# Contents

# Chapter 1

# The Real and Complex Number Systems

In this chapter, we talk about number systems. We begin our topic with rational number and shows rational number field has a shortcoming: rational number system does not have *least-upper bound property*.

We first talk about two topological concept: order set and least-upper bound property. Least-upper bound property implies greatest-lower bounded property (Theorem 1.11). We show that rational number system is an ordered set but does not have least-upper bound property.

In an algebraic view, number systems have an algebraic structure called field. Field $F$ is a set with two operations $+$ and $*$ and they follow distributive law. Under operation $+$, $F$ is a communitive group with identity element 0. Under operation $*$, $F$ is a communitive monoid, with all elements have inverse expect element 0. There are some basic arithmetic properties and rules under these two operations.

We combine the topological and algebraic structure and introduce the ordered field.

Now we back to our number system. We can extend rational number system to real number system and the real number system has least-upper bound property (Theorem 1.19). The proof is rather long and is therefore presented in the Appendix. By least-upper bound property of real number system, we give some properties of real number system:

1. There is a rational number between two real numbers (Theorem 1.20).

2. We can take the nth roots of positive reals (Theorem 1.21).

Then we give the concept of extended real number system and complex field. Extended real number system is no longer a field. The complex field can be defined as tuples with operation. So the image number $i$ has a clear meaning.

Finally we talk about the Euclidean spaces with inner product and the norm induced by inner product.

**Note 1** (Appendix)**.**

**Theorem 1.0.1** (Theorem 1.19)**.** *There exists an ordered field $R$ which has the least-upper-bound property.*

*Moreover, $R$ contains $\mathbb{Q}$ as a subfield.*

*Here we give the sketch of proof. We first construct the Dedekind cut $R$ from $\mathbb{Q}$. And shows that $R$ is an ordered field with the least-upper-bound property. Notices a cut may not has least upper bound in $\mathbb{Q}$, such as all rational number less than $\sqrt{2}$. For a rational number $r$, the rational cut $r^*$ which consists of all $p \in \mathbb{Q}$ such that $p < r$. And we can construct the isomorphism between $r^*$ and $r$. The second statement of the theorem is in isomorphic sense.*

*In this book the author does not prove a fact: any two ordered fields with the least-upper-bound property are isomorphic.*

**Remark 1.0.1.** *The statement in page 19 is not clear. It should be understand as follows: there is an integer $n$ such that $nw \in \alpha$ but $(n+1)w \notin \alpha$. $n$ can be negative.*

**Note 2.** *I wonder the relation between completeness of Dedekind cut and of Cauchy sequence. These are two different construction of real numbers. Also these construction invoke me the real number is the power set of rational number.*

# Chapter 2

# Basic Topology

In this chapter, we talk about point-set topology in $\mathbb{R}^n$. Before we talk about topological concept, we talk about set theory. One important observation for set $A$ being infinite set is there is a bijection from set $A$ to a proper subset of $A$. For infinite set, countable sets represent the smallest infinity since no uncountable set can be a subset of a countable set (Theorem 2.8). The countable union of countable sets are still countable.

Usually, the proof of all sequence whose elements are the digits 0 and 1 forming an uncountable set $A$ is to show that there is an element not match all listed sequences if we assume this set is countable. In this book author gives a direct proof: Let set of listed sequences be a countable subset of $A$, the "diagonal" element is always not in this countable subset. Thus every countable set of $A$ is a proper subset of $A$. If $A$ is countable, then $A$ is a proper subset of $A$. This is absurd. Thus $A$ is uncountable.

Now we begin our topic of point-set topology for $\mathbb{R}^n$. In this book, the definition of some concept like neighborhood, open set is different from general topology, since topology we talk about is induced by an usual metric. Here are two not obvious properties:

**Theorem 2.0.1** (Theorem 2.20)**.** *If $p$ is a limit point of a set $E$, then every neighborhood of $p$ contains infinitely many points of $E$.*

and its corollary: A finite point set has no limit points.

Some theorems are actually the definition of concept in general topology, such as

1. Neighborhood is an open set.
2. Set $A$ is open if and only if its complement set $A^c$ is closed.
3. Union of open set is open.

Then we talk about an important concept in topology: compactness. Compactness can be used for describe the space but open and closed not, since the properties of being open or closed depends on the space in which it is embedded. The theorems related to compactness are so many, but the proof of them is usually just a verification of definition. In this book we talk about compactness in metric space, however the following holds for a more general space, Hausdorff space.

1. Suppose $K \subset Y \subset X$. $K$ is compact relative to $X$ if and only if $K$ is compact relative to $Y$ (Theorem 2.33).

2. Compact subsets are closed (Theorem 2.34).

3. Closed subsets of compact sets are compact (Theorem 2.35).

Here is an important but simple theorem:

**Theorem 2.0.2** (Theorem 2.36). *$(K_\alpha)$ is a collection of compact subsets such that the finite intersection is nonempty, then $\cap_\alpha K_\alpha$ is nonempty.*

It is a trivial consequence of another equivalent definition of compactness: $A$ is compact if and only if any collection of closed subsets $(K_\alpha)$, finite intersection of $K_\alpha$ is nonempty implies $\cap_\alpha K_\alpha$ is nonempty. This definition is contrapositive of definition using finite cover.

Now we consider the one dimensional version of corollary of Theorem 2.36: For a sequence of decreasing interval $(I_n)$ in $\mathbb{R}^1$, $\cap_1^\infty I_n$ is not empty (Theorem 2.38). But the author uses a different proof idea. This different idea shows how the least-upper-bound property plays in infinite intersection for intervals in $\mathbb{R}^1$ and n-cells (closed cubes) in $\mathbb{R}^n$.

Here is an important property of space $\mathbb{R}^n$:

**Theorem 2.0.3** (Theorem 2.40). *Every n-cell is compact.*

The proof of Theorem 2.40 uses an important tool. It keeps separating n-cell and shows there is a point in sequences of decreasing sequence of n-cells. The existence of the point is guaranteed by compactness. But it can also be guaranteed by completeness. (Indeed, completeness and precompactness implies compactness, the separation steps actually show the $\mathbb{R}^n$ is precompact).

Theorem 2.41 generalizes Theorem 2.40, and gives the necessary condition for compactness in $\mathbb{R}^n$.

**Theorem 2.0.4.** *The following statements are equivalent for a set $E$ in $\mathbb{R}^n$:*

1. *$E$ is closed and bounded.*

2. *$E$ is compact.*

3. *Every infinite subset of $E$ has a limit point in $E$.*

The proof of Theorem 2.41, 3 implying 2, implicitly use the converse part of the sequence lemma (Lemma 21.2 in Munkres' Topology): Let $X$ be a metrizable space and $A \subset X$, then for any limit point $x$ of $A$, there is a sequence of points of $A$ that converges to $x$. If $X$ is not a metric space or is not metrizable, the sequence of convergent points can not be extracted. The proof also shows the limit point of this sequence is unique in metric space.

After compact sets, we study perfect sets. Here is one character of perfect sets: A non empty perfect sets in $\mathbb{R}^n$ is uncountable (Theorem 2.43). The proof given in this book is a little mysterious: $V_1$ is a neighborhood of $x_1$ but $V_n$ may be not a neighborhood of $x_n$ for $n > 1$. If we suppose a non empty perfect set is countable, we can construct a sequence of compact sets and shows the intersection contains no point. Thus we have proved Theorem 2.43

The Cantor set is an interesting example. It is compact since it is closed by intersection of closed set and obvious bounded. This intersection procure explains an usually misunderstanding: For finite $N$, $\cap_{n=1}^{N} E_n = E_N$ but expression $E_\infty$ has no meaning. We can only has a "legal" expression $\cap_{n=1}^{\infty} E_n$. Since every $E_n$ is closed (the index is always finite if we consider a particular index), we can state that $\cap_{n=1}^{\infty} E_n$ is always closed. It is also perfect. Since every neighborhood (segment $S$ contains $x$) of $x$ in Cantor set intersects other points $x_n$ (end point of $E_n$ contains $x$).

Finally we talk about connected sets. The definition of connected sets in this book is little different from general topology. Usually we say $A$ is connected if $A$ can not be written as union of two disjoint open set. And the statement $\bar{A} \cap B$ $A \cap \bar{B}$ are empty is an consequence of usual definition (Lemma 23.1 in Munkres' Topology). This is easy since $A = B^c$ and $B$ is open implies $A$ is closed. Thus $A$ and $B$ is both open and closed. The connected subsets of the line have a particularly simple structure: every point between two points in connected subsets is in connected subsets. This simple structure is by $\mathbb{R}^1$ being an ordered set.

# Chapter 3

# Numerical Sequences and Series

## 3.1 Limit of numerical sequences

In this chapter, we give some basic properties of limits and some convergence test of series. First we talk about limit in metric space. One conclusion is the converse part of the sequence lemma (Lemma 21.2 in Munkres' Topology). For a limit point $p$ of $E$, we can obtain a sequence in $E$ which converges to $p$ (Theorem 3.2 (d)). And the limit in metric space is unique (Theorem 3.2 (b)). Then we talk about algebraic operations on limit process in $\mathbb{R}^1$ and in $\mathbb{R}^n$, like addition, scalar multiplication, multiplication and division on convergent sequences.

Theorem 3.6 is a converse conclusion of Theorem 3.2 (d). Theorem 3.2 (d) says given a limit point of set $A$, we can obtain a convergent sequence from $A$. Theorem 3.6 guarantees the existence of the limit point for a sequence in compact set. The existence of the subsequence is also by the converse part of the sequence lemma (Lemma 21.2 in Munkres' Topology).

In Theorem 3.7, we show the subsequential limits form a closed set, which is important in definition of $\limsup$. Given a sequence $(s_n)$, let set $E$ contains all limits of subsequence. The idea of the proof of Theorem 3.7 is to draw a subsequence from each converge subsequence which the limit is converge to the limit point of $E$.

