Real Analysis

(Rudin, *Principles of Mathematical Analysis*)

Notes by Yucheng Wang

2022.5

Contents

1 Real and Complex Number System

1.1 Ordered Sets

Definition 1.1 (Order, Ordered Set)**.** *An **order** on set S is a relation denoted by $<$, with the following two properties:*

1. *For all $x, y \in S$, one and only one of the statements $x < y, x = y, y < x$ is true.*
2. *For all $x, y, z \in S$, if $x < y, y < z$ then $x < z$.*

*$x \leq y$ indicates $x < y$ or $x = y$. Sometimes we write $y > x$ in place of $x < y$, write $y \geq x$ in place of $x \leq y$. With such an order, the set S is called an **ordered set**.*

Definition 1.2 (Bound)**.** *Suppose S is an ordered set, and $E \subset S$. If there exists a $\beta \in S$ such that $x \leq \beta$ for every $x \in E$, then we say E is **bounded above**, and call β an **upper bound** of E.*

*Analogously, if there exists a $\alpha \in S$ such that $x \geq \alpha$ for every $x \in E$, then we say E is **bounded below**, and call α an **lower bound** of E.*

Definition 1.3 (Supremum, Infimum)**.** *Suppose S is an ordered set, and $E \subset S$ is bounded above. Suppose there exists an $\alpha \in S$ with the following properties:*

1. *α is an upper bound of E;*
2. *If $\gamma < \alpha$, then γ is not an upper bound of E.*

*Then α is called the **least upper bound (supremum)** of E, and we write $\alpha = \sup E$.*

Analogously, suppose there exists an $\alpha \in S$ with the following properties:

1. *α is a lower bound of E;*
2. *If $\gamma > \alpha$, then γ is not a lower bound of E.*

*Then α is called the **greatest lower bound (infimum)** of E, and we write $\alpha = \inf E$.*

Note that the definition above *supposes* but not *guarantees* that there exists supremum or infimum. The example below shows that even a bounded subset may not have supremum of infimum.

Example 1.1. *Let $A = \{x \in \mathbb{Q} : x^2 < 2\}$, which is clearly bounded above. It could be shown that it has no supremum in $\mathbb{Q}$.*

Definition 1.4 (Least-Upper-Bound Property)**.** *An ordered set S is said to have the **least-upper-bound property** (is **Dedekind complete**) if for all non-empty subset $E \subset S$, $\sup E$ exists in S whenever it is bounded above.*

Theorem 1.1. *Suppose S is an Dedekind complete ordered set, $B \subset S$ is non-empty and bounded below. Let L be the set of all lower bounds of B, then $\alpha = \sup L$ exists in S, and $\alpha = \inf B$.*

Proof. First note that L is non-empty since B is bounded below. Since L consists of exactly those $y \in S$ which satisfy $y \leq x$ for every $x \in B$, we see that every $x \in B$ is an upper bound of L. Thus L is bounded above, and thus $\sup L$ exists since S is Dedekind complete. Let $\alpha = \sup L$. Now we only need to show that $\alpha = \inf B$.

If $\gamma < \alpha$, then γ is not an upper bound of L, hence $\gamma \notin B$. Then we have $\alpha \leq x$ for every $x \in B$, and thus $\alpha \in L$.

If $\alpha < \beta$, then $\beta \notin L$, since α is an upper bound of L. Hence, we have shown that α is a lower bound of B (since $\alpha \in L$), but β is not if $\beta > \alpha$. Therefore, $\alpha = \inf B$. □

1.2 Fields and Real Field

Definition 1.5 (Field)**.** *A **field** is a set F with two operations, called **addition** and **multiplication**, which satisfy the following 'field axioms':*

1. *Axioms for addition:*
   - (a) *If $x, y \in F$, then their sum $x + y \in F$;*
   - (b) *Communtative law: $x + y = y + x$ for all $x, y \in F$;*
   - (c) *Associative law: $(x + y) + z = x + (y + z)$ for all $x, y, z \in F$;*
   - (d) *F contains element 0 such that $0 + x = x$ for every $x \in F$;*
   - (e) *To every $x \in F$ corresponds an element $-x \in F$ such that $x + (-x) = 0$.*
2. *Axioms for multiplication:*

    (a) *If $x, y \in F$, then their product $xy \in F$;*
    (b) *Commutative law: $xy = yx$ for all $x, y \in F$;*
    (c) *Associative law: $(xy)z = x(yz)$ for all $x, y, z \in F$;*
    (d) *F contains an element $1 \neq 0$ such that $1x = x$ for every $x \in F$;*
    (e) *If $x \in F$ and $x \neq 0$, then there exists an element $1/x \in F$ such that $x \cdot (1/x) = 1$.*
3. *The distributive law: $x(y + z) = xy + xz$ for all $x, y, z \in F$.*

Proposition 1.1. *The axioms for addition imply the following statements:*
1. *Cancellation law: If $x + y = x + z$, then $y = z$;*
2. *Uniqueness of 0: If $x + y = x$, then $y = 0$;*
3. *Uniqueness of $-x$: If $x + y = 0$, then $y = -x$;*
4. *$-(-x) = x$.*

Proof.
1. If $x + y = x + z$, then axioms for addition give
$$y = 0 + y = (-x + x) + y = -x + (x + y) = -x + (x + z) = (-x + x) + z = 0 + z = z.$$
2. It follows from (1) by taking $z = 0$.
3. If follows from (1) by taking $z = -x$.
4. If follows from (3) by replacing x by $-x$.

□

Proposition 1.2. *The axioms for multiplication imply the following statements:*
1. *Cancellation law: If $x \neq 0$ and $xy = xz$, then $y = z$;*
2. *Uniqueness of 1: If $x \neq 0$ and $xy = x$, then $y = 1$;*
3. *Uniqueness of $1/x$: If $x \neq 0$ and $xy = 1$, then $y = 1/x$;*
4. *If $x \neq 0$, then $1/(1/x) = x$.*

Proof. Analogous to the proof of Proposition 1.1. □

Proposition 1.3. *The field axioms imply the following statements:*
1. *$0x = 0$;*
2. *If $x \neq 0, y \neq 0$, then $xy \neq 0$;*
3. *$(-x)y = -(xy) = x(-y)$;*
4. *$(-x)(-y) = xy$.*

Proof.
1. Note that $0x + 0x = (0 + 0)x = 0x$. Hence Proposition 1.1(2) implies that $0x = 0$.
2. Assume $x \neq 0, y \neq 0$ but $xy = 0$. Then (1) gives
$$1 = (1/y)(1/x)xy = (1/y)(1/x)0 = 0$$
where has a contradiction. Hence (2) holds.
3. The first equality comes from
$$(-x)y + xy = (-x + x)y = 0y = 0,$$
combined with Proposition 1.1(3). The second equality is proved in the same way.
4. By (3) and Proposition 1.1(4) we have
$$(-x)(-y) = -[x(-y)] = -[-(xy)] = xy.$$

□

Definition 1.6 (Ordered Field)**.** *An **ordered field** is a field F which is also an ordered set such that*
1. *$x + y < x + z$ if $x, y, z \in F$ and $y < z$;*
2. *$xy > 0$ if $x, y \in F$, and $x, y > 0$.*

*We call x is **positive** if $x > 0$, or **negative** if $x < 0$.*

For example, $\mathbb{Q}$ is an ordered field.

Proposition 1.4. *The following statements are true in every ordered field.*

1. *If $x > 0$, then $-x < 0$, and vice versa;*
2. *If $x > 0$ and $y < z$, then $xy < xz$;*
3. *If $x < 0$ and $y < z$, then $xy > xz$;*
4. *If $x \neq 0$ then $x^2 > 0$. In particular, $1 = 1^2 > 0$;*
5. *If $0 < x < y$, then $0 < 1/y < 1/x$.*

Proof.

1. If $x > 0$, then $0 = -x + x > -x + 0$, so that $-x < 0$. If $x < 0$, then $0 = -x + x < -x + 0$, so that $-x > 0$.
2. Since $z > y$, we have $z - y > y - y = 0$, hence $x(z - y) > 0$, and therefore
$$xz = x(z - y) + xy > 0 + xy = xy.$$
3. By (1)(2) and Proposition 1.3(3),
$$-[x(z - y)] = (-x)(z - y) > 0,$$
so that $x(z - y) < 0$, hence $xz < xy$.
4. If $x > 0$, by definition we have $x^2 > 0$; if $x < 0$, then $-x > 0$ and hence $(-x)^2 > 0$. By Proposition 1.3(4) we have $(-x)^2 = x^2$.
5. If $y > 0$ and $v \leq 0$, then $yv \leq 0$. But $y \cdot (1/y) = 1 > 0$. Hence, $1/y > 0$. Analogously, $1/x > 0$. By multiplying both sides of the inequality $x < y$ by $(1/x)(1/y)$, we obtain $1/y < 1/x$.

□

Theorem 1.2. *There exists an ordered field $\mathbb{R}$ which has the least-upper-bound property. Moreover, $\mathbb{R}$ contains $\mathbb{Q}$ as a subfield.*

Proof. Omitted. □

Theorem 1.3.

1. *Archimedean property: If $x, y \in \mathbb{R}$, and $x > 0$, then there is a positive integer n such that $nx > y$.*
2. *Density of $\mathbb{Q}$: If $x, y \in \mathbb{R}$, and $x < y$, then there exists a $p \in \mathbb{Q}$ such that $x < p < y$.*

Proof.

1. Let $A = \{nx : n \in \mathbb{N}^+\}$. If false, then y would be an upper bound of A. Then A has a least upper bound in $\mathbb{R}$, denote $\alpha = \sup A$. Since $x > 0$, $\alpha - x < \alpha$, and $\alpha - x$ is not an upper bound of A. Hence, $\alpha - x < mx$ for some $m \in \mathbb{N}^+$. But then $\alpha < (m + 1)x$, which contradicts the fact that α is an upper bound of A.
2. Since $x < y$, we have $y - x > 0$, and (1) guarantees a positive number n such that
$$n(y - x) > 1.$$
Apply (1) again and we obtain positive integers m_1 and m_2 such that $m_1 > nx, m_2 > -nx$. Then
$$-m_2 < nx < m_1.$$
Hence, there is an integer m (with $-m_2 \leq m \leq m_1$) such that
$$m - 1 \leq nx < m.$$
Combining these inequalities, we have
$$nx < m \leq nx + 1 < ny.$$
Since $n > 0$, it follows that
$$x < \frac{m}{n} < y$$

where $p = m/n \in \mathbb{Q}$.

$\square$

Theorem 1.4. *For every real number $x > 0$ and every integer $n > 0$, there is one and only one real number y such that $y^n = x$.*

Proof. Omitted. $\square$

Definition 1.7 (Extended Real Number System)**.** *The* ***extended real number system*** *consists of the real field $\mathbb{R}$ and two symbols $+\infty$ and $-\infty$. We preserve the original order in $\mathbb{R}$ and define $-\infty < x < +\infty$ for every $x \in \mathbb{R}$.*

The extended real number system is not a field, but it is customary to make the following conventions:

1. If x is real, then
$$x + \infty = +\infty, x - \infty = -\infty, \frac{x}{+\infty} = \frac{x}{-\infty} = 0.$$
2. If $x > 0$, then $x \cdot (+\infty) = +\infty, x \cdot (-\infty) = -\infty$;
3. If $x < 0$, then $x \cdot (+\infty) = -\infty, x \cdot (-\infty) = +\infty$.

1.3 Complex Field

Definition 1.8 (Complex Number)**.** *A* ***complex number*** *is an ordered pair (a, b) of real numbers. Here 'ordered' means that (a, b) and (b, a) are distinct if $a \neq b$.*

Let $x = (a, b), y = (c, d)$. We define
$$x + y = (a + c, b + d),$$
$$xy = (ac - bd, ad + bc).$$

Theorem 1.5. *These definitions of addition and multiplication turn the set of all complex numbers into a field, with $(0, 0)$ and $(1, 0)$ in the role of 0 and 1.*

Proof. Let $x = (a, b), y = (c, d), z = (e, f)$ be any complex numbers. We verify axioms one by one.

1. Axioms for addition:
   (a) Closure: trivial.
   (b) Communtative law: $x + y = (a + c, b + d) = (c + a, d + b) = y + x$.
   (c) Associative law: $(x + y) + z = (a + c, b + d) + (e, f) = (a + c + e, b + d + f) = (a, b) + (c + e, d + f) = x + (y + z)$.
   (d) Zero element: $x + 0 = (a, b) + (0, 0) = (a, b) = x$.
   (e) Inverse element: put $-x = (-a, -b)$. Then $x + (-x) = (0, 0) = 0$.
2. Axioms for multiplication:
   (a) Closure: trivial.
   (b) Communtative law: $xy = (ac - bd, ad + bc) = (ca - db, da + cb) = yx$.
   (c) Associative law:
$$\begin{aligned}(xy)z &= (ac - bd, ad + bc)(e, f)\\ &= (ace - bde - adf - bcf, acf - bdf + ade + bce)\\ &= (a, b)(ce - df, cf + de) = x(yz).\end{aligned}$$
   (d) Unit element: $1x = (1, 0)(a, b) = (a, b) = x$.
   (e) Inverse element: If $x \neq 0$ then $(a, b) \neq (0, 0)$, which means at least one of the real numbers a, b is different from 0. Hence, $a^2 + b^2 > 0$ by Proposition 1.4(4). Define
$$\frac{1}{x} = \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\right).$$
   Then
$$x \cdot \frac{1}{x} = (a, b)\left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\right) = (1, 0) = 1.$$

3. Distributive law:
$$\begin{aligned} x(y+z) &= (a,b)(c+e,d+f) \\ &= (ac+ae-bd-bf, ad+af+bc+be) \\ &= (ac-bd, ad+bc) + (ae-bf, af+be) \\ &= xy+xz. \end{aligned}$$

□

Theorem 1.6. *For any real numbers a and b, we have*

$$(a,0)+(b,0)=(a+b,0), (a,0)(b,0)=(ab,0).$$

Proof. Trivial. □

Definition 1.9. $i=(0,1)$.

Theorem 1.7. $i^2=-1$.

Proof. $i^2=(0,1)(0,1)=(-1,0)=-1$. □

Theorem 1.8. *If a,b are real, then $(a,b)=a+bi$.*

Proof. $a+bi=(a,0)+(b,0)(0,1)=(a,0)+(0,b)=(a,b)$. □

Definition 1.10 (Conjugate, Real Part, Imaginary Part)**.** *If a,b are real and $z=a+bi$, then the complex number $\bar{z}=a-bi$ is called the **conjugate** of z. The numbers a and b are the **real part** and the **imaginary part** of z, respectively. We write $a=Re(z), b=Im(z)$.*

Theorem 1.9. *If z and w are complex, then*
1. $\overline{z+w}=\bar{z}+\bar{w}$;
2. $\overline{zw}=\bar{z}\cdot\bar{w}$;
3. $z+\bar{z}=2Re(z), z-\bar{z}=2iIm(z)$;
4. $z\bar{z}$ *is real and positive (except when $z=0$).*

Proof. Trivial. □

Definition 1.11 (Absolute Value)**.** *If z is a complex number, its absolute value $|z|$ is the non-negative square root of $z\bar{z}$, that is, $|z|=(z\bar{z})^{1/2}$.*

Theorem 1.10. *Let z and w be complex numbers. Then*
1. $|z|>0$ *unless* $z=0$. $|0|=0$;
2. $|\bar{z}|=|z|$;
3. $|zw|=|z||w|$;
4. $|Re(z)|\le|z|$;
5. $|z+w|\le|z|+|w|$.

Proof.
1. Trivial.
2. Trivial.
3. Let $z=a+bi, w=c+di$ with a,b,c,d real. Then

$$|zw|^2=(ac-bd)^2+(ad+bc)^2=(a^2+b^2)(c^2+d^2)=|z|^2|w|^2$$

   which means $|zw|^2=(|z||w|)^2$. Then (3) follows from the uniqueness assertion of Theorem 1.4.
4. Trivial.
5. Note that $\bar{z}w$ is the conjugate of $z\bar{w}$, so that $\bar{z}w+z\bar{w}=2Re(z\bar{w})$. Hence

$$\begin{aligned} |z+w|^2 &= (z+w)(\bar{z}+\bar{w}) = z\bar{z}+z\bar{w}+\bar{z}w+\bar{z}\bar{w} \\ &= |z|^2+2Re(z\bar{w})+|w|^2 \\ &\le |z|^2+2|z\bar{w}|+|w|^2 \\ &= |z|^2+2|z||w|+|w|^2 = (|z|+|w|)^2. \end{aligned}$$

□

Theorem 1.11 (Cauchy-Schwarz Inequality)**.** *If $a_1, ..., a_n$ and $b_1, ..., b_n$ are complex numbers, then*

$$\left|\sum_{j=1}^{n} a_j \bar{b}_j\right|^2 \le \sum_{j=1}^{n} |a_j|^2 \sum_{j=1}^{n} |b_j|^2.$$

Proof. Let $A = \sum |a_j|^2, B = \sum |b_j|^2, C = \sum a_j \bar{b}_j$. If $B = 0$, then $b_1 = ... = b_n = 0$ and the proof is trivial. Assume $B > 0$. By Theorem 1.9 we have

$$\begin{aligned}\sum |Ba_j - Cb_j|^2 &= \sum (Ba_j - Cb_j)(B\bar{a}_j - \overline{Cb_j}) \\ &= B^2 \sum |a_j|^2 - B\bar{C} \sum a_j \bar{b}_j - BC \sum \bar{a}_j b_j + |C|^2 \sum |b_j|^2 \\ &= B^2 A - B|C|^2 = B(AB - |C|^2).\end{aligned}$$

Since each term in the first sum is non-negative, we have

$$B(AB - |C|^2) \ge 0.$$

Since $B > 0$, it follows that $AB - |C|^2 \ge 0$ and we complete the proof. □

1.4 Euclidean Spaces

Definition 1.12 (Eucliedan Space)**.** *For each positive integer k, let $\mathbb{R}^k$ be the set of all ordered k-tuples*

$$x = (x_1, x_2, ..., x_k),$$

where $x_1, ..., x_k$ are real numbers, called the ***coordinates*** *of x. The elements of $\mathbb{R}^k$ are called points or vectors (especially when $k > 1$). If $y = (y_1, ..., y_n)$ and α is a real number, put*

$$x + y = (x_1 + y_1, ..., x_k + y_k),$$

$$\alpha x = (\alpha x_1, ..., \alpha x_k)$$

so that $x + y, \alpha x \in \mathbb{R}^k$. This defines addition of vectors, and multiplication of a vector by a real number (scalar). These two operations satisfy the commutative, associative, and distributive laws, and make $\mathbb{R}^k$ into a ***vector space over the real field****. The zero element of $\mathbb{R}^k$ is the point 0, all of whose coordinates are 0.*

We also define the ***inner product*** *of x and y by*

$$x \cdot y = \sum_{i=1}^{k} x_i y_i,$$

and the norm of x by

$$|x| = (x \cdot x)^{1/2} = \left(\sum_{i=1}^{k} x_i^2\right)^{1/2}.$$

The structure now defined is called ***Euclidean k-space****.*

Theorem 1.12. *Suppose $x, y, z \in \mathbb{R}^k$, and α is real. Then*

1. $|x| \ge 0$;
2. $|x| = 0$ *if and only if* $x = 0$;
3. $|\alpha x| = |\alpha||x|$;
4. $|x \cdot y| \le |x||y|$;
5. $|x + y| \le |x| + |y|$;
6. $|x - z| \le |x - y| + |y - z|$.

Note that (1)(2)(6) allow us to regard $\mathbb{R}^k$ as a metric space.

Proof.

1. Obvious.
2. Obvious.
3. Obvious.
4. It follows from Cauchy-Schwarz inequality (Theorem 1.11).
5. By (4) we have
$$\begin{aligned}|x+y|^2 &= (x+y)\cdot(x+y)\\ &= x\cdot x + 2x\cdot y + y\cdot y\\ &\le |x|^2 + 2|x||y| + |y|^2\\ &= (|x|+|y|)^2.\end{aligned}$$
6. It follows from (5) by replacing x by $x-y$, and y by $y-z$.

□

2 Basic Topology

2.1 Metric Spaces

Definition 2.1 (Metric Spaces)**.** *A set X whose elements is called **points**, is said to be a **metric space** if with any two points p, q of X, there is associated a real number $d(p, q)$ called the **distance** from p to q, such that*

1. *$d(p, q) > 0$ if $p \neq q$; $d(p, p) = 0$;*
2. *$d(p, q) = d(q, p)$;*
3. *$d(p, q) \leq d(p, r) + d(r, q)$ for any $r \in X$.*

*Any function d with these three properties is called a **distance function** or **metric**.*

In Euclidean spaces $\mathbb{R}^k$ we use Euclidean distance defined by

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

Definition 2.2 (Segments, Intervals, k-Cells)**.** *By the segment (a, b) we mean the set of all real numbers x such that $a < x < b$; by the interval $[a, b]$ we mean the set of all real numbers x such that $a \leq x \leq b$. Generally, if $a_i < b_i$ for $i = 1, ..., k$, the set $\{x = (x_1, ..., x_k) \in \mathbb{R}^k : a_i \leq x_i \leq b_i, i = 1, ..., k\}$ is called a k-**cell**. Thus a 1-cell is an interval, 2-cell is a rectangle, etc.*

Definition 2.3 (Ball, Neighborhood)**.** *If $x \in X$ and $r > 0$, the **(open) ball** B with center at x and radius r is defined as*

$$B_r(x) = \{y \in X : d(x, y) < r\}.$$

*Typically we also refer a ball of x as a **neighborhood** of x.*

Definition 2.4 (Interior, Boundary, Open Set, Bounded Set)**.** *Let X be a metric space*

1. *A point $x \in X$ is an **interior point** of $E \subseteq X$ if there is a neighborhood $N \subseteq X$ of x such that $N \subseteq E$. The set of all interior points of E is denoted E°.*
2. *A point $x \in X$ is a **boundary point** of $E \subseteq X$ if every its neighborhood contains point(s) of both E and E^c. The set of all boundary points of E is denoted ∂E.*
3. *$E \subseteq X$ is **open** if every point of E is an interior point of E; that is, for every point of E it has a neighborhood $N \subseteq X$ such that $N \subseteq E$.*
4. *$E \subseteq X$ is **bounded** if there is a real number $R > 0$ and a point $x \in X$ such that $d(x, y) < R$ for all $y \in E$; that is, $E \subseteq B_R(x)$.*

Theorem 2.1. *Every neighborhood is an open set.*

Proof. Consider a neighborhood $E = B_r(p)$ and let q be any point of E. Then there is a positive number h such that $d(p, q) = r - h$. To show that $B_q(h) \subseteq E$, note that for all $s \in B_q(h)$, we have $d(p, s) \leq d(p, q) + d(q, s) < r - h + h = r$, and thus $s \in E$. Therefore, q is an interior point of E, and hence E is open. □

Theorem 2.2. *If $E \subseteq X$ is open, then there exists a family of open balls $\{B_\alpha\}$ such that $E = \cup_\alpha B_\alpha$.*

Proof. For any $x \in E$, since E is open, there exists $r > 0$ such that $B_r(x) \subseteq E$. Let the family of open balls be $\{B_x = B_r(x)\}_{x \in X}$. Clearly, $\cup_x B_x \subseteq E$ since each $B_x \subseteq E$, and also $E \subseteq \cup_x B_x$ since all $y \in E$ is contained in B_y. □

Proposition 2.1. *$E \subseteq X$ is open if and only if $E = E^\circ$.*

Proof. Clearly $E^\circ \subseteq E$. To show that $E \subseteq E^\circ$, if $x \in E$ and E is open, then x is an interior point of E, and hence $x \in E^\circ$. □

Definition 2.5 (Limit Point, Isolated Point)**.** *A point p is a **limit point** of the set E if every neighborhood of p contains a point $q \neq p$ such that $q \in E$. The set of all limit points of E is denoted E'.*

*If $p \in E$ but p is not a limit point of E ($p \notin E'$), then p is called an **isolated point** of E.*

Theorem 2.3. *If p is a limit point of a set E, then every neighborhood of p contains infinitely many points of E.*

Proof. Suppose not, then there is a neighborhood N of p which contains only a finite number of points of E. Let $q_1, ..., q_n$ be those distinct points of $N \cap E$, and put

$$r = \min_{1 \leq m \leq n} d(p, q_m).$$

Then we have $r > 0$, and the neighborhood $B_r(p)$ contains no point q of E (such that $q \neq p$), and hence p is not a limit point of E, which has a contradiction. $\square$

Corollary 2.1. *A finite point set has no limit points.*

Definition 2.6 (Closure)**.** *The **closure** of $E \subseteq X$, denoted by $\bar{E}$, is the set of all points $x \in X$ such that every its neighborhood intersects E; that is, for all $r > 0$ we have $B_r(x) \cap E \neq \emptyset$.*

Proposition 2.2. *For any $E \subseteq X$, we have $E' \subseteq \bar{E}, \partial E \subseteq \bar{E}$ and $E \subseteq \bar{E}$.*

Proof. It immediately follows from definition of limit points, boundary points and closure. $\square$

Proposition 2.3. *$\bar{E} = E \cup E' = E \cup \partial E$. Hence, we have $\partial E = \bar{E} \backslash E^\circ$.*

Proof. By proposition above we have $E \cup E' \subseteq \bar{E}$ and $E \cup \partial E \subseteq \bar{E}$. Suppose $x \in \bar{E}$ but $x \notin E$. Since $x \in \bar{E}$, every neighborhood of x contains a point of E, and it cannot be $x \notin E$, so that $x \in E'$. Since every neighborhood of x also contains $x \in E^c$, so that $x \in \partial E$. $\square$

Definition 2.7 (Closed Set)**.** *The set $E \subseteq X$ is **closed** if one the following equivalent conditions hold:*
1. *$E = \bar{E}$;*
2. *$E' \subseteq E$;*
3. *$\partial E \subseteq E$;*
4. *E^c is open.*

Proof. The proposition above has shown the equivalence among (1)(2)(3). Then we only need to show

- (1)⇒(4): For any $x \in E^c$, we have $x \notin E$ and thus $x \notin \bar{E}$. It means there exists a neighborhood N of x which doesn't intersect E, i.e. $N \subseteq E^c$. Therefore, E^c is open.
- (4)⇒(1): Clearly $E \subseteq \bar{E}$. To show $\bar{E} \subseteq E$, if $x \notin E$, then $x \in E^c$ which is open. So there exists $r > 0$ such that $B_r(x) \subseteq E^c$, that is, $B_r(x) \cap E = \emptyset$. Hence, $x \notin \bar{E}$ and thus $\bar{E} \subseteq E$.

$\square$

Example 2.1. *Consider the following subsets of $\mathbb{C}$ (or equivalently, $\mathbb{R}^2$):*
1. *$\{z : |z| < 1\}$ is open, not closed, bounded;*
2. *$\{z : |z| \leq 1\}$ is not open, closed, bounded;*
3. *a finite set is not open, closed, bounded;*
4. *the set of all integers is not open, closed, not bounded;*
5. *$\{z : z = 1/n, n \in \mathbb{N}^+\}$ is not open, not closed, bounded;*
6. *The whole set $\mathbb{C}$ is open, closed, not bounded;*
7. *The segment (a, b) is not open, not closed, bounded.*

Theorem 2.4. *Let $\{E_\alpha\}$ be a (finite or infinite) collection of sets. Then*

$$\left(\bigcup_\alpha E_\alpha\right)^c = \bigcap_\alpha E_\alpha^c.$$

Proof. Let A and B be the LHS and RHS. If $x \in A$, then $x \notin E_\alpha$ for any α, which means $x \in E_\alpha^c$ for every α, so that $x \in \cap_\alpha E_\alpha^c = B$. Thus $A \subseteq B$.

