sequence of functions in $\mathscr{L}(\mu, E)$ with*

$$\sum_{i=1}^{\infty} \int_E |f_n| \, \mathrm{d}\mu < +\infty$$

Then,

1. *$\sum f_n$ converges absolutely a.e.on E and is integrable on E.*
2. $$\int_E \sum_{n=1}^{\infty} f_n \, \mathrm{d}\mu = \sum_{n=1}^{\infty} \int_E f_n \, \mathrm{d}\mu$$

15.4 Lebesgue integral vs Riemann integral

Theorem 15.28. *Let f be a bounded Riemann integrable function on $[a, b]$ with Riemann integral $\int_a^b f(x)\,\mathrm{d}x$. Then, $f \in \mathscr{L}(\mu_L, [a, b])$ and*

$$\int_a^b f\,\mathrm{d}x = \int_{[a,b]} f\,\mathrm{d}\mu_L$$

15.5 Radon-Nikodym theorem

Suppose $(X, \mathscr{F}, \mu)$ is a measure space and $f : X \to \overline{\mathbb{R}}$ is integrable. Define the **indefinite integer**, $f\dot{\mu} : \mathscr{F} \to \mathbb{R}$, as

$$f\dot{\mu}(E) = \int_E f\,\mathrm{d}\mu$$

Proposition 15.29.

1. *$f, g : X \to \overline{\mathbb{R}}$ are integerable, then $(f + g)\dot{\mu} = f\dot{\mu} + g\dot{\mu}$.*
2. *If $c \in \mathbb{R}$, then $(cf)\dot{\mu} = c(f\dot{\mu})$.*
3. *$f \geq 0$ if and only if $f\dot{\mu} \geq 0$.*
4. *$f\dot{\mu} = f^+\dot{\mu} - f^-\dot{\mu}$.*
5. *If $E \in \mathscr{F}$ and $\mu(E) = 0$, then $f\dot{\mu}(E) = 0$.*
6. *$f\dot{\mu}$ is countably additive.*

Therefore, $f\dot{\mu}$ is a finite sign measure and can be decomposed into

$$f\dot{\mu} = f^+\dot{\mu} - f^-\dot{\mu}$$

Definition: A measure ν is **absolutely continuous** relative to a measure μ, denoted by $v \ll u$ if

$$E \in \mathscr{F},\ \mu(E) = 0 \implies \nu(E) = 0$$

Hence, $f\dot{u}$ is absolutely continuous relative to μ.

Lemma 15.30. *Suppose ν is finite signed measure on $\mathscr{F}$ and $E \in \mathscr{F}$ such that $\nu(E) > 0$. Then, there exists a measurable subset $G \subset E$ such that $\nu_G \geq 0$ and $v(G) > 0$.*

Lemma 15.31. *Suppose μ and ν are two finite measure on $(X, \mathscr{F})$ such that $\nu \ll \mu$ and $\nu \neq 0$. Then, there exists a non-negative integrable function f such that $f\dot{\mu} \leq \nu$ and $f\dot{\mu}$.*

Theorem 15.32 (Radon-Nikodym theorem). *Suppose ν is finite signed measure and μ is a finite measure on $(X, \mathscr{F})$ such that $\nu \ll \mu$. Then, there exists an integrable function $f : X \to \overline{\mathbb{R}}$ such that $\nu = f\dot{\mu}$ and f is unique a.e..*

15.6 Fubini theorem

Let $(X, \mathscr{M}, \mu)$ and $(Y, \mathscr{N}, \nu)$ be two measure spaces. Let $X \times Y$ demote the space

$$X \times Y = \{(x, y) \,|\, x \in X, y \in Y\}$$

Definition: $A \times B \subset X \times Y$ is a **product set** if $A \in \mathscr{M}$ and $B \in \mathscr{M}$. The smallest σ-field in $X \times Y$ containing all product sets $A \times B$ is denoted by $\mathscr{M} \otimes \mathscr{N}$.

Definition: For $E \subset X \times Y$ and fix $x \in X$, let $E_x = \{y \in Y \,|\, (x, y) \in E\}$. E_x is called the x-slice of E.

Proposition 15.33. *IF $E \in \mathscr{M} \otimes \mathscr{N}$, then $E_x \in \mathscr{N}$.*

Corollary 15.34. *LEt $f : X \times Y \to \overline{\mathbb{R}}$ be meaurable with respect to $\mathscr{M} \otimes \mathscr{N}$. For fixed $x_0 \in X$, define $f_{x_0} : Y \to \mathbb{R}$ given by $f_{x_0}(y) = f(x_0, y)$. Then, for eah $x_0 \in X$, f_{x_0} is a measurable function on Y.*

Suppose X and Y are σ-finite. We now make a measure on $\mathscr{M} \otimes \mathscr{N}$ using μ and ν.