Then we talk about the Cauchy sequence. Let set $E_N$ consists of the points $p_N, p_{N+1}, p_{N+2}, \ldots$. The equivalent of $(p_n)$ is a Cauchy sequence is $\lim_{N\to\infty} \operatorname{diam} E_N = 0$.

Here is a theorem about limit of compact sets:

**Theorem 3.1.1** (Theorem 3.10)**.** *If $K_n$ is a sequence of compact sets in metric space $X$ such that $K_n \supset K_{n+1}$ and $\lim \operatorname{diam} K_n = 0$, then $\cap_1^\infty K_n$ consists of exactly one point.*

*Empty set is compact since every finite set is compact.* So the statement of

Theorem 3.10 is ambiguous since there are two definitions of diameter of empty set. We must exclude the case $K_n = \emptyset$. However, the author exclude the definition of diameter for empty set. Thus we can assume $K_n$ is not empty safely. Also, the metric space guarantees the $\cap_1^\infty K_n$ consists exactly one point. Theorem 3.11 gives the equivalence of Cauchy sequence and convergent sequence in $\mathbb{R}^n$ The proof of Theorem 3.11 uses boundedness and closed sets is compact for metric space. If we only concern the convergence of Cauchy sequence, we have concept "completeness" to describe the space that every Cauchy sequence is convergent. For boundedness and convergence, convergence implies boundedness, and the converse is true if we talk about monotonic sequence. The limit is given by least upper bound. The proof is easy if we using least upper bound

One important concept is upper and lower limits. For any real number sequence, the upper and lower limits always exist if we allow the $+\infty$ and $-\infty$ for limit. The definition of upper limit is the sup of all limits of subsequence and we denote as $\limsup_{n\to\infty} s_n = \sup E$, where $E$ contains all limits of subsequence. The lower limits are similarly defined. Theorem 3.17 says there actually exists a subsequence which the limit is $\limsup s_n$ (Theorem 3.17 (a)), and for any number $x > \limsup s_n$, $s_n < x$ for $n > N$ with large enough $N$ (Theorem 3.17 (b)). Then the uniqueness of $\limsup s_n$ is by these two properties.

The upper limit preserves the order. Precisely, $s_n \le t_n$ implies $\limsup s_n \le \limsup t_n$ (Theorem 3.19). We can prove by contradiction and by Theorem 3.17 (b). *Here is an useful trick. If we assume $a > b$, we can always find $c$ with $a > b > c$.* $\limsup s_n > c > \limsup t_n$ will leads a contradiction.

Finally author gives some special and important limits:

**Theorem 3.1.2** (Theorem 3.20)**.**

1. *If* $p > 0$, *then* $\lim_{n\to\infty} \frac{1}{n^p} = 0$.
2. *If* $p > 0$, *then* $\lim_{n\to\infty} \sqrt[n]{p} = 0$.
3. $\lim_{n\to\infty} \sqrt[n]{n} = 1$
4. *If* $p > 0$ *and* $\alpha$ *is real, then* $\lim_{n\to\infty} \frac{n^\alpha}{(1+p)^n} = 0$.
5. *If* $|x| < 1$, *then* $\lim_{n\to\infty} x^n = 0$

Many proofs use binomial theorem and "sandwich" rule.

## 3.2 Convergent criterion and some special series

One of most hardest and important part of this book is series and its convergent criterion. We can also use Cauchy criterion for series. It has some easy consequences:

1. If $\sum a_n$ converges, $\lim a_n = 0$.
2. Absolutely convergence implies convergence.

3. Comparison test: If $|a_n| < c_n$ for $n > N_0$, where $N_0$ is some fixed integer, then $\sum c_n$ converges implies $\sum a_n$ converges.

Here is another useful criterion of convergence. This theorem says a rather "thin" subsequence of $(a_n)$ determines the convergence or divergence of $\sum a_n$.

**Theorem 3.2.1** (Theorem 3.27). *Suppose $a_1 \geq a_2 \geq a_3 \geq a_4 \geq \cdots \geq 0$. Then the series $\sum a_n$ converges if and only if the series $\sum_{k=0}^{\infty} 2^k a_{2^k}$ converges.*

The proof of this theorem is by showing the partial sum $\sum_{k=0}^{n} 2^k a_{2^k} \leq 2\sum_{k=0}^{m} a_n$ with $m > 2^n$ and $\sum_{k=0}^{n} 2^k a_{2^k} \geq \sum_{k=0}^{m} a_n$ with $m < 2^n$. By this theorem, we can examine the convergence and divergence condition for series $\sum \frac{1}{n^p}$, $\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p}$ and more complex series involving $\log n$.

Before we continue our convergent criterion, we talk about the number $e$. We give it definition and shows it is the limit of an important sequence: $\lim_{n\to\infty}(1+\frac{1}{n})^n = e$. By the estimation of convergence speed, we can prove $e$ is irrational.

There are another two convergent tests frequently used: the root tests and the ratio tests:

**Theorem 3.2.2** (Root test). *Given $\sum a_n$, put $\alpha = \limsup \sqrt[n]{a_n}$. Then*

1. *If $\alpha < 1$, $\sum a_n$ converges.*
2. *If $\alpha > 1$, $\sum a_n$ diverges.*
3. *If $\alpha = 1$, the test gives no information.*

**Theorem 3.2.3** (Ratio test). *Given $\sum a_n$. Then*

1. *$\sum a_n$ converges if $\limsup \left|\frac{a_{n+1}}{a_n}\right| < 1$*
2. *$\sum a_n$ diverges if $\limsup \left|\frac{a_{n+1}}{a_n}\right| \geq 1$ for all $n > N_0$, where $N_0$ is a fixed number.*

The root tests are more powerful but the ratio tests are easy to apply. The relation of two tests is:

**Theorem 3.2.4.** *For any sequence $(c_n)$ of positive numbers,*

$$\liminf_{n\to\infty} \frac{c_{n+1}}{c_n} \leq \liminf_{n\to\infty} \sqrt[n]{c_n},$$

$$\limsup_{n\to\infty} \sqrt[n]{c_n} \leq \limsup_{n\to\infty} \frac{c_{n+1}}{c_n}$$

*Here is another useful trick. If we prove $\limsup a_n < \beta$ for any $\beta > \alpha$, we have $\limsup a_n < \alpha$. This is also another version of above trick: If we assume $a > b$, we can always find $c$ with $a > b > c$.*

Now we talk about convergent tests for power series and product of two series. For power series, we have radius of convergence of $\sum c_n z^n$. And the test is familiar with root tests and ratio tests. For product of two series, we first introduce the partial summation formula:

$$\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q-1} A_n(b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p \tag{3.2.1}$$

where $A_n = \sum_{k=0}^{n} a_k$. If $A_n$ is bounded and $b_n$ decrease monotonically to 0, then $\sum a_n b_n$ converges (Theorem 3.42). This implies alternating series with absolute value of each term decreasing monotonically to 0 converges. The strength of partial summation formula is that the comparison test is a test for absolute convergence, but the partial summation formula can be used for test for non-absolute convergence.

The Cauchy product $\sum c_n$ where $\sum_{k=0}^{n} a_k b_{n-k}$ is used for product of two series $(\sum a_n)(\sum b_n)$. And it's partial sum converges if one of series converges absolutely (Theorem 3.50). Theorem 3.51 shows the sum of Cauchy product is actually the product of two series if the sum are converges.

## 3.3 Rearrangements

Finally in this chapter we talk about rearrangements of series. The rearrangements of non-absolutely convergent series can be any value in extended real field. More precisely, the lim inf and lim sup of rearrangements can be any value in extended real field (Theorem 3.54). This theorem is due to Riemann. But the rearrangements of absolutely convergent series does not affect the limit.

If an infinite series converges, then the associative property holds (by $b_k = a_{k_1+1} + a_{k_1+2} + \cdots + a_{k_2}$ and considering $\sum b_k$). But a non converge series has no associative property (consider $\sum(-1)^n$).

## 3.4 Notes and Errata

**Note 3** (convergence of $1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} \ldots$)**.** *We use Cauchy criterion. The finite sum has associative property, thus*

$$\sum_{k=n}^{m} a_k = r_n + \sum_{k=3i}^{k=3j} \left(\frac{1}{4k-3} + \frac{1}{4k-1} - \frac{1}{2k}\right) + r_m$$

*where $3i$ is the first number greater than $n$ and $3j$ is the last number less than $m$. Notice:*

$$\sum_{k=3i}^{k=3j} \left(\frac{1}{4k-3} + \frac{1}{4k-1} - \frac{1}{2k}\right) = \sum_{k=3i}^{k=3j} \frac{8k-3}{2k(4k-3)(4k-1)} \leq \sum_{k=3i}^{k=3j} \frac{C}{k^2} \leq \sum_{k=n}^{k=m} \frac{C}{k^2}$$

*We know $|r_n|$, $|r_m|$ and $\sum_{k=n}^{k=m} \frac{C}{k^2}$ tends to 0 for $n, m > N$ for large enough $N$. So $|\sum_{k=n}^{m} a_k|$ tends to 0 for $n, m > N$ for large enough $N$.*

*Thus*

$$1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\frac{1}{9}+\frac{1}{11}-\frac{1}{6}\dots$$

*converges.*

**Note 4** (Simple rearrangement of the Alternating Harmonic Series)**.**

**Definition 3.4.1** (simple rearrangement)**.** *A simple rearrangement of a series is a rearrangement of the series in which the positive terms of the rearranged series occur in the same order as the original series and the negative terms occur in the same order.*

**Definition 3.4.2** (asymptotic density)**.** *If $\sum a_n$ is a simple rearrangement of the Alternating Harmonic Series, let $p_k$ be the number of positive terms in the first $k$ terms, $\{a_1, a_2, a_3, \dots, a_k\}$. The asymptotic density, $\alpha$, of the positive terms in the rearrangement is $\alpha = \lim_{k\to\infty} \frac{p_k}{k}$. if the limit exists.*

**Theorem 3.4.3.** *A simple rearrangement of the Alternating Harmonic Series converges to an extended real number if and only if $\alpha$, the asymptotic density of the positive terms in the rearrangement, exists.*