If $x \in B$, then $x \in E_\alpha^c$ for every α, which means $x \notin E_\alpha$ for any α, so that $x \notin \cup_\alpha E_\alpha$ and hence $x \in (\cup_\alpha E_\alpha)^c = A$. Thus $B \subseteq A$. Therefore, $A = B$. □

Theorem 2.5.
1. *For any collection $\{G_\alpha\}$ of open sets, $\cup_\alpha G_\alpha$ is open;*
2. *For any collection $\{F_\alpha\}$ of closed sets, $\cap_\alpha F_\alpha$ is closed;*
3. *For any finite collection $G_1, ..., G_n$ of open sets, $\cap_{i=1}^n G_i$ is open;*
4. *For any finite collection $F_1, ..., F_n$ of closed sets, $\cup_{i=1}^n F_i$ is closed.*

Proof.
1. If $x \in \cup_\alpha G_\alpha$, there are some α such that $x \in G_\alpha$. Since G_α is open, there exists a neighborhood N of x such that $N \subseteq G_\alpha$, and thus $N \subseteq \cup_\alpha G_\alpha$. Therefore, $\cup_\alpha G_\alpha$ is open.
2. It comes from (1) by replacing G_α with F_α^c, and using the conclusion of Theorem 2.4.
3. If $x \in \cap_{i=1}^n G_i$, then $x \in G_i$ for every i. Since G_i is open, there exists $r_i > 0$ such that $B_{r_i}(x) \subseteq G_i$. Let $r = \min_{i=1,...,n} r_i$ and $r > 0$, we have $B_r(x) \subseteq G_i$ for all i and hence $B_r(x) \subseteq \cap_{i=1}^n G_i$. Therefore, $\cap_{i=1}^n G_i$ is open.
4. It comes from (3) by replacing G_i with F_i^c and using the conclusion of Theorem 2.4.

□

Example 2.2. *The following examples show why finiteness in Theorem 2.5(3)(4) is necessary.*
1. *Suppose $G_n = (-1/n, 1/n)$ for $n = 1, 2, ...$ which are open subsets of $\mathbb{R}$. Then $\cap_{i=1}^n G_i = (-1/n, 1/n)$ is open for every n, while $\cap_{i=1}^\infty G_i = \{0\}$ is closed.*
2. *Suppose $F_n = [-1 + 1/n, 1 - 1/n]$ for $n = 1, 2, ...$ which are closed subsets of $\mathbb{R}$. Then $\cup_{i=1}^n F_i = [-1 + 1/n, 1 - 1/n]$ is closed for every n, while $\cap_{i=1}^\infty F_i = (-1, 1)$ is open.*

Theorem 2.6. *If X is a metric space and $E \subseteq X$, then*
1. *$\bar{E}$ is closed;*
2. *If $E \subseteq F$ and F is closed, then $\bar{E} \subseteq F$;*
3. *$\bar{E} = \cap_{E \subseteq F, F \text{ is closed}} F$; that is, $\bar{E}$ is the smallest closed set in X containing E.*

Proof.
1. If $x \notin \bar{E}$, then there exists a neighborhood N of x such that $N \cap E = \emptyset$. It means $\bar{E}^c$ is open and hence $\bar{E}$ is closed.
2. For any $x \in \bar{E}$ and any its neighborhood B, B contains point in E, and thus also contains point in F. Hence, $x \in \bar{F} = F$.
3. From (2) we have $\bar{E} \subseteq \cap F$. On the other hand, $\bar{E}$ itself is closed and $E \subseteq \bar{E}$, so that $\cap F \subseteq \bar{E}$.

□

Definition 2.8 (Relative Open/Closed)**.** *Suppose (X, d) is a metric space and $Y \subseteq X$. $E \subseteq Y$ is* ***open/closed relative to*** *Y if E is open/closed in the metric space (Y, d).*

Theorem 2.7. *Suppose $Y \subseteq X$. A subset E of Y is open relative to Y if and only if $E = Y \cap G$ for some open subset G of X.*

Proof. If E is open relative to Y, then for all $x \in E$, there exists $r > 0$ such that $d(x, y) < r, y \in Y$ implies $y \in E$, which means $[B_r(x)]_Y \equiv \{y \in Y : d(x, y) < r\} \subseteq E$. Let $G = \cup_{x \in E} B_r(x)$, then G is open in X and we have $G \cap Y = E$ since $[B_r(x)]_Y = B_r(x) \cap Y$ for all $x \in E$.

Suppose $E = Y \cap G$ where $G \subseteq X$ is open. Then for all $x \in E$, there exists $r > 0$ such that $B_r(x) \subseteq G$, so that $[B_r(x)]_Y = B_r(x) \cap Y \subseteq G \cap Y = E$. Therefore, E is open relative to Y. □

Corollary 2.2. *Suppose $Y \subseteq X$. A subset E of Y is closed relative to Y if and only if $E = Y \cap F$ for some closed subset F of X.*

Corollary 2.3. *If $E \subseteq Y \subseteq X$ is closed relative to Y, then E is closed relative to X if and only if $\bar{E} \subseteq Y$.*

Theorem 2.8. *Let E be a non-empty set of real numbers which is bounded above. Let $y = \sup E$. Then $y \in \bar{E}$. Hence, $y \in E$ if E is closed.*

Proof. (CONTINUE) □

2.2 Connected Sets

Definition 2.9 (Separate Sets, Connected Sets)**.** *Two subsets A, B of a metric space X are* ***separated*** *if both $A \cap \bar{B}$ and $\bar{A} \cap B$ are empty, that is, if no point of A lies in the closure of B, and no point of B lies in the closure of A.*

A set $E \subseteq X$ is ***connected*** *if E is not a union of two non-empty separeted sets.*

Proposition 2.4. *If $A \subseteq X$ is both closed and open, and $A \neq \emptyset, X$, then X is disconnected.*

Proof. It is because $X = A \cup A^c$, and $\bar{A} = A, \bar{A}^c = A^c$, and the fact that $A \cap A^c = \emptyset$. Note that A^c is non-empty since $A \neq X$. □

Lemma 2.1. *If A, B are separated, and $E \subseteq A \cup B$ is connected, then we have either $E \subseteq A$ or $E \subseteq B$ (but not both).*

Proof. Note that $E = (E \cap A) \cup (E \cap B)$, and $E \cap A$ and $E \cap B$ are separated. If both $E \cap A$ and $E \cap B$ are non-empty, then E is disconnected, where has a contradiction. □

Proposition 2.5. *If E, F are connected and $E \cap F \neq \emptyset$, then $E \cup F$ is connected.*

Proof. Suppose $E \cup F$ is disconnected, then we have $E \cup F = A \cup B$ where A, B are non-empty and separated. By lemma above, we have either $E \subseteq A$ or $E \subseteq B$, and either $F \subseteq A$ or $F \subseteq B$. If E, F are both subsets of A or B, then B or A will be empty and has a contradiction; if E and F belong to A and B respectively, then we have $E \cap F \subseteq A \cap B = \emptyset$ where also has a contradiction. □

Definition 2.10 (Connected Component)**.** *A* ***connected component*** *A of metric space X is a maximal connected subset of X such that (1) A is connected; (2) If $A \subseteq B \subseteq X$ and B is connected, then $A = B$.*

Proposition 2.6. *Connected components of X partition X; that is, they are pairwise disjoint, and their union is X.*

Proof. (CONTINUE) □

2.3 Compact Sets

Definition 2.11 (Open Cover)**.** *An* ***open cover*** *of a set E in a metric space X is a collection $\{G_\alpha\}$ of open subsets of X such that $E \subseteq \cup_\alpha G_\alpha$.*

Definition 2.12 (Compact Sets)**.** *A subset K of a metric space X is* ***compact*** *if every open cover of K contains a finite subcover. That is, for any open cover $\{G_\alpha\}$ of K, there are finitely many indices $\alpha_1, ..., \alpha_n$ such that $K \subseteq G_{\alpha_1} \cup ... \cup G_{\alpha_n}$.*

Theorem 2.9. *A compact set K in a metric space X is closed and bounded.*

Proof. To show that K is closed, we only need to show that K^c is open. If $x \notin K$, then for any $y \in K$, suppose $d(x, y) = r_y > 0$. By triangle inequality we have $B_{r_y/3}(x) \cap B_{r_y/3}(y) = \emptyset$. Then we can construct an open cover of K by $\{B_{r_y/3}(y)\}_{y \in X}$. Since K is compact, it has finite subcover $\{B_{r_{y_n}/3}(y_n)\}_{n=1}^N$. Let $r = \min_{1 \leq n \leq N} r_{y_n} > 0$, then we have $B_{r/3}(x) \subseteq K^c$. Therefore, K^c is open.

To show that K is bounded, choose any $x \in X$ and construct an open cover of K by $\{B_r(x)\}_{r>0}$. Since K is compact, it has a finite subcover $\{B_{r_n}(x)\}_{n=1}^N$. Let $R = \max_{1 \leq n \leq N} r_n$ and we have $K \subseteq B_R(x)$. □

Theorem 2.10. *Let $K \subseteq X$ be compact and $E \subseteq K$ be infinite. Then E has limit point in K.*

Proof. Suppose E has no limit point in K, then every $x \in K$ has a neighborhood B_x such that $B_x \cap E$ has at most one point. The collection $\{B_x\}_{x \in K}$ is an open cover of K, and by the compactness of K, it has a finite subcover of K (and also of E), but they only contain finite points of E, where has a contradiction. □

Theorem 2.11. *Suppose $K \subseteq Y \subseteq X$. Then K is compact relative to X if and only if K is compact relative to Y.*

Proof. If K is compact relative to Y, let $\{V_\alpha\}_{\alpha\in A}$ be an open cover of K and let $U_\alpha = V_\alpha \cap Y$ for every $\alpha \in A$. Then $\{U_\alpha\}_{\alpha\in A}$ is also an open cover of K, and by compactness of K (relative to Y) it has a finite subcover $\{U_i\}_{i=1}^n$. Since $K \subseteq \cup_{i=1}^n U_i \subseteq \cup_{i=1}^n V_i$, $\{V_i\}_{i=1}^n$ is also a finite subcover of K, and thus K is compact relative to X. □

Theorem 2.12. *If $E \subseteq K$ is closed and K is compact, then E is compact.*

Proof. Let $\{G_\alpha\}_{\alpha\in A}$ be an open cover of E, then $\{G_\alpha\}_{\alpha\in A} \cup \{E^c\}$ is an open cover of K. By the compactness of K, it has a finite subcover $\{G_i\}_{i=1}^n \cup \{E^c\}$. Hence, $\{G_i\}_{i=1}^n$ is a finite subcover of $\{G_\alpha\}_{\alpha\in A}$ containing E. Therefore, E is compact. □

Definition 2.13 (Finite Intersection Property)**.** *A family of subsets $\{E_\alpha\}_{\alpha\in A}$ of X has **finite intersection property (FIP)** if the intersection of any of its finite subcover $\{E_i\}_{i=1}^n$ is non-empty.*

Theorem 2.13. *K is compact if and only if for all families of closed sets $\{E_\alpha \subseteq K\}_{\alpha\in A}$ with FIP, their intersection $\cap_{\alpha\in A} E_\alpha$ is non-empty.*

Proof. Suppose K is compact, and there exists $\{E_\alpha\}_{\alpha\in A}$ which are closed sets of K with FIP, and $\cap_{\alpha\in A} E_\alpha = \emptyset$. Then take $G_\alpha = E_\alpha^c$, and $\{G_\alpha\}_{\alpha\in A}$ becomes an open cover of K since $\cup_{\alpha\in A} G_\alpha \supseteq K$ (which follows $\cap_{\alpha\in A} E_\alpha = \emptyset$). However, $\{E_\alpha\}_{\alpha\in A}$ has FIP implies $\cap_{\alpha\in B} E_\alpha$ is non-empty for any finite subset $B \subseteq A$, that is, $\cup_{\alpha\in B} G_\alpha$ cannot cover K for any finite subset $B \subseteq A$, so that $\{G_\alpha\}_{\alpha\in A}$ has no finite subcover, where has a contradiction.

Suppose for all families of closed subsets $\{E_\alpha\}_{\alpha\in A}$ of K with FIP, their intersection is non-empty. If K is not compact, there exists an open cover $\{G_\alpha\}_{\alpha\in A}$ with no finite subcover. Take $E_\alpha = K\backslash G_\alpha$, then $\{E_\alpha\}_{\alpha\in A}$ are closed subsets of K with FIP and non-empty intersection.

(CONTINUE) □

Corollary 2.4. *If $\{K_\alpha\}_{\alpha\in A}$ is a family of compact subsets of X, and it has FIP, then their intersection $\cap_{\alpha\in A} K_\alpha$ is non-empty.*

Proof. Choose a member K_0 from $\{K_\alpha\}_{\alpha\in A}$ and let $G_\alpha = K_\alpha^c$. If there is no point of K_0 belongs to every K_α, then the sets $\{G_\alpha\}_{\alpha\in A}$ is an open cover of K_0. By the compactness of K_0, it has a finite subcover $\{G_i\}_{i=1}^n$ such that $K_0 \subset \cup_{i=1}^n G_i$. However, it implies that $K_0 \cap (\cap_{i=1}^n K_i) = K_0 \cap (\cup_{i=1}^n G_i)^c$ is empty, which contradicts the FIP. □

Corollary 2.5. *If $\{K_n\}$ is a sequence of non-empty compact sets such that $K_n \supset K_{n+1}$, then $\cap_{n=1}^\infty K_n$ is not empty.*

Theorem 2.14. *Suppose $a < b$ are real numbers. Then the interval $[a, b] \subseteq \mathbb{R}$ is compact.*

Proof. Suppose not, then there is an open cover $\{G_\alpha\}_{\alpha\in A}$ of $[a, b]$ with no finite subcover. Take $c = (a + b)/2$ be the center point, then either $[a, c]$ or $[c, b]$ has no finite subcover and denote it $[a_1, b_1]$. Repeating this procedure we could define a sequence of intervals $\{[a_n, b_n]\}_{n=1}^\infty$ such that

1. $[a_n, b_n] \supseteq [a_{n+1}, b_{n+1}]$;
2. $|b_n - a_n| = 2^{-n}|b - a|$;
3. $[a_n, b_n]$ cannot be covered by any finite elements of $\{G_\alpha\}_{\alpha\in A}$.

Let $x = \sup_n a_n$; it exists since $a_n \le x \le b$ for all n. Then we have $[a_n, b_n] \subseteq B_{2^{1-n}|b-a|}(x)$ for all n. Since open sets $\{G_\alpha\}_{\alpha\in A}$ covers x, there exists $r > 0$ such that $B_r(x) \subseteq G_\alpha$ for some $\alpha \in A$. Hence, when n is great enough such that $2^{1-n}|b - a| < r$, we have $[a_n, b_n] \subseteq B_{2^{1-n}|b-a|}(x) \subseteq B_r(x) \subseteq G_\alpha$, where has a contradiction. □

Theorem 2.15. *Every k-cell $\{(x_1, ..., x_k) : a_i \le x_i \le b_i, i = 1, ..., k\} \subseteq \mathbb{R}^k$ is compact.*

Proof. Omitted. □

Theorem 2.16. *For a set $E \subseteq \mathbb{R}^k$ the following three properties are equivalent:*

1. E is closed and bounded;
2. E is compact;

3. Every infinite subset of E has a limit point in E.
The equivalence of (1)(2) is known as the ***Heine-Borel Theorem****.*

Proof. Theorem 2.9 has shown that (2)⇒(1), and Theorem 2.10 has shown that (2)⇒(3). Then we only need to show

1. (1)⇒(2): Since E is bounded, then it is a subset of k-cell $\{(x_1, ..., x_k) \in \mathbb{R}^k : -R \leq x_i \leq R, i = 1, ..., k\}$ for some $R > 0$, which is compact (by Theorem 2.15). Thus E is compact since it is a closed subset of a compact set (Theorem 2.12).
2. (3)⇒(1): (CONTINUE)

□

Theorem 2.17 (Weierstrass)**.** *Every bounded infinite subset of $\mathbb{R}^k$ has a limit point in $\mathbb{R}^k$.*

Proof. Since it is bounded, it is a subset of $\{(x_1, ..., x_k) \in \mathbb{R}^k : -R \leq x_i \leq R, i = 1, ..., k\}$ for some $R > 0$, which is compact. Then, it has a limit point since it is an infinite subset of a compact set (Theorem 2.10). □

Definition 2.14 (Dense)**.** *In metric space (X, d), $E \subseteq X$ is* ***dense*** *in X if $\bar{E} = X$.*

For example, $\mathbb{Q}$ is dense in $\mathbb{R}$.

3 Numerical Sequences and Series

3.1 Convergence

Definition 3.1 (Sequence)**.** *A **sequence** in X is a mapping $p : \mathbb{N} \to X$, denoted by (p_n) or $(p_0, p_1, ...)$.*

Definition 3.2 (Convergence)**.** *The sequence (p_n) in metric space (X, d) **converges** to $p \in X$, if for all $\varepsilon > 0$, there exists N such that for all $n \geq N$, we have $d(p_n, p) < \varepsilon$. We write $p_n \to p$ or $\lim_{n\to\infty} p_n = p$.*

*Sequence (p_n) is **divergent** if it is not convergent.*

Note that the convergence of a sequence may depends on the space it belongs to. For example, the sequence $(1, 1.4, 1.41, 1.414, ...)$ (the first decimals of $\sqrt{2}$) converges to $\sqrt{2}$ in $\mathbb{R}$, but divergent in $\mathbb{Q}$.

Proposition 3.1. *$x_n \to x$ if and only if every neighborhood of x contains all but finitely many x_n.*

Theorem 3.1. *If $E \subseteq X$ and p is a limit point of E, then there is a sequence (p_n) in E such that $p_n \to p$.*

Proof. Since p is a limit point of E, for all $n = 1, 2, ...$ there exists a point $p_n \in E$ such that $d(p_n, p) < 1/n$. Then it could be shown that such sequence (p_n) converges to p: for all $\varepsilon > 0$, choose N so that $N\varepsilon > 1$, then $d(p_n, p) < \varepsilon$ for all $n \geq N$. □

Theorem 3.2. *$E \subseteq X$ is closed if and only if for every sequence (x_n) in E such that $x_n \to x$, we have $x \in E$.*

Proof. If (x_n) in E converges to $x \in E$, then every neighborhood of x contain some x_n, which means $x \in \bar{E}$. Since E is closed, we have $x \in E = \bar{E}$.

Conversely, for all $x \in \bar{E}$, choose $x_n \in B_{1/n}(x) \cap E$ for all $n = 1, 2, ...$ (such x_n exists since x is a limit point of E) and we have $x_n \to x$. By $x \in E$ we have $\bar{E} \subseteq E$, which means $E = \bar{E}$ and thus E is closed. □

Theorem 3.3. *The limit of a sequence is unique (if exists).*

Proof. Suppose $x_n \to p$ and $x_n \to q$; we only need to show that $d(p, q) = 0$. The convergence means that for all $\varepsilon > 0$, there exists N, M such that for all $n \geq N$ we have $d(x_n, p) < \varepsilon/2$, and for all $n \geq M$ we have $d(x_n, p) < \varepsilon/2$. Hence, for all $n \geq \max\{N, M\}$, we have $d(p, q) \leq d(x_n, p) + d(x_n, q) < \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we must have $d(p, q) = 0$. □

Definition 3.3 (Boundedness)**.** *A sequence (x_n) is **bounded** if its range $\{x_n : n \in \mathbb{N}\}$ is bounded.*

Theorem 3.4. *Any convergent sequence is bounded.*

Proof. Suppose $x_n \to x$, then there exists N such that for all $n \geq N$ we have $d(x_n, x) < 1$. Hence,

$$\sup d(x_n, x) \leq \sup\{d(x_0, x), d(x_1, x), ..., d(x_N, x), 1\} < \infty$$

since the set in the equation above is finite. □

Proposition 3.2. *If $(a_n), (b_n)$ are sequences such that (1) $a_n = b_n$ for all but finitely many n, or (2) $(b_n) = (a_N, a_{N+1}, ...)$ for some N, then either they both diverge, or they have the same limit.*

3.2 Cauchy Convergence

Definition 3.4 (Cauchy Convergence)**.** *The sequence (x_n) is **Cauchy** if for all $\varepsilon > 0$, there exists N such that for all $n, m > N$, we have $d(x_n, x_m) < \varepsilon$.*

Theorem 3.5. *Every convergent sequence in Cauchy.*

Definition 3.5 (Cauchy Completeness)**.** *A space (X, d) is **Cauchy complete** if every Cauchy sequence in X converges in X.*

Theorem 3.6. *Compact sets are Cauchy complete.*

Proof. Suppose (x_n) is Cauchy in a compact set K. If there exists $y \in K$ such that $x_n = y$ for infinitely many n, then $x_n \to y$ since (x_n) is Cauchy.

If not, let $E_N = \{x_n : n \geq N\}$ for $N = 1, 2, ...$ Then we know that every E_N is infinite, $\bar{E}_N \subseteq K$ (closed set of compact set is also compact), and clearly has FIP. Hence there exists $x \in \cap_N \bar{E}_N$.

To show that $x_n \to x$, for all $\varepsilon > 0$, choose M such that for all $n, m > M$ we have $d(x_n, x_m) < \varepsilon/2$. Meanwhile, $x \in \bar{E}_M$ implies that the neighborhood $B_{\varepsilon/2}(x)$ contains some x_n. Hence, $d(x, x_m) < \varepsilon$ for all $m > M$ and thus $x_n \to x$. □

Corollary 3.1. *$\mathbb{R}^n$ is complete.*

Proof. (CONTINUE) □

Definition 3.6 (Cauchy Completion)**.** *If $X \subseteq Y$ where X is a metric space, Y is Cauchy complete, and X is dense in Y, then we call Y the **Cauchy completion** of X.*

Theorem 3.7. *For metric space (X, d), its Cauchy completion always exists and unique up to isomorphism.*

3.3 Subsequence

Definition 3.7 (Subsequence)**.** *Let (x_n) be a sequence in metric space (X, d), and (n_k) be a strictly inreasing sequence in $\mathbb{N}$. Then $(x_{n_k})_{k \in \mathbb{N}}$ is a **subsequence** of (x_n). If $x_{n_k} \to y$ as $k \to \infty$, then we call y a **subsequential limit** of (x_n).*

Proposition 3.3. *$x_n \to x$ if and only if every subsequence of (x_n) converges to x.*

Proof. Since $n_k \to \infty$ as $k \to \infty$, we have $x_n \to x \Rightarrow x_{n_k} \to x$; If $x_n \nrightarrow x$, then for some $\varepsilon > 0$, we can construct a subsequence (x_{n_k}) by letting $k \in A = \{n \in \mathbb{N} : x_n \notin B_\varepsilon(y)\}$ which is infinite, and then $x_{n_k} \nrightarrow x$. □

Proposition 3.4. *If (x_n) is Cauchy, and some subsequences (x_{n_k}) converge to x, then $x_n \to x$.*

Proof. For any $\varepsilon > 0$, since (x_n) is Cauchy, there exists N such that for all $n, m > N$, we have $d(x_n, x_m) < \varepsilon/2$. Take $n_k > N$ such that $d(x_{n_k}, x) < \varepsilon/2$, then for all $m > N$, we have $d(x, x_m) < d(x_{n_k}, x_m) + d(x_{n_k}, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. □

Theorem 3.8. *If x is a limit point of the range of (x_n), then x is also a subsequential limit of (x_n).*

Proof. We can construct such a subsequence by taking $x_{n_k} \in B_{1/k}(x)$ for every $k = 1, 2, ...$, while making sure that $n_k > n_{k-1}$. Such subsequence exists since every neighborhood of x contains some points in (x_n). □

However, the converse doesn't hold. A counter-example is $(1, 1/2, 1, 1/3, 1, 1/4, ...)$ which has no limit, 2 subsequential limits (0 and 1), but only 1 limit point (0).

Corollary 3.2. *If (x_n) is a sequence in a compact space K, then (x_k) has a convergent subsequence.*

Proof. If the range of (x_n) is finite, then clearly it has convergent (and indeed constant) subsequence and thus subsequential limit. If the range is infinite, we can show it by the fact that an infinite subset of a compact set has limit point (Theorem 2.10), and thus has subsequential limit by the theorem above. □

Theorem 3.9 (Bolzano-Weierstrass)**.** *Any bounded sequence in $\mathbb{R}^n$ has convergent subsequences.*

Proof. It follows from the corollary above, and the fact that a bounded subset in $\mathbb{R}^n$ lies in a compact subset of $\mathbb{R}^n$ (Heine-Borel, Theorem 2.16). □

Theorem 3.10. *For any sequence (x_n), the set of all subsequential limits is closed.*

Proof. Suppose (x_n) is a sequence, $\{x_{n_k^l} = (x_{n_0^l}, x_{n_1^l}, x_{n_2^l}, ...)\}_{l\in\mathbb{N}}$ is an infinite set of subsequences of (x_n), where $x_{n_k^l} \to y_l (k \to \infty)$ and $y_l \to y(l \to \infty)$. Then we only need to show that y is also a sequential limit of (x_n).

Choose $n_0 = n_0^0$; given n_{l-1}, take $n_l \in \{n_k^l\}_{k\in\mathbb{N}}$ such that $n_l > n_{l-1}$, and $d(x_{n_l}, y_l) < 2^{-l}$. By triangle inequality, we have

$$d(x_{n_l}, y) \le d(x_{n_l}, y_l) + d(y_l, y) < 2^{-l} + d(y_l, y).$$

For all $\varepsilon > 0$, there exists N such that $2^{-N} < \varepsilon/2$ and $d(y_l, y) < \varepsilon/2$ for all $l > N$, hence $d(x_{n_l}, y) < \varepsilon$ for all $l > N$. Therefore, $x_{n_l} \to x$ as $l \to \infty$ and x is a sequential limit of (x_n). ☐

Definition 3.8 (Monotonicity)**.** *Let (a_n) be a sequence in $\mathbb{R}$. If $a_n \le a_{n+1}$ for all n, we call (a_n) **(monotone) increasing**; if $a_n \ge a_{n+1}$ for all n, we call (a_n) **(monotone) decreasing**. (a_n) is **monotone** if it is either increasing or decreasing.*

Theorem 3.11. *A monotone sequence has a limit if and only if it is bounded.*

Proof. Suppose (a_n) is increasing and bounded. Let $\alpha = \sup range(a_n) \in \mathbb{R}$. For all $\varepsilon > 0$, there exists N such that $a_N > \alpha - \varepsilon$, and $\alpha - \varepsilon < a_n \le \alpha$ for all $n > N$, hence $a_n \to \alpha$. The case of decreasing bounded sequences is analogous. The converse is shown in Theorem 3.4. ☐

Definition 3.9 (Tend to Infinity)**.** *For a sequence (x_n) in $\mathbb{R}$, we say $x_n \to \infty$ if for all R, there exists N such that for all $n > N$, we have $x_n \ge R$; we say $x_n \to -\infty$ if $x_n \le -R$.*

Note that a sequence is not bounded if and only if it has some subsequence that tends to infinity.

3.4 Limit Superior and Inferior

Definition 3.10 (Limit Superior/Inferior)**.** *For a sequence (x_n) in $\mathbb{R}$, define its **limit superior** and **limit inferior** as*

$$\limsup x_n = \lim_{n\to\infty} \sup E_n$$

$$\liminf x_n = \lim_{n\to\infty} \inf E_n$$

where $E_N = \{x_n : n \ge N\}$ is the range of (x_n) excluding first N terms.

