Lecture 1 Real Analysis

Xuan Li*

Math Camp CU 2017

Updated on 08/18/2017

1 Sets

The notion of **set** or **collection** is a primitive notion that we take in its everyday meaning. A set is characterized by what objects it contains; we call them its **elements** or **members**.

- Usually, we use lowercase letters for elements and CAPITAL letters for sets.
- To say that the object x **belongs to**, or **is an element of** the set X, we note $x \in X$.
- When describing a set through an exhaustive listing of its elements, we use curly brackets:

  $X = \{1, \pi, \text{apple}, \text{Economics}, \text{Columbia University}\}$.
- The curly brackets can also be used to define a set through the properties that its elements satisfy:

  $X = \{\text{integers } n \text{ such that } n^2 = 4\}$
- We allow for the possibility that a set contains no element. We call this set the **empty set**, and note it $\emptyset$.

1.1 Inclusion

We say that A **is included in** B, or that A **is a subset of** B, and note $A \subseteq B$ (or $A \subset B$) if every member of A is a member of B: $x \in A \Rightarrow x \in B$. Any set is a subset of itself.

- We denote $A = B$ iff $A \subseteq B$ and also $B \subseteq A$.
- We denote $A \subsetneqq B$ iff $A \subset B$ but $B \subset A$ is not true. In this case, we say that A is a **proper subset** of B.
- The inclusion is **transitive** : if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.
- The set of all subsets of a set X is called the **power set** of X, and noted $P(X)$ or 2^X.

*This lecture note is largely based on the math camp materials from Xingye Wu, who taught this math camp in 2016. I also refer to notes written by Stephane Dupraz, who taught in 2014 and 2015, and Keshav Dogra, who taught in 2011, 2012, and 2013. If you find an error or typo, please send me an email at xl2404@columbia.edu.

1.2 Union and Intersection

The **intersection** of two sets A and B is the set of all elements that belong to both A and B[1]:

$$A \cap B := \{x : x \in A \text{ and } x \in B\}$$

- If no element lies in both A and B, $A \cap B$ is still defined: defined to be the empty set $A \cap B = \emptyset$. We say that A and B are **disjoint**.
- The intersection is commutative: $A \cap B = B \cap A$.
- The intersection is associative: $(A \cap B) \cap C = A \cap (B \cap C)$.
- $A \cap B = A$ iff $A \subseteq B$.

The **union** of two sets A and B, noted $A \cup B$ is the set of all elements that belong to either A or B (possibly both):

$$A \cup B := \{x : x \in A \text{ or } x \in B\}$$

- The union is commutative: $A \cup B = B \cup A$.
- The union is associative: $(A \cup B) \cup C = A \cup (B \cup C)$.
- $A \cup B = A$ iff $B \subseteq A$.

More generally, we can talk about the union of any collection of sets, possibly infinite. Let Θ be a set (potentially infinite), and let $\{A_\theta\}_{\theta \in \Theta}$ be a collection[2] of sets indexed by $\theta \in \Theta$. That is, for each $\theta \in \Theta$, A_θ is a set in the collection. There are as many sets in the collection $\{A_\theta\}_{\theta \in \Theta}$ as there are elements in Θ.

Generalize the notation we defined previously to allow infinite union/intersection:

$$\bigcup_{\theta \in \Theta} A_\theta := \{x : \exists\ \theta \in \Theta \text{ s.t. } x \in A_\theta\}$$
$$\bigcap_{\theta \in \Theta} A_\theta := \{x : x \in A_\theta, \forall\ \theta \in \Theta\}$$

Notice that when the index set Θ has only two elements θ_1 and θ_2, the notations above reduces to pairwise union/intersection we defined at the beginning.

- The intersection is distributive wrt. the union:
  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
  $B \cap \left(\bigcup_{\theta \in \Theta} A_\theta\right) = \bigcup_{\theta \in \Theta} (B \cap A_\theta)$.
- The union is distributive wrt. the intersection:
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
  $B \cup \left(\bigcap_{\theta \in \Theta} A_\theta\right) = \bigcap_{\theta \in \Theta} (B \cup A_\theta)$.

[1](The notation ":=" is for defining a new notation. The left-hand side is the new notation, and the right-hand side is its meaning.)

[2]Essentially, a collection of sets is a set of sets. We use the term "a collection of sets" to avoid confusion.

1.3 Difference and complement

Given two sets A and B, the **set difference** $A - B$ or $A \backslash B$ is the set of all x that belong to A but not to B:

$$B \backslash A := \{x \in B : x \notin A\}$$

- Notice that in the definition above A does not need to be a subset of B.

Let U be the set of all objects that are relevant to the current question, and A be a subset of U. Define the **complement** of A as

$$A^c := U \backslash A$$

- Notice that the notation A^c makes sense only when the "universe" U has no ambiguity in the question. Whenever U currently used is ambiguous, we should be explicit about U by writing $U \backslash A$ instead of A^c.
- We note $x \notin A$ for $x \in A^c$ ($x \in U$ implicitly).
- $(A^c)^c = A$.