*Moreover, the sum of a rearrangement with asymptotic density $\alpha$ is $\ln 2 + \frac{1}{2}\ln(\frac{\alpha}{1-\alpha})$*

**Note 5** (proof of Theorem 3.54)**.** *We choose real-valued sequences $(a_n)$, $(b_n)$ such that $a_n \to a$, $b_n \to b$, $a_n < b_n$, $b_1 > 0$.*

*But no such restriction that $b_2 > a_1$, so the second step may not be achieved, except we mean $m_2$ here is greater or equal to 0. If $b_2 < a_1$, we just let $m_2 = 0$, which means no $P_n$ term added. If the last terms are $P_{m_n}$, then $m_n > 0$. And we have $x_n - P_{m_n} < b_n$ by construction of $x_n$ ($m_n$ is the smallest integers such that $x_n > b_n$ and $m_n > 0$ guarantees there is a $P_{m_n}$ in $x_n$). Then we have $x_n - b_n < P_{m_n}$. and $x_n > b_n$ guarantees $|x_n - b_n| < P_{m_n}$.*

*For another inequality, $y_n < a_n$ and $y_n + Q_{k_n} > a_n$ implies $Q_{k_n} > a_n - y_n$ and thus $Q_{k_n} > |a_n - y_n| = |y_n - a_n|$. Since $P_n \to 0$ and $Q_n \to 0$, we see that $x_n \to b$, $y_n \to a$*

*Let $x'_{m_n+1} = x'_{m_n+2} = \dots = x'_{m_{n+1}} = x_n$, $x'_{k_n+1} = x'_{k_n+2} = \dots = x'_{k_{n+1}} = x_{n-1}$. We have the partial sum $s_n$ satisfies $s_n < x'_n$ for all $n$. Thus by Theorem 3.19, $\limsup s_n \le \limsup x'_m = \lim x'_m = b$. The $\liminf$ part are almost the same.*

# Chapter 4

# Continuity

We can define the limit under a function $\lim_{x \to p} f(x) = q$ either using neighborhood language:

$$\forall \epsilon > 0, \exists \delta > 0, 0 < d_X(x,p) < \delta \Rightarrow d_Y(f(x) - q) < \epsilon$$

or sequence language:

$$\lim_{n \to \infty} f(p_n) = q \quad \forall (p_n) \ s.t. \ p_n \neq p, \lim_{n \to \infty} p_n = p$$

. This limit operation (or we can call it topological operation), is communitive with arithmetic operation on function like addition and multiplication. If we give the value of function at the limit point, we can talk about continuous functions. Notice if point $p$ is an isolated point, every function is continuous at $p$ by definition.

There are some equivalent characterizations of continuous function:

1. $f$ is continuous.
2. For every subset $A$, one has $f(\bar{A}) \subset \overline{f(A)}$.
3. For every open subset $V$, $f^{-1}(V)$ is open.
4. For every closed subset $V$, $f^{-1}(V)$ is closed.

By above characterizations, we can easily prove that the composition of continuous functions are continuous function. Also,addition and multiplication preserves continuity by properties of limit operation.

Recall we have two identity: $f(f^{-1}(E)) \subset E$ and $f^{-1}(f(E)) \supset E$. This is useful when prove continuous function maps compact set to compact set. By the compact of image and least upper bound property of $\mathbb{R}^1$, $f$ can attain the maximum and minimum value. Also, a continuous map from compact metric space to another metric space is a homeomorphism (continuous bijection with continuous inversion).

**Remark 4.0.1.** *The first inclusion is equality if $f$ is surjective and the second inclusion is equality if $f$ is injective.*

The counterexample for $E$ not compact on $\mathbb{R}^1$ (not closed or not bounded) is interesting (Theorem 4.20)

1. Continuous function but not bounded on a bounded set $E$ (hence $E$ is not closed): $f(x) = \frac{1}{x-x_0}$ with $x_0$ is a limit point not in $E$.

2. Continuous function but not uniformly continuous on a bounded set $E$ (hence $E$ is not closed): $f(x) = \frac{1}{x-x_0}$ with $x_0$ is a limit point not in $E$.

3. Continuous function but not attained maximum on bounded set $E$ (hence $E$ is not closed): $g(x)\frac{1}{1+(x-x_0)^2}$ with $x_0$ is a limit point not in $E$.

4. For a non-bounded set $E$, $f(x) = x$ and $h(x) = \frac{x^2}{1+x^2}$ are counterexample for boundedness and maximum. But there is no example for uniformly continuous for any non compact and unbounded set $E$. Indeed, if $E$ is the set of all integers, any function on $E$ is uniformly continuous.

If $E$ is not compact. there is example, a function $f$ is not a homeomorphism even if $f$ is bijective and continuous and $f(E)$ is compact ($f(x) = (\cos x, \sin x)$).

Another topological property related to continuous function is continuous function maps connected set to connected set. One consequence of this property is intermediate values theorem. However, intermediate value property does not implies continuous ($\sin \frac{1}{x}$ with 0 at $x = 0$).

Monotonic functions has many good properties. The right and left limit of points always exist (Theorem 4.29). In proof of Theorem 4.29, we show for any $\epsilon > 0$, there exists $\delta > 0$ s.t. $|f(t) - A| < \epsilon$ with $x - \delta < t < x$. However, no hint shows the $\delta$ tends to 0. But if $(t_n) \subset (x - \delta, x)$, then $A - \epsilon < \liminf f(t_n) \leq \limsup f(t_n) < A + \epsilon$. Thus for any converge sequence $(t_n)$ less than $x$, $A - \epsilon < \lim(f_n) < A + \epsilon$. Hence $f(x-) = A$.

An interesting example for monotonic function is in Remark 4.31. It construct a monotonic function with countable discontinuous point s which is dense in $(a, b)$. The construction is by an absolutely convergent series and all rationals less than a particular value $x$. Formally:

$$f(x) = \sum_{x_n < x} c_n$$

where $\sum c_n$ converges absolutely. The function constructed can be either left continuous or right continuous.

# Chapter 5

# Differentiation

The steps for introducing differentiation is same as for continuity:

1. definition, notice the derivative of end points are left (right) derivative:

$$f'(x) = \lim_{t \to x} \frac{f(t) - f(x)}{t - x} \quad t \neq x$$

2. arithmetic operation under differentiation, like addition, multiplication and division.

3. composition under differentiation: $(f \circ g(y))' = f'(g(y))g'(y)$.

Next we talk about mean value theorem. The frequently used form is

$$f(b) - f(a) = (b - a)f'(x) \quad x \in (a, b)$$

There is also an intermediate value theorem for derivative of $f$. But we need not assume the continuity for derivative. The proof is by construct function $g(x) = f(x) - \lambda x$ and shows there is a maximum (minimum) value for $g(x)$ and thus $g'(x) = 0$. Notice by definition of derivative, $g'(a) < 0$ implies $g(x_1) < g(a)$ for some $a < x_1 < b$. An consequence of this intermediate value theorem is that $f'$ cannot have any simple discontinuities.

The proof of L'Hospital's rule uses density of $\mathbb{R}$ a lot. Notice to apply L'Hospital's rule, one of following condition needs to be satisfied:

1. $f(x) \to 0$ and $g(x) = 0$ as $x \to a$

2. $g(x) \to +\infty$ or $g(x) \to -\infty$ as $x \to a$

In the end of this chapter we talk about higher order derivative and Taylor's theorem. To make $f^{(n)(x)}$ exists at $x$. $f^{(n-1)}(x)$ must exist in a neighborhood of $x$. And to make proceeding statement true, $f^{(n-2)}(x)$ must be differentiable in that neighborhood.

For vector-valued functions, the derivative exists if the norm of different between quotient and a value tends to 0:

$$\lim_{t\to x}\left|\frac{\mathbf{f}(t)-\mathbf{f}(x)}{t-x}-\mathbf{f}'(x)\right|=0$$

The arithmetic operation under differentiation still holds. But the mean value theorem and L'Hospital's rule fails. However, a mean value inequality holds for vector-valued functions: $|\mathbf{f}(b)-\mathbf{f}(a)|\le(b-a)\,|\mathbf{f}'(x)|$

## 5.1 Notes and Errata

**Note 6** (additional material for proof of Theorem 5.13)**.** *We give analogous proof for $g(x)\to-\infty$ as $x\to a$.*

*We already have:*

1. *There is a point $c\in(a,b)$, s.t. $\frac{f'(x)}{g'(x)}<r$ for all $x\in(a,c)$.*
2. *If $a<x<y<c$, then there is a point $t\in(x,y)$ s.t. $\frac{f(x)-f(y)}{g(x)-g(y)}=\frac{f'(t)}{g'(t)}$ and by above statement, $\frac{f(x)-f(y)}{g(x)-g(y)}<r$.*

*Now suppose $g(x)\to-\infty$ as $x\to a$. We can choose a point $c_1\in(a,y)$ s.t. $g(x)<g(y)$ and $g(x)<0$ if $a<x<c_1$. Thus $\frac{f(x)-g(y)}{g(x)}>0$ and the following steps are the same as in this book.*

# Chapter 6

# The Riemann-Stieltjes Integral

The definition of Riemann Integral is based on partition $P$ of interval $[a, b]$. When upper and lower Riemann integral meets, we say the Riemann integral exists. There is an important criterion of existence of Riemann integral, but we postpone it until we talk about Lebesgue integral.

Given a monotonically increasing function $\alpha(x)$, if we use the increment value $\alpha(x_i) - \alpha(x_{i+1})$ between two points $x_i$ and $x_{i+1}$ instead of the difference $x_{i+1} - x_i$, we can define Riemann-Stieltjes integral similarly.

Now we give a criterion of existence of Riemann integral.

**Theorem 6.0.1.** *$f \in \mathscr{R}(\alpha)$ on $[a, b]$ if and only if for every $\epsilon > 0$ there is a partition $P$ such that*

$$U(P, f, \alpha) - L(P, f, \alpha) < \epsilon$$

By theorem 6.0.1, we can show the following functions are R-S integrable.