Usually we say $\limsup x_n = \infty$ or $\liminf x_n = -\infty$ if (x_n) is not bounded above or below.

Note that

1. the sequence $(\sup E_n)$ is decreasing, while $(\inf E_n)$ is increasing. Since the values are extended to extended reals $\mathbb{R} \cup \{\pm\infty\}$, the limit superior and inferior always exist.
2. if (x_n) converges, then we have $\limsup x_n = \liminf x_n = \lim x_n$.

Proposition 3.5. $\limsup x_n$ *(if finite) is the least number α such that for all $\varepsilon > 0$, there exists N such that for all $n > N$, we have $x_n < \alpha + \varepsilon$.*

Theorem 3.12. *Let (x_n) be a sequence and y a subsequential limit of (x_n). Then we have*

$$\liminf x_n \le y \le \limsup x_n.$$

Proof. Suppose $y \le \limsup x_n$ doesn't hold, so that $y - \limsup x_n = \varepsilon > 0$. Since $\limsup x_n = \lim_{N\to\infty} \sup\{x_n : n > N\}$, there exists N such that $\sup\{x_n : n > N\} < \limsup x_n + \varepsilon/2 = y - \varepsilon/2$. Hence, $B_{\varepsilon/2}(y)$ doesn't contain any x_n such that $n > N$, which means y cannot be a sequential limit. Analogously we can show that $y \ge \liminf x_n$. ☐

Corollary 3.3. *For any sequence (x_n), we have $\liminf x_n \le \limsup x_n$.*

Proof. If (x_n) is bounded, then by Bolzano-Weierstrass theorem, it has subsequential limit y, then by theorem above we complete the proof. If (x_n) is not bounded above(below), then there exists some subsequence tends to $+\infty(-\infty)$, so that $\limsup x_n = +\infty(\liminf x_n = -\infty)$ and we complete the proof. ☐

3.5 Series

Definition 3.11 (Series). *Given sequence (x_n) in $\mathbb{R}^d$, for $N \leq M \in \mathbb{N}$, denote*

$$\sum_{n=N}^{M} x_n = x_N + x_{N+1} + ... + x_M.$$

*We call $S_N = \sum_{n=0}^{N} x_n$ the **partial sum** of (x_n), and we associate a sequence (S_n) with (x_n).*

*We call $\sum_{n=0}^{\infty} x_n$ a **series**; we say $\sum_{n=0}^{\infty} x_n = \lim_{n\to\infty} S_n$ if (S_n) converges; call $\sum_{n=0}^{\infty} x_n$ diverges if (S_n) diverges.*

Recall that $\mathbb{R}^d$ is Cauchy complete; so that $\sum_{n=0}^{\infty} x_n$ converges if and only if (S_n) is Cauchy.

Theorem 3.13. *If the series $\sum x_n$ converges, then $x_n \to 0$.*

Proof. Since (s_n) is Cauchy, for all $\varepsilon > 0$ there exists N such that for all $n > N$, we have

$$|x_{n+1}| = |S_{n+1} - S_n| \leq \varepsilon$$

so that $x_n \to 0$. ☐

Note that the converse does not hold. A typical counter-example is that the harmonic series 1+1/2+1/3+... diverges.

Theorem 3.14 (Comparison Test).

1. *If $|a_n| \leq c_n$ for all $n \geq N_0$, and $\sum c_n$ converges, then $\sum a_n$ converges.*
2. *If $a_n \geq d_n \geq 0$ for all $n \geq N_0$, and $\sum d_n$ diverges, then $\sum a_n$ diverges.*

Proof. Suppose $\sum c_n$ converges, so that for all $\varepsilon > 0$, there exists $N \geq N_0$ such that for all $m \geq n \geq N$ we have $\sum_{k=n}^{m} c_k \leq \varepsilon$ by Cauchy criterion. Then (1) follows from

$$\left|\sum_{k=n}^{m} a_k\right| \leq \sum_{k=n}^{m} |a_k| \leq \sum_{k=n}^{m} c_k \leq \varepsilon.$$

(2) follows from (1), for if $\sum a_n$ converges, $\sum d_n$ must converge. ☐

Theorem 3.15 (Root Test). *Given series $\sum a_n$, let $\alpha = \limsup \sqrt[n]{|a_n|}$. Then*

1. *If $\alpha < 1$, $\sum a_n$ converges;*
2. *If $\alpha > 1$, $\sum a_n$ diverges;*
3. *If $\alpha = 1$, this test cannot tell if $\sum a_n$ converges or diverges.*

Proof.

1. If $\alpha < 1$, there exists N such that for all $n > N$, we have $\sqrt[n]{|a_n|} < \beta$, where $\alpha < \beta < 1$. Then we have $|a_n| < \beta^n$ for all $n > N$. The series $\sum \beta^n$ converges since $0 < \beta < 1$. Hence, $\sum a_n$ converges by the comparison test.
2. If $\alpha > 1$, there is a sequence (n_k) such that $\sqrt[n_k]{|a_{n_k}|} \to \alpha$. Hence, $|a_n| > 1$ for infinitely many n, so that $a_n \nrightarrow 0$ and thus $\sum a_n$ diverges.
3. Consider two series $\sum 1/n$ and $\sum 1/n^2$; for each of them $\alpha = 1$, but the first diverges, the second converges.

☐

Theorem 3.16 (Ratio Test). *Given series $\sum a_n$, it*

1. *converges if* $\limsup |a_{n+1}/a_n| < 1$*;*
2. *diverges if $|a_{n+1}/a_n| \geq 1$ for all $n \geq n_0$.*

Proof.

1. If $\limsup |a_{n+1}/a_n| < 1$, then there exists N such that for all $n > N$, we have $|a_{n+1}/a_n| < \beta$ where $\beta < 1$. So that
$$|a_{N+p}| < \beta|a_{N+p-1}| < ... < \beta^p|a_N|, \forall p = 1, 2, ...$$

which implies

$$|a_n| < (|a_N|\beta^{-N})\beta^n$$

for all $n > N$. Hence, $\sum a_n$ converges by the comparison test.

2. If $|a_{n+1}| \geq |a_n|$ for all $n \geq n_0$, we have $a_n \nrightarrow 0$ and thus $\sum a_n$ diverges.

□

Theorem 3.17 (Comparison of Root and Ratio Tests)**.** *If the series* $\sum_{n=1}^{\infty} a_n$ *passes ratio test, then it will also pass root test. Hence, the root test is always better to detect convergent series.*

Proof. Suppose the ratio test shows convergence of $\sum_{n=1}^{\infty} a_n$, that is, $\limsup a_{n+1}/a_n = \lambda < 1$. Then for all $\varepsilon > 0$, there exists N such that for all $n = 1, 2, ...$, we have $a_{N+n} \leq (\lambda+\varepsilon)^n a_N$. Hence

$$\limsup \sqrt[n]{a_n} \leq \lim_{m\to\infty} [(\lambda+\varepsilon)^m a_N]^{1/(N+m)} = \lambda + \varepsilon.$$

Therefore, $\limsup \sqrt[n]{a_n} \leq \limsup a_{n+1}/a_n$, and the inequality holds in some cases (shown below). □

Example 3.1. *Consider the series* $1/2 + 1/3 + 1/2^2 + 1/3^2 + 1/2^3 + 1/3^3 + ...$ *Note that*

1. $a_{n+1}/a_n = (2/3)^k$ *or* $(1/2)(3/2)^k$*, so that* $\limsup a_{n+1}/a_n = \infty$*;*
2. $\sqrt[n]{a_n} \approx 1/\sqrt{3}$ *or* $1/\sqrt{2}$.

Hence, this series fails to pass the ratio test, while passes the root test (and thus convergent).

Definition 3.12 (Absolute Convergence)**.** *The series* $\sum a_n$ ***converges absolutely*** *if the series* $\sum |a_n|$ *converges.*

Theorem 3.18. *If* $\sum a_n$ *converges absolutely, then* $\sum a_n$ *converges.*

Note that the converse does not hold. For example, the series $\sum(-1)^n/n$ converges while the harmonic series $\sum 1/n$ diverges.

Theorem 3.19. *If* $\sum a_n = A, \sum b_n = B$*, then* $\sum(a_n + \lambda b_n) = A + \lambda B$ *for any fixed* λ*.*

Proof. It is because the partial sums satisfy

$$\sum_{n=0}^{k}(a_n + \lambda b_n) = \sum_{n=0}^{k} a_n + \lambda \sum_{n=0}^{k} b_n$$

By letting $k \to \infty$ we complete the proof. □

Definition 3.13 (Product of Series)**.** *Given two series* $\sum a_n, \sum b_n$*, we put*

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}$$

and call $\sum c_n$ *the* ***product*** *of the two given series.*

Theorem 3.20. *Suppose* $\sum a_n$ *converges absolutely,* $\sum a_n = A, \sum b_n = B$*, and* $c_n = \sum_{k=0}^{n} a_k b_{n-k}$*, then* $\sum c_n = AB$*.*

Proof. Put

$$A_n = \sum_{k=0}^{n} a_k, \quad B_n = \sum_{k=0}^{n} b_k, \quad C_n = \sum_{k=0}^{n} c_k, \quad \beta_n = B_n - B.$$

Then we have

$$\begin{aligned}C_n &= a_0b_0 + (a_0b_1 + a_1b_0) + ... + (a_0b_n + a_1b_{n-1} + ... + a_nb_0)\\ &= a_0B_n + a_1B_{n-1} + ... + a_nB_0\\ &= a_0(B+\beta_n) + a_1(B+\beta_{n-1}) + ... + a_n(B+\beta_0)\\ &= A_nB + a_0\beta_n + a_1\beta_{n-1} + ... + a_n\beta_0.\end{aligned}$$

Put $\gamma_n = \alpha_0\beta_n + ... + \alpha_n\beta_0$. Since $A_nB \to AB$, then we only need to show that $\gamma_n \to 0$.

Put $\alpha = \sum_{n=0}^{\infty} |a_n|$ (recall that $\sum a_n$ converges absolutely). Since $B_n \to B$, for all $\varepsilon > 0$, there exists N such that for all $n \geq N$, we have $|\beta_n| = |B_n - B| \leq \varepsilon$. Then for $n \geq N$, we have

$$\begin{aligned}|\gamma_n| &\leq |\beta_0 a_n + ... + \beta_N a_{n-N}| + |\beta_{N+1}a_{n-(N+1)} + ... + \beta_n a_0| \\ &\leq |\beta_0 a_n + ... + \beta_N a_{n-N}| + \varepsilon\alpha.\end{aligned}$$

Keeping N fixed, by letting $n \to \infty$, we get

$$\limsup |\gamma_n| \leq \varepsilon\alpha$$

since $a_k \to 0$ as $k \to \infty$. Since $\varepsilon > 0$ is arbitrary, we have $\gamma_n \to 0$. □

Theorem 3.21. *If the series $\sum a_n, \sum b_n, \sum c_n$ converge to A, B, C, and $c_n = a_0b_n + ... + a_nb_0$, then $C = AB$.*

Theorem 3.22 (Alternating Series Test). *Suppose (a_n) is a non-negative monotone decreasing sequence such that $a_1 \geq a_2 \geq ...$ and $\lim a_n = 0$. Then $\sum_{n=1}^{\infty}(-1)^{n+1}a_n = a_1 - a_2 + a_3 - a_4 + ...$ converges.*

Proof. The partial sum of even number of terms is

$$S_{2N} = (a_1 - a_2) + (a_3 - a_4) + ...$$

in which every term is non-negative (since (a_n) is decreasing), so that the sequence $(S_{2N})_N$ is non-negative and increasing. On the other hand, the partial sum of odd number of terms is

$$S_{2N+1} = a_1 - (a_2 - a_3) - (a_4 - a_5) - ...$$

in which every term (except for a_1) is non-positive, so that the sequence $(S_{2N+1})_N$ is decreasing. Also, for all N we have

$$0 \leq S_{2N} \leq S_{2N+1} \leq a_1$$

so that both sequences converges. Moreover, since

$$\lim_{N\to\infty} |S_{2N+1} - S_{2N}| = \lim_{N\to\infty} a_{2N+1} = 0,$$

so that both sequences have the same limit $\lim_{N\to\infty} S_N$. Therefore, the series $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ converges. □

Definition 3.14 (Rearrangement of Series). *Suppose $k : \mathbb{N}^+ \to \mathbb{N}^+$ is a bijection. Then we say that $\sum a'_n = \sum a_{k(n)}$ is a* ***rearrangement*** *of $\sum a_n$.*

Theorem 3.23 (Riemann Rearrangement). *Suppose $\sum a_n$ is a series of real numbers which converges but not absolutely. Suppose $-\infty \leq \alpha \leq \beta \leq +\infty$. Then there exists a rearrangement $\sum a'_n = \sum a_{k(n)}$ with partial sums s'_n such that*

$$\liminf s'_n = \alpha, \qquad \limsup s'_n = \beta.$$

Proof. Let

$$p_n = \frac{|a_n| + a_n}{2}, \qquad q_n = \frac{|a_n| - a_n}{2}$$

so that $p_n - q_n = a_n, p_n + q_n = |a_n|$ and $p_n, q_n \geq 0$ for all n. The series $\sum p_n, \sum q_n$ must both diverge: they cannot both be convergent since $\sum(p_n + q_n) = \sum |a_n|$ diverges; if one of them converges while the other diverges, $\sum(p_n - q_n) = \sum a_n$ will diverge and bring contradiction.

Let $P_1, P_2, P_3, ...$ denote the non-negative terms of $\sum a_n$ (in the order in which they occur), and Q_1, Q_2, Q_3 be the absolute values of negative terms of $\sum a_n$ (also in their original order). The series $\sum P_n, \sum Q_n$ differ from $\sum p_n, \sum q_n$ only by zero terms, and are therefore divergent.

Now we manage to construct sequences $(m_n), (k_n)$ such that the series

$$P_1 + ... + P_{m_1} - Q_1 - ... - Q_{k_1} + P_{m_1+1} + ... + P_{m_2} - Q_{k_1+1} - ... - Q_{k_2} + ...$$

which clearly is a rearrangement of $\sum a_n$, satisfies $\liminf s'_n = \alpha, \limsup s'_n = \beta$.

To construct such rearrangement, choose real-valued sequences $(\alpha_n), (\beta_n)$ such that $\alpha_n \to \alpha, \beta_n \to \beta$, $\alpha < \beta_n$ and $\beta_1 > 0$. The construction takes following steps:

1. Let m_1 be the smallest integer such that $P_1 + ... + P_{m_1} > \beta_1$;
2. Let k_1 be the smallest integer such that $P_1 + ... + P_{m_1} - Q_1 - ... - Q_{k_1} < \alpha_1$;
3. Let m_2 be the smallest integer such that $P_1 + ... + P_{m_1} - Q_1 - ... - Q_{k_1} + P_{m_1+1} + ... + P_{m_2} > \beta_2$;
4. Let k_2 be the smallest integer such that $P_1 + ... + P_{m_1} - Q_1 - ... - Q_{k_1} + P_{m_1+1} + ... + P_{m_2} - Q_{k_1+1} - ... - Q_{k_2} < \alpha_2$;
5. Repeat previous steps to find $m_3, k_3, ...$

Such integers $m_1, k_1, m_2, k_2, ...$ could be found since $\sum P_n$ and $\sum Q_n$ diverges.

If x_n, y_n denote the partial sums of rearrangement whose last terms are $P_{m_n}, -Q_{k_n}$, then we have

$$|x_n - \beta_n| \le P_{m_n}, \qquad |y_n - \alpha_n| \le Q_{k_n}.$$

Since $P_n \to 0, Q_n \to 0$ as $n \to \infty$ (as $\sum a_n$ converges), we have $x_n \to \beta, y_n \to \alpha$.

Finally, it is clear that no number less than α or greater than β can be a sequential limit of the partial sums of the rearrangement. □

Theorem 3.24. *If $\sum a_n$ is a series of complex numbers which converges absolutely, then every rearrangement of $\sum a_n$ converges, and they all converge to the same sum.*

Proof. Let $\sum a_{k(n)} = \sum a'_n$ be a rearrangement with partial sums s'_n. Since $\sum |a_n|$ converges, for all $\varepsilon > 0$, there exists N such that for all $m \ge n \ge N$, we have $\sum_{i=n}^{m} |a_i| \le \varepsilon$.

Choose p such that integers $1, 2, ..., N$ are all contained in the set $k(1), ..., k(p)$. Then for all $n > p$, we have $|s_n - s'_n| \le \varepsilon$ since $a_1, ..., a_N$ are contained in both s_n and s'_n. Therefore, (s'_n) converges to the same sum as (s_n). □

3.6 Some Important Series

Theorem 3.25. *For $r \in \mathbb{R}$, the series $\sum_{n=0}^{\infty} r^n = 1/(1-r)$ if $|r| < 1$, diverges if $|r| \ge 1$.*

Proof. The partial sum is given by

$$S_n = 1 + r + ... + r^n = \frac{1 - r^{n+1}}{1 - r}.$$

If $|r| \ge 1$, then $|r^n| \ge 1$ for all n, so that (S_n) diverges; if $|r| < 1$, then $r^{n+1} \to 0$, so that $S_n \to 1/(1-r)$. □

Theorem 3.26. *For $p \in \mathbb{R}$, the series $\sum_{n=1}^{\infty} 1/n^p$ converges if $p > 1$, diverges if $p \le 1$.*

Proof. Suppose $p > 1$. Then the partial sum satisfies

$$\sum_{n=1}^{2^{N+1}} \frac{1}{n^p} \le 1 + \frac{1}{2^p} + \frac{1}{2^p} + \frac{1}{4^p} + \frac{1}{4^p} + \frac{1}{4^p} + \frac{1}{4^p} + ... = \sum_{n=0}^{N} \left(\frac{2}{2^p}\right)^n$$

which converges.

Suppose $p = 1$. Then the partial sum satisfies

$$\sum_{n=1}^{2^N - 1} \frac{1}{n} \ge \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + ... = \frac{N}{2}$$

which clearly diverges, so that the harmonic series diverges.

Suppose $p < 1$, then the series clearly diverges since $1/n^p \ge 1/n$ for all n. □

Theorem 3.27. *Suppose (a_n) is monotone decreasing and $a_n \ge 0$ for all n. Then the series $\sum a_n$ converges if and only if the series*

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + ...$$

converges.

Proof. We only need to show that the partial sums are bounded. Let the partial sums be $s_n = a_1 + ... + a_n$ and $t_k = a_1 + 2a_2 + ... + 2^k a_{2^k}$.

For $n < 2^k$, we have

$$\begin{aligned} s_n &\leq a_1 + (a_2 + a_3) + ... + (a_{2^k} + ... + a^{2^{k+1}-1}) \\ &\leq a_1 + 2a_2 + ... + 2^k a_{2^k} = t_k \end{aligned}$$

so that $s_n \leq t_k$. On the other hand, for $n > 2^k$, we have

$$\begin{aligned} s_n &\geq a_1 + a_2 + (a_3 + a_4) + ... + (a_{2^{k-1}+1} + ... + a_{2^k}) \\ &\geq \frac{1}{2}a_1 + a_2 + 2a_4 + ... + 2^{k-1}a_{2^k} = \frac{1}{2}t_k \end{aligned}$$

so that $2s_n \geq t_k$. Hence, the sequences (s_n) and (t_k) are either both bounded or both unbounded. Therefore we complete the proof. □

Theorem 3.28. *If $p > 1$, the series $\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p}$ converges; if $p \leq 1$, this series diverges.*

Proof. Note that the sequence $(1/n \log n)$ decreases. Then by theorem above, it leads us to the series

$$\sum_{k=1}^{\infty} 2^k \frac{1}{2^k (\log 2^k)^p} = \sum_{k=1}^{\infty} \frac{1}{(k \log 2)^p} = \frac{1}{(\log 2)^p} \sum_{k=1}^{\infty} \frac{1}{k^p}$$

which converges if and only if $p > 1$, diverges if and only if $p \leq 1$. □

Definition 3.15 (e)**.** $e = \sum_{n=0}^{\infty} 1/n!$ *(here $0! = 1$).*

Note that such series converges, since the sequence of partial sums ($s_n = \sum_{k=0}^{n} 1/k!$) is monotone increasing, and it is bounded:

$$\begin{aligned} s_n &= 1 + 1 + \frac{1}{1 \times 2} + \frac{1}{1 \times 2 \times 3} + ... + \frac{1}{1 \times 2 \times ... \times n} \\ &< 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + ... + \frac{1}{2^{n-1}} < 3. \end{aligned}$$

Theorem 3.29. $\lim_{n\to\infty}(1 + 1/n)^n = e$.

Proof. Let $s_n = \sum_{k=0}^{n} 1/k!$ and $t_n = (1 + 1/n)^n$. By binomial theorem,

$$\begin{aligned} t_n &= 1 + \frac{n!}{(n-1)!1!}\frac{1}{n} + \frac{n!}{(n-2)!2!}\frac{1}{n^2} + ... + \frac{n!}{1!(n-1)!}\frac{1}{n^n} \\ &= 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \frac{1}{3!}\left(1 - \frac{2}{n}\right) + ... + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) ... \left(1 - \frac{n-1}{n}\right). \end{aligned}$$

Hence $t_n \leq s_n$, so that $\limsup t_n \leq e$.

Next, if $n \geq m$,

$$t_n \geq 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + ... + \frac{1}{m!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) ... \left(1 - \frac{m-1}{n}\right).$$

Keeping m fixed, by letting $n \to \infty$, we have

$$\liminf t_n \geq 1 + 1 + \frac{1}{2!} + ... + \frac{1}{m!}$$

so that $s_m \leq \liminf t_n$. Letting $m \to \infty$ we obtain $e \leq \liminf t_n$. Therefore, $\lim_{n\to\infty} t_n = e$. □

Theorem 3.30. *e is irrational.*

Proof. Since $e = \sum_{n=0}^{\infty} 1/n!$, let the partial sums be $s_n = \sum_{k=0}^{n} 1/k!$, then we have

$$\begin{aligned} e - s_n &= \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + ... \\ &< \frac{1}{(n+1)!}\left[1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + ...\right] = \frac{1}{n!n}. \end{aligned}$$

So that

$$0 < e - s_n < \frac{1}{n!n}.$$

Suppose e is rational, then $e = p/q$ where p, q are positive integers. Then we have

$$0 < q!(e - s_q) < \frac{1}{q}.$$

Since

$$q!s_q = q!\left(1 + 1 + \frac{1}{2!} + ... + \frac{1}{q!}\right)$$

is an integer, $q!(e - s_q)$ is also an integer. However, $1/q \in (0, 1]$ since $q \geq 1$, where has a contradiction. □

Definition 3.16 (Power Series)**.** *Given a sequence (c_n) of complex numbers, the series $\sum_{n=0}^{\infty} c_n z^n$ is called a **power series**, where z is a complex number.*

Theorem 3.31. *Given the power seires $\sum c_n z^n$, put*

$$\alpha = \limsup \sqrt[n]{|c_n|}, \qquad R = \frac{1}{\alpha}$$

(If $\alpha = 0$, $R = +\infty$; if $\alpha = +\infty, R = 0$). Then $\sum c_n z^n$ converges absolutely if $|z| < R$, diverges if $|z| > R$. R is called the radius of convergence of $\sum c_n z^n$.

Proof. Put $a_n = c_n z^n$ and apply the root test:

$$\limsup \sqrt[n]{|a_n|} = |z| \limsup \sqrt[n]{|c_n|} = \frac{|z|}{R}.$$

□

However, at the boundary $|z| = R$ one cannot simply assert the convergence or divergence of the power series. Indeed, any situation may be the case at the boundary.

Example 3.2.
1. *The power series $\sum z^n$ has radius $R = 1$ and diverges whenever $|z| = 1$.*
2. *The power series $\sum z^n/n^2$ has radius $R = 1$ and converges whenever $|z| = 1$.*
3. *The power series $\sum z^n/n$ has radius $R = 1$, diverges when $z = 1$, but converges conditionally for other $|z| = 1$.*

3.7 p-Norm, $\mathbb{R}^{\mathbb{N}}$ and $\mathbb{R}^{\infty}$

Definition 3.17 (p-Norm)**.** *For $p \in [1, \infty)$, the p-**norm** on $\mathbb{R}^n$ is given by*

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

It also defines a metric on $\mathbb{R}^n$ by $d(x, y) = \|x - y\|_p$. When $p = \infty$ define the ∞-norm as

$$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p = \max_{1\leq i\leq n} |x_i|.$$

Theorem 3.32. *For any $x \in \mathbb{R}^d$ and $1 \le p \le q \le \infty$,*

$$\|x\|_q \le \|x\|_p \le d^{\frac{1}{p}-\frac{1}{q}}\|x\|_q.$$

This means that in finite dimensional vector space $\mathbb{R}^d$, all p-norms share the identical topological structure with typical 2-norm.

Definition 3.18 ($\mathbb{R}^{\mathbb{N}}$)**.** $\cup_{n=1}^{\infty}\mathbb{R}^n = \mathbb{R}^{\mathbb{N}}$.

Note that infinite sequences are NOT contained in $\mathbb{R}^{\mathbb{N}}$; all elements in $\mathbb{R}^{\mathbb{N}}$ are of finite dimension.

Now consider $(\mathbb{R}^{\mathbb{N}}, p)$, the space $\mathbb{R}^{\mathbb{N}}$ with p-norm. The following proposition shows that different p-norms have different topological structure in $\mathbb{R}^{\mathbb{N}}$.

Proposition 3.6. *Let $p \in [1,\infty], q \in \mathbb{R}$. Then in $(\mathbb{R}^{\mathbb{N}}, p)$ the sequence $(a_n) = ((1, 1/2^q, ..., 1/n^q))_{n\in\mathbb{N}}$ is (1) Cauchy if $pq > 1$; (2) unbounded if $pq \le 1$; (3) bounded but not Cauchy if $pq = \infty \cdot 0$.*

Proof. For $p \in [1,\infty)$ we have

$$\|a_n\|_p = \left(\sum_{k=1}^{n} \frac{1}{k^{pq}}\right)^{1/q}$$

so that for $m > n$,

$$\|a_m - a_n\|_p = \left(\sum_{k=n+1}^{m} \frac{1}{k^{pq}}\right)^{1/q}.$$

Hence, (a_n) is Cauchy if and only if the sequence $(1/n^{pq})$ converges, that is, if and only if $pq > 1$.

For $p = \infty$ we have

$$\|a_n\|_\infty = \begin{cases} 1 & q \ge 0 \\ n^{-q} & q < 0 \end{cases}$$

so that (a_n) is bounded if and only if $q \ge 0$. Also, for $m > n$,

$$\|a_m - a_n\|_\infty = \max\left\{\frac{1}{n^p}, \frac{1}{m^p}\right\}.$$

Hence, (a_n) is Cauchy if and only if $1/n^p \to 0$, that is, if and only if $p > 0$. □

Theorem 3.33. *For $p \in [1,\infty]$, the space $(\mathbb{R}^{\mathbb{N}}, p)$ is not Cauchy complete.*

Example 3.3. *Let $p = 2, a_n = (1, 1/2, ...1/n)$. Consider whether the Cauchy sequence (a_n) is convergent in $\mathbb{R}^{\mathbb{N}}$: for any $x \in \mathbb{R}^d \subseteq \mathbb{R}^{\mathbb{N}}$ and $m > d$, we have*

$$\|x - a_m\|_2^2 = \|x - a_d\|_2^2 + \frac{1}{(d+1)^2} + ... + \frac{1}{m^2}.$$

So that $x \notin B_\varepsilon(a_m)$ for all $\varepsilon < 1/(d+1)^2$. Hence, $a_n \nrightarrow x$ and (a_n) diverges.