Proposition 1.1 (De Morgan's law)**.** *Let $\{A_\theta\}_{\theta \in \Theta}$ be a collection of subsets of U. We have:*

(1) $\left(\bigcup_{\theta \in \Theta} A_\theta \right)^c = \bigcap_{\theta \in \Theta} A_\theta^c$

(2) $\left(\bigcap_{\theta \in \Theta} A_\theta \right)^c = \bigcup_{\theta \in \Theta} A_\theta^c$

Proof. (1) $\subset$:

Take any $x \in \left(\bigcup_{\theta \in \Theta} A_\theta \right)^c = U \backslash \left(\bigcup_{\theta \in \Theta} A_\theta \right)$. By definition, we have $x \in U$, but $x \notin \bigcup_{\theta \in \Theta} A_\theta$.

Therefore $x \notin A_\theta$ for any $\theta \in \Theta$. Then $x \in A_\theta^c$ for any $\theta \in \Theta$, which implies $x \in \bigcap_{\theta \in \Theta} A_\theta^c$.

$\supset$:

Take any $x \in \bigcap_{\theta \in \Theta} A_\theta^c$. By definition, we have $x \in A_\theta^c$ for any $\theta \in \Theta$. Therefore $x \in U$ and $x \notin A_\theta$ for any $\theta \in \Theta$.

So we have $x \in U$ and $x \notin \bigcup_{\theta \in \Theta} A_\theta$, which implies $x \in \left(\bigcup_{\theta \in \Theta} A_\theta \right)^c$.

(2) By (1), we have

$$\left(\bigcap_{\theta \in \Theta} A_\theta \right)^c = \left(\bigcap_{\theta \in \Theta} (A_\theta^c)^c \right)^c = \left(\left(\bigcup_{\theta \in \Theta} A_\theta^c \right)^c \right)^c = \bigcup_{\theta \in \Theta} A_\theta^c$$

■

1.4 Cartesian Product

Define the **Cartesian product** of set A and B as

$$A \times B := \{(a, b) : a \in A \text{ and } b \in B\}$$

- The notation (a, b) is for an ordered pair. The pair is ordered in the sense that (a, b) is different from (b, a). In words, Cartesian product of A and B is the set of all ordered pairs whose first element is taken from A and whose second element is taken from B. Notice that $A \times B \neq B \times A$.

We can also talk about the Cartesian produce of multiple sets $A_1 \times A_2 \times \cdots \times A_n$, or simply $\prod_{i=1}^{n} A_i$, defined as

$$A_1 \times A_2 \times \cdots \times A_n := \{(a_1, a_2, \ldots, a_n) : a_i \in A_i, \forall\ n = 1, 2, \ldots, n\}$$

- If all A_i's are the same, order does not matter and we can denote $A \times A \times \cdots \times A$ as A^n.

2 Relations

Mathematically, a **(binary) relation** R **from** A **to** B is a subset of $A \times B$. Each ordered pair $(a, b) \in A \times B$ either has this relation R (if $(a, b) \in R$) or does not have this relation R (if $(a, b) \notin R$). We can also use the notation aRb to stand for $(a, b) \in R$.

For example, think of the set P of professors, the set S of students, and the "advisorship" relation between professors and students. For each professor-student pair $(p, s) \in P \times S$, professor p either advises student s or not. If p advises s, we put the pair (p, s) into the “advisorship" relation R, and finally we will obtain R as a set of professor-student pairs, i.e. a subset of $P \times S$. Also notice that in the "advisorship" relation, a professor is allowed advice multiple or no student, and a student is allowed to have multiple or no advisor.[3]

For a relation R from A to B, its inverse is a relation from B to A, defined as

$$R^{-1} := \{(b, a) \in B \times A : (a, b) \in R\}$$

Notice that we can always invert a relation to get another relation (in contrast to functions as we will see later).

A relation R from A to A itself is also called a **relation on** A. For example, the inclusion relation $\subset$ of sets can be viewed as a relation on the set of all sets.

Definition 2.1. *Let R be a relation on the set X.*

(1) Relation R is ***reflexive*** *iff*[4] *xRx for any $x \in X$.*

(2) Relation R is ***transitive*** *iff for any $x, y, z \in X$ s.t. xRy and yRz, we have xRz.*

(3) Relation R is ***anti-symmetric*** *iff for any $x, y \in X$ s.t. xRy and yRx, we have $x = y$.*[5]

(4) Relation R is ***complete*** *iff for any $x, y \in X$, either xRy or yRx.*

(5) Relation R is ***symmetric*** *iff for any $x, y \in X$ s.t. xRy, we have yRx.*

2.1 Orders

Definition 2.2. *A relation $\leq$ on X is a* ***pre-order*** *iff $\leq$ is reflexive and transitive. If $\leq$ is a pre-order, we often note $a < b$ when $a \leq b$ and $a \neq b$.*

[3] It is also possible to talk about multilateral relations $R \subset A_1 \times A_2 \times \cdots \times A_n$. This generalized concept will be useful if we want to model, for example, which professor teaches which course on which day, in which case (p, c, d) is in R iff professor p teaches course c on day d. Throughout this math camp, however, we only use binary relations.

[4] To state a definition, the convention is in fact to use "if" instead of "iff" when it actually means "iff". I simply use "iff" to avoid this ambiguity.

[5] By $x = y$, we mean x and y are the same element in X.

Definition 2.3. *A relation $\sim$ on a set X, is an* ***equivalence relation*** *on X iff $\sim$ is reflexive, symmetric and transitive.*

Note that an equivalence relation is a pre-order that also satisfies the symmetry axiom. But the name *pre-order* comes from the fact that if we add the property of antisymmetry, we have a (partial) order.