1. $f$ is continuous on $[a, b]$
2. $f$ is monotonic on $[a, b]$ and $\alpha$ is continuous on $[a, b]$
3. $f$ is bounded on $[a, b]$, f has only finitely many points of discontinuity on $[a, b]$ and $\alpha$ is continuous at every point at which f is discontinuous.
4. $f = \phi(g)$ where $\phi$ is continuous and $g$ is R-S integrable.

Just like continuity and differentiation, there are some arithmetic properties of R-S integral, like addition, scalar multiplication and multiplication, and there are some basic inequalities involve R-S integration (theorem 6.12 and theorem 6.13). But we do not mention them here.

Stieltjes process is somewhat more flexibility than original Riemann integral: If $\alpha(x)$ is a pure step function, then the integral reduces to a finite or infinite series. If $\alpha(x)$ has an integrable derivative, i.e., $\alpha'(x) \in \mathscr{R}$, then the R-S integral reduces to an ordinary Riemann integral.

The final important theorem for R-S integral is change of variable:

**Theorem 6.0.2.** *Suppose $\phi$ is a strictly increasing continuous function that maps an interval $[A, B]$ onto $[a, b]$. Suppose $\alpha$ is monotonically increasing on $[a, b]$ and $f \in \mathscr{R}(\alpha)$ on $[a, b]$. Define $\beta(y) = \alpha(\phi(y))$ and $g(y) = f(\phi(y))$. Then $g \in \mathscr{R}(\beta)$ and*

$$\int_A^B g d\beta = \int_a^b f d\alpha$$

Take $\alpha(x) = x$ and assume $\phi' \in \mathscr{R}$. Then

$$\int_a^b f(x)dx = \int_A^B f(\phi(y))\phi'(y)dy$$

Now we talk about connection between integration and differentiation. Let

$$F(x) = \int_a^x f(t)dt$$

Then $F(x)$ is continuous. Furthermore, $F(x)$ is differentiable at $x_0$ which is continuous point of $f(x)$. $F'(x_0) = f(x_0)$.

If we assume $F(x)$ is differentiable and $F'(x) = f(x)$ then the fundamental theorem of calculus and integration by parts hold.

Finally we talk about R-S integral of vector-valued functions and rectifiable curves. Many analogues of theorems for R-S integral holds for vector-valued functions. We call a continuous mapping $\gamma$ of an interval $[a, b]$ into $\mathbb{R}^n$ a curve in $\mathbb{R}^n$. By partition, we can define the length of $\gamma$:

$$\Lambda(\gamma) = \sup \Lambda(P, \gamma) = \sup \sum_{i=1}^{n} |\gamma(x_i) - \gamma(x_{i-1})|$$

If $\Lambda(\gamma) < \infty$, we say that $\gamma$ is rectifiable. For continuously differentiable curves $\gamma$. Then length of $\gamma$ can be given as Riemann integral:

$$\Lambda(\gamma) = \int_a^b |\gamma'(t)| \, dt$$

# Chapter 7

# Sequences and Series of Functions

The main problem this chapter focuses is when the limit processes can be interchanged with series, integral and differentiation. Another problem is what family of functions is dense in continuous function space on compact set.

We first define the limit function $f(x)$ of sequence $(f_n(x))$ as pointwise limit of $(f_n(x))$. We give some examples in book, which show the limit process can not be interchanged carelessly.

## 7.1 Uniform convergence

To avoid the interchange problem, we introduce a new concept, uniform convergence:

**Definition 7.1.1.** *A sequence of functions $(f_n(x))$ converges uniformly on $E$ to a function $f$ if for every $\epsilon > 0$ there is an integer $N$ such that $n > N$ implies:*

$$|f_n(x) - f(x)| \leq \epsilon$$

*for all $x \in E$.*

A very convenient test for uniform convergence is Weierstrass test. It says if $|f_n(x)| < M_n$ and $\sum M_n$ converges, then $\sum f_n(x)$ converges uniformly.

Under uniform convergence, we have following interchange statement:

1. If $f_n \to f$ uniformly, then $\lim_{t\to x} \lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} \lim_{t\to x} f_n(t)$.

2. If $(f_n(x))$ are continuous and $f_n \to f$ uniformly, then $f$ is continuous. The converse is true if $(f_n(x))$ are on compact set $K$ and $(f_n(x))$ is decreasing sequence.

3. If $(f_n(x))$ are R-S integrable and $f_n \to f$ uniformly, then $f$ is R-S integrable and $\int_a^b f d\alpha = \lim_{n\to\infty} \int_a^b f_n d\alpha$.

4. If $(f_n(x))$ are differentiable, $(f_n(x))$ converges for some point $x_0$ and $(f'_n(x))$ converges uniformly, then $f_n(x) \to f(x)$ uniformly and $f'(x) = \lim_{n\to\infty} f'_n(x)$.

By the second statement, we claim $C(K)$ is complete space with supremum norm: $\|f\| = \sup |f(x)|$.

We know for any bounded infinite sequence in $\mathbb{R}^n$, we can extract a convergent subsequence. This is by the relation between boundedness and compact for $\mathbb{R}^n$. We will talk about similar properties for functions, but first we need to specify what convergence we consider. We will consider pointwise convergence and uniform convergence.

By diagonal process, given a sequence of pointwise bounded functions $(f_n(x))$ on $E$, we can extract a subsequence which converges on a countable subset of $E$. But we can not extract a subsequence which converges on $E$ even if $(f_n(x))$ is uniformly bounded.

We want to find a type a convergence which helps to make a subset of $C(K)$ 'compact'.

## 7.2 Continuous function space

To solve the problem at the end of section 1, we need a more stronger continuity for family of functions, called equicontinuity.

**Definition 7.2.1.** *A family $\mathscr{F}$ of complex functions $f$ defined on a set $E$ in a metric space $X$ is said to be equicontinuous on $E$ if for every $\epsilon > 0$ there exists a $\delta > 0$ s.t.*

$$|f(x) - f(y)| < \epsilon$$

*whenever $d(x, y) < \delta$, $x \in E$, $y \in E$ and $f \in \mathscr{F}$.*

It is clear that every member of an equicontinuous family is uniformly continuous.

Here is a criterion for equicontinuity: If $(f_n(x))$ is a sequence of continuous function on a compact set $K$, and $f_n(x) \to f(x)$ uniformly, then $(f_n(x))$ is equicontinuous on $K$.

Now we give the 'relative compactness' of a family of functions.

**Theorem 7.2.2.** *If $K$ is compact, if $(f_n(x)) \subset C(K)$, and if $(f_n(x))$ is pointwise bounded and equicontinuous on $K$, then*

1. *$(f_n(x))$ is uniformly bounded on $K$.*
2. *$(f_n(x))$ contains a uniformly convergent subsequence ($(f_n(x))$ is 'relative compact').*

Actually, the converse is also true, and this result is Ascoli theorem.

The next topic of this section is the density subset of a continuous function space on compact space $K$. This result is Stone-Weierstrass theorem.

First we show the Weierstrass' result: for a function $f \in C(K)$, there is a sequence of polynomial $(P_n)$ that converges to $f$ uniformly. In other words, the space of polynomial functions on $K$ is dense in $C(K)$.

To generalize the Weierstrass' result, we need to isolate some properties of the polynomials.

**Definition 7.2.3** (algebra). *A family $\mathscr{A}$ of complex functions defined on a set $E$ is said to be an algebra if for $f, g \in \mathscr{A}$:*

1. $f + g \in \mathscr{A}$
2. $fg \in \mathscr{A}$
3. $cf \in \mathscr{A}$ *for all* $c \in \mathbb{C}$

**Definition 7.2.4** (seperate points). *Let $\mathscr{A}$ be a family of functions on a set $E$. Then $\mathscr{A}$ is said to separate points on $E$ if to every pair of distinct points $x_1, x_2 \in E$, there is a function $f \in \mathscr{A}$ such that $f(x_1) \neq f(x_2)$.*

**Definition 7.2.5** (vanishes at no points). *Let $\mathscr{A}$ be a family of functions on a set $E$. Then $\mathscr{A}$ vanishes at no points on $E$ if to every point $x \in E$, there is a function $f \in \mathscr{A}$ such that $f(x) \neq 0$.*

**Definition 7.2.6** (lattice). *Let $\mathscr{A}$ be a family of real functions on $\mathbb{R}$. Then $\mathscr{A}$ is called a lattice if to every pair of functions $f_1, f_2$, $\max(f_1, f_2)$ and $\min(f_1, f_2)$ is in $\mathscr{A}$.*

**Remark 7.2.1.** *A closed subalgebra of $C(K)$ is a lattice.*

Now we give the Stone's generalization of the Weierstrass theorem:

**Theorem 7.2.7** (Stone-Weierstrass theorem, real case). *Let $\mathscr{A}$ be an algebra of real continuous functions on a compact set $K$. If $\mathscr{A}$ separates points on $K$ and if $\mathscr{A}$ vanishes at no point of $K$, then $\mathscr{A}$ is dense in $C(K)$.*

This theorem does not hold for complex functions. We need an extra condition on $\mathscr{A}$ for complex case: $\mathscr{A}$ is self-adjoint. This means for every $f \in \mathscr{A}$, $\bar{f}$ is in $\mathscr{A}$.

**Theorem 7.2.8** (Stone-Weierstrass theorem, complex case). *Let $\mathscr{A}$ be a self-adjoint algebra of complex continuous functions on a compact set $K$. If $\mathscr{A}$ separates points on $K$ and if $\mathscr{A}$ vanishes at no point of $K$, then $\mathscr{A}$ is dense in $C(K)$.*

# Chapter 8

# Some Special Functions

## 8.1 Power series or analytic functions

The analytic functions is of the form:

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n$$

Although author restricts himself to real values of $x$, many conclusions still hold on complex values of $x$.