Definition 3.19 (Sequence Space)**.** *The* ***sequence space*** *$\mathbb{R}^\infty$ is the set of all (infinite) sequences in $\mathbb{R}$.*

We claim that $\mathbb{R}^{\mathbb{N}} \subseteq \mathbb{R}^\infty$, by identifying $(x^1, ..., x^d) \in \mathbb{R}^{\mathbb{N}}$ with $(x^1, ..., x^d, 0, 0, ...) \in \mathbb{R}^\infty$.

Definition 3.20 (ℓ^p Space)**.** *For $p \in [1,\infty]$ and $a = (a^1, a^2, ...) \in \mathbb{R}^\infty$, define its p-norm as*

$$\|a\|_p = \left(\sum_{n=1}^{\infty} (a^n)^p\right)^{1/p}, \quad 1 \le p < \infty$$

and

$$\|a\|_\infty = \sup_n |a^n|.$$

*Define ℓ^p **space** as*

$$\ell^p = \{a \in \mathbb{R}^\infty : \|a\|_p < \infty\}$$

and it is also a metric space with $d(x, y) = \|x - y\|_p$.

Example 3.4. *For any $a = (a^n) \in \mathbb{R}^\infty$,*

1. *$\|a\|_1 = \sum_{n=1}^\infty |a^n|$, so that ℓ^1 is the set of all absolute convergent sequences.*
2. *$\|a\|_\infty = \sup_n |a^n|$, so that ℓ^∞ is the set of all bounded sequences.*

Proposition 3.7. *For any $p \in [1, \infty]$, $\mathbb{R}^\mathbb{N} \subseteq \ell^p \subseteq \mathbb{R}^\infty$.*

Proposition 3.8. *For any $p \in [1, \infty)$ and $a = (a^n) \in \mathbb{R}^\infty$, define $a_n = (a^1, ..., a^n, 0, ...) \in \mathbb{R}^\mathbb{N} \subseteq \mathbb{R}^\infty$, then $a \in \ell^p$ if and only if $a_n \to a$ in ℓ^p.*

Proof. It is because

$$a \in \ell^p \Leftrightarrow \sum |a^n|^p \text{ converges} \Leftrightarrow \sum_{n=N}^\infty |a^n|^p \to 0 (N \to \infty) \Leftrightarrow a_n \to a.$$

□

Theorem 3.34. *For $p \in [1, \infty)$, ℓ^p is the Cauchy completion of $\mathbb{R}^\mathbb{N}$ with respect to the p-norm.*

Proof. By proposition above we know that $\mathbb{R}^\mathbb{N}$ is dense in ℓ^p. Then we only need to show that ℓ^p is Cauchy complete. Here we only prove the ℓ^1 case.

Suppose $(a_n) = ((a_n^1, a_n^2, ...))_n$ is Cauchy in ℓ^1. First we show that the sequences of each fixed component $(a_n^k \in \mathbb{R})_n$ is convergent, since it is Cauchy:

$$|a_n^k - a_m^k| \leq \sum_{k=1}^\infty |a_n^k - a_m^k| = \|a_n - a_m\|_1 \to 0 (n, m \to \infty)$$

and $\mathbb{R}$ is Cauchy complete. Hence, we denote $a^k = \lim_{n\to\infty} a_n^k$ and $a = (a^1, a^2, ...)$.

To ensure that $a \in \ell^1$, since (a_n) is Cauchy, there exists R such that $\|a_n\|_1 = \sum_k |a_n^k| \leq R$ for all n. Hence, for all K we have

$$\sum_{k=1}^K |a^k| = \sum_{k=1}^K \left|\lim_{n\to\infty} a_n^k\right| = \lim_{n\to\infty} \sum_{k=1}^K |a_n^k| \leq R$$

and by letting $K \to \infty$ we have $\sum_k |a^k| \leq R$. Therefore, $a \in \ell^1$.

Finally, we show that $a_n \to a$. Since (a_n) is Cauchy, for all $\varepsilon > 0$, there exists N such that for all $n, m > N$ we have

$$\|a_n - a_m\|_1 = \sum_{k=1}^\infty |a_n^k - a_m^k| < \varepsilon.$$

Thus, for all K we also have $\sum_{k=1}^K |a_n^k - a_m^k| < \varepsilon$. By letting $m \to \infty$ we have

$$\sum_{k=1}^K |a_n^k - a^k| \leq \varepsilon.$$

Therefore,

$$\|a_n - a\|_1 = \sum_{k=1}^\infty |a_n^k - a^k| = \sup_K \sum_{k=1}^K |a_n^k - a^k| \leq \varepsilon$$

and $a_n \to a$. □

4 Continuous Function

Definition 4.1 (Basic Concepts of Functions). *Let $f : X \to Y$ be a **function**. We call X the **domain** of f, and Y the **codomain** of f.*

*For $E \subseteq X$, call $f(E) = \{f(x) : x \in E\}$ the **image** of E under f. $f(X)$ is called the **image** of f.*

*For $F \subseteq Y$, call $f^{-1}(F) = \{x \in X : f(x) \in F\}$ the **pre-image (inverse image)** of F.*

Example 4.1.

1. *Identity mapping $id : X \to X, x \mapsto x$;*
2. *Constant mapping $c : X \to Y, x \mapsto c \in Y$;*
3. *Quadratic functions $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$, $g : \mathbb{Q} \to \mathbb{Q}, g(x) = x^2$, and $h : \mathbb{Q} \to \mathbb{R}, h(x) = x^2$. We also write $h = f|_{\mathbb{Q}}$ as the **restriction** of f to $\mathbb{Q}$.*

Definition 4.2 (Limit in Function). *Let X, Y be metric spaces, $f : X \to Y$, $E \subseteq X$ and $p \in E'$. We say*

$$\lim_{x \to p} f(x) = y$$

if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in E$ and $x \neq p$, we have

$$d_X(x, p) < \delta \Rightarrow d_Y(f(x), y) < \varepsilon.$$

Note that the limit cannot be well-defined for $p \notin E'$.

Proposition 4.1. *Suppose $f : E \to Y$ where $E \subseteq X, p \in E'$. Then $\lim_{x\to p} f(x) = y$ if and only if for all sequences (x_n) in E (and $x_n \neq x$ for any n) such that $\lim_{n\to\infty} x_n = p$, we have $\lim_{n\to\infty} f(x_n) = y$.*

Proof.

- $\Rightarrow$ Take $\varepsilon > 0$; since $\lim_{x\to p} f(x) = y$, there exists $\delta > 0$ such that $d_X(x, p) < \delta \Rightarrow d_Y(f(x), y) < \varepsilon$. Moreover, since $x_n \to p$, there exists $N > 0$ such that for all $n > N$, we have $d_X(x_n, p) < \delta$, and thus $d_Y(f(x_n), y) < \varepsilon$ for all $n > N$. Therefore, $f(x_n) \to y$.
- $\Leftarrow$ Suppose $\lim_{x\to p} f(x) \neq y$, then there exists $\varepsilon > 0$ such that for all δ, there exists $x \in B_\delta(p), x \neq p$ such that $d_Y(f(x), y) \geq \varepsilon$. Then for all $n \in \mathbb{N}_+$, we can choose $x_n \in B_{1/n}(p), x_n \neq p$ such that $d_Y(f(x_n), y) \geq \varepsilon$, so that $f(x_n) \nrightarrow y$ where has a contradiction.

$\square$

Corollary 4.1. *The limit in function is unique; if $\lim_{x\to p} f(x) = a$ and $\lim_{x\to p} f(x) = b$, then $a = b$.*

To incorporate the concept of infinite limits and limits and infinity, we can extend the concept of limit in function. Here we define that for any $c \in \mathbb{R}$, the set $(c, +\infty)$ is a neighborhood of $+\infty$, and $(-\infty, c)$ is a neighborhood of $-\infty$.

Definition 4.3 (Limit in Function). *Let f be a real function defined on E. We say that $f(t) \to A$ as $t \to x$ where $A, x \in \mathbb{R} \cup \{\pm\infty\}$, if for every neighborhood U of A, there exists a neighborhood V of x such that $f(t) \in U$ for all $t \in V \cap E, t \neq x$ (and $V \cap E$ is non-empty).*

4.1 Continuous Function

Definition 4.4 (Continuous Function). *Let X, Y be metric spaces, $f : X \to Y, E \subseteq X$ and $p \in E$. We say that the function f is **continuous** at point p if for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in X$, we have $d_X(x, p) < \delta \Rightarrow d_Y(f(x), f(p)) < \varepsilon$.*

*If f is **continuous** at any point $p \in E$, we say that f is continuous on E; f is **continuous** if it is continuous on X.*

Proposition 4.2. *If p is an isolated point of X, then f is continuous at p.*

Proposition 4.3. *$f : X \to Y$ is continuous at p if and only if for all sequences (x_n) in X such that $x_n \to p$, we have $f(x_n) \to f(p)$.*

Theorem 4.1. *$f : X \to Y$ is continuous if and only if for all open subsets $U \subseteq Y$, its inverse $f^{-1}(U) \subseteq X$ is also open in X.*

Proof.
- $\Leftarrow$: For all $p \in X, \varepsilon > 0$, define $A = f^{-1}(B_\varepsilon(f(p))) \subseteq X$. If property holds, then $A \subseteq X$ is open and $p \in A$, so that there exists $\delta > 0$ such that $B_\delta(p) \subseteq A$; that is, $d_X(x, p) < \delta \Rightarrow d_Y(f(x), f(p)) < \varepsilon$, so that f is continuous at p.
- $\Rightarrow$: Consider a open subset $U \subseteq Y$ and any $p \in f^{-1}(U)$: since U is open, there exists $\varepsilon > 0$ such that $B_\varepsilon(f(p)) \subseteq U$. Since f is continuous, there exists $\delta > 0$ such that $d_X(x, p) < \delta \Rightarrow d_Y(f(x), f(p)) < \varepsilon$. Hence, $f(B_\delta(p)) \subseteq B_\varepsilon(f(p)) \subseteq U$ and $B_\delta(p) \subseteq f^{-1}(U)$. Since $p \in f^{-1}(U)$ is arbitrary, $f^{-1}(U)$ is open.

$\square$

Corollary 4.2. *$f : X \to Y$ is continuous if and only if the pre-images of all closed subsets of Y are closed.*

Theorem 4.2. *Composition of continuous functions is also continuous. Suppose $f : X \to Y, g : Y \to Z$ are continuous, then $g \circ f : X \to Z$ is also continuous.*

Proof. It could be proved in three ways using different definitions of continuity which are actually equivalent.

1. For all $\varepsilon > 0, p \in X$, since g is continuous, there exists $\eta > 0$ such that
$$d_Y(f(p), y) < \eta \Rightarrow d_Z(g \circ f(p), g(y)) < \varepsilon.$$
Moreover, since f is continuous, there exists $\delta > 0$ such that
$$d_X(p, x) < \delta \Rightarrow d_Y(f(p), y) < \eta \Rightarrow d_Z(g \circ f(p), g \circ f(x)) < \varepsilon.$$
2. Let $x_n \to p \in X$, then $f(x_n) \to f(p) \in Y$ since f is continuous, and $g \circ f(x_n) \to g \circ f(x) \in Z$ since g is continuous.
3. Let $U \subseteq Z$ be open, then $g^{-1}(U) \subseteq Y$ is open since g is continuous, and $f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U) \subseteq X$ is open since f is continuous.

$\square$

Proposition 4.4. *If $\lim_{x \to p} f(x) = y$ for $f : E \to Y$ and $p \in E'$, and $g : Y \to Z$ is continuous, then $\lim_{x \to p} g \circ f(x) = g(y)$.*

Theorem 4.3. *$+ : \mathbb{R}^2 \to \mathbb{R}$ is continuous.*

Proof. Take a sequence $((x_n, y_n))_n$ in $\mathbb{R}$ such that $(x_n, y_n) \to (x, y)$, then
$$+((x_n, y_n)) = x_n + y_n \to x + y = +((x, y)).$$

$\square$

Lemma 4.1. *Let $f, g : E \to \mathbb{R}, E \subseteq X$ and $p \in E'$. Let $(f, g) : E \to \mathbb{R}^2, x \mapsto (f(x), g(x))$, then*
$$\lim_{x \to p} (f, g)(x) = \left(\lim_{x \to p} f(x), \lim_{x \to p} g(x) \right).$$
Also, (f, g) is continuous if and only if f, g are both continuous.

Theorem 4.4. *For $f, g : X \to \mathbb{R}$ and $p \in X'$, we have*
$$\lim_{x \to p} f(x) + g(x) = \lim_{x \to p} f(x) + \lim_{x \to p} g(x).$$

Proof. It is because
$$\begin{aligned}\lim_{x \to p} f(x) + g(x) &= \lim_{x \to p} +((f, g)(x)) = + \left(\lim_{x \to p} (f, g)(x) \right) \\ &= + \left(\left(\lim_{x \to p} f(x), \lim_{x \to p} g(x) \right) \right) = \lim_{x \to p} f(x) + \lim_{x \to p} g(x).\end{aligned}$$

□

Proposition 4.5. *In $\mathbb{R}^n$, the addition, scalar multiplicatoin, and dot product are continuous; in $\mathbb{C}$ and $\mathbb{R}$, the multliplication is continuous, and division is continuous if the denominator is non-zero.*

Consequently, the properties (continuity, limit, etc.) of $f, g : X \to \mathbb{R}$ can be inherited to $f + g, f \cdot g, f/g$.

4.2 Continuity and Compactness

Definition 4.5 (Boundedness)**.** *A function is **bounded** if its image is bounded.*

Theorem 4.5. *If function $f : K \to Y$ is continuous and K is compact, the image $f(K)$ is also compact in Y.*

Proof. Let $\{G_\alpha\}_\alpha$ be an open cover of $f(K)$. First we show that $\{f^{-1}(G_\alpha)\}_\alpha$ is also an open cover of K: (1) Each $f^{-1}(G_\alpha)$ is open since f is continuous; (2) For any $x \in K$, $f(x) \in f(K)$ so that $f(x) \in f(G_\alpha)$ for some α, so that $x \in f^{-1}(G_\alpha)$.

Since K is compact, $\{f^{-1}(G_\alpha)\}_\alpha$ has finite subcover $\{f^{-1}(G_n)\}_{n=1}^N$. For any $y \in f(K)$, we have $y = f(x)$ for some $x \in K$, and $x \in f^{-1}(G_n)$ for some n, so that $y \in G_n$ and thus $\{G_n\}_{n=1}^N$ is a finite subcover of K. □

Corollary 4.3. *If function $f : K \to \mathbb{R}$ is continuous and K is compact,*

1. *$f(K)$ is bounded;*
2. *there exists $a, b \in K$ such that $f(a) = \inf_{x\in K} f(x), f(b) = \sup_{x\in K} f(x)$;*
3. *if $f(x) > 0$ for all $x \in K$, then $\inf_{x\in K} f(x) > 0$.*

Consider a continuous function $f : K \to Y$ with compact domain K; if $E \subseteq K$ is closed, it is also compact, so that $f(E) \subseteq Y$ is compact, which implies $f(E)$ is closed in Y. However, it does not necessarily hold for open sets.

Proposition 4.6. *Let $f : X \to Y$ be surjective and K be compact. If $U \subseteq K$ is open, then $f(U)$ is also open; if f is bijective, then $f^{-1} : Y \to K$ is also continuous.*

4.3 Uniform Continuity

Definition 4.6 (Uniform Continuity)**.** *A function $f : X \to Y$ is **uniformly continuous** if for all $\varepsilon > 0$, there exists $\delta > 0$ such that $d_X(x, y) < \delta \Rightarrow d_Y(f(x), f(y)) < \varepsilon$.*

Recall that the definition of continuity only requires that $f : X \to Y$ is continuous at all $x \in X$, so that for a fixed $\varepsilon > 0$, the chosen δ may vary with $x \in X$; however, δ must also be fixed for all $x \in X$ by definition of uniform continuity.

Theorem 4.6. *Let $f : X \to Y$ be uniformly continuous, and (x_n) be Cauchy in X, then $(f(x_n))$ is also Cauchy in Y.*

Proof. By definition of uniform continuity, for all $\varepsilon > 0$, there exists $\delta > 0$ such that $d_X(a, b) < \delta \Rightarrow d_Y(f(a), f(b)) < \varepsilon$. Since (x_n) is Cauchy, there exists N such that $d_X(x_n, x_m) < \delta \Rightarrow d_Y(f(x_n), f(x_m)) < \varepsilon$ for all $m, n > N$, so that $(f(x_n))$ is also Cauchy. □

Example 4.2. *To show that the uniform continuity matters, suppose $f(x) = 1/x$ and $(x_n) = (1/n)$ which is Cauchy, then $(f(x_n)) = (n)$ which is not Cauchy.*

Theorem 4.7. *If $f : K \to Y$ is continuous and K is compact, then f is uniformly continuous.*

Proof. For all $\varepsilon > 0$ and all $x \in K$, we define r_x such that

$$B_{r_x}(x) \subseteq f^{-1}\left(B_{\varepsilon/2}(f(x))\right).$$

Since $\{B_{r_x/2}(x)\}_{x\in K}$ is an open cover of K, it has finite subcover $\{B_{r_{x_n}/2}(x_n)\}_{n=1}^N$. Take $\delta = \min_n r_{x_n}/2$; if $d_X(x, y) < \delta$, there exists some x_n such that $d_X(x_n, y) \le d_X(x_n, x) + d_X(x, y) < r_{x_n}$. Hence,

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

□

Theorem 4.8. *Suppose $f : E \to Y$ where $E \subseteq X$ and $p \in E'$. If f is uniformly continuous and Y is complete, then $\lim_{x\to p} f(x)$ exists.*

Proof. Suppose the sequence (a_n) in E converges to p and $a_n \neq p$ for all n. Then (a_n) is Cauchy and so is $(f(a_n))$, so that $f(a_n) \to y \in Y$ since Y is complete. To show that y does not depend on the choice of sequence (a_n), suppose another sequence (b_n) in E converges to p and $b_n \neq p$ for all n. Then the mixed sequence $(a_0, b_0, a_1, b_1, ...)$ is also Cauchy, and so is $(f(a_0), f(b_0), f(a_1), f(b_1), ...)$. Hence, there must be $f(b_n) \to y$ as well. □

Theorem 4.9. *Let $E \subseteq X, f : E \to Y$ be uniformly continuous, and Y is complete. Then there exists an extension $\bar{f} : \bar{E} \to Y$ which is continuous.*

Proof. For $p \in \bar{E}\backslash E$, define $\bar{f}(p) = \lim_{x\to p} f(x)$ (the existence of limit is shown in theorem above). Now we only to show the continuity of $\bar{f}$ at $p \in \bar{E}\backslash E$, that is, for all $\varepsilon > 0$, there exists $\delta > 0$ such that $d(x, y) < \delta \Rightarrow d(f(x), f(y)) < \varepsilon$ for any $x, y \in \bar{E}\backslash E$.

Take two sequences $(x_n), (y_n)$ in E such that $x_n \to x, y_n \to y$, then by definition we have $f(x_n) \to \bar{f}(x), f(y_n) \to \bar{f}(y)$. Since f is uniformly continuous, for all $\varepsilon > 0$, there exists $\delta > 0$ such that $d(x_n, y_n) < 3\delta \Rightarrow d(f(x_n), f(y_n)) < \varepsilon/3$.

Suppose $d(x, y) < \delta$; we only need to show that $d(\bar{f}(x), \bar{f}(y)) < \varepsilon$. Since $x_n \to x, y_n \to y$, there exists N such that for all $n > N$, we have $d(x_n, x) < \delta, d(y_n, y) < \delta$, then by triangle inequality,

$$d(x_n, y_n) < d(x_n, x) + d(x, y) + d(y, y_n) < 3\delta$$

so that $d(f(x_n), f(y_n)) < \varepsilon/3$.

Since $f(x_n) \to \bar{f}(x), f(y_n) \to \bar{f}(y)$, there exists $M > N$ such that for all $n > M$, we have $d(f(x_n), \bar{f}(x)) < \varepsilon/3, d(f(y_n), \bar{f}(y)) < \varepsilon/3$. Hence, by triangle inequality,

$$d(\bar{f}(x), \bar{f}(y)) < d(\bar{f}(x), f(x_n)) + d(f(x_n), f(y_n)) + d(f(y_n), \bar{f}(y)) < \varepsilon$$

and we complete the proof. □

4.4 Continuity and Connectedness

Theorem 4.10. *If $f : X \to Y$ is continuous and X is connected, then the image $f(X) \subseteq Y$ is also connected.*

Proof. Suppose $f(X)$ is disconnected, so that $f(X) = A \cup B \subseteq Y$, where $A, B \subseteq Y$ are non-empty and separated (i.e. $\bar{A} \cap B = A \cap \bar{B} = \emptyset$). Then $X \subseteq f^{-1}(A) \cup f^{-1}(B)$.

It could be shown that $f^{-1}(A), f^{-1}(B)$ are separated: since $f^{-1}(A) \subseteq f^{-1}(\bar{A})$ which is closed due to the continuity of f, we have $\overline{f^{-1}(A)} \subseteq f^{-1}(\bar{A})$, then

$$\overline{f^{-1}(A)} \cap f^{-1}(B) \subseteq f^{-1}(\bar{A}) \cap f^{-1}(B) = \emptyset$$

since $\bar{A} \cap B = \emptyset$. Analogously, $f^{-1}(A) \cap \overline{f^{-1}(B)} = \emptyset$.

However, since $X \subseteq f^{-1}(A) \cup f^{-1}(B)$ is connected, there must be either $X \subseteq f^{-1}(A)$ or $X \subseteq f^{-1}(B)$, which implies $f(X) \subseteq A$ or $f(X) \subseteq B$, and contradicts the non-emptyness of B or A. Therefore, $f(X)$ is connected. □

Corollary 4.4 (Intermediate Value Theorem)**.** *If $f : X \to Y$ is continuous and X is connected, and there exist $a, b \in X$ such that $f(a) < r < f(b)$, then there exists $x \in X$ such that $f(x) = r$.*

However, IVT (and connected domain) cannot imply continuity. An example is

$$f(x) = \begin{cases} \sin(1/x) & x > 0 \\ 0 & x = 0 \end{cases}$$

It satisfies IVT but is not continuous at $x = 0$.

Theorem 4.11. *E is connected if and only if any function $f : E \to \{0, 1\}$ (with discrete metric) which is continuous, we have $f(E) = \{0\}$ or $\{1\}$.*

Proof. If E is connected and f is continuous, by IVT $f(E)$ is connected, so that $f(E) = \{0\}$ or $\{1\}$.

To show the converse, suppose E is disconnected and $E = A \cup B$ which are disjoint, closed and non-empty. Take

$$f(x) = \begin{cases} 0 & x \in A \\ 1 & x \in B \end{cases}$$

Since f is continuous and $\{0, 1\}$ has four open subsets $\emptyset, \{0\}, \{1\}, \{0, 1\}$, we have $f^{-1}(\{0\}) = A, f^{-1}(\{1\}) = B$ must be open, where has a contradiction. Therefore, E is connected. ☐

A more common notion of connectedness is to use the idea of **paths**.

Definition 4.7 (Path Connectedness)**.** *A path in X is a continuous function $f : [0, 1] \to X$, from point $f(0) = a$ to point $f(1) = b$. A set E is* ***path connected*** *if for all $a, b \in E$, there exists a path $f : [0, 1] \to E$ such that $f(0) = a, f(1) = b$.*

With the theorem above, now we can show that path connectedness implies connectedness.

Theorem 4.12. *Path connectedness implies connectedness.*

Proof. If E is disconnected, we have $E = A \cup B$ where A, B are separated. Then pick $a \in A, b \in B$. If f is a path from a to b, since A, B are open and f is continuous, $f^{-1}(A), f^{-1}(B) \subseteq [0, 1]$ are non-empty and open. However, $[0, 1]$ is impossible to separate into two (non-empty) open subsets. ☐

However, the converse does not hold. A well-known example is given below.

Example 4.3 (Topologists' sine curve)**.** *Consider the set*

$$\{(x, y) : y = \sin(1/x), x > 0\} \cup \{(0, y) : -1 \leq y \leq 1\}.$$

It could be shown that such set is connected but not path connected.

4.5 Discontinuities

Definition 4.8 (Limit from Above/Below)**.** *For function $f : \mathbb{R} \to \mathbb{R}$ and $p \in \mathbb{R}$, take $g = f|_{(-\infty,p)}, h = f|_{(p,+\infty)}$, define the* ***limit from above/below*** *by*

$$\lim_{x \to p^+} f(x) = \lim_{x \to p} h(x), \quad \lim_{x \to p^-} f(x) = \lim_{x \to p} g(x).$$

Note that f is continuous at p if and only if

$$\lim_{x \to p^+} f(x) = \lim_{x \to p^-} f(x) = f(p).$$

Usually we categorize three kinds of continuities:

1. Removable discontinuity: $\lim_{x \to p^+} f(x) = \lim_{x \to p^-} f(x) \neq f(p)$. A simple example is

$$f(x) = \begin{cases} x^2 & x \neq 0 \\ 1 & x = 0 \end{cases}$$

2. Jump discontinuity: $\lim_{x \to p^+} f(x) \neq \lim_{x \to p^-} f(x)$. A simple example is $sgn(x)$ at $x = 0$.
3. Essential discontinuity: cases other than (1)(2); that is, at least one of $\lim_{x \to p^+} f(x)$ and $\lim_{x \to p^-} f(x)$ does not exist. A simple example is $f(x) = 1/x$ at $x = 0$.

The following examples show that the number of discontinuities is not limited.

Example 4.4.

1. *Suppose $f(x)$ is discontinuous at $x = 0$, then the function $g(x) = f(x - a) + f(x - b) + f(x - c)$ is discontinuous at $x = a, b, c$. This means we can construct functions with finitely many discontinuities.*

2. *The well-known **Dirichlet function** is given by*

$$1_{\mathbb{Q}} = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

Since $\mathbb{Q}$ is dense in $\mathbb{R}$, the Dirichlet function is discontinuous everywhere.

3. *We can construct function $f(x) = 1_{\mathbb{Q}}(x)\sin(nx)$ so that it is continuous only at $x \in \mathbb{Z}$. That is, we can construct functions which is discontinuous everywhere except some countably many points.*
4. *The well-known **Thomae's function** is given by*

$$f(x) = \begin{cases} 1/q & x = p/q \in \mathbb{Q}, p, q \in \mathbb{Z} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

where p/q is the most reduced form of $x \in \mathbb{Q}$. It could be shown that f is continuous at irrational points but discontinuous at rational points. That is, we can construct functions with countably many discontinuities.

5. *The **Conway Base-13 Function** is discontinuous everywhere but still satisfies IVT. It is defined as follows: for every $x \in \mathbb{R}$, write it in base 13 using the alphabet $\{0, 1, ..., 9, +, -, .\}$. If from some point onwards, the expansion is of form $+x_1x_2...x_n.y_1y_2...$ or $-x_1x_2...x_n.y_1y_2...$, then $f(x) = x_1...x_n.y_1y_2...$ or $-x_1x_2...x_n.y_1y_2...$; else $f(x) = 0$.*

Definition 4.9 (Monotonicity)**.** *For ordered sets X, Y, the function $f : X \to Y$ is **monotone increasing/decreasing** if $a \geq b$ implies $f(a) \geq / \leq f(b)$. In both cases f is **monotone**.*

Here we only focus on $f : \mathbb{R} \to \mathbb{R}$.