Definition 2.4. *A relation $\leq$ on X is a* ***partial order*** *iff $\leq$ is reflexive, transitive, and anti-symmetric. In this case, we call $(X, \leq)$ a* ***partially ordered set****, or a* ***poset****.*

Definition 2.5. *A relation $\leq$ on X is a* ***total order*** *(or* ***linear order****) iff $\leq$ is complete, transitive, and anti-symmetric. In this case, we call $(X, \leq)$ a* ***totally ordered set****.*

Notice that both total order and partial order requires anti-symmetry, which rules out ties. This is in contrast to the economic concept of rational preference relation, which only requires completeness and transitivity, and therefore allows indifference between two alternatives.

It is easy to verify that the set inclusion relation $\subseteq$ is reflexive, transitive, anti-symmetric, but not complete.

Because completeness implies reflexiveness, a totally ordered set is a special case of partially ordered set. I'm going to state the following concepts in terms of poset to maintain the generality. In most applications, we will be working with total orders. For example, on the set $\mathbb{R}$ of real numbers, we can verify that the naturally defined relation

$$\leq := \left\{(x, y) \in \mathbb{R}^2 : y - x \text{ is nonnegative}\right\}$$

is a total order.

Based on a partial order $\leq$, we can define a few more relations for convenience:

- $<$: $x < y$ iff $x \leq y$ and $y \nleq x$
- $\geq$: inverse of $\leq$
- $>$: inverse of $<$

2.2 Upper/Lower Bound, Maximum/Minimum, and Supremum/Infimum

Definition 2.6. *Let $(X, \leq)$ be a poset, and let $A \subset X$.*

(1) $u \in X$ is an ***upper bound*** *of A iff $u \geq x$, $\forall\ x \in A$. If such u exists, we say that the set A is bounded from above.*

(2) $l \in X$ is a ***lower bound*** *of A iff $l \leq x$, $\forall\ x \in A$. If such l exists, we say that the set A is bounded from below.*

(3) $x^* \in A$ is a ***maximum*** *of A iff x^* is an upper bound of A.*

(4) $x_* \in A$ is a ***minimum*** *of A iff x_* is a lower bound of A.*

A maximum/minimum of A must be an element in A, and at the same time an upper/lower bound of A. Clearly, maximum/minimum of a set A may not exist, but when it exists, it must be unique due to anti-symmetry (exercise). So it makes sense to talk about "the" maximum/minimum if it exists, and we denote them as $\max A$ and $\min A$.

According to the definition, if A is the empty set, then any $u \in X$ is an upper bound, since the requirement ($u \geq x$ for any $x \in A$) is void.

A set may have many upper/lower bounds in general. However, there is a particular upper/lower bound of special interest.

Definition 2.7. *Let $(X, \leq)$ be a poset, and let $A \subset X$.*
*(1) $u \in X$ is the **least upper bound**, or **supremum**, of A iff*
a) u is an upper bound of A, and
b) $u \leq v$ for any upper bound v of A.
*(1) $l \in X$ is the **greatest lower bound**, or **infimum**, of A iff*
a) l is a lower bound of A, and
b) $l \geq m$ for any lower bound m of A.

By definition, the supremum is the minimum of the set of upper bounds, and the infimum is the maximum of the set of lower bounds. Therefore they are unique if exist, and so it makes sense to talk about "the" supremum/infimum. We denote them as $\sup A$ and $\inf A$.

Proposition 2.8. *(1) If a maximum exists, then it is the least upper-bound.*
(2) If a minimum exists, then it is the greatest lower-bound.

Notice that supremum/infimum may not exist even when upper/lower bounds exist. For example, let $X = \mathbb{R}\backslash\{0\}$ and $A = \{x \in \mathbb{R} : x < 0\}$. In the partially ordered set $(X, \leq)$, the set A has an upper bound (1 for example), but there is no least upper bound. We say that a poset $(X, \leq)$ has the **least upper bound (l.u.b.) property** iff any nonempty subset of X bounded from above has a least upper bound. As we have shown, the set $X = \mathbb{R}\backslash\{0\}$ endowed with the natural order $\leq$ does not have the l.u.b. property. However, the whole set of real numbers $\mathbb{R}$ endowed with $\leq$ has the l.u.b. property, due to the construction of real numbers.

3 Functions

Functions are special cases of relations.

Definition 3.1. *A relation f from X to Y is a **function** iff*
(1) $\forall\ x \in X$, $\exists\ y \in Y$ s.t. $(x, y) \in f$, and
(2) $\forall\ x \in X$ and $y_1, y_2 \in Y$ s.t. $(x, y_1) \in f$ and $(x, y_2) \in f$, we have $y_1 = y_2$.

Requirement (1) is an existence statement, and (2) is a uniqueness statement. Together they require that for each x, there exists one and only one y s.t. (x, y) has the relation f. That is, a function f is a special relation in which one $x \in X$ corresponds to one and only one $y \in Y$. As a result, it is unambiguous to use the notation $f(x)$ to denote the unique element $y \in Y$ s.t. $(x, y) \in f$. We call $f(x)$ the value of f evaluated at x. However, we are silent about how many x's a y may correspond to.