We look at some conclusions when $f(x) = \sum_{n=0}^{\infty} c_n x^n$ converges for $|x| < R$:

1. $\sum_{n=0}^{\infty} c_n x^n$ converges uniformly on $|x| < R - \epsilon$ (Theorem 8.1).
2. $f(x)$ is continuous and differentiable for $|x| < R$, and $f'(x) = \sum_{n=1}^{\infty} n c_n x^{n-1}$ for $|x| < R$ (Theorem 8.1).
3. If $f(x)$ converges at $x = R$, then $f(x)$ continuous at $x = R$ (Theorem 8.2). This can be used to give another proof of formula for product of two series.
4. $f(x)$ has another representation, $f(x) = \sum_{n=1}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$. And it converges in $|x-a| < R - |a|$ (Theorem 8.4). Note the new representation may converge in a larger interval. This gives a method for analytic continuation.

If two power series converges in the same region $\Omega$, and they coincide in some sequence of distinct points with limit point in $\Omega$, then they coincide in region $\Omega$ (Theorem 8.5). This theorem is particularly used to extend power series like exponential function from $\mathbb{R}$ to $\mathbb{C}$.

## 8.2 Exponential, logarithmic and trigonometric functions

We define

$$E(x) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

By showing $E(z+w) = E(z)E(w)$ and $E(1) = e$, we have $E(p) = e^p$ for all positive rational number. Then we define $E(x) = \sup e^p$ for $p < x$ with $p \in \mathbb{Q}$. Now we have $E(x)$ defined on all real number $x$. Using $E(z)E(-z) = E(0) = 1$, we have $E(x)$ defined on $\mathbb{R}$.

The logarithm function $L(y)$ on $\mathbb{R}^+$ is defined by

$$E(L(y)) = y$$

By differentiating and integrating above equation, we have integral formula for $L(y)$:

$$L(y) = \int_1^y \frac{dx}{x}$$

Using $x^n = E(nL(x))$ and the same procedure as in exponential function, we have $x^\alpha = e^{\alpha \log x}$ for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^+$.

The trigonometric functions are defined as:

$$C(x) = \frac{1}{2}(E(ix) + E(-ix))$$

and

$$S(x) = \frac{1}{2i}(E(ix) - E(-ix))$$

The interesting part in author's constructions is he define $\pi$ as the smallest number which make $C(\frac{\pi}{2}) = 0$. He also shows by this definition, $E(z + 2\pi i) = E(z)$ for complex $z$. And he proves $C(x)$ and $S(x)$ have period $2\pi$. To make his argument meaningful, he finally shows the length of unit circle is $2\pi$ with his definition of $\pi$, and $C(x)$ and $S(x)$ are identical with usual definition of $\cos x$ and $\sin x$.

## 8.3 Fourier series

We define Fourier series of function $f$ as:

$$\sum_{n=-\infty}^{\infty} c_n e^{inx}$$

where $c_n$ called Fourier coefficients of $f$:

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx$$

Notice $(e^{inx})$ is an orthonormal system of functions on $[-\pi, \pi]$.

In general, given any orthonormal system of functions $(\phi_n(x))$ on $[a, b]$, consider two partial sums $s_n = \sum_{m=1}^n b_m \phi_m(x)$ and $t_n = \sum_{m=1}^n \gamma_m \phi_m(x)$ where

$$b_m = \int_a^b f(x)\phi_m(x)dx$$

We have

$$\int_a^b |f - s_n|^2 \, dx \le \int_a^b |f - t_n|^2 \, dx$$

and equality holds if and only if $\gamma_m = c_m$.

From above inequality, we have Bessel inequality for all orthonormal $(\phi_n)$:

$$\sum_{n=1}^{\infty} |b_n|^2 \le \int_a^b |f(x)|^2 \, dx$$

and Riemann-Lebesgue lemma: $\lim c_n = 0$.

Now we come back to the topic on Fourier series. First is pointwise convergence. By Riemann-Lebesgue lemma, partial sum of Fourier series converges to $f(x)$ at point $x$, where $|f(x+t) - f(x)| \le M\,|t|$ for all small enough $t$. The second is partial sum of Fourier series of $f$ converges uniformly to continuous $f$. Since $(e^{inx})$ is orthonormal, this is easy by Stone-Weierstrass theorem and Bessel inequality. Finally by uniform convergence of partial sum of Fourier series of $f$ and $C([-\pi, \pi])$ being dense in $\mathscr{R}$, we have Parseval's theorem:

**Theorem 8.3.1** (Parseval's theorem)**.** *Suppose $f$ and $g$ are Riemann-integrable functions with period $2\pi$, and*

$$f(x) \sim \sum_{-\infty}^{\infty} c_n e^{inx} \quad g(x) \sim \sum_{-\infty}^{\infty} \gamma_n e^{inx}$$

*Then partial sum $\sum_{-n}^{n} c_n e^{inx}$ converges to $f$ in 2-norm. And*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx = \sum_{-\infty}^{\infty} c_n \bar{\gamma}_n$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 \, dx = \sum_{-\infty}^{\infty} |c_n|^2$$

## 8.4 The gamma function

We define gamma function as:

$$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt \quad (0 < x < \infty)$$

Gamma function $\Gamma(x)$ has following properties and these properties characterize $\Gamma(x)$ completely:

1. $\Gamma(x)$ is a positive function on $(0, \infty)$.
2. $\Gamma(x+1) = x\Gamma(x)$.
3. $\Gamma(1) = 1$.
4. $\log \Gamma$ is convex.

This theorem is called Bohr-Mollerup theorem. From the proof of this theorem we have the relation:

$$\Gamma(x) = \lim_{n\to\infty} \frac{n!n^x}{x(x+1)\cdots(x+n)}$$

There is also a simple approximate expression called Stirling's formula for $\Gamma(x+1)$ when $x$ is large:

$$\lim_{x\to\infty} \frac{\Gamma(x+1)}{(\frac{x}{e})^x\sqrt{2\pi x}} = 1$$

Finally we introduce beta function which is related to gamma function:

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

## 8.5 Details of proof

A function $f$ called convex function if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. We first prove that for $x < y < z$. Then:

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

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

*Proof.* Let $y = \lambda x + (1-\lambda)z$. Then $f(y) \leq \lambda f(x) + (1-\lambda)f(z)$. Now we have:

$$\frac{f(y)-f(x)}{y-x} \leq \frac{\lambda f(x) + (1-\lambda)f(z) - f(x)}{\lambda x + (1-\lambda)z - x} \leq \frac{f(z)-f(x)}{z-x}$$

and

$$\frac{f(z)-f(y)}{z-y} \geq \frac{f(z) - (\lambda f(x) + (1-\lambda)f(z))}{z - (\lambda x + (1-\lambda)z)} \geq \frac{f(z)-f(x)}{z-x}$$

□

By above inequality, $\log n \leq \frac{\phi(n+1+x)-\phi(n+1)}{x} \leq \log(n+1)$ holds.

**Note 7** (Proof of Theorem 8.20)**.** *For fixed y:*

$$\begin{aligned} B(\frac{x_1}{p} + \frac{x_2}{q}, y) &= \int_0^1 t^{\frac{x_1-1}{p}+\frac{x_2-1}{q}}(1-t)^{y-1}dt \\ &= \int_0^1 t^{\frac{x_1-1}{p}}(1-t)^{\frac{y-1}{p}} t^{\frac{x_2-1}{q}}(1-t)^{\frac{y-1}{q}} dt \end{aligned}$$

*Using Holder's inequality, we show* $\log B(x,y)$ *is convex for fixed* $y$*.*

# Chapter 9

# Functions of Several Variables

## 9.1 Linear operator on finite linear space

First the author gives some definitions and properties of Euclidean space $\mathbb{R}^n$. These are just reviews of linear algebra so we do not repeat here. Then we talk about linear operator in $L(\mathbb{R}^n, \mathbb{R}^m)$. We define the norm of $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ as:

$$\|A\| = \sup_{|x| \leq 1} |Ax|$$

By norm of linear operator, we give a important character of invertible linear operator in $\mathbb{R}^n$.

**Theorem 9.1.1.** *Let $\Omega$ be the set of all invertible linear operators on $\mathbb{R}^n$.*

1. *If $A \in \Omega$, $B \in L(\mathbb{R}^n)$, and $\|B - A\|\|A^{-1}\| < 1$, then $B \in \Omega$.*
2. *$\Omega$ is open and the mapping $A \mapsto A^{-1}$ is a homeomorphism on $\Omega$.*

This theorem can be easily extend to the space bounded linear operator on a Banach space $B(E)$:

**Theorem 9.1.2.** *Given a Banach space $E$, Let $B(E)$ be the space bounded linear operator on $E$ and the $GL(E)$ be the set of all invertible linear operators on $E$.*

1. *If $A \in GL(E)$, $B \in B(E)$, and $\|B - A\|\|A^{-1}\| < 1$, then $B \in GL(E)$.*
2. *$GL(E)$ is open and the mapping $A \mapsto A^{-1}$ is a homeomorphism on $GL(E)$.*

There are two varieties of the first statement:

1. Given $A \in GL(E)$, $B \in B(E)$, if $\|A^{-1}B - I\| < 1$ then $B$ has left inverse.

2. Given $A \in GL(E)$, if $\|A\| < 1$, then $I - A \in GL(E)$.

To prove the above two statements, we use power series.

Using Schwarz inequality we can show:

$$\|A\| \le (\sum_{i,j} a_{i,j}^2)^{\frac{1}{2}}.$$

Replace $A$ by $B-A$ and $a_{i,j}$ by continuous functions $a_{i,j}(p)$ we can see mapping $p \mapsto A_p$ is a continuous mapping.

## 9.2 Differentiation

Let $\mathbf{f}$ be a mapping from an open set $E$ in $\mathbb{R}^n$ into $\mathbb{R}^m$. First we talk about the differential of $\mathbf{f}$ or total derivative of $\mathbf{f}$ at $x$. We write $\mathbf{f}'(x) = A$ where $A$ is a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$ which satisfies:

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

An useful equivalent definition of $\mathbf{f}'(x)$ is

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\mathbf{h} + \mathbf{r}(\mathbf{h})$$

where $\mathbf{r}(\mathbf{h})$ satisfies $\lim_{\mathbf{h}\to 0} \frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|} = 0$. Or $|\mathbf{r}(\mathbf{h})| = \epsilon(\mathbf{h})\,|\mathbf{h}|$ where $\epsilon(\mathbf{h}) \to 0$ as $\mathbf{h} \to 0$. In other words, *continuously differentiable transformation behave locally very much like their derivatives.*

For fixed $x$, $\mathbf{f}'(x)$ is a linear transformation in $L(\mathbb{R}^n, \mathbb{R}^m)$. $\mathbf{f}'$ is a function from $E$ into $L(\mathbb{R}^n, \mathbb{R}^m)$.