Theorem 4.13. *Monotone functions can only have jump discontinuities (or no discontinuity).*

Proof. For any p, consider the sequences $\{f(p - 1/n)\}_n$ and $\{f(p + 1/n)\}_n$; they are both monotone bounded by $f(p)$, so that they are both convergent. Denote $f(p - 1/n) \to f(p^-)$ and $f(p + 1/n) \to f(p^+)$.

To show that $\lim_{x \to p^-} f(x) = f(p^-)$, since $f(p - 1/n) \to f(p^-)$, for all ε there exists N such that $|f(p - 1/n) - f(p^-)| < \varepsilon$ for all $n > N$, so take $\delta = 1/(N + 1)$ and we have $|x - p| < \delta \Rightarrow |f(x) - f(p^-)| < \varepsilon$. Analogously, $\lim_{x \to p^+} f(x) = f(p^+)$.

Finally, to show that $f(p^-), f(p), f(p^+)$ are ordered, it follows from the fact that $f(p-1/n), f(p), f(p+1/n)$ are ordered for all n. □

Theorem 4.14. *Monotone functions can be discontinuous only at countably many points.*

Proof. Suppose f is monotone increasing and E is the set of jump discontinuous points. For $p \in E$, define

$$I_p = \left(\lim_{x \to p^-} f(x), \lim_{x \to p^+} f(x) \right)$$

as the jump interval. Clearly each I_p is non-empty open interval, and I_p, I_q are disjoint if $p \neq q$. However, $\mathbb{R}$ cannot contain uncountably many non-empty disjoint open intervals, since each of the intervals contains at least one rational number. □

Except for the restriction given by the theorem above, the discontinuities of monotone functions can be arbitrary.

Example 4.5. *Let $\sum a_n$ be a convergent series, and $\{p_n\}$ be a countable collection of points in $\mathbb{R}$ (suppose $p_1 < p_2 < ...$). Then define $f(x) = \sum_{n:p_n<x} a_n$. It is an increasing function, but discontinuous at each p_n:*

$$\lim_{x \to p_n^-} f(x) = \sum_{k=1}^{n-1} a_k, \quad \lim_{x \to p_n^+} f(x) = \sum_{k=1}^{n} a_k.$$

5 Differentiation

5.1 Derivative of Real Function

Definition 5.1 (Derivative)**.** *For function $f : \mathbb{R} \to \mathbb{R}$ and $x \in [a, b]$, define (provided this limit exists)*

$$f'(x) = \lim_{t \to x} \frac{f(t) - f(x)}{t - x} = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

where $t \neq x, h \neq 0$.

*Thus, we associate with the function f a function f' whose domain is the set of points x at which the limit exists. f' is called the **derivative** of f.*

*If f' is defined at the point x, we say that f is **differentiable** at x; if f' is defined at every point of a set $E \subseteq \mathbb{R}$, we say that f is differentiable on E.*

Example 5.1.

1. *Suppose*

   $$f(x) = \begin{cases} x^2 & x \in \mathbb{Q} \\ -x^2 & x \notin \mathbb{Q} \end{cases}$$

   then

   $$\frac{f(0 + h) - f(0)}{h} = \begin{cases} h & h \in \mathbb{Q} \\ -h & h \notin \mathbb{Q} \end{cases}$$

   so that $f'(0) = 0$.
2. *The well-known Cantor staircase function is continuous and has derivative 0 (except for the Cantor set), but not constant.*

Theorem 5.1. *Let f be defined on $[a, b]$. If f is differentiable at a point $x \in [a, b]$, then f is continuous at x.*

Proof. It is because as $t \to x$,

$$f(t) - f(x) = \frac{f(t) - f(x)}{t - x}(t - x) \to f'(x) \cdot 0 = 0.$$

□

Theorem 5.2. *Suppose f, g are defined on $[a, b]$ and are differentiable at a point $x \in [a, b]$. Then $f + g, fg, f/g$ (if $g(x) \neq 0$) are differentiable at x, and*

1. *$(f + g)'(x) = f'(x) + g'(x)$;*
2. *$(fg)'(x) = f'(x)g(x) + f(x)g'(x)$;*
3. *$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$.*

Proof.

1. $$\begin{aligned}(f + g)'(x) &= \lim_{h \to 0} \frac{[f(x + h) + g(x + h)] - [f(x) + g(x)]}{h} \\ &= \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} + \frac{g(x + h) - g(x)}{h} = f'(x) + g'(x).\end{aligned}$$
2. $$\begin{aligned}(fg)'(x) &= \lim_{h \to 0} \frac{f(x + h)g(x + h) - f(x)g(x)}{h} \\ &= \lim_{h \to 0} \frac{f(x + h)g(x + h) - f(x)g(x + h)}{h} + \frac{f(x)g(x + h) - f(x)g(x)}{h} \\ &= \lim_{h \to 0} f'(x)g(x + h) + f(x)g'(x) = f'(x)g(x) + f(x)g'(x).\end{aligned}$$

3.

$$\left(\frac{f}{g}\right)'(x) = \lim_{h\to 0}\frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h}$$
$$= \lim_{h\to 0}\frac{1}{g(x+h)g(x)}\frac{f(x+h)g(x) - f(x)g(x+h)}{h}$$
$$= \frac{1}{[g(x)]^2}[f'(x)g(x) - f(x)g'(x)].$$

□

Theorem 5.3 (Chain Rule)**.** *Suppose f is continuous on $[a,b]$, $f'(x)$ exists at some point $x \in [a,b]$, g is defined on an interval I which contains the range of f, and g is differentiable at the point $f(x)$. Then $h = f \circ g$ is differentiable at x, and*

$$h'(x) = g'(f(x))f'(x).$$

Proof. Let $y = f(x)$. By the definition of the derivative, we have

$$f(t) - f(x) = (t-x)[f'(x) + u(t)]$$

$$g(s) - g(y) = (s-y)[g'(y) + v(s)]$$

where $t \in [a,b], s \in I$, and $u(t) \to 0, v(s) \to 0$ as $t \to x, s \to y$. Let $s = f(t)$. Then we obtain

$$\begin{aligned} h(t) - h(x) = g(f(t)) - g(f(x)) &= [f(t) - f(x)][g'(y) + v(s)] \\ &= (t-x)[f'(x) + u(t)][g'(y) + v(s)] \end{aligned}$$

so that when $t \neq x$,

$$\frac{h(t) - h(x)}{t-x} = [g'(y) + v(s)][f'(x) + u(t)]$$

As $t \to x$, we have $s \to y$; by the continuity of f, the RHS tends to $g'(y)f'(x)$. □

Example 5.2. *Consider the function*

$$f(x) = \begin{cases} x\sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Using the chain rule we have

$$f'(x) = \sin(1/x) + x\cos(1/x)(-1/x^2) = \sin(1/x) - \cos(1/x)/x$$

which does not have limit at $x = 0$, so that $f'(0)$ does not exist.

However, consider the function

$$g(x) = \begin{cases} x^2\sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Using the chain rule we have

$$g'(x) = 2x\sin(1/x) + x^2\cos(1/x)(-1/x^2) = 2x\sin(1/x) - \cos(1/x)$$

which does not have limit at $x = 0$, but by definition it could be shown that $g'(0) = 0$.

5.2 Mean Value Theorems

Theorem 5.4. *Let $f : U \to \mathbb{R}$ be monotone increasing(decreasing), and differentiable at $p \in U$. Then $f'(p) \geq (\leq) 0$.*

Proof. Since f is monotone increasing, we have $f(p+h) \geq f(p)$ when $h > 0$, and $f(p+h) \leq f(p)$ when $h < 0$. Hence, $\frac{f(p+h)-f(p)}{h} \geq 0$ for all $h \neq 0$ and $f'(p) \geq 0$. □

Definition 5.2 (Local Maximum/Minimum). *Let f be a real function defined on a metric space X. We say that f has a **local maximum(minimum)** at a point $p \in X$ if there exists $\delta > 0$ such that for all $q \in X$, we have $d(p,q) < \delta \Rightarrow f(q) \leq (\geq) f(p)$.*

Theorem 5.5. *Let f be defined on $[a,b]$; if f has a local maximum at a point $x \in (a,b)$, and $f'(x)$ exists, then $f'(x) = 0$.*

Proof. Since f has a local maximum at $x \in (a,b)$, choose δ so that $d(x,t) < \delta \Rightarrow f(t) \leq f(x)$, and $a < x - \delta < x < x + \delta < b$. If $x - \delta < t < x$, we have

$$\frac{f(x) - f(t)}{x - t} \geq 0.$$

Letting $t \to x$, we have $f'(x) \geq 0$. If $x < t < x + \delta$, we have

$$\frac{f(x) - f(t)}{x - t} \leq 0.$$

Letting $t \to x$, we have $f'(x) \leq 0$. Hence, $f'(x) = 0$. □

Theorem 5.6 (Rolle's Theorem). *If $f : [a,b] \to \mathbb{R}$ is continuous and differentiable on (a,b), and $f(a) = f(b)$, then there is a point $x \in (a,b)$ such that $f'(x) = 0$.*

Proof. Since $[a,b]$ is compact, there exist $x \in [a,b]$ at which f attains its maximum, and also one of its local maximum; and $y \in [a,b]$ at which f attains its minimum, and also one of its local minimum. If x or $y \in (a,b)$, by Theorem 5.5 we have $f'(x) = 0$ or $f'(y) = 0$.

If x, y are both endpoints, we have $f(a) = f(b) = \sup_{x \in [a,b]} f(x) = \inf_{x \in [a,b]} f(x)$, so that f is constant and $f'(x) = 0$ for all $x \in (a,b)$. □

Corollary 5.1 (Mean Value Theorem). *If f is a real continuous real function on $[a,b]$ which is differentiable on (a,b), then there is a point $x \in (a,b)$ such that*

$$f(b) - f(a) = (b-a)f'(x).$$

Proof. Let $g(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x-a)$, then it follows from applying Rolle's theorem to g. □

Corollary 5.2. *Suppose f is differentiable on (a,b), then*

1. *f is monotone increasing if and only if $f'(x) \geq 0$ for all $x \in (a,b)$.*
2. *f is constant if and only if $f'(x) = 0$ for all $x \in (a,b)$.*
3. *f is monotone decreasing if and only if $f'(x) \leq 0$ for all $x \in (a,b)$.*

5.3 Lipschitz Function

Definition 5.3 (Lipschitz Function). *A function $f : X \to Y$ is **Lipschitz** if there exists $M \geq 0$ such that for all $x, y \in X$, we have*

$$d_Y(f(x), f(y)) \leq M d_X(x,y).$$

*We call M the **Lipschitz constant** for f.*

Note that Lipschitz function is uniformly continuous: for all $\varepsilon > 0$, take $\delta = \varepsilon/M$ so that $d_X(x,y) < \delta \Rightarrow d_Y(f(x), f(y)) < \delta M = \varepsilon$.

Theorem 5.7. *Suppose $f : U \to \mathbb{R}$ is differentiable on a connected set U. If there exists $M \geq 0$ such that $|f'(x)| \leq M$ for all $x \in U$, then f is a Lipschitz function with constant M.*

Note that the converse does not hold; a counter-example is $f(x) = |x|$, which is Lipschitz with constant 1, but not differentiable at $x = 0$.

Proof. For all $a, b \in U$ (suppose $a < b$), since $-M \leq f'(x) \leq M$ for all $x \in (a,b)$, by MVT we have

$$\frac{f(b) - f(a)}{b - a} \leq M \Rightarrow f(b) - f(a) \leq M(b-a)$$

and

$$\frac{f(b)-f(a)}{b-a} \geq -M \Rightarrow f(b)-f(a) \geq -M(b-a).$$

Hence, $|f(b)-f(a)| \leq M|b-a|$ and f is Lipschitz with constant M. □

Theorem 5.8. *If f is differentiable at p, then there exists $M \geq 0, r > 0$ such that*

$$|f(x)-f(p)| \leq M|x-p|, \quad \forall x \in B_r(p).$$

We can say that f is 'locally Lipschitz' at p.

Note that the converse does not hold; the counter-example still can be $f(x) = |x|$ at $x = 0$.

Proof. Since f is differentiable at p, $\frac{f(x)-f(p)}{x-p} \to f'(p)$ as $x \to p$. Take $r > 0$ such that

$$|x-p| < r \Rightarrow \left|\frac{f(x)-f(p)}{x-p} - f'(p)\right| < 1.$$

Then we have

$$\frac{|f(x)-f(p)|}{|x-p|} \leq |f'(p)| + 1.$$

Hence,

$$|f(x)-f(p)| \leq (|f'(p)|+1)|x-p|$$

and the Lipschitz constant is given by $M = |f'(p)| + 1$. □

Definition 5.4 (Best Linear Approximation)**.** *$a + b(x-p)$ is called the **best linear approximation** of $f(x)$ at p if*

$$\lim_{x \to p} \frac{f(x) - [a + b(x-p)]}{x-p} = 0.$$

Equivalently,

$$f(x) = a + b(x-p) + o(x-p).$$

Theorem 5.9. *Suppose $f : U \to \mathbb{R}$ is continuous and $p \in U$. Then $a + b(x-p)$ is the best linear approximation of f at p if and only if f is differentiable at p, and*

$$a = f(p), \quad b = f'(p).$$

5.4 Higher Derivatives and Taylor's Theorem

Definition 5.5 (Higher Derivatives)**.**

$$f^{(n)}(x) = \frac{d}{dx} f^{(n-1)}(x) = \frac{d^n}{dx^n} f(x).$$

Definition 5.6 (Taylor Polynomial)**.** *Suppose $f : U \to \mathbb{R}$ is n-times differentiable at $p \in U$, then the n-degree **Taylor polynomial** for f at p is a degree-n polynomial*

$$T_p^n(f)(x) = f(p) + f'(p)(x-p) + \ldots + \frac{f^{(n)}(p)}{n!}(x-p)^n.$$

Lemma 5.1. *Suppose $f : [a,b] \to \mathbb{R}$ is n-times differentiable on $[a,b]$, and $f^{(i)}$ is continuous on $[a,b]$ for all $i = 0, 1, \ldots, n-1$. Suppose $f(b) = 0$, $f(a) = f'(a) = \ldots = f^{(n-1)}(a) = 0$. Then there exists $c \in (a,b)$ such that $f^{(n)}(c) = 0$.*

Proof. It could be shown by iterating Rolle's theorem. By $f(a) = f(b) = 0$, there is some $c_1 \in (a,b)$ such that $f'(c_1) = 0$; by $f'(a) = f'(c_1) = 0$, there is some $c_2 \in (a, c_1)$ such that $f''(c_2) = 0$; ... By iterating this process, we finally find some $c_n \in (a, c_{n-1}) \subset (a,b)$ such that $f^{(n)}(c_n) = 0$. □

Theorem 5.10 (Taylor's Theorem)**.** *Suppose $f : [a,b] \to \mathbb{R}$ is n-times differentiable on $[a,b]$, and suppose $a < \alpha < \beta < b$. Then there exists $z \in (\alpha, \beta)$ such that*

$$f(\beta) = T_\alpha^{n-1}(f)(\beta) + \frac{f^{(n)}(z)}{n!}(\beta - \alpha)^n.$$

Proof. Let

$$g(x) = f(x) - T_\alpha^{(n-1)}(f)(x) - M(x-\alpha)^n.$$

Then $g^{(i)}$ is continuous on $[\alpha, \beta]$ for all $i = 0, 1, ..., n-1$, and $g(\alpha) = g'(\alpha) = ... = g^{(n-1)}(\alpha) = 0$, and $g(\beta) = 0$ for some M.

To find M, by the lemma above, there exists $z \in (\alpha, \beta)$ such that

$$g^{(n)}(z) = f^{(n)}(z) - Mn! = 0.$$

Hence, we have $M = f^{(n)}(z)/n!$ and therefore

$$g(\beta) = f(\beta) - T_\alpha^{(n-1)}(f)(\beta) - \frac{f^{(n)}(z)}{n!}(x-\alpha)^n = 0.$$

□

Corollary 5.3.

1. If $f'' \geq 0$ on $[a,b]$, then $f(x) \geq f(p) + f'(p)(x-p)$ for all $x, p \in [a,b]$;
2. If $|f^{(n+1)}| \leq C$ on $[a,b]$, then $|f(x) - T_p^n(f)(x)| \leq C|x-p|^{n+1}$.

Theorem 5.11. *Suppose $f : U \to \mathbb{R}$ is n-times differentiable at $p \in U$, and $Q : \mathbb{R} \to \mathbb{R}$ is an n-degree polynomial. Then*

$$\lim_{x \to p} \left| \frac{f(x) - Q(x)}{(x-p)^n} \right| = 0$$

if and only if $Q = T_p^n(f)$. Therefore,

$$f(x) = f(p) + f'(p)(x-p) + \frac{f''(p)}{2}(x-p)^2 + ... + \frac{f^{(n)}(p)}{n!}(x-p)^n + o(|x-p|^n).$$

Proof. Suppose $Q = T_p^n(f)$. By L'Hospital's rule we have

$$\lim_{x \to p} \frac{f(x) - T_p^n(f)(x)}{(x-p)^n} = \lim_{x \to p} \frac{f'(x) - T_p^{n-1}(f')(x)}{n(x-p)^{n-1}} = ... = \lim_{x \to p} \frac{f^{(n-1)}(x) - f^{(n-1)}(p) - f^{(n)}(p)(x-p)}{n!(x-p)} = 0$$

since $f^{(n-1)}$ is differentiable at p. Note that

$$\frac{f(x) - Q(x)}{(x-p)^n} = \frac{f(x) - T_p^n(f)(x)}{(x-p)^n} + \frac{T_p^n(f)(x) - Q(x)}{(x-p)^n}.$$

We have shown that the first term on RHS converges to zero. So that it converges to zero if and only if the second term on RHS converges to zero, and

$$\frac{T_p^n(f)(x) - Q(x)}{(x-p)^n} = \frac{a_0 + a_1(x-p) + ... + a_n(x-p)^n}{(x-p)^n} \to 0 (x \to p)$$

if and only if $a_0 = a_1 = ... = a_n = 0$. Therefore, $Q = T_p^n(f)$. □

Corollary 5.4 (L'Hospital's Rule)**.** *If f, g are 2-times differentiable and $f(p) = g(p) = 0$, $g'(x) \neq 0$ for all x close to p, then*

$$\lim_{x \to p} \frac{f(x)}{g(x)} = \lim_{x \to p} \frac{f'(p)}{g'(p)}.$$

Proof. By theorem above we have

$$\lim_{x\to p}\frac{f(x)}{g(x)}=\lim_{x\to p}\frac{f(p)+f'(p)(x-p)+o(|x-p|)}{g(p)+g'(p)(x-p)+o(|x-p|)}=\lim_{x\to p}\frac{f'(p)(x-p)+o(|x-p|)}{g'(p)(x-p)+o(|x-p|)}=\lim_{x\to p}\frac{f'(p)}{g'(p)}.$$

□

Definition 5.7 (Taylor Series)**.** *The* ***Taylor series*** *for infinite-time differentiable function f at p is the power series*

$$T_p(f)(x)=\sum_{n=0}^{\infty}\frac{f^{(n)}(p)}{n!}(x-p)^n.$$

Definition 5.8 (Analytic Function)**.** *For infinite-time differentiable function f and point p, if there exists $R>0$ such that*

$$T_p(f)(x)=f(x),\quad \forall x\in B_R(p),$$

then we say that f is ***analytic*** *at p.*

By the definition, an analytic function is infinite-time differentiable, but the converse does not hold.

Example 5.3. *Consider*

$$f(x)=\begin{cases}e^{-1/x} & x>0\\ 0 & x\le 0\end{cases}$$

It could be shown that $f^{(n)}(0)=0$ for all $n=0,1,\ldots$ (so that f is infinite-time differentiable), and the Taylor series of f at 0 is $T_0(f)=0$. However, $f(x)\neq 0$ for any $x>0$, so that f is not analytic at 0.

5.5 Vector-Valued Function

A vector-valued function $f:\mathbb{R}\to\mathbb{R}^n$ can be represented by functions mapping to each component: $f(x)=(f_1(x),f_2(x),\ldots,f_n(x))'$. Then clearly $f'(x)=(f_1'(x),f_2'(x),\ldots,f_n'(x))'$ (if exists).

The MVT does not hold for vector-valued function; a counter-example is to consider $f(x)=(\cos x,\sin x)'$, we have $f(0)=f(2\pi)=0$ but $f'(x)\neq 0$ for all $x\in\mathbb{R}$.

The relationship between bounded derivative and Lipschitz property still holds.

Theorem 5.12. *Suppose $f:[a,b]\to\mathbb{R}^n$ is continuously differentiable on $[a,b]$, and $\|f'\|\le M$, then*

$$\|f(b)-f(a)\|\le M|b-a|.$$

Proof. We cannot use MVT directly to f because it does not hold for vector-valued functions. To apply MVT, let $v\in\mathbb{R}^n$ and consider $v\cdot f:[a,b]\to\mathbb{R}$; we have $(v\cdot f)'=v\cdot f'$. Then by MVT we have

$$\left|\frac{v\cdot(f(b)-f(a))}{b-a}\right|\le\sup_{x\in[a,b]}(v\cdot f')(x)\le\|v\|M$$

so that

$$\frac{v}{\|v\|}(f(b)-f(a))\le M(b-a).$$

Take $v=f(b)-f(a)$ and we complete the proof. □

The relationship between derivative and best linear approximation still holds. Indeed, if we extend the domain so that $f:\mathbb{R}^n\to\mathbb{R}^m$, we can define the derivative f' by the best linear approximation: For matrix $A\in\mathbb{R}^{m\times n}$, we call $A=f'(x)$ if

$$f(x)-f(p)-A(x-p)=o(\|x-p\|).$$

6 Integrals

6.1 Darboux Integral and Riemann-Stieltjes Integral

Definition 6.1 (Partition). *A **partition** of an interval $[a,b]$ is a finite subset $P \subseteq [a,b]$ with $a,b \in P$. We write*

$$P = \{a = x_0 \leq x_1 \leq x_2 \leq ... \leq x_n = b\}.$$

*We denote $\Delta_i = x_i - x_{i-1}$ so that $b - a = \Delta_1 + \Delta_2 + ... + \Delta_n$. For $t_i \in [x_{i-1}, x_i], i = 1, 2, ...,$ we say that P with $\{t_i\}_{i=1}^n$ is a **tagged partition** of $[a,b]$.*

Given function $f : [a,b] \to \mathbb{R}$ and a partition $P = \{x_0, x_1, ..., x_n\}$, now we denote

$$m_i = \inf_{x\in[x_{i-1},x_i]} f(x), \quad M_i = \sup_{x\in[x_{i-1},x_i]} f(x).$$

So that for all tags $\{t_i\}_{i=1}^n$ we have

$$\sum_{i=1}^n m_i\Delta_i \leq \sum_{i=1}^n f(t_i)\Delta_i \leq \sum_{i=1}^n M_i\Delta_i$$

Definition 6.2 (Darboux Integral). *For $f : [a,b] \to \mathbb{R}$ and partition of $[a,b]$, the **upper/lower Darboux sums** are defined as*

$$U(f,p) = \sum_{i=1}^n M_i\Delta_i, \quad L(f,p) = \sum_{i=1}^n m_i\Delta_i.$$

*The **upper/lower Darboux integrals** are defined as*

$$\overline{\int_a^b} f = \inf_P U(f,p), \quad \underline{\int_a^b} f = \sup_P L(f,p).$$

If the upper and lower Darboux integral are equal (to C), we say that f is integrable (written $f \in \mathscr{R}$) and denote $\int_a^b f = C$.

Definition 6.3 (Riemann-Stieltjes Integral). *Given a monotone function $\alpha : [a,b] \to \mathbb{R}$, for function $f : [a,b] \to \mathbb{R}$ and partition P, denote $\alpha_i = \alpha(x_i), \forall x_i \in P$ and define*

$$U(f,p,\alpha) = \sum_{i=1}^n M_i(\alpha_i - \alpha_{i-1}), \quad L(f,p,\alpha) = \sum_{i=1}^n m_i(\alpha_i - \alpha_{i-1}).$$

*Then the **upper/lower Stieltjes integrals** are defined as*

$$\overline{\int_a^b} f d\alpha = \inf_P U(f,p,\alpha), \quad \underline{\int_a^b} f d\alpha = \sup_P L(f,p,\alpha).$$

*If the upper and lower Stieltjes integral are equal (to C), we say that $f \in \mathscr{R}(\alpha)$ and call $\int_a^b f = C$ the **Riemann-Stieltjes integral**.*

Proposition 6.1. *Suppose $f : [a,b] \to \mathbb{R}$ is bounded, α is increasing, P, Q are two partitions of $[a,b]$ such that $P \subseteq Q$ (Q is 'finer' than P), then we have*

$$U(f,P,\alpha) \geq U(f,Q,\alpha), \quad L(f,P,\alpha) \leq L(f,Q,\alpha).$$

Proof. Essentially, it is because $A \subseteq B$ can imply $\inf B \leq \inf A$ and $\sup B \geq \sup A$. Then for any $x_{i-2} < x_{i-1} < x_i$ in the interval $[a,b]$, we have

$$\inf_{x\in[x_{i-2},x_i]} f(x) \leq \inf_{x\in[x_{i-2},x_{i-1}]} f(x), \inf_{x\in[x_{i-1},x_i]} f(x);$$

$$\sup_{x\in[x_{i-2},x_i]} f(x) \geq \sup_{x\in[x_{i-2},x_{i-1}]} f(x), \sup_{x\in[x_{i-1},x_i]} f(x).$$

□

Corollary 6.1. *Suppose $f : [a,b] \to \mathbb{R}$ is bounded, α is increasing, P, Q are any two partitions of $[a,b]$, then*

$$L(f,P,\alpha) \leq U(f,Q,\alpha).$$

Proof. If $P = Q$, then it is trivial; if not, we can use their common refinement $P \cup Q$ so that

$$L(f,P,\alpha) \leq L(f,P\cup Q,\alpha) \leq U(f,P\cup Q,\alpha) \leq U(f,Q,\alpha).$$

□

Proposition 6.2. *Suppose $f : [a,b] \to \mathbb{R}$ is bounded, α is increasing, then*

1. $\underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha$.
2. *$f \in \mathscr{R}(\alpha)$ if and only if for all $\varepsilon > 0$, there exists a partition P such that*

$$U(f,P,\alpha) - L(f,P,\alpha) = \sum_{i=1}^{n} (M_i - m_i)(\alpha_i - \alpha_{i-1}) < \varepsilon.$$

Proof.