We use the notation $f : X \to Y$ to denote a function from X to Y, and we call X the **domain** of f, and Y the **codomain** of f. A function is also called a **mapping**.

3.1 Image and Inverse Image

Define the **image** of a set $S \subset X$ under f as

$$f(S) := \{f(x) : x \in S\}$$

i.e. the set of elements y in Y s.t. there exists $x \in S$ with $f(x) = y$.

Define the **inverse image** of a set $T \subset Y$ under f as

$$f^{-1}(T) := \{x \in X : f(x) \in T\}$$

i.e. the set of elements x in X s.t. $f(x)$ is in T.

The image of the whole domain X under f, $f(X)$, is called the **range** of f. Notice that $f(X)$ may be a proper subset of the codomain Y, since there may exist y's that correspond to no x.

We have the following results.

Claim 3.2. *Let $f: X \to Y$, $S_1, S_2 \subset X$ and $T_1, T_2 \subset Y$.*

(1a) $S_1 \subset S_2 \;\Rightarrow\; f(S_1) \subset f(S_2)$

(1b) $T_1 \subset T_2 \;\Rightarrow\; f^{-1}(T_1) \subset f^{-1}(T_2)$

(2a) $f(S_1 \cup S_2) = f(S_1) \cup f(S_2)$

(2b) $f^{-1}(T_1 \cup T_2) = f^{-1}(T_1) \cup f^{-1}(T_2)$

(3a) $f(S_1 \cap S_2) \subset f(S_1) \cap f(S_2)$

(3b) $f^{-1}(T_1 \cap T_2) = f^{-1}(T_1) \cap f^{-1}(T_2)$

(4a) No result for $f(S^c)$ and $(f(S))^c$

(4b) $f^{-1}(T^c) = \left(f^{-1}(T)\right)^c$ [6]

Proof. (1a), (1b), (2a), (3b), and (4b) are left as exercises.

(2b)

$\subset$:

Take any $x \in f^{-1}(T_1 \cup T_2)$. By definition of inverse image, we have $f(x) \in T_1 \cup T_2$. Then either $f(x) \in T_1$ or $f(x) \in T_2$. By definition of inverse image again, either $x \in f^{-1}(T_1)$ or $x \in f^{-1}(T_2)$. Therefore, $x \in f^{-1}(T_1) \cup f^{-1}(T_2)$.

$\supset$:

Take any $x \in f^{-1}(T_1) \cup f^{-1}(T_2)$. Then either $x \in f^{-1}(T_1)$ or $x \in f^{-1}(T_2)$. That is, either $f(x) \in T_1$ or $f(x) \in T_2$. Therefore, $f(x) \in T_1 \cup T_2$, which implies $x \in f^{-1}(T_1 \cup T_2)$.

(3a)

Take any $y \in f(S_1 \cap S_2)$. By definition of image, there exists $x \in S_1 \cap S_2$ s.t. $f(x) = y$. Therefore we have $x \in S_1$ and $x \in S_2$, and thus $y = f(x) \in f(S_1)$ and $y = f(x) \in f(S_2)$. So $y \in f(S_1) \cap f(S_2)$ ■

3.2 Composite Functions

Definition 3.3. *Let $f: X \to Y$ and $g: Y \to Z$. Then the **composite function** of f and g, denoted as $g \circ f$, is a function from X to Z s.t. for each $x \in X$, the value of the function $g \circ f$ is defined as $(g \circ f)(x) := g(f(x))$.*

The composite function $g \circ f$ first applies f and maps x to $f(x)$ in Y, and then applies g and maps $f(x)$ to $g(f(x))$ in Z. Notice that the composite operator is associative, i.e. $(h \circ g) \circ f = h \circ (g \circ f)$, since for each $x \in X$, the two functions both map x to $h(g(f(x)))$.[7] As a result, we can use notations like $h \circ g \circ f$ with no ambiguity. However, the composite operator is not commutative, and so we cannot switch the order of h, g, and f.

3.3 Injections, Surjections, and Bijections

Definition 3.4. *Consider a function $f: X \to Y$.*

*(1) f is an **injective function**, or **injection**, iff $\forall\, x_1, x_2 \in X$ s.t. $f(x_1) = f(x_2)$, we have $x_1 = x_2$.*

*(2) f is a **surjective function**, or **surjection**, iff $f(X) = Y$*

*(3) f is a **bijective function**, or **bijection**, iff f is both injective and surjective.*

[6] Notice that in the LHS the complement is taken in Y, while in the RHS the complement is taken in X. The equality really means $f^{-1}(Y \backslash T) = X \backslash f^{-1}(T)$.

[7] Two functions are the same iff they have the same domain and codomain, and for each element in the domain, they maps it to the same element in the codomain.

In words, an injection requires that each y corresponds to at most one x (uniqueness), a surjection requires that each y corresponds to at least one x (existence), and a bijection requires that each y corresponds to exactly one x.

Some books use the term "*one-to-one* function" for injections, "function from X *onto* Y" for surjections, and "*one-to-one correspondence*" for bijections. We are going to stick to our terminologies.

If we view a function f as a relation from X to Y, we can think of its inverse relation f^{-1} from Y to X. Clearly, f^{-1} may fail to be a function. It is not difficult to see that f^{-1} is a function iff f is a bijection. That is the reason why we also call bijections **invertible functions**, since their inverse is still a function.