If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $x \in \mathbb{R}^n$ and we define $A(x) = Ax$, then $A'(x) = A$.

Another type of derivative for $\mathbf{f}(x) = (f_1(x), f_2(x), \ldots, f_m(x))$ is partial derivative:

$$\frac{\partial f_i}{\partial x_j}(x) = \lim_{t\to 0} \frac{f_i(x + te_j) - f_i(x)}{t}$$

Notice even if partial derivatives exist at a $x$, $\mathbf{f}(x)$ may not continuous at $x$. Even if $\mathbf{f}(x)$ is continuous at $x$ and partial derivatives exist at $x$, $\mathbf{f}'(x)$ may not exist. However, if $\mathbf{f}'(x)$ exists, then partial derivatives exist at $x$ and:

$$\mathbf{f}'(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \ldots & \frac{\partial f_1}{\partial x_n}(x) \\ \ldots & \ldots & \ldots \\ \frac{\partial f_m}{\partial x_1}(x) & \ldots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$$

Given a mapping $f$ from an open set $E$ in $\mathbb{R}^n$ into $\mathbb{R}$. There is also a type of derivative called directional derivative, defined as:

$$\lim_{t\to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{t} = (\nabla f)(\mathbf{x}) \cdot \mathbf{u}$$

Given differentiable $\mathbf{f}$ which maps a *convex* open set $E$ in $\mathbb{R}^n$ in to $\mathbb{R}^m$, there is a mean value inequality:

$$|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \le \|\mathbf{f}'\|\,|\mathbf{b} - \mathbf{a}|$$

## 9.3 The inverse function theorem and the implicit function theorem

Before we talk about the inverse function theorem and the implicit function theorem, we give a fixed point theorem which is useful in proof of the inverse function theorem:

**Theorem 9.3.1.** *If $X$ is complete metric space with metric $d$ and $\phi$ satisfies $d(\phi(x), \phi(y)) \leq cd(x, y)$ where $c < 1$, then there exists one and only one $x$ such that $\phi(x) = x$.*

Now we give the inverse function theorem:

**Theorem 9.3.2.** *Suppose $\mathbf{f}$ is a $C^1$ mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^n$ and $\mathbf{f}'(a)$ is invertible at some point $a$.. Then there exists an open set $U$ and $V$ s.t. $a \in U$ $f(a) \in V$, and $\mathbf{f}$ is a homeomorphism from $U$ to $V$. Also, the inverse of $\mathbf{f}$ is also in $C^1$.*

The key idea in this proof is constructing function $\phi(x) = x + A^{-1}(y - f(x))$ and showing $\phi(x)$ has a fixed point in an open set $U$. One interesting corollary of above theorem is that if $\mathbf{f}'(x)$ is invertible at $x$ in open set $E$, then $\mathbf{f}$ is an open mapping of $E$ to $\mathbb{R}^n$.

Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Given $A \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$, we define:

$$A_x h = A(h, 0), \quad A_y k = A(0, k), \quad A(h, k) = A_x h + A_y k$$

where $h \in \mathbb{R}^n$ and $k \in \mathbb{R}^m$.

The implicit function theorem is:

**Theorem 9.3.3.** *Let $\mathbf{f}$ be a $C^1$ mapping of an open set $E \subset \mathbb{R}^{n+m}$ into $\mathbb{R}^n$, s.t. $\mathbf{f}(a, b) = 0$ for some point $(a, b) \in E$. Let $A = \mathbf{f}'(a, b)$ and assume $A_x$ is invertible. Then:*

*There exists open sets $U \subset \mathbb{R}^{n+m}$ and $W \subset \mathbb{R}^m$, with $(a, b) \in U$ and $b \in W$, having the following property:*

1. *To every $y \in W$ there is a unique $x$ such that $(x, y) \in U$ and $\mathbf{f}(x, y) = 0$.*
2. *If this $x$ is defined to be $g(y)$, then $g$ is a $C^1$ mapping of $W$ into $\mathbb{R}^n$, $g(b) = a$ and $f(g(y), y) = 0$ for $y \in W$.*
3. *$g'(b) = -(A_x)^{-1} A_y$.*

The first statement claims that for $y \in W$, there is a 'curve' satisfies $f(x, y) = 0$ for $(x, y) \in U$. The second and third statements claim we can have the 'explicit' representation for 'curve' $x = g(y)$ and this curve has derivative at point $b$.

To prove this theorem, we define $F(x, y) = (f(x, y), y)$. Once we show $F(x, y)$ is 1-1 mapping of $E$ into $\mathbb{R}^{n+m}$, we consider curve defined the preimage of the 'vertical' line $(0, y)$. This gives the $x = g(y)$ corresponding to $y$ that makes $f(x, y) = 0$. Since $F(x, y) = F(g(y), y)$ is 1-1 mapping and $F'(a, b)$ is invertible, we consider the function $G$ which inverts $F$. By inverse function theorem, $G(f(x, y), y) = (x, y)$ is $C^1$ function, which is obvious in $C^1$ on the 'curve' $f(x, y) = 0$. That means $G(0, y) = (g(y), y)$ is in $C^1$ and the first component is in $C^1$ consequently.

## 9.4 The rank theorem

The inverse function theorem and the implicit function theorem are only concerned with function $f$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ and $f'(x)$ is invertible at some point $a$. A more general theorem is concerned with function $f$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $f'(x)$ has only rank r.

**Theorem 9.4.1.** *Suppose $m$, $n$, $r$ are nonnegative integers, $m \geq r$, $n \geq r$, $F$ is a $C^1$ mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, and $F'(x)$ has rank $r$ for every $x \in E$.*

*Fixed $a \in E$, put $A = F'(a)$, let $Y_1$ be subspace $A(\mathbb{R}^n)$ and $P$ is projection in $\mathbb{R}^m$ whose range is $Y_1$. Let $Y_2$ be the null space of $P$.*

*Then there are open sets $U$ and $V$ in $\mathbb{R}^n$, with $a \in U$, $U \subset E$, and there is a 1-1 $C^1$ mapping $H$ of $V$ onto $U$ (whose inverse is also in $C^1$) such that*

$$F(H(x)) = Ax + \phi(Ax) \quad (x \in V)$$

*where $\phi$ is a $C^1$ mapping of the open set $A(V) \subset Y_1$ into $Y_2$.*

This theorem illustrates again the fact that the local behavior of a continuously differentiable mapping $F$ near a point $x$ is similar to that of the linear transformation $F'(x)$. Consider the following example:

Given a $C^1$ mapping $F$ that maps $E \subset \mathbb{R}^n$ into $\mathbb{R}^n$ and $F'(x)$ is invertible for all $x \in E$. choose $a \in E$ s.t. there is an open ball $B(a, h)$ in $E$, we have:

$$F(a + h) - F(a) = F'(a)h + r(h)$$

Now consider the mapping $G$ inverts $F$ and an inverse linear transformation $T$ , then:

$$F(G(T(a + h))) - F(G(T(a))) = Th = F'(a)h + r(h)$$

Let $h \to 0$, we see $T \to F'(a)$. The mapping $H$ in this example can be considered as $G(F'(a))$.

So what does the theorem told us? If the range of $F'(x)$ is $\mathbb{R}^m$, we can first map $x$ to $H(x)$, then $F \circ H$ can be described as linear transform $F'(a)$. $H$ changes as your chosen $a$ changes. In other words, after a homeomorphism mapping $x \to H^{-1}(x)$, any function $F$ with $F'(x)$ being surjective can be considered as a linear mapping locally in $H^{-1}(U)$.

Things changes if $F'(a)$ is not surjective. $F'(x)$ can not fill the space $\mathbb{R}^m$, so we fill the rest part by the function $\phi$.

## 9.5 Derivatives of higher order and differentiation of Integrals

We actually talk about higher order partial derivatives, we have mean value theorem on second partial derivative:

$$f(a+h,b+k)-f(a+h,b)-f(a,b+k)+f(a,b)=hk\frac{\partial^2 f}{\partial x_1\partial x_2}(x,y)$$

where $(x,y)$ is in the interior of rectangle having $(a,b)$ and $(a+h,b+k)$ as opposite vertices.

The next topic is the problem when the equation

$$\frac{d}{dt}\int_a^b \phi(x,t)d\alpha(x)=\int_a^b \frac{\partial\phi}{\partial t}(x,t)d\alpha(x)$$

is true. The sufficient condition is when $\phi(x,t)$ is in $\mathscr{R}(\alpha)$ for all $t$ and $\frac{\partial\phi}{\partial t}$ is continuous w.r.t $t$.

For Lebesgue integral, we pass the limit in integral sign by dominated convergence theorem.

# Chapter 10

# Integration of Differential Forms

This chapter is the hardest part of this book. First we introduce the integration on n-cell:

$$\int_{\mathbb{I}^n} f(x)dx$$

This integration is assigned by integrating in each dimension one by one.

Another tools in integration on higher dimension is change of variables:

$$\int_{\mathbb{R}^n} f(y)dy = \int_{\mathbb{R}^n} f(T(x)) \left| J_T(x) \right| dx$$

where $T$ is a 1-1 $C^1$ mapping and $J_T(x) \neq 0$ on support of $f(y)$.