1. By Corollary 6.1 we know that $L(f,P,\alpha) \leq U(f,Q,\alpha)$ for all partitions P, Q. Hence,

$$\underline{\int_a^b} f d\alpha = \sup_P L(f,P,\alpha) \leq \inf_Q U(f,Q,\alpha) = \overline{\int_a^b} f d\alpha.$$

2. Denote $U = \{U(f,P,\alpha)\}, L = \{L(f,P,\alpha)\} \subseteq \mathbb{R}$; by Corollary 6.1 we have $l \leq u$ for any $l \in L, u \in U$. Hence, $\inf U(f,P,\alpha) \geq \sup L(f,P,\alpha)$.
   If this inequality hold, there must be $\inf_P U(f,P,\alpha) = \sup_P L(f,P,\alpha)$; if not, take $\varepsilon < \inf_P U(f,P,\alpha) - \sup_P L(f,P,\alpha)$ and then there is contradiction. Hence, $\inf_P U(f,P,\alpha) = \sup_P L(f,P,\alpha)$ and $f \in \mathscr{R}(\alpha)$.
   If $f \in \mathscr{R}$, we have $\int f d\alpha = \inf_P U(f,P,\alpha) = \sup_P L(f,P,\alpha)$. For all $\varepsilon > 0$, take partition P such that $U(f,P,\alpha) < \int f d\alpha + \varepsilon/2$, and partition Q such that $L(f,Q,\alpha) > \int f d\alpha - \varepsilon/2$. Then consider their common partition $P \cup Q$: by Proposition 6.1 we have

$$\begin{aligned} U(f,P\cup Q,\alpha) - L(f,P\cup Q,\alpha) &= \left|U(f,P\cup Q,\alpha) - \int f d\alpha\right| + \left|L(f,P\cup Q,\alpha) - \int f d\alpha\right| \\ &< \left|U(f,P,\alpha) - \int f d\alpha\right| + \left|L(f,Q,\alpha) - \int f d\alpha\right| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

□

Theorem 6.1. *Suppose $f : [a,b] \to \mathbb{R}$ is continuous, then $f \in \mathscr{R}(\alpha)$.*

Proof. Let $L = \alpha(b) - \alpha(a)$. Since f is uniformly continuous (note that $[a,b] \subseteq \mathbb{R}$ is compact), for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - y| < \delta \quad \Rightarrow \quad |f(x) - f(y)| < \frac{\varepsilon}{L}.$$

Then we consider a partition P such that each $\Delta_i < \delta$, so that

$$U(f,P,\alpha) - L(f,P,\alpha) \leq \sum_{i=1}^{n} \frac{\varepsilon}{L}(\alpha_i - \alpha_{i-1}) = \frac{\varepsilon}{L} L = \varepsilon$$

and $f \in \mathscr{R}(\alpha)$. □

Example 6.1. *The topologist's sine curve*

$$f(x) = \begin{cases} \sin(1/x) & x > 0 \\ 0 & x = 0 \end{cases}$$

is integrable on $[0,1]$ *even though it is not continuous at 0. To show this, first note that it is integrable on* $[\varepsilon, 1]$ *for all* $\varepsilon \in (0,1]$ *(due to the continuity). By Proposition 6.2(2), take partition* $P = \{0, \varepsilon, x_2, ..., x_n = 1\}$ *such that*

$$U(f, \{\varepsilon, x_2, ..., x_n\}) - \int_\varepsilon^1 f \le \varepsilon,$$

$$\int_\varepsilon^1 f - L(f, \{\varepsilon, x_2, ..., x_n\}) \le \varepsilon.$$

Note that the additional Riemann sum on interval $[0, \varepsilon]$ *can be between* $-\varepsilon$ *and* ε*. Hence,*

$$U(f,P) \le \int_\varepsilon^1 f + 2\varepsilon, \quad L(f,P) \ge \int_\varepsilon^1 f - 2\varepsilon$$

so that

$$U(f,P) - L(f,P) \le 4\varepsilon.$$

Since $\varepsilon > 0$ *is arbitrary,* f *is integrable on* $[0,1]$*.*

6.2 Properties of Integral

Proposition 6.3. $\int_a^b f d\alpha$ *is linear in* f *and linear in* α*. That is,*

1. *If* $f, g \in \mathscr{R}(\alpha)$ *and* $\lambda \in \mathbb{R}$*, then* $f + \lambda g \in \mathscr{R}(\alpha)$*.*
2. *If* $f \in \mathscr{R}(\alpha), \mathscr{R}(\beta)$ *and* $\lambda > 0$*, then* $f \in \mathscr{R}(\alpha + \lambda\beta)$*.*
3. $\int_a^c f d\alpha = \int_a^b f d\alpha + \int_b^c f d\alpha$ *(if any two exist).*

Theorem 6.2. *Let* $g : [a,b] \to \mathbb{R}$ *be bounded,* $f : \mathbb{R} \to \mathbb{R}$ *be continuous, and* α *be increasing. If* $g \in \mathscr{R}(\alpha)$*, then* $f \circ g \in \mathscr{R}(\alpha)$*.*

Proof. Suppose f is uniformly continuous (since g is bounded). Take ε, δ such that

$$\sup_{x,y \in [x_{i-1}, x_i]} g(x) - g(y) < \delta \quad \Rightarrow \quad \sup_{x,y \in [x_{i-1}, x_i]} f \circ g(x) - f \circ g(y) < \varepsilon.$$

Note that

$$U(g,P,\alpha) - L(g,P,\alpha) \ge \sum_{i: M_i - m_i > \delta} \delta(\alpha_i - \alpha_{i-1})$$

so that (where $K = \sup_{x \in [a,b]} f \circ g(x) - \inf_{x \in [a,b]} f \circ g(x)$)

$$\begin{aligned} U(f \circ g, P, \alpha) - L(f \circ g, P, \alpha) &\le \varepsilon[\alpha(b) - \alpha(a)] + \sum_{i: M_i - m_i > \delta} K(\alpha_i - \alpha_{i-1}) \\ &\le \varepsilon[\alpha(b) - \alpha(a)] + \frac{K}{\delta}[U(g,P,\alpha) - L(g,P,\alpha)] \end{aligned}$$

holds for all partitions P. Hence, take P to make the second term on RHS sufficiently small, such that

$$U(f \circ g, P, \alpha) - L(f \circ g, P, \alpha) \le \varepsilon[\alpha(b) - \alpha(a) + 1]$$

and we complete the proof. □

Corollary 6.2. *If* $f, g \in \mathscr{R}(\alpha)$*, then* $f^2, g^2, f \pm g, fg \in \mathscr{R}(\alpha)$*. (Note that* $fg = (1/4)[(f+g)^2 - (f-g)^2]$*.)*

Theorem 6.3 (Monotonicity of Integral). *If α is increasing, $f, g \in \mathscr{R}(\alpha)$ and $f \leq g$, then*

$$\int f d\alpha \leq \int g d\alpha.$$

Proof. Essentially, it is because for any interval of any partition,

$$\sup_{x \in [x_{i-1}, x_i]} f(x) \leq \sup_{x \in [x_{i-1}, x_i]} g(x), \qquad \inf_{x \in [x_{i-1}, x_i]} f(x) \leq \inf_{x \in [x_{i-1}, x_i]} g(x).$$

□

Proposition 6.4. *Suppose $f, g \in \mathscr{R}(\alpha)$ where α is increasing, then*

1. $|\int f d\alpha| \leq \int |f| d\alpha$;
2. $\int fg d\alpha \leq (\int f^2 d\alpha \int g^2 d\alpha)^{1/2}$.

Proof.

1. Since $-|f| \leq f \leq |f|$, we have $-\int |f| d\alpha \leq \int f d\alpha \leq \int |f| d\alpha$. Note that $|f|$ is integrable because $|\cdot|$ is continuous.
2. Note that $fg \in \mathscr{R}(\alpha)$. By Cauchy-Schwarz we have

$$fg = (\lambda f)(g/\lambda) \leq \frac{\lambda^2 f^2}{2} + \frac{g^2}{2\lambda^2}, \quad \forall \lambda > 0$$

so that

$$\int fg d\alpha \leq \inf_{\lambda > 0} \left[\frac{\lambda^2}{2} \int f^2 d\alpha + \frac{1}{2\lambda^2} \int g^2 d\alpha \right] = \sqrt{\int f^2 d\alpha \int g^2 d\alpha}$$

□

Theorem 6.4 (Change of Variables). *Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is continuous and strictly increasing, $f, \alpha : [\phi(a), \phi(b)] \to \mathbb{R}$, $f \in \mathscr{R}(\alpha)$. Suppose $g = f \circ \phi, \beta = \alpha \circ \phi$, then $g \in \mathscr{R}(\beta)$ and*

$$\int_a^b g d\beta = \int_{\phi(a)}^{\phi(b)} f d\alpha.$$

Proof. For any partition $P = \{a, x_1, ..., x_{n-1}, b\}$ of $[a, b]$, define $\phi(P) = \{\phi(a), \phi(x_1), ..., \phi(x_{n-1}), \phi(b)\}$ as a partition of $[\phi(a), \phi(b)]$. ϕ is an one-to-one mapping from partitions of $[a, b]$ to partitions of $[\phi(a), \phi(b)]$, since ϕ is continuous and strictly increasing. Hence,

$$U(f, \phi(P), \alpha) = U(g, P, \beta), \quad L(f, \phi(P), \alpha) = L(g, P, \beta).$$

□

Theorem 6.5. *Suppose $f, \alpha : [a, b] \to \mathbb{R}$, $\alpha' : (a, b) \to \mathbb{R}$ is continuous. Then $f \in \mathscr{R}(\alpha)$ if and only if $f\alpha' \in \mathscr{R}$ and*

$$\int f d\alpha = \int f \alpha'.$$

Proof. For all $\varepsilon > 0$, take partition $P = \{x_0, x_1, ..., x_n\}$ such that $U(f, P, \alpha') - L(f, P, \alpha') < \varepsilon$. The MVT implies that for each interval $[x_{i-1}, x_i]$, there is some $t_i \in [x_{i-1}, x_i]$ such that

$$\alpha_i - \alpha_{i-1} = \alpha'(t_i) \Delta_i.$$

Also for all $s_i \in [x_{i-1}, x_i]$, we have

$$\sum_{i=1}^n |\alpha'(s_i) - \alpha'(t_i)| \Delta_i < \varepsilon.$$

Let $M = \sup |f(x)|$, then by inequality above,

$$\left|\sum_{i=1}^{n} f(s_i)(\alpha_i - \alpha_{i-1}) - \sum_{i=1}^{n} f(s_i)\alpha'(s_i)\Delta_i\right| \leq M\varepsilon.$$

On one hand, taking supremum over s_i on the second term on LHS gives

$$\sum_{i=1}^{n} f(s_i)(\alpha_i - \alpha_{i-1}) \leq U(f\alpha', P) + M\varepsilon.$$

Then by taking supremum over s_i on LHS, we have

$$U(f, P, \alpha) \leq U(f\alpha', P) + M\varepsilon.$$

On the other hand, taking supremum on the first term on LHS and then the other gives

$$U(f\alpha', P) \leq U(f, P, \alpha) + M\varepsilon.$$

Thus,

$$|U(f, P, \alpha) - U(f\alpha', P)| < M\varepsilon.$$

Since it also holds for any refinement of P, we conclude that

$$\left|\overline{\int_a^b} f d\alpha - \overline{\int_a^b} f\alpha'\right| < M\varepsilon.$$

Note that $\varepsilon > 0$ is arbitrary. Hence

$$\overline{\int_a^b} f d\alpha = \overline{\int_a^b} f\alpha'.$$

The equality of the lower integrals follows from the same way. □

Theorem 6.6.

1. *Let $\alpha(x) = 1_{x>p}$. If f is continuous at p, then $f \in \mathscr{R}(\alpha)$ and $\int f d\alpha = f(p)$.*
2. *Suppose $c_n \geq 0$ for $n = 1, 2, \ldots$ and $\sum c_n$ converges, $\{s_n\}$ is a sequence of distinct points in (a, b), and $\alpha(x) = \sum_{n=1}^{\infty} c_n 1_{x>s_n}$, and f is continuous. Then*

$$\int_a^b f d\alpha = \sum_{n=1}^{\infty} c_n f(s_n).$$

Proof.

1. Consider the partiton $P = \{x_0, x_1, x_2, x_3\}$ where $a = x_0 < x_1 = p < x_2 < x_3 = b$. Then clearly we have

$$U(f, P, \alpha) = \sup_{x \in [x_1, x_2]} f(x), \quad L(f, P, \alpha) = \inf_{x \in [x_1, x_2]} f(x).$$

   Hence, by letting $x_2 \to x_1 = p$ we have $U(f, P, \alpha), L(f, P, \alpha) \to f(p)$.
2. For all $\varepsilon > 0$, take N such that $\sum_{n=N+1}^{\infty} c_n < \varepsilon$. Decompose α into $\alpha = \alpha_1 + \alpha_2$, where

$$\alpha_1(x) = \sum_{n=1}^{N} c_n 1_{x>s_n}, \quad \alpha_2(x) = \sum_{n=N+1}^{\infty} c_n 1_{x>s_n}.$$

   Then by (1) we have

$$\int_a^b f d\alpha_1 = \sum_{n=1}^{N} c_n f(s_n), \quad \left|\int_a^b f d\alpha_2\right| \leq M\varepsilon$$

where $M = \sup |f(x)|$. Hence,

$$\left| \int_a^b f d\alpha - \sum_{n=1}^{N} c_n f(s_n) \right| \leq M\varepsilon.$$

By letting $N \to \infty$ we complete the proof.

□

6.3 Bounded Variation

Definition 6.4 (Total Variation)**.** *For function $f : [a, b] \to \mathbb{R}$ and partition P of $[a, b]$, the **variation** of f over P is*

$$V(f, P) = \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|.$$

*The **total variation** of f is $TV(f) = \sup_P V(f, P)$. If $TV(f)$ is finite, f has **bounded variation** (f is BV).*

Theorem 6.7 (Jordan Decomposition)**.** *A function f is BV if and only if it can be decomposed into the sum of increasing and decreasing functions.*

Proof. First note that if $f : [a, b] \to \mathbb{R}$ is monotone, then clearly $TV(f) = |f(a) - f(b)|$. Hence, if $f = f_1 + ... + f_n$ where $f_1, ..., f_n$ are monotone (and BV), we have $TV(f) \leq TV(f_1) + ... + TV(f_n)$ so that f is BV.

Suppose f is BV, consider $f_x = f|_{[a,x]}, x \in [a, b]$; then $TV(f_x)$ is increasing in x. Then given $\alpha < \beta$, choose any partition P and we have

$$f(\beta) - f(\alpha) = \sum [f(x_i) - f(x_{i-1})] \leq \sum |f(x_i) - f(x_{i-1})| = TV(f_\beta) - TV(f_\alpha).$$

Rearranging terms gives

$$f(\beta) - TV(f_\beta) \leq f(\alpha) - TV(f_\alpha)$$

so that $f(x) - TV(f_x)$ is decreasing. Hence, f can be decomposed into

$$f(x) = TV(f_x) + [f(x) - TV(f_x)]$$

in which the first term on RHS is increasing, and the second term is decreasing. □

Proposition 6.5. *Suppose α is BV, $f \in \mathscr{R}(\alpha)$. Then for all $\varepsilon > 0$, there exist some tagged partition P such that*

$$\left| \int f d\alpha - \sum f(t_i)(\alpha_i - \alpha_{i-1}) \right| < \varepsilon.$$

Proof. By Jordan decomposition, $\alpha = \alpha^+ - \alpha^-$ which are both increasing. By linearity of integral we have $\int f d\alpha = \int f d\alpha^+ + \int f d\alpha^-$ and $\mathscr{R}(\alpha) = \mathscr{R}(\alpha^+) \cap \mathscr{R}(\alpha^-)$. Hence,

$$\left| \int f d\alpha - \sum f(t_i)(\alpha_i - \alpha_{i-1}) \right| \leq \left| \int f d\alpha^+ - \sum f(t_i)(\alpha_i^+ - \alpha_{i-1}^+) \right| + \left| \int f d\alpha^- - \sum f(t_i)(\alpha_i^- - \alpha_{i-1}^-) \right|.$$

Since α^+, α^- are increasing, there exists partition P such that both two terms on RHS are smaller than $\varepsilon/3$. □

6.4 Fundamental Theorem of Calculus

Theorem 6.8 (FToC, Version 1)**.** *If $f \in \mathscr{R}$ on $[a, b]$, and let $F(x) = \int_a^x f, x \in [a, b]$, then F is continuous. Further, if f is continuous at $p \in [a, b]$, then $F'(p) = f(p)$.*

Note that the continuity of f is necessary; for example, consider $f(x) = sgn(x)$ and thus $F(x) = |x|$; F is not differentiable at 0 since f is discontinuous at 0.

Proof. Since f is bounded (as $f \in \mathscr{R}$), $|f(x)| \le M$ for some $M > 0$. Then

$$|F(x) - F(y)| = \left| \int_a^x f - \int_a^y f \right| = \left| \int_y^x f \right| \le M|x - y|$$

and F is Lipschitz (thus continuous). Moreover, we can obtain a linear approximation equation

$$F(x) - F(p) = \int_p^x f = \int_p^x f(p) + \int_p^x [f - f(p)] = f(p)(x - p) + E$$

Hence, we only need to show that if f is continuous at p, then $E = o(|x - p|)$.

Since f is continuous, for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - p| < \delta \quad \Rightarrow \quad |f(x) - f(p)| < \varepsilon$$

so that $|E| \le \varepsilon |x - p|$. Since ε can be arbitrarily small, we complete the proof. $\square$

Theorem 6.9 (FToC, Version 2)**.** *If $f \in \mathscr{R}$ and $f' \in \mathscr{R}$ on $[a, b]$, then*

$$f(b) = f(a) + \int_a^b f'.$$

Proof. For any partition P of $[a, b]$, by MVT there exist tags $\{t_i\}$ such that

$$\sum f'(t_i)\Delta_i = \sum f(x_i) - f(x_{i-1}) = f(b) - f(a).$$

Hence, $L(f', P) \le f(b) - f(a) \le U(f', P)$ for any partition P. Since f' is integrable, we must have $\int_a^b f' = f(b) - f(a)$. $\square$

6.5 Integration by Parts

Theorem 6.10 (Integration by Parts)**.** *If $f, g : [a, b] \to \mathbb{R}$ and $f', g' : [a, b] \to \mathbb{R}$ are integrable, then*

$$\int_a^b fg' = fg|_a^b - \int_a^b f'g.$$

Proof. It follows from the product rule and FToC:

$$fg|_a^b = \int_a^b (fg)' = \int_a^b f'g + \int_a^b fg'.$$

$\square$

Theorem 6.11 (Generalized Integration by Parts)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is differentiable and α is BV. If $f \in \mathscr{R}(\alpha)$ and $\alpha f' \in \mathscr{R}$, then*

$$\int_a^b f d\alpha = f\alpha|_a^b - \int_a^b \alpha f'.$$

Proof. If α is increasing, for any partition we have (denote $f_i = f(x_i)$)

$$\begin{aligned} f\alpha|_a^b &= \sum_i (f_i\alpha_i - f_{i-1}\alpha_{i-1}) \\ &= \sum_i f_i(\alpha_i - \alpha_{i-1}) + \sum_i (f_i - f_{i-1})\alpha_{i-1} \\ &= \sum_i f_i(\alpha_i - \alpha_{i-1}) + \sum_i \alpha_{i-1} f'(t_i)\Delta_i \end{aligned}$$

in which the last equality follows from MVT, where $t_i \in [x_{i-1}, x_i]$. By definition of integrals, we have

$$f\alpha|_a^b = \sum_i f_i(\alpha_i - \alpha_{i-1}) + \sum_i \alpha_{i-1} f'(t_i)\Delta_i \to \int_a^b f d\alpha + \int_a^b \alpha f'$$

as partition P becomes finer.

For the general α, decompose it into $\alpha = \alpha^+ - \alpha^-$ and similar result follows from applying the argument to them respectively. □

Indeed, the generalized integration by parts can include the special case of FToC: let $\alpha = 1_{[l,r]}$ where $[l, r] \subseteq [a, b]$; if f is integrable, we have

$$\int_a^b f d\alpha = f\alpha|_a^b - \int_a^b \alpha f' = -\int_l^r f'$$

Meanwhile, we know that $\int_a^b f d\alpha = f(l) - f(r)$, so that

$$\int_l^r f' = -\int_a^b f d\alpha = f(r) - f(l)$$

which exactly refers to the FToC (Version 2; Theorem 6.9).

7 Sequence of Functions

7.1 Pointwise Convergence

Definition 7.1 (Function Space)**.** *For metric spaces X, Y, denote $\mathscr{F}(X,Y)$ the set of all functions $f : X \to Y$. If Y is a vector space, then $\mathscr{F}(X,Y)$ is also a vector space. Denote $\mathscr{F}(X) = \mathscr{F}(X,\mathbb{R})$.*

Definition 7.2 (Pointwise Convergence)**.** *For a sequence (f_n) of functions in $\mathscr{F}(X,Y)$, it* ***converges (pointwisely)*** *to $f : X \to Y$ if $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$. We denote it by $f_n \to f$.*

Note that a sequence (x_n) of (real) numbers can be viewed as a function $f : \mathbb{N} \to \mathbb{R}$, so that $\mathbb{R}^\infty = \mathscr{F}(\mathbb{N},\mathbb{R})$. Here pointwise convergence $a_n \to a$ means $a_n^k \to a^k (n \to \infty)$ for all $k = 0, 1, 2, ...$.

Example 7.1 (Problems of Pointwise Convergence, in $\mathscr{F}(\mathbb{N},\mathbb{R})$)**.**

1. *Let $a_n = (1, ..., 1, 0, 0, ...)$ with first n arguments being 1, others being 0. Then $a_n \to a = (1, 1, ...)$ since $\lim_{n\to\infty} a_n^k = 1$ for all k. However,*

$$\lim_{k\to\infty} a_n^k = 0, \forall n \quad \lim_{k\to\infty} a^k = 1 \quad \Rightarrow \quad \lim_{n\to\infty} \lim_{k\to\infty} a_n^k \neq \lim_{k\to\infty} \lim_{n\to\infty} a_n^k$$

2. *Let $e_n = (0, ..., 0, 1, 0, ...)$ with the n-th argument being 1, others being 0. Then $e_n \to e = 0$ since $\lim_{n\to\infty} e_n^k = 0$ for all k. However,*

$$\sum_k e_n^k = 1, \forall n \quad \sum_k e^k = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \sum_k e_n^k \neq \sum_k \lim_{n\to\infty} e_n^k$$

$$\|e_n\|_p = 1, \forall n \quad \|e\|_p = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \|e_n\|_p \neq \left\| \lim_{n\to\infty} e_n \right\|_p$$

These examples show that pointwise convergence does not preserve limits, sums and ℓ^p norms.

Example 7.2 (Problems of Pointwise Convergence, in $\mathscr{F}(\mathbb{R},\mathbb{R})$)**.**

1. *Let*

$$f_n(x) = \frac{1}{(1+x^2)^n}, n = 1, 2, ...$$

They are all uniformly continuous (continuous, Lipschitz, and differentiable) since $|f_n'| \leq n$. However, this sequence pointwisely converges to

$$f_n(x) \to f(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases}$$

which is discontinuous (not Lipschitz, and not differentiable).

2. *Let*

$$f_n(x) = \frac{1}{n}\sin(nx), n = 1, 2, ...$$

whose graph oscillates between $[-1/n, 1/n]$ with wavelength $2\pi/n$, so that $f_n' \approx (2/n)/(\pi/n) = 2/\pi$ which does not depend on n. However, $f_n \to f = 0$ and $f' = 0$.

3. *Let*

$$f_n(x) = \begin{cases} 2^n & 2^{-n} \leq x \leq 2 \times 2^{-n} \\ 0 & \textit{otherwise} \end{cases}, n = 1, 2, ...$$

so that $\int_0^1 f_n = 1$ for all n. However, $f_n \to f = 0$ and $\int_0^1 f = 0$.

4. *Let*

$$f_n(x) = \begin{cases} 1 & x = p/q \in \mathbb{Q}, q \leq n \\ 0 & \textit{otherwise} \end{cases}, n = 1, 2, ...$$

so that $f_n(x) = 1$ for finitely many $x \in [0,1]$ and $f_n \in \mathscr{R}$ for all n. However, $f_n \to 1_\mathbb{Q} \notin \mathscr{R}$.

These examples show that pointwise convergence does not preserve continuity, differentiablity (and derivatives) and integrability (and integrals).

7.2 Uniform Convergence

Definition 7.3 (Uniform Convergence)**.** *For a sequence (f_n) of functions in $\mathscr{F}(X,Y)$, it **uniformly converges** to f (denoted $f_n \rightrightarrows f$) if for all $\varepsilon > 0$, there exists N such that for all $n \geq N$,*

$$d(f_n(x), f(x)) < \varepsilon, \quad \forall x \in X.$$

Note that the difference from pointwise convergence is that ε does not depend on x.

Proposition 7.1. *$f_n \rightrightarrows f$ if and only if*

$$\sup_{x\in X} d(f_n(x), f(x)) \to 0 \quad (n \to \infty)$$

Similar with Cauchy property, define the 'uniformly Cauchy' property in sequence of functions as follows: for all $\varepsilon > 0$, there exists N such that

$$d(f_n, f_m) = \sup_{x\in X} d_Y(f_n(x), f_m(x)) < \varepsilon, \quad \forall n, m > N.$$

Theorem 7.1 (Cauchy Criterion)**.** *Suppose Y is Cauchy complete, (f_n) is 'uniformly Cauchy' in $\mathscr{F}(X,Y)$. Then $f_n \rightrightarrows f$ for some $f \in \mathscr{F}(X,Y)$.*

Proof. For any fixed $x \in X$, the sequence $(f_n(x))_n$ is Cauchy since $d(f_n(x), f_m(x)) \leq \sup_{x\in X} d(f_n(x), f_m(x))$ for all n, m and $x \in X$. Since Y is Cauchy complete, the limit of $(f_n(x))$ exists and and let $f(x) \equiv \lim_{n\to\infty} f_n(x)$. Now we only need to show that $f_n \rightrightarrows f$.

For all $\varepsilon > 0$, there exists N such that $\sup_{x\in X} d(f_n(x), f_m(x)) < \varepsilon$ for all $n, m > N$. Then

$$d(f_n(x), f(x)) = \lim_{m\to\infty} d(f_n(x), f_m(x)) \leq \varepsilon, \quad \forall x \in X, \forall n > N.$$

Therefore, $f_n \rightrightarrows f$ as $\varepsilon > 0$ is arbitrarily small. $\square$

Corollary 7.1 (Weierstrass M-Test)**.** *Suppose (f_n) is a sequence in $\mathscr{F}(X)$ such that $\sup_{x\in X} |f_n(x)| \leq M_n$ and $\sum_n M_n < \infty$. Then $\sum_n f_n$ converges uniformly.*

Proof. Note that

$$\left| \left(\sum_{k=1}^{m} f_k \right)(x) - \left(\sum_{k=1}^{n} f_k \right)(x) \right| = |f_{n+1}(x) + ... + f_m(x)| \leq M_{n+1} + ... + M_m$$

so that (M_n) is Cauchy implies that (f_n) is 'uniformly Cauchy'. $\square$

Theorem 7.2. *Suppose $E \subseteq X$, $f_n, f : E \to Y$ and $p \in E'$. If $f_n \rightrightarrows f$ and $\lim_{x\to p} f_n(x) = L_n \in Y$ for every n, then (L_n) is Cauchy and*

$$\lim_{n\to\infty} L_n = \lim_{x\to p} f(x)$$

if the limit of (L_n) exists. In other words, we have

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

Proof. To show that (L_n) is Cauchy, note that

$$d(L_n, L_m) \leq d(L_n, f_n(x)) + \sup_{x\in E} d(f_n(x), f_m(x)) + d(L_m, f_n(x)).$$

For the second term on RHS, since (f_n) converges uniformly, for all $\varepsilon > 0$, pick N such that $\sup_{x\in E} d(f_n(x), f_m(x)) < \varepsilon/3$ for all $n, m > N$. For the first and third term on RHS, there exists $x \in X$ such that $d(L_n, f_n(x)), d(L_m, f_m(x)) < \varepsilon/3$. Hence,

$$d(L_n, L_m) < \varepsilon, \quad \forall n, m \geq N$$

and (L_n) is Cauchy.