3.4 Monotonic Functions

When both the domain and the codomain are ordered sets, we can talk about monotonicity of a function. A monotonic function is simply a function that preserve, or inverse, the order.

Definition 3.5. *Let $(X, \leq_X)$ and $(Y, \leq_Y)$ be posets, and consider a function from X to Y.*

(1) f is ***weakly increasing*** *iff $x \leq_X x'$ implies $f(x) \leq_Y f(x')$.*
(2) f is ***weakly decreasing*** *iff $x \leq_X x'$ implies $f(x) \geq_Y f(x')$.*
(3) f is ***strictly increasing*** *iff $x <_X x'$ implies $f(x) <_Y f(x')$.*
(4) f is ***strictly decreasing*** *iff $x <_X x'$ implies $f(x) >_Y f(x')$.*

4 Numbers

The set $\mathbb{N}$ of **natural numbers** consists of a starting element 1, the successor of 1 (denoted as 2), the successor of the successor of 1 (denoted as 3), and so on. Shortly put, $\mathbb{N} := \{1, 2, 3, \ldots\}$. However, be aware that the notation is not universal. Some books define natural numbers $\mathbb{N} := \{0, 1, 2, \ldots\}$, but we are going to stick to $\mathbb{N} := \{1, 2, 3, \ldots\}$.

The set $\mathbb{Z}$ of **integers** consists of natural numbers and their negative counterparts, as well as a neutral element denoted as 0, i.e. $\mathbb{Z} := \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

The set $\mathbb{Q}$ of **rational numbers** consists of ordered pairs (m, n) of integers, and treats (m, n) and (m', n') as the same element iff $m \cdot n' = m' \cdot n$. $\mathbb{Q} := \{m/n : m.n \in \mathbb{Z}, m, ncoprime\}$.[8]

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$$

4.1 Real numbers

We will not define/construct the real line. Instead, we state the key property of the real line:

Axiom 4.1. $(\mathbb{R}, \leq)$ *has the least upper bound property.*

$(\mathbb{Q}, \leq)$ does not have the least upper bound property. See Chapter 1 of Rudin's book for a detailed discussion[9]

[8] m, n coprime if the only positive integer that divides both m and n is 1.

[9] In some sense, the set $\mathbb{R}$ of **real numbers** *completes* the set $\mathbb{Q}$ of rational numbers. There are several equivalent ways to construct real numbers. Some people use the l.u.b. property directly as a part of the definition of real numbers, while some people construct real numbers without explicitly using the l.u.b. property, but later derive it as a property of real numbers. All we need to know is that $(\mathbb{R}, \leq)$ has the l.u.b. property.

4.2 Complex numbers

The set of complex numbers $\mathbb{C}$ is at its core just the set $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, the set of all ordered pairs of $\mathbb{R}$. Instead of noting $z = (a, b) \in \mathbb{R}^2$, we note $z = a + ib$, where i is the **imaginary unit**. We call a the **real part**, noted $Re(z)$, and b the **imaginary part** of $a + ib$, noted $Im(z)$. We see the real line as the subset of $\mathbb{C}$ whose numbers have imaginary part equal to zero. The set really becomes $\mathbb{C}$ once we define operations on it: an addition and a multiplication, like on $\mathbb{R}$. The operations are defined so that they generalize the operations on $\mathbb{R}$.

- The addition: $(a + ib) + (c + id) = (a + b) + i(b + d)$.
- The multiplication : we define $i^2 = -1$ so that $(a+ib)(c+id) = (ac-bc)+i(bc+ad)$.

We define the **conjugate** of a complex number $z = a + ib$ to be $\bar{z} = a - ib$.
We define the **modulus** of a complex number $z = a + ib$ to be $|z| = \sqrt{a^2 + b^2}$. The modulus extends the notion of absolute value $|x| = \max(x, -x)$ on $\mathbb{R}$.
We have that $z\bar{z} = |z|^2$.

5 Countability and Cardinality

5.1 Countability of Sets

It is natural to use the number of elements in a set to get an idea about the size of it. However, this approach does not work if the set has infinitely many elements, and we need a more sophisticated approach.

Definition 5.1. *A set X is **countably infinite** iff there exists a bijection between $\mathbb{N}$ and X.*

The next claim states that countably infinite sets are the "smallest" infinite sets.

Claim 5.2. *If X is countably infinite, then any infinite subset $Y \subset X$ is also countably infinite.*

Proof. * Because X is countably infinite, by definition there exists bijection $f : \mathbb{N} \to X$. Consider the inverse image of Y under f, $f^{-1}(Y)$, which is an infinite set of natural numbers. Let n_1 be the smallest number in $f^{-1}(Y)$, n_2 be the second smallest number in $f^{-1}(Y)$... Define the function $g : \mathbb{N} \to Y$ as $g(i) := f(n_i)$ for each $i \in \mathbb{N}$. By construction, g is a bijection from Y to $\mathbb{N}$, and so Y is countably infinite. ∎

Definition 5.3. *A set X is **countable** iff it is either finite or countably infinite. A set X is **uncountable** iff it is not countable.*

Intuitively, uncountable sets are infinite sets which we can not find a bijection from $\mathbb{N}$ to. Therefore, they are considerably "larger" than countably infinite sets.