## 10.1 Differential forms

There are lots of definitions in topic of differential forms. First we introduce the k-surface:

**Definition 10.1.1.** *Suppose $E$ is an open set in $\mathbb{R}^n$. A k-surface in $E$ is a $C^1$ mapping $\Phi$ from a compact set $D \subset \mathbb{R}^k$ in to $E$.*

The 'k' indicates the dimension of parameter of $\Phi$. Roughly speaking, the theory of differential form is the study of integration of surface. The definition of differential form makes this idea concrete:

**Definition 10.1.2.** *Suppose $E$ is an open set in $\mathbb{R}^n$. A differential form of order $k \geq 1$ in $E$ is a function $\omega$, symbolically represented by the sum:*

$$\omega = \sum a_{i_1,\ldots,i_k}(x) dx_{i_1} \wedge \cdots dx_{i_k}$$

*which assigns to each k-surface $\Phi$ in $E$ a number $\omega(\Phi) = \int_\Phi \omega$ according to the rule:*

$$\int_\Phi \omega = \int_D \sum a_{i_1,\dots,i_k}(\Phi(u)) \frac{\partial(x_{i_1},\dots,x_{i_1})}{\partial(u_1,\dots,u_k)} du$$

*where $D$ is the parameter domain of $\Phi$ and $\frac{\partial(x_{i_1},\dots,x_{i_1})}{\partial(u_1,\dots,u_k)}$ is the Jacobian of mapping $(u_1,\dots,u_k) \mapsto (\phi_{i_1}(u),\dots \phi_{i_k}(u))$. We say $\omega = 0$ if $\int_\Phi \omega$ for every k-surface $\Phi$.*

There are some elementary properties of k-form but we will not mention them here. We only talk about two operations of k-form, the multiplication and differentiation.

Suppose $\omega = \sum_I b_I(x)dx_I$ and $\lambda = \sum_J c_J(x)dx_J$ are p-form and q-form. The product of $\omega$ and $\lambda$ is defined to be:

$$\omega \wedge \lambda = \sum_{I,J} b_I(x)c_J(x)dx_I \wedge dx_J$$

Notice the product of 0-form with p-form can be interchanged:

$$f\omega = \omega f = \sum_I f(x)b_I(x)dx_I$$

The differentiation of k-form $\omega$ is assigning a (k+1)-form $d\omega$ on $\omega$. When $\omega$ is a 0-form, in other words, $\omega$ is a real function $f$, we define:

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x)dx_i$$

When $\omega = \sum b_I(x)dx_I$ is a k-form, we define:

$$d\omega = \sum_I (db_I) \wedge dx_I = \sum_I (\sum_{i=1}^n \frac{\partial b_I}{\partial x_i}(x)dx_i) \wedge dx_I = \sum_I \sum_{i=1}^n \frac{\partial b_I}{\partial x_i}(x)dx_i \wedge dx_I$$

This two operations on differential form are a little different from differentiation on functions:

1. If $\omega$ and $\lambda$ are k-form and p-form. Then $d(\omega\wedge\lambda) = (d\omega)\wedge\lambda + (-1)^k \omega \wedge d\lambda$.
2. $d^2\omega = 0$.

## 10.2 Change of variables in differential form

Suppose $E$ is an open set in $\mathbb{R}^n$, $T = (t_1(x), t_2(x), \dots, t_m(x))$ is a $C^1$ mapping of $E$ into a open set $V \subset \mathbb{R}^m$, and $\omega = \sum_I b_I(y)dy_I$ is a k-form in $V$. We define $\omega_T$ as:

$$\omega_T = \sum_I b_I(T(x))dt_{i_1} \wedge \dots \wedge dt_{ik}$$

Given a k-surface $\Phi$ with parameter domain $D$, we define a k-surface $\Delta$ in $\mathbb{R}^k$ by $\Delta(u) = u$. we have change of domain of integration:

$$\int_\Phi \omega = \int_\Delta \omega_\Phi$$

More generally, we can change of domain of integration between two k-surfaces under a $C^1$ mapping $T$:

$$\int_{T\Phi} \omega = \int_\Phi \omega_T$$

where $T$ maps k-surface $\Phi$ in $\mathbb{R}^n$ to k-surface $T\Phi$ in $\mathbb{R}^m$.

## 10.3 Simplexes and chains

We define oriented affine k-simplex

$$\sigma = [p_0, p_1, \dots p_k]$$

to be the k-surface wither parameter domain $Q^k$ which is given by the affine mapping:

$$\sigma(\alpha_1 e_1 + \cdots + \alpha_k e_k) = p_0 + \sum_{i=1}^{k} \alpha_i (p_i - p_0)$$

The $Q^k$ called the standard simplex, which is the set of all $u \in \mathbb{R}^k$ of the form $u = \sum_{i=1}^k \alpha_i e_i$ such that $\alpha_i \geq 0$ and $\sum a_i \leq 1$. When $k = 0$, the oriented 0-simplex is defined to be a point with a sign attached. We write $\sigma = +p_0$ or $\sigma = -p_0$. And integration of real function on this simplex $\sigma = \epsilon p_0$ is defined to be:

$$\int_\sigma f = \epsilon f(p_0)$$

An affine k-chain $\Gamma$ is a collection of finitely many oriented affine k-simplexes $\sigma_1, \dots \sigma_r$. And we define the integration of k-form $\omega$ on $\Gamma$ to be:

$$\int_\Gamma \omega = \sum_{i=1}^{r} \int_\sigma \omega$$

and symbolically we write:

$$\Gamma = \sigma_1 + \cdots + \sigma_r = \sum_{i=1}^{r} \sigma_i$$

We write $\Gamma = 0$ if $\int_\Gamma \omega = 0$ for all $\omega$.

For $k \geq 1$, given an affine k-simplex $\sigma = [p_0, p_1, \dots p_k]$, we define the boundary of $\sigma$ to be the affine (k-1)-chain:

$$\partial\sigma = \sum_{j=0}^{k} (-1)^j [p_0, \dots, p_{j-1}, p_{j+1}, \dots, p_k]$$

After the affine case, now we can consider the general case. Let $T$ be a $C^2$ mapping from $E \subset \mathbb{R}^n$ to $V \subset \mathbb{R}^m$. We consider the composite mapping $\Phi = T\sigma$, which is a k-surface in $V$. We call $\Phi$ an oriented k-simplex. The k-chain $\Psi = \sum \Phi_i$ is defined to be:

$$\int_{\Psi} \omega = \sum_{i=1}^{r} \int_{\Phi_i} \omega$$

And the boundary $\partial\Phi$ of k-simplex $\Phi$ is (k-1)-chain:

$$\partial\Phi = T(\partial\sigma)$$

And the boundary $\partial\Psi$ of k-chain $\Psi = \sum \Phi_i$ is (k-1)-chain:

$$\partial\Psi = \sum \partial\Phi_i$$

## 10.4 Stoke's theorem, closed and exact forms

The Stoke's theorem is:

**Theorem 10.4.1.** *If $\Psi$ is a k-chain and if $\omega$ is a (k-1)-form, then*

$$\int_{\Psi} d\omega = \int_{\partial\Psi} \omega$$

Given a k-form $\omega$ in an open set $E \subset \mathbb{R}^n$, if there is a (k-1)-form $\lambda$ s.t. $\omega = d\lambda$, then $\omega$ is said to be exact in $E$. If $d\omega = 0$, then $\omega$ is said to be closed. Since $d^2\omega = 0$, every exact form is closed. The converse is true if $E$ is convex open set (Poincare's lemma). The convex property guarantees us to use mean value inequality, more precisely, $f'(x) = 0$ on $E$ implies $f(x)$ is constant on $E$ if $E$ is convex. This statement is important in proof of Poincare's lemma.

Given a open set $E$, which satisfies all closed form on $E$ is exact form. We map $E$ onto $U$ by a 1-1 $C^2$ mapping $T$ which $T^{-1}$ is also in $C^2$. Then every closed k-form in $U$ is exact in $U$ (Theorem 10.40). We call $E$ and $U$ are $C^2$-equivalent.

Thus not only on convex open set closed forms are exact s, but also on any open set is $C^2$-equivalent to a convex set they are.

## 10.5 Vector analysis

In this section we talk some examples and theorems related to differential forms in $\mathbb{R}^3$.

Let $F = F_1e_1 + F_2e_2 + F_3e_3$ be a continuous mapping of an open set $E \subset \mathbb{R}^3$. We associated $F$ with following differential forms and functions:

1. $\lambda_F = F_1dx + F_2dy + F_3dz$

2. $\omega_F = F_1 dy \wedge dz + F_2 dz \wedge dx + F_3 dx \wedge dz$

3. $\nabla \cdot F = D_1 F_1 + D_2 F_2 + D_3 F_3$

4. $\nabla \times F = (D_2 F_3 - D_3 F_2)e_1 + (D_3 F_1 - D_1 F_3)e_2 + (D_1 F_2 - D_2 F_1)e_3$

And associated a real function $u$ with its gradient:

$$\nabla u = (D_1 u)e_1 + (D_2 u)e_2 + (D_3 u)e_3$$

By easy computation, the following equations:

$$F = \nabla u, \quad \nabla \times F = 0, \quad F = \nabla \times G, \quad \nabla \cdot F = 0$$

are equivalent to:

$$\lambda_F = du, \quad d\lambda_F = 0, \quad \omega_F = d\lambda_G, \quad d\omega_F = 0$$

By using the equations of differential form, we can prove:

1. If $F = \nabla u$, then $\nabla \times F = 0$

2. If $F = \nabla \times G$, then $\nabla \cdot F = 0$

Then converse is true if $E$ is $C^2$ equivalent to a convex set.

Now we consider the relation between integrations of above differential forms and functions.

Given a 1-surface (curve) $\gamma$ in an open set $E \subset \mathbb{R}^3$ with parameter interval $[0, 1]$, we define the unit vector $\mathbf{t}(u)$ to be $\gamma'(u) = |\gamma'(u)|\, \mathbf{t}(u)$ and the element of arc length $ds$ to be $|\gamma'(u)|\, du$. We have:

$$\int_\gamma \lambda_F = \int_0^1 F(\gamma(u)) \cdot \mathbf{t}(u))ds$$

And we denote the second integration as $\int_\gamma (F \cdot \mathbf{t})ds$.