Denote $\lim_{n\to\infty} L_n = L$. To show that $f(x) \to L$ as $x \to p$, note that

$$d(f(x), L) \le \sup_{x\in E} d(f(x), f_n(x)) + d(f_n(x), L_n) + d(L_n, L), \quad \forall n.$$

Taking $n \to \infty$, the first and third term on RHS converge to zero; then by taking $x \to p$, the second term on RHS converge to zero (note that the first and third term on RHS do not depend on x). Therefore, $f(x) = L = \lim_{n\to\infty} L_n$. $\square$

Corollary 7.2. *Suppose $f_n : X \to Y$ are continuous, $f_n \rightrightarrows f$, then f is also continuous.*

Definition 7.4. *Suppose X, Y are metric spaces and Y is complete. Denote $C^0(X, Y)$ the set of bounded continuous functions $f : X \to Y$ with metric (called 'uniform distance')*

$$d_0(f, g) = \sup_{x\in X} d_Y(f(x), g(x)).$$

Note that

1. Convergence in C^0 is equivalent with uniform convergence.
2. $C^0 \subseteq \mathscr{R}$.
3. C^0 is a complete metric space. The completeness can help to show the existence of some continuous functions.

Example 7.3 (Peano Curve)**.** *Peano curve is a surjective continuous function $f : [0,1] \to [0,1] \times [0,1]$ defined by the limit of sequence (f_n). Details are omitted.*

By construction, (f_n) is Cauchy since

$$|f_n(x) - f_{n-1}(x)| \le 2\sqrt{2} \cdot 3^{-n}$$

so that

$$d_0(f_n, f_m) \le \sum_{i=n+1}^{m} 2\sqrt{2} \cdot 3^{-i} \to 0 \quad (n, m \to \infty).$$

Hence, $f_n \rightrightarrows f$ and f is continuous. It is surjective because (1) its image is closed since the domain is compact; (2) its image is dense in $[0,1] \times [0,1]$. Therefore, its image is exactly $[0,1] \times [0,1]$.

Proposition 7.2. *Integration $C^0([a,b]) \to \mathbb{R}$ is continuous.*

Proof. Notice that $C^0 \subseteq \mathscr{R}$, so we only need to show that $f_n \rightrightarrows f$ (in $C^0([a,b])$) implies $\int f_n \to \int f$. It is because

$$\left|\int f_n - \int f\right| = \left|\int_a^b (f_n - f)\right| \le \sup_{x\in[a,b]} (f_n(x) - f(x))(b - a) \to 0 \quad (n \to \infty).$$

$\square$

Theorem 7.3. *Suppose α is BV, $f_n \rightrightarrows f$ and $f_n \in \mathscr{R}(\alpha)$ for all n. Then $f \in \mathscr{R}(\alpha)$ and $\int f_n d\alpha \to \int f d\alpha$.*

Proof. Without loss of generality, suppose α is monotone increasing. For all $\varepsilon > 0$, there exists n such that $\sup_{x\in[a,b]} |f_n(x) - f(x)| < \varepsilon$. Then take partition P such that $U(f_n, P, \alpha) - L(f_n, P, \alpha) < \varepsilon$ (since $f_n \in \mathscr{R}(\alpha)$). Then we have

$$\sup_{x\in[x_{i-1},x_i]} f(x) < \sup_{x\in[x_{i-1},x_i]} f_n(x) + \varepsilon, \qquad \inf_{x\in[x_{i-1},x_i]} f(x) > \inf_{x\in[x_{i-1},x_i]} f_n(x) - \varepsilon.$$

Hence, ($M_i^n \equiv \sup_{x\in[x_{i-1},x_i]} f_n(x), m_i^n \equiv \inf_{x\in[x_{i-1},x_i]} f_n(x)$)

$$\begin{aligned}
U(f,P,\alpha) - L(f,P,\alpha) &= \sum_i (M_i - m_i)(\alpha_i - \alpha_{i-1}) \\
&\le \sum_i (M_i^n - m_i^n + 2\varepsilon)(\alpha_i - \alpha_{i-1}) \\
&= U(f_n,P,\alpha) - L(f_n,P,\alpha) + 2\varepsilon(\alpha(b) - \alpha(a)) \\
&\le \varepsilon + 2\varepsilon(\alpha(b) - \alpha(a))
\end{aligned}$$

and f is thus integrable. □

Example 7.4 (Integrability of Thomae's Function)**.** *Recall the Thomae's function*

$$f(x) = \begin{cases} 1/q & x = p/q \in \mathbb{Q}, p, q \in \mathbb{Z} \\ 0 & x \notin \mathbb{Q} \end{cases}$$

Indeed, it is the limit of sequence (f_n) *given by*

$$f_n(x) = \begin{cases} 1/q & x = p/q \in \mathbb{Q}, p, q \in \mathbb{Z}, q \le n \\ 0 & x \notin \mathbb{Q} \end{cases}$$

and the convergence is uniform as $\sup_{x\in[0,1]} |f_n(x) - f(x)| = 1/n \to 0 (n \to \infty)$*.*

Since each f_n *is integrable and* $\int f_n = 0$*, the Thomae's function* f *is also integrable and* $\int f = 0$*.*

However, even uniform convergence cannot preserve differentiability.

Example 7.5. *Recall Example 7.2(2); let* $f_n(x) = \frac{1}{n}\sin(nx)$ *and* $f_n'(x) = \cos(nx)$*. It could be shown that* $f_n \rightrightarrows f = 0$*, but* $\{f_n'(x)\}$ *does not converge.*

Lemma 7.1. *Suppose* f_n *are differentiable,* $f_n \to f$ *pointwise, and* $f_n' \rightrightarrows f'$*. Then for all* p*,*

$$\Phi_n(x) = \frac{f_n(x) - f_n(p)}{x - p} \rightrightarrows \Phi(x) = \frac{f(x) - f(p)}{x - p}.$$

Proof. Note that for all x,

$$\Phi_n(x) - \Phi_m(x) = \frac{(f_n - f_m)(x) - (f_n - f_m)(p)}{x - p} = (f_n - f_m)'(\alpha)$$

for some α. Since (f_n') is Cauchy, $(f_n - f_m)'(\alpha) \to 0$ as $n, m \to \infty$, and (Φ_n) is also Cauchy. Then taking $m \to \infty$ gives for all x,

$$|\Phi_n(x) - \Phi(x)| \le \sup_x \lim_{m\to\infty} |\Phi_n(x) - \Phi_m(x)| \to 0 \quad (n \to \infty).$$

□

Theorem 7.4. *Suppose* $f_n : [a,b] \to \mathbb{R}$ *are differentiable,* $f_n \to f$ *pointwise, and* $f_n' \rightrightarrows g$*, then* $f' = g$ *(if* f' *exists).*

Proof. Define Φ_n, Φ as the lemma above; we only need to show that $\Phi(x) \to g(p)$ as $x \to p$ (for all p). Note that

$$|\Phi(x) - g(p)| \le |\Phi(x) - \Phi_n(x)| + |\Phi_n(x) - f_n'(p)| + |f_n'(p) - g(p)|, \quad \forall n, \forall x.$$

Since $\Phi_n(x) \rightrightarrows \Phi(x)$ (by lemma above), the first term on RHS converges to 0 as $n \to \infty$ (regardless of x); since $f_n' \rightrightarrows g$, the third term on RHS also converges to 0 as $n \to \infty$ (regardless of p); the second term on RHS converges to 0 as $x \to p$. □

Finally, the following example shows that the limit of 'uniformly Cauchy' sequence of differentiable functions may be nowhere differentiable.

Example 7.6 (Weierstrass function). *The Weierstrass function can be constructed as follows:*

1. *Pick a bounded, oscillating and periodic function, such as* $T(x) = \sin x$.
2. *Let* $f_n(x) = \alpha^n T(\beta^n x)$, *where* $\alpha < 1, \beta > 1$ *and* $\alpha\beta \geq 1$. *Note that* $\sup|f_n(x)| = \alpha^n \to 0$ *but* $\sup|f_n'(x)| = (\alpha\beta)^n \to \infty$ *(if* $\alpha\beta > 1$*) as* $n \to \infty$.
3. *Take* $f(x) = \sum_{n=1}^{\infty} f_n(x)$. *It converges uniformly (by Weierstrass M-test) but* $\sum_{n=1}^{\infty} f_n'(x)$ *diverges for all* x.

7.3 Equicontinuity

Definition 7.5 (Modulus of Continuity). *A* ***modulus of continuity*** *is a function* $\omega : (0, \infty) \to [0, \infty]$ *such that* $\lim_{x\to 0} \omega(x) = 0$. *Function* f *has modulus of continuity at* p *if there exists* ω_p *such that*

$$d_X(x, p) < \delta \Rightarrow d_Y(f(x), f(p)) < \omega_p(\delta), \quad \forall p > 0.$$

Proposition 7.3. f *is continuous at* p *if and only if it has a modulus of continuity at* p.

By definition of modulus of continuity, it is easy to see that

1. f is uniformly continuous if and only if its modulus of continuity is independent of p; that is, f has a modulus of continuity ω at all p.
2. f is Lipschitz with constant L if and only if it has a modulus of continuity $\omega(\delta) = L\delta$ at all p.

Definition 7.6 (Equicontinuity). *A collection of functions* $F \subseteq \mathscr{F}(X, Y)$ *is* ***equicontinuous*** *if for all* $p \in X$, *there exists* ω_p *such that for all* $f \in F$, ω_p *is a modulus of continuity of* f *at* p.

In other words, for all $\varepsilon > 0$, *there exists* $\delta > 0$ *such that for all* $x, p \in X$, *we have* $d_X(x, p) < \delta \Rightarrow d_Y(f(x), f(p)) < \varepsilon$ *for all* $f \in F$.

Theorem 7.5. *If* $f_n \to f$ *pointwise, and* (f_n) *is equicontinuous, then* f *is continuous.*

Proof. Suppose ω_p is a common modulus of continuity of (f_n) at p. Then if $|x - p| < \delta$, we have

$$|f(x) - f(p)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(p)| + |f_n(p) - f(p)| \leq |f(x) - f_n(x)| + \omega_p(\delta) + |f_n(p) - f(p)|, \quad \forall n.$$

For all $\varepsilon > 0$, since $\omega_p(\delta) \to 0$ as $\delta \to 0$, there exists $\delta_0 > 0$ such that $\omega_p(\delta) < \varepsilon/3$ for all $\delta < \delta_0$; since $f_n \to f$ pointwise, there exists N such that $|f(x) - f_n(x)|, |f(p) - f_n(p)| < \varepsilon/3$ for all $n > N$. Hence,

$$|x - p| < \delta_0 \quad \Rightarrow \quad |f(x) - f(p)| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon$$

and f is continuous. □

7.4 Bounded Continuous Function Space C^0

Consider the compactness in function space C^0.

Example 7.7. *Compact set in* C^0 *must be closed and bounded, but the converse may not hold. To see this, take* $f(x) : [0, +\infty) \to \mathbb{R}$ *such that (1)* $f(0) = 0$, *(2)* $f(1) = 1$ *which is the global maximum of* f, *and (3)* $\lim_{x\to\infty} f(x) = 0$. *Then let* $f_n(x) = f(nx)$ *(so that the global maximum of* f_n *is at* $1/n$ *and* $f_n(1/n) = 1$*).*

Clearly, $f_n \in C^0$ *for all* n, *and* $f_n \to 0$ *pointwise, so that* $(f_n) \subseteq \overline{B_1(0)}$. *However, the convergence is not uniform, and any subsequence of* (f_n) *cannot be uniformly convergent. Therefore,* $\overline{B_1(0)}$ *is not compact. (Recall Corollary 3.2: any sequence of a compact set has convergent subsequence.)*

Proposition 7.4. *If* $f_n \rightrightarrows f$ *in* $C^0(X, Y)$, *then for all* $p \in X$, *there exists a modulus of continuity* ω_p *such that* f *and all* f_n *have* ω_p *at* p. *That is,* $(f_n) \cup \{f\}$ *is equicontinuous.*

Proof. For any given $p \in X$, pick a modulus of continuity $\omega_{p,n}$ at p for each f_n, and ω_p at p for f. Then define

$$\bar{\omega}_p(r) = \sup\{\omega_p(r), \omega_{p,n}(r)\}_{n\in\mathbb{N}}, \quad \forall r.$$

Then for any function $g \in (f_n) \cup \{f\}$, we have

$$d(x, p) < \delta \quad \Rightarrow \quad d(g(x), g(p)) < \bar{\omega}_p(\delta).$$

That is, $\bar{\omega}_p(r)$ is a commonly modulus of continuity for all f_n and f if it can be legitimated.

To show that $\bar{\omega}_p(r)$ is a legitimate modulus of continuity, we only need to show that $\bar{\omega}_p(r) \to 0$ as $r \to 0$. For all $\varepsilon > 0$, since $f_n \rightrightarrows f$, there exists N such that for all $n > N$,

$$d(f_n(x), f_n(p)) < d(f(x), f(p)) + \varepsilon.$$

So by selecting $\omega_{p,n}(r) \le \omega_p(r) + \varepsilon, \forall r$ for all $n > N$, we have

$$\bar{\omega}_p(r) \le \max\{\omega_{p,0}(r), ..., \omega_{p,n}(r), \omega_p(r) + \varepsilon\}$$

so that $\lim_{r\to 0} \bar{\omega}_p(r) \le \varepsilon$. Therefore, $\lim_{r\to 0} \bar{\omega}_p(r) = 0$ as $\varepsilon > 0$ can be arbitrarily small. □

Theorem 7.6. *If $F \subseteq C^0$ is compact, then*

1. *F is uniformly bounded; that is, there exists $M > 0$ such that $|f(x)| \le M$ for all $x \in X$ and all $f \in F$.*
2. *F is equicontinuous.*

Proof. (1) follows from the boundedness of F. To show (2), take a modulus of continuity $\omega_{p,f}$ of each $f \in F$ and $p \in X$. Then define

$$\omega_p(r) = \sup_{f\in F} \omega_{p,f}(r), \quad \forall p, r.$$

Clearly, if $\omega_p(r)$ is a modulus of continuity, it is a common modulus of continuity for all $f \in F$.

Now we only need to show that $\omega_p(r)$ is a legitimate modulus of continuity, that is, $\omega_p(r) \to 0$ as $r \to 0$. If not, then there exists $(f_n) \subseteq F$ such that $\liminf \omega_{p,f_n}(1/n) = \alpha > 0$. However, by proposition above, (f_n) will have no convergent subsequence (CONTINUE), which contradicts the compactness of F. Therefore, $\omega_p(r)$ is a modulus of continuity. □

Theorem 7.7 (Arzela-Ascoli)**.** *Suppose K is compact, $F \subseteq C^0(K, \mathbb{R})$. If*

1. *F is bounded in C^0;*
2. *F is equicontinuous,*

then any sequence in F has a convergent subsequence in C^0. Hence, if F is also closed, the limit of convergent subsequence is in F, so that F is compact.

Proof. Let $(f_k)_{k\in\mathbb{N}}$ be a sequence in F. Since F is equicontinuous, we can find a common modulus of continuity ω_p for F at any $p \in K$. Since K is compact (any open cover of K has finite subcover), for all $k \in \mathbb{N}$, there exists a finite collection $\{p_{i,k}, \delta_{i,k}\}_{i=1}^{n_k}$ such that

$$\omega_{p_{i,k}}(\delta_{i,k}) < \frac{1}{k}, \forall i = 1, ..., n_k \quad \text{and} \quad K \subseteq \bigcup_{i=1}^{n_k} B_{\delta_{i,k}}(p_{i,k}).$$

Take $E_k = \{p_{i,k}\}_{i=1}^{n_k}$. Since E_k is finite and $\cup_{k=1}^{\infty} E_k$ is countable, there exists $A \subseteq \mathbb{N}$ such that $(f_k(p))_{k\in A}$ converges for all $p \in \cup_{k=1}^{\infty} E_k$. (CONTINUE)

Finally, we only need to show that $(f_k)_{k\in A}$ (as a subsequence of $(f_k)_{k\in\mathbb{N}}$) is uniformly convergent (so that it is convergent in C^0). For all $\varepsilon > 0$, for sufficiently large k such that $1/k < \varepsilon$, so that $\omega_p(\delta_p) \le \varepsilon$ for all $p \in E_k$. By construction above, for all $x \in K$ there exists $p \in E$ such that $x \in B_{\delta_p}(p)$. Hence, for sufficiently large $m, n \in A$, we have

$$|f_n(x) - f_m(x)| \le |f_n(p) - f_m(p)| + |f_n(x) - f_n(p)| + |f_m(x) - f_m(p)| \le 3\varepsilon$$

since (1) $x \in B_{\delta_p}(p), \omega_p(\delta_p) \le \varepsilon$ implies

$$|f_n(x) - f_n(p)| \le \omega_p(|x - p|) \le \omega_p(\delta_p) \le \varepsilon$$

and (2) the convergence of $(f_k(p))_{k\in A}$ implies $|f_n(p) - f_m(p)| \le \varepsilon$ for sufficiently large $n, m \in A$.

Therefore, $(f_k)_{k\in A}$ is convergent as $\varepsilon > 0$ can be arbitrarily small. □

Theorem 7.8 (Stone-Weierstrass)**.** *Suppose $f : [a, b] \to \mathbb{R}$ is continuous. Then there exists a sequence of polynomials (P_n) such that $P_n \rightrightarrows f$.*

Hence, the set of all polynomials on $[a, b]$ is dense in $C^0([a, b])$.

Note that (1) comparing to Taylor series, the function f needs not be differentiable; (2) the coefficients of P_n may vary with n.

Definition 7.7 (Algebra)**.** *In function space $\mathscr{F}(X, \mathbb{R})$, an* ***algebra*** *$A \subseteq \mathscr{F}(X, \mathbb{R})$ is a collection of functions such that for all $f, g \in A$ and $\lambda \in \mathbb{R}$, we have $f + g, fg, \lambda f \in A$; moreover, any constant function $C \in A$.*

A collection $F \subseteq \mathscr{F}(X, \mathbb{R})$ of functions ***separates points*** *if for any two distinct points $x, y \in X$, there exists $f \in F$ such that $f(x) \neq f(y)$.*

Lemma 7.2. *For any $N > 0$, there exists a sequence of polynomials (P_n) that uniformly converge to $|x|$ on $[-N, N]$.*

Theorem 7.9 (Generalized Stone-Weierstrass)**.** *Suppose K is compact in a metric space, $A \subseteq C^0(K)$ is an algebra which separate points, then A is dense in $C^0(K)$.*

Proof. Pick two distinct points $x, y \in K$, then there exists $g \in A$ such that $g(x) \neq g(y)$. Let $\bar{g} = \frac{g - g(y)}{g(x) - g(y)} \in A$, so that $\bar{g}(x) = 1, \bar{g}(y) = 0$. Then consider $f\bar{g} + f(1 - \bar{g}) \in A$; we have $f\bar{g} + f(1 - \bar{g}) = f$ at $x, y \in X$.

For any $x \in K$ and $\varepsilon > 0$, for all $y \in K$, there exists $g_y \in A$ such that (1) $g_y(x) = f(x), g_y(y) = f(y)$ (see above); (2) g_y is continuous, so that there exists δ_y such that $|f(z) - g_y(z)| < \varepsilon$ for all $z \in B_{\delta_y}(y)$.

Since K is compact, there exists $\{y_i, \delta_i\}_{i=1}^n$ such that $K \subseteq \cup_{i=1}^n B_{\delta_i}(y_i)$. Consider the corresponding $\{g_{y_i}\}_{i=1}^n$; there exists $\bar{g}_x \in A$ such that $\bar{g}_x \in B_\varepsilon\left(\max_{1 \leq i \leq n} g_{y_i}\right)$, so that $\bar{g}_x(x) \in B_\varepsilon(f(x))$ and $\bar{g}_x(z) > f(z) - 2\varepsilon$ for all $z \in K$.

Moreover, we can find $\{\xi_i\}_{i=1}^m$ such that $f(z) + 2\varepsilon > \min_{1 \leq i \leq m} \bar{g}_{\xi_i}(z)$ for all $z \in K$. Hence, we finally take $g \in A$ such that $g \in B_\varepsilon\left(\min_{1 \leq i \leq m} \bar{g}_{\xi_i}\right)$, so that $g \in B_{3\varepsilon}(f)$ and we complete the proof. $\square$

8 Functions of Several Variables

8.1 Linear Transformations

Definition 8.1 (Vector Space). *A non-empty set $X \subset \mathbb{R}^n$ is a **vector space** if $x + y \in X$ and $cx \in X$ for all $x, y \in X$ and scalars c.*

*If $x_1, ..., x_k \in \mathbb{R}^n$ and $c_1, ..., c_k$ are scalars, the vector $c_1x_1 + ... + c_kx_k$ is called a **linear combination** of $x_1, ..., x_k$. If $S \subset \mathbb{R}^n$ and E is the set of all linear combinations of elements of S, we say that S **spans** E, or that E is the **span** of S, denoted $E = span(S)$.*

*A set consisting of vectors $x_1, ..., x_k$ is **independent** if the relation $c_1x_1 + ... + c_kx_k = 0$ implies $c_1 = ... = c_k = 0$. Otherwise, $\{x_1, ..., x_k\}$ is **dependent**.*

*If a vector space X contains an independent set of r vectors but contains no independent set of $r+1$ vectors, we say that X has **dimension** r, denoted $dim(X) = r$. The set consisting of 0 alone is a vector space. Its dimension is defined to be 0.*

*An independent subset of a vector space X which spans X is called a **basis** of X. If $B = \{x_1, ..., x_r\}$ is a basis of X, then every $x \in X$ has a unique representation of the form $x = \sum_j c_jx_j$. The numbers $c_1, ..., c_r$ are called the **coordinates** of x with respect to the basis B. The **standard basis** of $\mathbb{R}^n$ is $\{e_1, ..., e_n\}$, where e_j is the vector in $\mathbb{R}^n$ whose j-th coordinate is 1 and other coordinates are all 0.*

Theorem 8.1. *Let $r \in \mathbb{N}_+$. If a vector space X is spanned by a set of r vectors, then $dim(X) \le r$.*

Proof. If not, there is a vector space X which contains an independent set $Q = \{y_1, ..., y_{r+1}\}$ and which is spanned by a set S_0 consisting of r vectors.

Now we construct $S_1, ..., S_n$ by induction. Suppose a set S_i $(i = 0, ..., r-1)$ has been constructed which spans X and which consists of $y_1, ...y_i$ and a certain collection of $r - i$ members of S_0, say $x_1, ..., x_{r-i}$. Since S_i spans X, y_{i+1} is in the span of S_i, and there are scalars $a_1, ..., a_{i+1}, b_1, ..., b_{r-i}$ with $a_{i+1} = 1$ such that

$$\sum_{j=1}^{i+1} a_jy_j + \sum_{k=1}^{r-i} b_kx_k = 0.$$

If all b_k's were 0, the independence of $Q = \{y_1, ..., y_{r+1}\}$ implies all a_j's will be 0, a contradiction. Hence, some $x_k \in S_i$ is a linear combination of the other members of $T_i = S_i \cup \{y_{i+1}\}$. Remove this x_k from T_i and call the remaining set S_{i+1}. Then S_{i+1} spans the same set as T_i, namely X. In this way we guarantee that S_{i+1} satisfies the hypothesis above.

Starting with S_0, we thus construct sets $S_1, ..., S_r$. Finally $S_r = \{y_1, ..., y_r\}$ and it spans X. However, Q is independent, and hence y_{r+1} is not in the span of S_r. This contradiction completes the proof. ☐

Corollary 8.1. *$dim(\mathbb{R}^n) = n$.*

Proof. Since $\{e_1, ..., e_n\}$ spans $\mathbb{R}^n$, the theorem above shows $dim(\mathbb{R}^n) \le n$. Since $\{e_1, ..., e_n\}$ is independent, $dim(\mathbb{R}^n) \ge n$. ☐

Theorem 8.2. *Suppose X is a vector space with $dim(X) = n$.*

1. *A set E of n vectors in X spans X if and only if E is independent.*
2. *X has a basis, and every basis consists of n vectors.*
3. *If $1 \le r \le n$ and $\{y_1, ..., y_r\}$ is an independent set in X, then X has a basis containing $\{y_1, ..., y_r\}$.*

Proof.

1. Suppose $E = \{x_1, ..., x_n\}$. Since $dim(X) = n$, the set $\{x_1, ..., x_n, y\}$ is dependent for every $y \in X$. If E is independent, it follows that y is in $span(E)$ and hence $span(E) = X$. If E is dependent, one of its members can be removed without changing the $span(E)$, and hence $span(E) \neq X$.
2. Since $dim(X) = n$, X contains an independent set of n vectors, and (1) shows that every such set is a basis of X. The second statement then follows from Theorem 8.1.
3. Let $\{x_1, ..., x_n\}$ be a basis of X. The set

   $$S = \{y_1, ..., y_r, x_1, ..., x_n\}$$

   spans X and is dependent, since it has more than n vectors. The proof of Theorem 8.1 shows that one of the x_i's is a linear combination of the other members of S. If this x_i is removed from S, the remaining set still spans X. This process can be repeated r times and leads to a basis of X containing $\{y_1, ..., y_r\}$ by (1).

$\square$

Definition 8.2 (Linear Transformation)**.** *A mapping A of a vector space X into a vector space Y is a* ***linear transformation*** *if*

$$A(x_1 + x_2) = Ax_1 + Ax_2, \quad A(cx) = cAx$$

for all $x, x_1, x_2 \in X$ and scalars c.

Linear transformations of X into X are often called ***linear operators*** *on X.*

Definition 8.3 (Inverse of Linear Operator)**.** *If A is a linear operator on X which (i) is one-to-one (injective), and (ii) maps X onto X, we say that A is* ***invertible****, and we can define an operator A^{-1} on X by requiring that $A^{-1}(Ax) = x$ for all $x \in X$.*

It is easy to see that $A(A^{-1}x) = x$ for all $x \in X$, and A^{-1} is also linear. The following theorem shows that condition (i) and (ii) are equivalent in finite-dimensional vector spaces.

Theorem 8.3. *A linear operator A on a finite-dimensional vector space X is one-to-one if and only if the range of A is all of X.*

Proof. Let $\{x_1, ..., x_n\}$ be a basis of X. The linearity of A shows that its range $\mathscr{R}(A)$ is the span of the set $Q = \{Ax_1, ..., Ax_n\}$. It follows from Theorem 8.2(1) that $\mathscr{R}(A) = X$ if and only if Q is independent. Therefore, we only need to prove it is true if and only if A is one-to-one.

Suppose A is one-to-one and $\sum_i c_i Ax_i = 0$. Then $A(\sum_i c_i x_i) = 0$ and hence $\sum_i c_i x_i = 0$, and hence $c_1 = ... = c_n = 0$. Therefore, Q is independent.

Conversely, suppose Q is independent and $A(\sum_i c_i x_i) = 0$. Then $\sum_i c_i Ax_i = 0$ and hence $c_1 = ... = c_n = 0$. Thus, $Ax = 0$ if and only if $x = 0$. As a result, $Ax = Ay$ implies $A(x - y) = 0$ and $x - y = 0$. It means A is one-to-one. $\square$

Definition 8.4 (Sum and Product of Linear Transformations)**.** *Let $L(X, Y)$ be the set of all linear transformations of the vector space X into the vector space Y. We write $L(X)$ instead of $L(X, X)$.*

If $A_1, A_2 \in L(X, Y)$ and if c_1, c_2 are scalars, define $c_1A_1 + c_2A_2$ by

$$(c_1A_1 + c_2A_2)x = c_1A_1x + c_2A_2x, \quad x \in X.$$

It is then clear that $c_1A_1 + c_2A_2 \in L(X, Y)$.