Theorem 5.4. *The countable union of countable sets is countable.*

Proof. * We need to extend the previous proof to an infinitely countable union. Consider an infinitely countable collection of set $\{S_1, S_2, ...\}$. For each set S_j, since S_j is countable, we can index its argument as $(x_1^j, x_2^j, x_3^j, ...)$. Place each sequence $(x_1^j, x_2^j, x_3^j, ...)$ as the j^{th} column of a matrix with infinitely many rows. The problem is that this time, the matrix has also infinitely many columns, so we cannot follow our previous proof. However, Georg Cantor, the late XIXth-century mathematician who founded set theory, got a smart idea: snake along the diagonals to define the sequence $(x_1^1, x_2^1, x_1^2, x_1^3, x_2^2, x_3^1, ...)$. This way, we count all the arguments of $\bigcup_n S_n$. (Again, some elements may repeat themselves; just drop them). ∎

Using this proposition, it is easy to show that $\mathbb{Q}$ is countable.

Corollary 5.5. $\mathbb{Q}$ *is countable.*

Proof. * Partition $\mathbb{Q}$ by the value of the denominators of fractions: $\mathbb{Q} = \bigcup_{q\in\mathbb{N},q\neq 0} Q_p$ with $Q_p = \{\frac{p}{q}, p \in \mathbb{N}\}$. Each Q_p is countable, so $\mathbb{Q}$ is a countable union of countable sets. ■

Proposition 5.6. *The cartesian product of finitely many countable sets is countable.*

Proof. * It is enough to prove it for the cartesian product of two sets; the result will then follow by induction. The proof is a corollary of the result on the countable union of countable sets. Let S and T be two countable sets. Note $(x_1, x_2, ...)$ the elements of S. Write the cartesian product as $S \times T = \bigcup_n T_n$, where $T_n = \{(x_n, y), y \in T\}$. Each T_n is countable, so $S \times T$ is a countable union of countable sets. ■

Claim 5.7. $\mathbb{R}$ *is uncountable.*

5.2 Distinguishing between infinities*

Definition 5.8. * *Let S and T two sets.*

- *S and T have* ***equal cardinality*** *if there exists a bijection between them.*
- *S has* ***higher cardinality*** *than T if there exists an injection from T onto S.*
- *S has* ***strictly higher cardinality*** *than T if it has a higher but not equal cardinality than T.*

We can prove Claim 5.7 by showing it has strictly higher cardinality than $\mathbb{N}$.

Proof. * First le's show that $S := \{0,1\}^{\mathbb{N}}$ (an element of S is a sequence of 0 and 1) is uncountable. By contradiction, assume it is countable. Then there exists a bijection from $\mathbb{N}$ to S; call it $(s^1, s^2, ..., s^n, ...)$, where each s^j is an element of S, i.e. a sequence of 0 and 1. Place the sequence s^j as the j^{th} column of an infinite matrix. Now, consider the diagonal of this matrix: it is a sequence of 0 and 1, that is an element of S; call it d. Define the sequence $s^* \in S$ such that for all i, s^*_i is the complement of d_i ($s^*_i = 0$ if $d_i = 1$, and conversely). But s^* is different from all $s^j, j \in \mathbb{N}$ since $s^j_j \neq s^*_j$, which means $s^* \notin S$, a contradiction.

Then we prove $\mathbb{R}$ is in bijection with $\{0,1\}^{\mathbb{N}}$ is in two steps. First we show that $\mathbb{R}$ is in bijection with the interval $(0,1)$, then that $(0,1)$ is in bijection with $\{0,1\}^{\mathbb{N}}$.

- For the first step, note that $x \mapsto \tan(\pi x - \frac{\pi}{2})$ works.
- For the second step, the idea is to use the binary expression of a real number: any real $x \in (0,1)$ can be written as $x = \sum_{n=1}^{\infty} a_n 2^{-n}$, so that the sequence $(a_n)_n \in \{0,1\}^{\mathbb{N}}$ characterizes x. There is a small difficulty because some numbers have actually two such representations: for instance $2^{-1} = \sum_{n=2}^{\infty} 2^{-n}$ are the same number. But we just need to agree not to use the second representation for such numbers.

Therefore, $\mathbb{R}$ is uncountable.

■

Proposition 5.9. * *$\mathbb{R}^n$ (and so $\mathbb{C}$) has the same cardinality as $\mathbb{R}$.*

Proof. * By induction, it is enough to prove that $\mathbb{R}^2$ has the same cardinality as $\mathbb{R}$. Since $\mathbb{R}$ is in bijection with $S = \{0,1\}^{\mathbb{N}}$, $\mathbb{R}^2$ is in bijection with S^2. Hence it is enough to find a bijection between S^2 and S. Consider the function $f : S^2 \to S$ that associates to the couple $((a_n)_n, (b_n)_n)$ the interlaced sequence $(a_1, b_1, a_2, b_2, ...)$. It is a bijection. ■

6 Metric Spaces

Definition 6.1. *Let X be a set. A function $d : X^2 \to \mathbb{R}_+$ is a **distance function**, or **metric**, on X iff it satisfies the following properties*
(1) $d(x, y) = 0$ iff $x = y$,
(2) Symmetry: $d(x, y) = d(y, x)$, $\forall\ x, y \in X$, and
(3) Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$, $\forall\ x, y, z \in X$
*If d is a metric on X, then the couple (X, d) is called a **metric space**.*

Elements of a metric space are often called **points**. Note that the distance function d is an inseparable part of a metric space. (X, d_1) and (X, d_2) are two different metric spaces if d_1 and d_2 are two different distance functions. We can denote a metric space simply as X only when there is no ambiguity regarding what distance function is used.