Given a 2-surface $\Phi$ in an open set $E \subset \mathbb{R}^3$ with parameter domain $D$, we define

$$N(u, v) = \frac{\partial(y, z)}{\partial(u, v)}e_1 + \frac{\partial(z, x)}{\partial(u, v)}e_2 + \frac{\partial(x, y)}{\partial(u, v)}e_3$$

where $(x, y, z) = \Phi(u, v)$. And we define the unit vector $\mathbf{n}(u, v)$ to be $N(u, v) = |N(u, v)|\, \mathbf{n}(u, v)$ and the element of area $dA$ to be $|N(u, v)|\, dudv$. We have:

$$\int_\Phi \omega_F = \int_D F(\Phi(u, v)) \cdot \mathbf{n}(u, v))dA$$

And we denote the second integration as $\int_\Phi (F \cdot \mathbf{n})dA$.

By Stoke's formula, we have:

1. Stoke's formula: $\int_\Phi (\nabla \times F) \cdot \mathbf{n} dA = \int_{\partial\Phi} (F \cdot \mathbf{t})ds$

2. The divergence theorem: $\int_\Omega (\nabla \cdot F)dV = \int_{\partial\Omega} (F \cdot \mathbf{n})dA$

# Chapter 11

# The Lebesgue Theory

This chapter is about Lebesgue theory of measure and integration. Our final goal in this chapter is to introduce the Lebesgue integration. We first introduce the measurable set and the measure on it. Then measurable functions are defined on measurable set. Finally we combine measurable set and measurable function to introduce the Lebesgue integration.

## 11.1 measure on sets

A measure (or a set function) is defined on a family of subset of a given space. First we specify what family of subsets we care about.

We care about two type of family of subsets: the ring and the $\sigma$-ring. A family $\mathscr{R}$ of subsets is called ring if it is closed under union operation and set minus operation. A ring $\mathscr{R}$ is called a $\sigma$-ring if it is closed under countable union operation.

A set function $\phi$ on $\mathscr{R}$ is a map from $\mathscr{R}$ to $\overline{\mathbb{R}}$. We call $\phi$ is additive if $A \cap B = 0$ implies $\phi(A \cap B) = \phi(A) + \phi(B)$ and we call $\phi$ is countably additive (or $\sigma$-additive) if $A_j \cap A_i = 0$ implies $\phi(\cap_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \phi(A)$.

Countably additive set function $\phi$ on a ring $\mathscr{R}$ is continuous in following sense:

1. $A_i$ is a sequence of increasing sets, $A_1 \subset A_2 \subset \cdots$, $A \in \mathscr{R}$ and $A = \cup_{i=1}^{\infty} A_i$. Then $\phi(A_n) \to \phi(A)$.
2. $A_i$ is a sequence of decreasing sets, $A_1 \supset A_2 \supset \cdots$, $A \in \mathscr{R}$ and $A = \cap_{i=1}^{\infty} A_i$. If there is a number n s.t. $\phi(A_n) < \infty$, then $\phi(A_n) \to \phi(A)$.

## 11.2 Lebesgue measure and extension theorem

We define the elementary set (or n-cube) $I$ in $\mathbb{R}^n$ as product of interval $[a_i, b_i]$, $[a_i, b_i)$, $(a_i, b_i]$ and $(a_i, b_i)$. Let $\mathscr{E}$ be the set containing all elementary sets and their finite disjoint union as its elements. $\mathscr{E}$ is a ring but not a $\sigma$-ring.

Now we study the measure $m$ on $\mathscr{E}$. We define the measure $m$ on $\mathscr{E}$ by $m(I) = \prod_{i=1}^{n}(b_i - a_i)$. This measure is regular, which means for any set $A \in \mathscr{E}$, there is an open set $G$ and closed set $F$ such that $F \subset A \subset G$ and

$$\phi(G) - \epsilon \leq \phi(A) \leq \phi(F) + \epsilon$$

or

$$\phi(G - A) \leq \epsilon \quad and \quad \phi(A - F) \leq \epsilon$$

We can extend the regular set function $m$ on $\mathscr{E}$ to a countably additive set function $m'$ on a $\sigma$-ring which contains $\mathscr{E}$. For extension we means $m'(A) = m(A)$ for all $A \in \mathscr{E}$.

To construct $m'$, first we define $m^*(E)$ for all $E \in \mathbb{R}^n$ to be:

$$m^*(E) = \inf \sum_{i=1}^{\infty} m(A_n)$$

where $A_n$ are open elementary sets in $\mathscr{E}$ and $E \subset \cup_{i=1}^{\infty} A_i$. $m^*$ has subadditivity, which means $m^*(E) \leq \sum_{i=1}^{\infty} m^*(E_i)$ with $E \subset \cup_{i=1}^{\infty} E_i$. But by regularity of $m$, we can show $m^*(A) = m(A)$ for all $A \in \mathscr{E}$.

Now we give the extension theorem of regular measure $m$ on ring $\mathscr{E}$. Let $\mathscr{R}_F$ be the set contains all elements in $\mathscr{E}$ and their limit. And let $\mathscr{R}$ be the set contains all elements in $\mathscr{R}_F$ and their countable union. Then $\mathscr{R}$ is a $\sigma$-ring and $m^*$ restrict on $\mathscr{R}$ is countably additive. We say a sequence of set $A_n$ converges to set $A$ if their Hausdorff distance $d(A_n, A) \to 0$.

A remark here is that the $\sigma$-ring $\mathscr{R}$ is not the smallest $\sigma$-ring contains all open set. The element in smallest $\sigma$-ring containing all open set is called Borel set, and this $\sigma$-ring is denoted as $\mathscr{B}$. The element in $\mathscr{R}$ is the union of a Borel set and a set of measure zero.

## 11.3 Measurable functions

A function defined on measurable space $X$ with values in $\overline{\mathbb{R}}$ is measurable if the set

$$\{x | f(x) > a\}$$

is measurable for every real number $a$.

Measurable functions are closed under following arithmetic operation:

1. If $f$ is measurable, then $|f|$ is measurable.
2. If $f_n$ is a sequence of measurable function, then $\sup f_n(x)$ and $\limsup f_n(x)$ is measurable.
3. If $f$ is measurable, then $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$ are measurable.
4. If $f$ and $g$ are measurable, and $F$ is continuous on $\mathbb{R}^2$, then $F(f(x), g(x))$ are measurable.
5. If $f$ and $g$ are measurable, then $f + g$ and $fg$ are measurable.

## 11.4 Lebesgue integration

In the following discussion, we use $\mu$ to represent the Lebesgue measure. First we define the Lebesgue integration for simple function $s(x) = \sum_{i=1}^{n} c_i \chi_{E_i}(x)$ to be:

$$\int_E s d\mu = \sum_{i=1}^{n} c_i \mu(E \cap E_i)$$

and the integration for non-negative measurable function $f$ to be

$$\int_E f d\mu = \sup \int_E s d\mu$$

where $0 \leq s \leq f$. For measurable function $f$, we define the integration of $f$ to be:

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

if at least one integration in right hide side is finite.

There are three important theorems in Lebesgue integration:

1. Monotone convergence theorem: Given an increasing sequence of non-negative measurable function $f_n$ and its pointwise limit $f$, we have

$$\lim \int_E f_n d\mu = \int_E f d\mu$$

2. Fatou's lemma: Given a sequence of non-negative measurable function $f_n$, we have

$$\int_E \liminf f_n d\mu \leq \liminf \int_E f_n d\mu$$

3. Dominated convergence theorem: Given a sequence of measurable function $f_n$, the pointwise limit of $f_n$ exists, $f_n(x) \to f(x)$. If $|f_n(x)| \leq g(x)$ for a measurable function $g$ with all $n$, we have

$$\lim \int_E f_n d\mu = \int_E f d\mu$$

Most properties of Riemann integration are also satisfied for Lebesgue integration. We do not repeat them here. But there is one special property satisfied by Lebesgue integration, which is we can partition the domain of integration for measurable set.

**Theorem 11.4.1.** *If $A = \cup A_n$ and $A_i \cap A_j = \emptyset$, then:*

$$\int_A f d\mu = \sum_{i=1}^{\infty} \int_{A_i} f d\mu$$

On bounded interval $[a, b]$ in $\mathbb{R}$, if $f$ is Riemann integrable, then $f$ is Lebesgue integrable, and the Lebesgue integration and Riemann integration coincides. Furthermore, for a bounded function on $[a, b]$, $f$ is Riemann integrable if and only if $f$ is discontinuous on a countable set.

## 11.5 $L^2$ function

we say a measurable function is in $L^2(X)$ if the integration $\int_X |f|^2 \, d\mu < \infty$ and we define the norm of $f$ to be $\|f\| = \left( \int_X |f|^2 \, d\mu \right)^{\frac{1}{2}}$.

$\{f_n\}$ is called a Cauchy sequence in $L^2(X)$ if there exists a number $N$ such that $\|f_n - f_m\| \leq \epsilon$ for all $m, n > N$. $L^2(X)$ is complete space, which means Cauchy sequence $f_n$ converges to a measurable function $f$ and $f \in L^2(X)$.

An orthonormal set $\{\phi_n\}$ is said to be complete if for $f \in L^2$, $\int_X f \bar{\phi}_n d\mu = 0$ implies $\|f\| = 0$. Given a complete orthonormal set, for any $L^2$ function $f$, we define

$$c_n = \int_X f \bar{\phi}_n d\mu$$

We have:

$$\int_X |f|^2 \, d\mu = \sum_{i=1}^{\infty} |c_n|^2$$

Conversely, given a sequence of number $c_n$ and $\sum |c_n|^2$ converges, the function $f$ defined to be:

$$\sum_{i=1}^{\infty} c_n \phi_n$$

is a $L^2$ function (Riesz-Fischer theorem).

Hence every complete orthonormal set induces a 1-1 correspondence between the functions $f \in L^2$ and the sequence $\{c_n\}$ for which $\sum_{i=1}^{\infty} |c_n|^2$ converges. Thus the function space $L^2$ is isometric to the sequence space $\ell^2$.