If X, Y, Z are vector spaces, $A \in L(X, Y)$ and $B \in L(Y, Z)$, we define their product BA to be the composition of A and B:

$$(BA)x = B(Ax), \quad x \in X.$$

Then $BA \in L(X, Z)$.

Definition 8.5 (Norm)**.** *For $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, define the* ***norm*** *$\|A\|$ of A to be the supremum of all numbers $|Ax|$, where x ranges over all vectors in $\mathbb{R}^n$ with $|x| \le 1$.*

Theorem 8.4.

1. *If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, then $\|A\| < \infty$ and A is a uniformly continuous mapping of $\mathbb{R}^n$ into $\mathbb{R}^m$.*
2. *If $A, B \in L(\mathbb{R}^n, \mathbb{R}^m)$ and c is a scalar, then*

$$\|A + B\| \le \|A\| + \|B\|, \quad \|cA\| = |c|\|A\|.$$

With the distance between A and B defined as $\|A - B\|$, $L(\mathbb{R}^n, \mathbb{R}^m)$ is a metric space.

3. *If $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ and $B \in L(\mathbb{R}^m, \mathbb{R}^k)$, then*

$$\|BA\| \le \|B\|\|A\|.$$

Proof.

1. Let $\{e_1, ..., e_n\}$ be a standard basis in $\mathbb{R}^n$ and suppose $x = \sum_i c_i e_i$, $|x| \leq 1$, so $|c_i| \leq 1$ for every $i = 1, ..., n$. Then
$$|Ax| = \left|\sum_i c_i e_i\right| \leq \sum_i |c_i||Ae_i| \leq \sum |Ae_i|,$$
so that
$$\|A\| \leq \sum_i |Ae_i| < \infty.$$
Next, since $|Ax - Ay| \leq \|A\||x - y|$ for $x, y \in \mathbb{R}^n$, A is uniformly continuous.
2. The first inequality follows from
$$|(A + B)x| = |Ax + Bx| \leq |Ax| + |Bx| \leq (\|A\| + \|B\|)|x|,$$
and the second inequality is proved analogously. The triangle inequality follows from
$$\|A - C\| = \|(A - B) + (B - C)\| \leq \|A - B\| + \|B - C\|.$$
3. It follows from
$$|(BA)x| = |B(Ax)| \leq \|B\||Ax| \leq \|B\|\|A\||x|.$$

□

Theorem 8.5. *Let Ω be the set of all invertible linear operators in $\mathbb{R}^n$.*
1. *If $A \in \Omega, B \in L(\mathbb{R}^n)$, and*
$$\|B - A\|\|A^{-1}\| < 1,$$
then $B \in \Omega$.
2. *Ω is an open subset of $L(\mathbb{R}^n)$, and the mapping $A \mapsto A^{-1}$ is continuous on Ω.*

Proof. Put $\|A^{-1}\| = 1/\alpha$ and $\|B - A\| = \beta$, then $\beta < \alpha$. For every $x \in \mathbb{R}^n$,
$$\alpha|x| = \alpha|A^{-1}Ax| \leq \alpha\|A^{-1}\||Ax| = |Ax| \leq |(A - B)x| + |Bx| \leq \beta|x| + |Bx|,$$
so that
$$(\alpha - \beta)|x| \leq |Bx|, \quad \forall x \in \mathbb{R}^n.$$

Since $\alpha - \beta > 0$, it shows that $Bx \neq 0$ if $x \neq 0$, and hence B is one-to-one (injective), and by Theorem 8.3 B is also surjective, and hence B is invertible.

Replacing x with $B^{-1}y$, then the resulting inequality
$$(\alpha - \beta)|B^{-1}y| \leq |BB^{-1}y| = |y|$$
shows that $\|B^{-1}\| \leq (\alpha - \beta)^{-1}$. Therefore, the identity
$$B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1},$$
combined with Theorem 8.4(3) implies that
$$\|B^{-1} - A^{-1}\| \leq \|B^{-1}\|A - B\|\|A^{-1}\| \leq \frac{\beta}{\alpha(\alpha - \beta)} \to 0, \quad \beta \to 0.$$
This establishes the continuity since $\beta \to 0$ as $B \to A$. □

Definition 8.6 (Matrix)**.** *Suppose $\{x_1, ..., x_n\}$ and $\{y_1, ..., y_m\}$ are bases of vector spaces X and Y, respectively. Then every $A \in L(X, Y)$ determines a set of numbers $\{a_{ij}\}_{1 \leq i \leq m, 1 \leq j \leq n}$ such that*
$$Ax_j = \sum_{i=1}^{m} a_{ij} y_i, \quad j = 1, ..., n.$$

We visualize these numbers in a rectangular array of m rows and n columns, called an $m \times n$ ***matrix****:*

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Sometimes we write

$$A = [Ax_1, ..., Ax_n],$$

since the coordinates a_{ij} of the vector $Ax_j = [a_{1j}, ..., a_{mj}]^\top$ appear in the j-th column of A. The vectors Ax_j are therefore called the ***column vectors*** *of A.*

8.2 Differentiation

Definition 8.7. *Suppose E is an open set in $\mathbb{R}^n$, f maps E into $\mathbb{R}^m$, and $x \in E$. If there exists a linear transformation A of $\mathbb{R}^n$ into $\mathbb{R}^m$ such that*

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

then we say that f is ***differentiable*** *at x, and we write $f'(x) = A$.*

If f is differentiable at every $x \in E$, we say that f is ***differentiable*** *in E.*

Theorem 8.6 (Uniqueness of Derivative)**.** *Suppose E and f are as in Definition 8.7, $x \in E$, and the limit holds with $A = A_1$ and with $A = A_2$, then $A_1 = A_2$.*

Proof. Let $B = A_1 - A_2$, then the inequality

$$|Bh| \leq |f(x+h) - f(x) - A_1 h| + |f(x+h) - f(x) - A_2 h|$$

shows that $|Bh|/|h| \to 0$ as $h \to 0$. For fixed $h \neq 0$, it follows that

$$\frac{|B(th)|}{|th|} \to 0, \quad t \to 0.$$

The linearity of B shows that the result is independent of t, thus $Bh = 0$ for every $h \in \mathbb{R}^n$. Hence $B = 0$. ☐

Theorem 8.7 (Chain Rule)**.** *Suppose E is an open set in $\mathbb{R}^n$, f maps E into $\mathbb{R}^m$, f is differentiable at $x_0 \in E$, g maps an open set containing $f(E)$ into $\mathbb{R}^k$, and g is differentiable at $f(x_0)$. Then the mapping $g \circ f$ of E into $\mathbb{R}^k$ is differentiable at x_0, and*

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

Proof. Put $y_0 = f(x_0), A = f'(x_0), B = g'(y_0)$, and define

$$u(h) = f(x_0 + h) - f(x_0) - Ah,$$

$$v(k) = g(y_0 + k) - g(y_0) - Bk,$$

for all $h \in \mathbb{R}^n$ and $k \in \mathbb{R}^m$ for which $f(x_0 + h)$ and $g(y_0 + k)$ are defined. Then by the definition of derivatives,

$$|u(h)| = \varepsilon(h)|h|, \quad |v(k)| = \eta(k)|k|,$$

where $\varepsilon(h) \to 0$ as $h \to 0$ and $\eta(k) \to 0$ as $k \to 0$.

Given h, put $k = f(x_0 + h) - f(x_0)$. Then

$$(g \circ f)(x_0 + h) - (g \circ f)(x_0) - BAh = g(y_0 + k) - g(y_0) - BAh = B(k - Ah) + v(k) = Buh(H) + v(k).$$

Hence, for $h \neq 0$ we have

$$\frac{|(g \circ f)(x_0 + h) - (g \circ f)(x_0) - BAh|}{|h|} \leq \|B\|\varepsilon(h) + [\|A\| + \varepsilon(h)]\eta(k).$$

Let $h \to 0$, then $\varepsilon(h) \to 0$. Also, then $k \to 0$ and thus $\eta(k) \to 0$. Therefore, $(g \circ f)'(x_0) = BA$. $\square$

Definition 8.8 (Partial Derivatives)**.** *Suppose f maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$. Let $\{e_1, ..., e_n\}$ and $\{u_1, ..., u_m\}$ be the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. The* ***components*** *of f are real functions $f_1, ..., f_m$ defined by*

$$f(x) = \sum_{i=1}^{m} f_i(x) u_i, \quad x \in E.$$

For $x \in E, i = 1, ..., m, j = 1, ..., n$, we define

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

provided the limit exists. The notation $\frac{\partial f_i}{\partial x_j}$ is often used in place of $D_j f_i$, and it is called a ***partial derivative****.*

Theorem 8.8. *Suppose f maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, and f is differentiable at a point $x \in E$. Then the partial derivatives $(D_j f_i)(x)$ exist, and*

$$f'(x)e_j = \sum_{i=1}^{m} (D_j f_i)(x) u_i, \quad j = 1, ..., n.$$

Proof. Fix j. Since f is differentiable at x,

$$f(x + te_j) - f(x) = f'(x)(te_j) + r(te_j)$$

where $|r(te_j)|/t \to 0$ as $t \to 0$. The linearity of $f'(x)$ shows that

$$\lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = f'(x)e_j.$$

If we represent f in terms of its components, then it becomes

$$\lim_{t \to 0} \sum_{i=1}^{m} \frac{f_i(x + te_j) - f_i(x)}{t} u_i = f'(x)e_j.$$

It follows that each quotient in this sum has a limit as $t \to 0$, so that each $(D_j f_i)(x)$ exists and the equation follows from the definition of partial derivatives. $\square$

Definition 8.9 (Gradient, Directional Derivative)**.** *Let f be a real-valued differentiable function with an open domain $E \subset \mathbb{R}^n$. Associated with each $x \in E$, the* ***gradient*** *of f at x is defined by*

$$(\nabla f)(x) = \sum_{i=1}^{n} (D_i f)(x) e_i.$$

Fix $x \in E$ and let $u \in \mathbb{R}^n$ be a unit vector. The limit

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

is called the ***directional derivative*** *of f at x in the direction of the unit vector u.*

Theorem 8.9. *Suppose f maps a convex open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$, f is differentiable in E, and there is a real number M such that*

$$\|f'(x)\| \leq M, \quad \forall x \in E.$$

Then

$$|f(b) - f(a)| \leq M|b - a|, \quad \forall a, b \in E.$$

Proof. Fix $a, b \in E$. Define
$$\gamma(t) = (1-t)a + tb, \quad t \in \mathbb{R}$$
provided that $\gamma(t) \in E$. Since E is convex, $\gamma(t) \in E$ for $t \in [0, 1]$. Put
$$g(t) = f(\gamma(t)).$$
Then using the chain's rule,
$$g'(t) = f'(\gamma(t))\gamma'(t) = f'(\gamma(t))(b-a),$$
so that
$$|g'(t)| \le \|f'(\gamma(t))\||b-a| \le M|b-a|, \quad \forall t \in [0,1].$$
Hence,
$$|f(b) - f(a)| = |g(1) - g(0)| \le M|b-a|.$$
□

Corollary 8.2. *If, in addition, $f'(x) = 0$ for all $x \in E$, then f is constant.*

Definition 8.10 (Continuous Differentiability)**.** *A differentiable mapping f of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$ is* ***continuously differentiable*** *in E if f' is a continuous mapping of E into $L(\mathbb{R}^n, \mathbb{R}^m)$.*

More explicitly, for every $x \in E$ and $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $y \in E$,
$$|x - y| < \delta \quad \Rightarrow \quad \|f'(y) - f'(x)\| < \varepsilon.$$

If this is so, we also say that f is a $\mathscr{C}'$-mapping, denoted $f \in \mathscr{C}'(E)$.

Theorem 8.10. *Suppose f maps an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^m$. Then $f \in \mathscr{C}'(E)$ if and only if the partial derivatives $D_j f_i$ exist and are continuous on E for $i = 1, ..., m, j = 1, ..., n$.*

Proof. Suppose $f \in \mathscr{C}'(E)$, then f' is continuous. By the definition of partial derivatives,
$$(D_j f_i)(x) = (f'(x)e_j) \cdot u_i, \quad \forall i, j, \forall x \in E.$$
Hence,
$$(D_j f_i)(y) - (D_j f_i)(x) = [(f'(y) - f'(x))e_j] \cdot u_i.$$
Since $|u_i| = |e_j| = 1$, it follows that
$$|(D_j f_i)(y) - (D_j f_i)(x)| \le |(f'(y) - f'(x))e_j| \le \|f'(y) - f'(x)\|.$$
Hence, $D_j f_i$ is continuous.

Conversely, it suffices to consider the case $m = 1$. Fix $x \in E$ and $\varepsilon > 0$. Since E is open, there is an open ball $S \subset E$ centered at x with radius r, and the continuity of functions $D_j f$ shows that r can be chosen so that
$$|(D_j f)(y) - (D_j f)(x)| < \frac{\varepsilon}{n}, \quad \forall y \in S, \forall j = 1, ..., n.$$
Suppose $h = \sum h_j e_j$ and $|h| < r$. Let $v_0 = 0$ and $v_k = h_1 e_1 + ... + h_k e_k$ for $k = 1, ..., n$. Then
$$f(x+h) - f(x) = \sum_{j=1}^{n} [f(x + v_j) - f(x + v_{j-1})].$$
Since $|v_k| < r$ for every k and S is convex, $x + v_{j-1}, x + v_j \in S$. Since $v_j = v_{j-1} + h_j e_j$, the mean value theorem shows that
$$f(x + v_j) - f(x + v_{j-1}) = h_j (D_j f)(x + v_{j-1} + \theta_j h_j e_j)$$
for some $\theta_j \in (0, 1)$, and it differs from $h_j (D_j f)(x)$ by less than $|h_j|\varepsilon/n$. It follows that
$$\left| f(x+h) - f(x) - \sum_{j=1}^{n} h_j (D_j f)(x) \right| \le \frac{1}{n} \sum_{j=1}^{n} |h_j| \varepsilon \le |h| \varepsilon$$

for all h such that $|h| < r$.

This says that f is differentiable at x, and $f'(x)$ is the linear function which assigns the number $\sum h_j(D_jf)(x)$ to the vector $h = \sum h_j e_j$. Hence, the matrix $f'(x)$ consists of the row $(D_1f)(x), ..., (D_nf)(x)$. Since $D_1f, ..., D_nf$ are continuous functions on E, $f \in \mathscr{C}'(E)$. □

8.3 Inverse Function and Implicit Function Theorem

Definition 8.11 (Contraction)**.** *Let X be a metric space with metric d. If φ maps X into X and if there is a number $c < 1$ such that*

$$d(\varphi(x), \varphi(y)) \leq cd(x, y), \quad \forall x, y \in X,$$

then φ is a ***contraction*** *of X into X.*

Theorem 8.11 (Contraction Principle)**.** *If X is a complete metric space, and φ is a contraction of X into X, then there exists a unique $x \in X$ such that $\varphi(x) = x$.*

Proof. Pick $x_0 \in X$ arbitrarily, and define $\{x_n\}$ recursively by setting

$$x_{n+1} = \varphi(x_n), \quad n = 0, 1, 2, ...$$

Then for $n \geq 1$ we have

$$d(x_{n+1}, x_n) = d(\varphi(x_n), \varphi(x_{n-1})) \leq cd(x_n, x_{n-1}).$$

Hence, induction gives

$$d(x_{n+1}, x_n) \leq c^n d(x_1, x_0), \quad n = 0, 1, 2, ...$$

For $n < m$, it follows that

$$d(x_n, x_m) \leq \sum_{i=n+1}^{m} d(x_i, x_{i-1}) \leq (c^n + ... + c^{m-1})d(x_1, x_0) \leq \frac{c^n}{1-c} d(x_1, x_0) \to 0, \quad n \to \infty.$$

Thus, $\{x_n\}$ is Cauchy. Since X is complete, $x_n \to x$ for some $x \in X$.

Finally, since φ is a construction, φ is continuous on X and

$$\varphi(x) = \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

□

Theorem 8.12 (Inverse Function)**.** *Suppose f is a $\mathscr{C}'$-mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^n$, $f'(a)$ is invertible for some $a \in E$, and $b = f(a)$. Then*

1. *There exists open sets U and V in $\mathbb{R}^n$ such that $a \in U, b \in V$, f is one-to-one on U, and $f(U) = V$.*
2. *If g is the inverse of f, defined in V by*

$$g(f(x)) = x, \quad x \in U,$$

then $g \in \mathscr{C}'(V)$.

Proof.

1. Put $f'(a) = A$ and choose λ so that

$$\lambda \|A^{-1}\| = 1/2.$$

Since f' is continuous at a, there is an open ball $U \subset E$ centered at a such that

$$\|f'(x) - A\| < \lambda, \quad \forall x \in U.$$

We associate to each $y \in \mathbb{R}^n$ a function φ defined by

$$\varphi(x) = x + A^{-1}(y - f(x)), \quad x \in E.$$

Note that $f(x) = y$ if and only if x is a fixed point of φ.
Since $\varphi'(x) = I - A^{-1}f'(x) = A^{-1}(A - f'(x))$, by above we have

$$\|\varphi'(x)\| < \|A^{-1}\|\lambda = 1/2, \quad \forall x \in U.$$

Hence, by Theorem 8.9 we have

$$|\varphi(x_1) - \varphi(x_2)| \leq \frac{1}{2}|x_1 - x_2|, \quad \forall x_1, x_2 \in U$$

It follows that φ has at most one fixed point in U, so that $f(x) = y$ for at most one $x \in U$. Therefore, f is one-to-one (injective) in U.
Next, put $V = f(U)$ and pick $y_0 \in V$. Then $y_0 = f(x_0)$ for some $x_0 \in U$. Let B be an open ball centered at x_0 with radius $r > 0$, so that its closure $\bar{B} \subset U$. Now we show that $y \in V$ whenever $|y - y_0| < \lambda r$, which proves that V is open.
Fix y with $|y - y_0| < \lambda r$. By the definition of φ,

$$|\varphi(x_0) - x_0| \leq |A^{-1}(y - y_0)| < \|A^{-1}\|\lambda r = \frac{r}{2}.$$

Given this, it could be shown that φ is a contraction of $\bar{B}$ into itself: if $x \in \bar{B}$, then

$$|\varphi(x) - x_0| \leq |\varphi(x) - \varphi(x_0)| + |\varphi(x_0) - x_0| < \frac{1}{2}|x - x_0| + \frac{r}{2} \leq r,$$

so that $\varphi(x) \in B$. Since $\bar{B}$ is complete as a closed subset of $\mathbb{R}^n$, by the contraction principle φ has a fixed point $x \in \bar{B}$, and $f(x) = y$. Thus, $y \in f(\bar{B}) \subset f(U) = V$.

2. Pick $y \in V$ and $y + k \in V$, then there exist $x \in$U and $x + h \in U$ such that $y = f(x), y + k = f(x + h)$. By the definition of φ,

$$\varphi(x + h) - \varphi(x) = h + A^{-1}[f(x) - f(x + h)] = h - A^{-1}k.$$

By the contraction inequality above, $|h - A^{-1}k| \leq (1/2)|h|$. Hence, $|A^{-1}k| \geq (1/2)|h|$, and

$$|h| \leq 2\|A^{-1}\||k| = \lambda^{-1}|k|.$$

By Theorem 8.5, $f'(x)$ has an inverse denoted by T. Since

$$g(y + k) - g(y) - Tk = h - Tk = -T[f(x + h) - f(x) - f'(x)h],$$

the inequality above implies

$$\frac{|g(y + k) - g(y) - Tk|}{|k|} \leq \frac{\|T\|}{\lambda} \frac{|f(x + h) - f(x) - f'(x)h|}{|h|}$$

As $k \to 0$, $h \to 0$ and RHS $\to 0$. Hence, $g'(y) = T$, and

$$g'(y) = [f'(g(y))]^{-1}.$$

Finally, note that g is a continuous mapping of V onto U, f' is a continuous mapping of U into the set Ω of all invertible elements of $L(\mathbb{R}^n)$, and inversion is a continuous mapping of Ω onto Ω by Theorem 8.5. Combining these results gives $g \in \mathscr{C}'(V)$.

$\square$

Theorem 8.13. *If f is a $\mathscr{C}'$-mapping of an open set $E \subset \mathbb{R}^n$ into $\mathbb{R}^n$ and $f'(x)$ is invertible for every $x \in E$, then $f(W)$ is an open subset of $\mathbb{R}^n$ for every open set $W \subset E$. In other words, f is an open mapping of E into $\mathbb{R}^n$.*

If $x = (x_1, ..., x_n) \in \mathbb{R}^n$ and $y = (y_1, ..., y_m) \in \mathbb{R}^m$, write

$$(x, y) = (x_1, ..., x_n, y_1, ..., y_n) \in \mathbb{R}^{n+m}.$$

Every $A \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$ can be split into two linear transformations A_x and A_y defined by

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

for $h \in \mathbb{R}^n, k \in \mathbb{R}^m$. Then $A_x \in L(\mathbb{R}^n), A_y \in L(\mathbb{R}^m, \mathbb{R}^n)$, and

$$A(h, k) = A_x h + A_y k.$$

Theorem 8.14 (Implicit Function, Linear)**.** *Let $A \in L(\mathbb{R}^{n+m}, \mathbb{R}^n)$ and if A_x is invertible, then there corresponds to every $k \in \mathbb{R}^m$ a unique $h \in \mathbb{R}^n$ such that $A(h, k) = 0$. This h can be computed from k by the formula*

$$h = -(A_x)^{-1} A_y k.$$

Theorem 8.15 (Implicit Function)**.** *Let f be a $\mathscr{C}'$-mapping of an open set $E \subset \mathbb{R}^{n+m}$ into $\mathbb{R}^n$, such that $f(a, b) = 0$ for some point $(a, b) \in E$. Put $A = f'(a, b)$ and assume that A_x is invertible. Then there exist open setes $U \subset \mathbb{R}^{n+m}$ and $W \subset \mathbb{R}^m$ with $(a, b) \in U$ and $b \in W$, such that every $y \in W$ correspond a unique x such that*

$$(x, y) \in U, \quad f(x, y) = 0.$$

If this x is defined to be $g(y)$, then g is a $\mathscr{C}'$-mapping of W into $\mathbb{R}^n$, $g(b) = a$, $f(g(y), y) = 0$ for all $y \in W$, and

$$g'(b) = -(A_x)^{-1} A_y.$$

Proof. □

8.4 Derivatives of Higher Order

Definition 8.12 (Second-Order Partial Derivatives)**.** *Suppose f is a real function defined in an open set $E \subset \mathbb{R}^n$, with partial derivatives $D_1 f, ..., D_n f$. If the functions $D_i f$ are themselves differentiable, then the* ***second-order partial derivatives*** *of f are defined by*

$$D_{ij} f = D_i D_j f, \quad i, j = 1, ..., n.$$

If all these functions $D_{ij} f$ are continuous in E, we say that f is of class $\mathscr{C}''$ in E, written $f \in \mathscr{C}''(E)$.

The following two theorems are stated for real functions of two variables.

Theorem 8.16. *Suppose f is defined in an open set $E \subset \mathbb{R}^2$, and $D_1 f, D_2 f$ exist at every point of E. Suppose $Q \subset E$ is a closed rectangle with sides parallel to the coordinate axes, having (a, b) and $(a+h, b+k)$ as opposite vertices. Put*

$$\Delta(f, Q) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b).$$

Then there is a point (x, y) in the interior of Q such that

$$\Delta(f, Q) = hk(D_{21} f)(x, y).$$

Proof. It follows from applying mean value theorem twice. Put $u(t) = f(t, b + k) - f(t, b)$. Then there is an $x \in (a, a + h)$ and a $y \in (b, b + k)$ such that

$$\Delta(f, Q) = u(a + h) - u(a) = hu'(x) = h[(D_1 f)(x, b + k) - (D_1 f)(x, b)] = hk(D_{21} f)(x, y).$$

□

Theorem 8.17 (Young)**.** *Suppose f is defined in an open set $E \subset \mathbb{R}^2$, $D_1 f, D_{21} f, D_2 f$ exist at every point of E, and $D_{21} f$ is continuous at some point $(a, b) \in E$. Then $D_{12} f$ exists at (a, b) and*

$$(D_{12} f)(a, b) = (D_{21} f)(a, b).$$

Proof. Put $A = (D_{21} f)(a, b)$. Choose $\varepsilon > 0$. Since $D_{21}(f)$ is continuous at (a, b), if Q is a rectangle in Theorem 8.16, and if h and k are sufficiently small, we have

$$|A - (D_{21} f)(x, y)| < \varepsilon, \quad \forall (x, y) \in Q.$$

Thus,

$$\left|\frac{\Delta(f,Q)}{hk} - A\right| < \varepsilon.$$

Fix h and let $k \to 0$. Since $D_2 f$ exists in E, this inequality implies that

$$\left|\frac{(D_2 f)(a+h,b) - (D_2 f)(a,b)}{h} - A\right| \le \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, and this inequality holds for all sufficiently small $h \neq 0$, it follows that $(D_{12} f)(a,b) = A$. □

Corollary 8.3. *$D_{21} f = D_{12} f$ if $f \in \mathscr{C}''(E)$.*

8.5 Differentiation of Integrals

This subsection is about the condition under which the integral and differentiation of a function of two variables can be interchanged, that is,

$$\frac{d}{dt}\int_a^b \varphi(x,t)dx = \int_a^b \frac{\partial \varphi}{\partial t}(x,t)dx.$$

Theorem 8.18. *Suppose $\varphi(x,t)$ is defined for $x \in [a,b], t \in [c,d]$. Suppose α is an increasing function on $[a,b]$. Suppose*

1. *$\varphi \in \mathcal{R}(\alpha)$ as a function of x for every $t \in [c,d]$.*
2. *φ_2 is continuous in t on (c,d) uniformly with respect to x, that is, for every $s \in (c,d)$, for any $\varepsilon > 0$ there is a $\delta > 0$ such that*

$$|\varphi_2(x,t) - \varphi_2(x,s)| < \varepsilon, \quad \forall x \in [a,b], \forall t \in (s-\delta, s+\delta).$$

Define

$$f(t) = \int_a^b \varphi(x,t)d\alpha(x).$$

Then $\varphi_t \in \mathbb{R}(\alpha)$ as a function of x, $f'(s)$ exists, and

$$f'(s) = \int_a^b \varphi_t(x,s)d\alpha(x).$$

Proof. Consider the difference quotients

$$\psi(x,t) = \frac{\varphi(x,t) - \varphi(x,s)}{t-s}$$

for $0 < |t-s| < \delta$. By mean value theorem for each (x,t) there is a number u between s and t such that

$$\psi(x,t) = \varphi_2(x,u).$$

Hence, condition (2) implies that

$$|\psi(x,t) - \varphi_2(x,s)| < \varepsilon, \quad \forall x \in [a,b], 0 < |t-s| < \delta.$$

which implies $\psi \rightrightarrows \varphi_2(x,s)$ (as functions of x) on $[a,b]$. Note that

$$\frac{f(t) - f(s)}{t-s} = \int_a^b \psi(x,t)d\alpha(x).$$

Since $\varphi \in \mathcal{R}(\alpha)$ as a function of x for every t, the conclusion follows from the fact that uniform convergence preserves integral (Theorem 7.3). □