As an example, the set $\mathbb{R}^k$ can be endowed with a natural metric, the **Euclidean distance function** d_2, defined as

$$d_2(x, y) := \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2}$$

for any $x, y \in \mathbb{R}^k$. This distance function is natural in the sense that it is consistent with how we understand "distance" in our real life where $k = 3$. One can easily verify that d_2 satisfies properties (1) (2) (3)[10]. The metric space $(\mathbb{R}^k, d_2)$ is often called a Euclidean space.

Examples of other metrics on $\mathbb{R}^k$:

(1) d_n metric:

$$d_n(x, y) := \sqrt[n]{\sum_{i=1}^{k} |x_i - y_i|^n}$$

where k can be any positive integer[11]. This subsumes the Euclidean distance d_2 as a special case. Notice that when $k = 1$, all d_n's reduce to the same absolute distance function $d(x, y) := |x - y|$.

(2) discrete metric:

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

Given a metric space (X, d) and a subset $S \subset X$, we can restrict ourselves to the subset S to get a smaller metric space, still using the distance function defined on the larger space X. Formally, define a new distance function $d|_S : S^2 \to \mathbb{R}_+$ as $d|_S(x, y) := d(x, y)$ for any $x, y \in S$. That is, $d|_S$ is the original distance function d restricted in the subset S. Clearly, $d|_S$ is a valid metric on S, and so $(S, d|_S)$ is a valid metric space. Sometimes, we write the metric space $(S, d|_S)$ as (S, d) for simplicity, but keep in mind that rigorously speaking the new metric d in (S, d) has a different domain from the original metric.

7 Basic Topology

Definition 7.1. *Let (X, d) be a metric space. The **open ball** centered at $x \in X$ with radius $r > 0$ is defined as the set*

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

[10] The triangle inequality can be shown using Cauchy-Schwarz inequality.
[11] The triangle inequality can be shown using Minkowski inequality.

Keep in mind that an open ball $B_r(x)$ depends both on the whole space X and the distance function d. For example, in the metric space $(\mathbb{R}, d_2)$, the open ball $B_1(0) = (-1, 1)$; however, in $(\mathbb{R}_+, d_2)$, the open ball $B_1(0) = [0, 1)$. If we use the discrete metric d, then in $(\mathbb{R}, d)$ the open ball $B_1(0) = \{0\}$. Therefore, when we write down the notation for an open ball like $B_1(0)$, we have to be clear about which metric space we are working with.

Definition 7.2. *Let (X, d) be a metric space and S be a subset of X. The set S is said to be **bounded** iff there exists $x \in X$ and $r > 0$ s.t. $B_r(x) \supset S$.*

That is, a set S is bounded iff we can bound it using an open ball.

7.1 Open sets and closed sets

Definition 7.3. *Let (X, d) be a metric space, and S a subset of X.*

*A point $x \in X$ is an **interior point** of S iff $\exists\ r > 0$ s.t. $B_r(x) \subset S$. The set of interior points of S is denoted as $int(S)$.*

*The set S is an **open set** iff $S \subset int(S)$, i.e. all points in S are interior points.*

Clearly, any interior point of S is a point in S, and therefore, $S \subset int(S)$ is equivalent to $S = int(S)$.

By definition, the empty set $\emptyset$ is open, because $int(\emptyset) = \emptyset$. Also, the whole space X is also open, because $int(X) = X$. Keep in mind that whether a set is open depends on the metric space. For example, $[0, 1)$ is not an open set in $(\mathbb{R}, d_2)$, since 0 is not an interior point. However, $[0, 1)$ is an open set in $(\mathbb{R}_+, d_2)$. The point 0 becomes an interior point of $[0, 1)$ because an open ball centered at 0 is now $B_r(0) = [0, r)$ instead of $(-r, r)$.

As its name suggests, an open ball is open.

Claim 7.4. *In metric space (X, d), any open ball is an open set.*

Proof. Take any open ball $B_r(x)$ in the metric space, and take any point $z \in B_r(x)$. Let $\varepsilon := r - d(z, x)$.

First, because $z \in B_r(x)$, we have $d(z, x) < r$ and thus $\varepsilon > 0$.

Second, take any $y \in B_\varepsilon(z)$, we have

$$d(y, x) \leq d(y, z) + d(z, x) < \varepsilon + d(z, x) = r$$

and therefore $y \in B_r(x)$. ∎

Note that open intervals are open in $(\mathbb{R}, d_2)$, because they are special cases of open balls.

The next proposition is an important property of open sets. It states that an arbitrary union of open sets is open, and that a finite intersection of open sets is also open.

Proposition 7.5. *In metric space (X, d):*

(1) Let $\{E_\alpha\}_{\alpha \in A}$ be an arbitrary family of open sets (potentially uncountably many of them). Then their union $\bigcup\limits_{\alpha \in A} E_\alpha$ is also open.

(2) Let $\{E_i\}_{i=1}^n$ be a finite family of open sets. Then their intersection $\bigcap\limits_{i=1}^n E_i$ is also open.