Definition: Let Z ba set and let $\mathscr{S}$ be a collection of subsets of Z. $\mathscr{S}$ is called a λ-system if the following three properties hold.

λ1. $Z \in \mathscr{S}$.

λ2. If $E_1 \subset E_2 \subset \ldots$ is an increasing sequence with each $E_n \in \mathscr{S}$, then

$$\bigcup_{i=1}^{\infty} E_n \in \mathscr{S}$$

λ3. If $E, F \in \mathscr{S}$ and $E \subset F$, then $F - E \in \mathscr{S}$.

Definition: Let $\mathscr{P}$ be a collection of subsets of Z. $\mathscr{P}$ is called a π-system if the following property holds.

π1. If $A, B \in \mathscr{P}$, then $A \cap B \in \mathscr{P}$.

Theorem 15.35 (Dynkin π-λ theorem). *If $\mathscr{S}$ is a λ-system and $\mathscr{P}$ is a π-system with $\mathscr{P} \subset \mathscr{S}$, then the smallest σ-field containing $\mathscr{P}$, $\sigma(\mathscr{P})$ is contained in $\mathscr{S}$.*

Proposition 15.36. *If $E \in \mathscr{M} \otimes \mathscr{N}$ and $\phi_E : X \to \mathbb{R}$ is defined by $\phi_E(x) = \nu(E_x)$, then ϕ_E is measurable.*

Definition: Let $E \in \mathscr{M} \otimes \mathscr{N}$ and define

$$\pi'(E) = \int_X \phi_E(x) \,\mathrm{d}\mu$$

to be the product measure on E.

Proposition 15.37. *π' is a measure.*

Suppose instead of x-slices we used y-slices and denoted the measure by π''.

Theorem 15.38 (Fubini, version 1).

$$\pi' = \pi''$$

Definition: The measure $\pi' = \pi''$ is denoted by $\mu \times \nu$ and is called the product measure on $\mathscr{M} \otimes \mathscr{N}$.

Example 15.3. Let $X = Y = \mathbb{R}$ and $\mathscr{M} = \mathscr{N} = \mathscr{B}(\mathbb{R})$. Also, let $\mu = \nu = \mu_L$ the Lebesgue measure on $\mathbb{R}$. We claim that, $\mathscr{M} \otimes \mathscr{N} = \mathscr{B}(\mathbb{R}^2)$ and $\mu \times \nu = \mu_L^2$, the Borel sets and Lebesgue measure in $\mathbb{R}^2$.

Example 15.4. We can show that

$$\mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m) = \mathscr{B}\big(\mathbb{R}^{m+n}\big)$$

and $\mu_L^m \times \mu_L^n = \mu_L^{m+n}$.

Theorem 15.39 (Fubini, version 2). *Let $f : X \times Y \to \mathbb{R}$ be a non-negative measurable function. Then*

1. *For each $x_0 \in X$, $f(x_0, y)$ is a measurable function of y.*
2. *For each $y_0 \in X$, $f(x, y_0)$ is a measurable function of x.*
3. *$\int_Y f(x,y)\,\mathrm{d}\nu$ is a measurable function of x.*
4. *$\int_X f(x,y)\,\mathrm{d}\mu$ is a measurable function of y.*
5. $$\int_{X\times Y} f(x,y)\,\mathrm{d}\mu\times\nu = \int_X\int_Y f(x,y)\,\mathrm{d}\nu\,\mathrm{d}\mu = \int_Y\int_X f(x,y)\,\mathrm{d}\mu\,\mathrm{d}\nu$$

Theorem 15.40 (Fubini, version 2). *Let $f : X \times Y \to \mathbb{R}$ be an integrable function. Then*

1. *For almost all $x \in X$, $f(x, y)$ is a integrable function of y.*
2. *For almost all $y \in X$, $f(x, y)$ is a integrable function of x.*
3. *$\int_Y f(x,y)\,\mathrm{d}\nu$ is equal a.e.to an integrable function on X.*
4. *$\int_X f(x,y)\,\mathrm{d}\mu$ is equal a.e.to an integrable function on Y.*
5. $$\int_{X\times Y} f(x,y)\,\mathrm{d}\mu\times\nu = \int_X\int_Y f(x,y)\,\mathrm{d}\nu\,\mathrm{d}\mu = \int_Y\int_X f(x,y)\,\mathrm{d}\mu\,\mathrm{d}\nu$$

15.7 Random variables, expectation values, and indepedence