Proof. (1) Take any $x \in \bigcup\limits_{\alpha \in A} E_\alpha$, we need to find $r > 0$ s.t. $B_r(x) \subset \bigcup\limits_{\alpha \in A} E_\alpha$.

By definition of union, $\exists\ \hat{\alpha} \in A$ s.t. $x \in E_{\hat{\alpha}}$. Because $E_{\hat{\alpha}}$ is open, we can find $r > 0$ s.t. $B_r(x) \subset E_{\hat{\alpha}}$.

This is an r we need to find because $B_r(x) \subset E_{\hat{\alpha}} \subset \bigcup_{\alpha \in A} E_\alpha$.

(2) Take any $x \in \bigcap_{i=1}^{n} E_i$, we need to find $r > 0$ s.t. $B_r(x) \subset \bigcap_{i=1}^{n} E_i$.

By definition of intersection, $x \in E_i$ for any $i = 1, 2, \ldots, n$. For each i, because E_i is open, $\exists\ r_i > 0$ s.t. $B_{r_i}(x) \subset E_i$.

Let $r := \min\{r_1, r_2, \ldots, r_n\}$, and this is an r we need to find. First, clearly $r > 0$. Second, $B_r(x) \subset B_{r_i}(x) \subset E_i$ for any i, and therefore $B_r(x) \subset \bigcap_{i=1}^{n} E_i$. ∎

Note that an infinite intersection of open sets may not be open. For example, consider $E_n = (-1/n, 1/n)$, and we have $\bigcap_{n=1}^{+\infty} E_n = \{0\}$.

Now let's move on to closed sets.

Definition 7.6. *Let (X, d) be a metric space, and S a subset of X.*

*A point $x \in X$ is a **limit point** of S iff $(B_r(x) \setminus \{x\}) \cap S \neq \emptyset$, $\forall\ r > 0$. The set of limit points of S is denoted as S'.*

*The set S is a **closed set** iff $S \supset S'$, i.e. S contains all of its limit points.*

The condition "$(B_r(x) \setminus \{x\}) \cap S \neq \emptyset$, $\forall\ r > 0$" states that the open ball $B_r(x)$ with the center removed always contains some points in the set S, no matter how small the radius r is. That is, a point x is a limit point of S iff we can use points in S to approximate x arbitrarily well (the point x itself may be a point in S, but we are not allowed to use x to approximate itself).

Notice that not every point in S is necessarily a limit point, and so $S \supset S'$ is *not* equivalent to $S = S'$. For example in $(\mathbb{R}, d_2)$, the "isolated" point 2 in the set $S = [0, 1] \cup \{2\}$ is not a limit point of S. The set S is indeed closed by definition, since $S' = [0, 1]$ which is a proper subset of S.

By definition, the empty set $\emptyset$ is closed, because $\emptyset' = \emptyset$. Also, the whole space X is also closed, because a limit point of X, by definition, must be a point in X in the first place. Keep in mind that whether a set is closed depends on the metric space. For example, $(0, 1]$ is not a closed set in $(\mathbb{R}, d_2)$, since it does not contain its limit point 0. However, $(0, 1]$ is a closed set in $(\mathbb{R}_{++}, d_2)$. The point 0 is no longer a limit point of $(0, 1]$, since it is not even a point because it is not in the metric space.

The following proposition establishes two characterizations of closed sets.

Proposition 7.7. *Let (X, d) be a metric space, and S a subset of X. Then the following three statements are equivalent.*

(1) S is a closed set.

(2) (sequential definition) For any sequence (x_n) in S convergent to some point $x \in X$, we have $x \in S$.

(3) (topological definition) S^c is an open set.

A corollary of (1) $\Leftrightarrow$ (3) is that S is an open set iff S^c is a closed set. Simply put, the complement of an open set is closed, and the complement of a closed set is open.

Proof. It is sufficient to prove (1) $\Rightarrow$ (2), (2) $\Rightarrow$ (3), and (3) $\Rightarrow$ (1).

(1) $\Rightarrow$ (2):

Take any sequence (x_n) in S convergent to $x \in X$, WTS $x \in S$.

Suppose that $x \notin S$, WTS $x \in S'$.

Take any $r > 0$. Because $x_n \to x$, $\exists\, n$ s.t. $x_n \in B_r(x)$. Because $x_n \in S$ but $x \notin S$, we know that $x_n \neq x$, and thus $x_n \in B_r(x) \setminus \{x\}$. Therefore, $x_n \in (B_r(x) \setminus \{x\}) \cap S$, which implies $(B_r(x) \setminus \{x\}) \cap S \neq \emptyset$.

So we have shown that $x \in S'$.

Because S is closed, we have $x \in S' \subset S$, which contradicts the hypothesis $x \notin S$ we started with.

Therefore, it must be the case that $x \in S$.

(2) $\Rightarrow$ (3):

Take any $x \in S^c$. We want to find $r > 0$ s.t. $B_r(x) \subset S^c$.

Suppose that we cannot find such r, then for any $n \in \mathbb{N}$, we have $B_{1/n}(x) \not\subset S^c$. Then for each n, there exists $x_n \in B_{1/n}(x)$ s.t. $x_n \in S$. Clearly, $x_n \to x$ because for any $\varepsilon > 0$, we can let N be some number greater than $1/\varepsilon$, and then for any $n > N$, we have $d(x_n, x) < 1/n < 1/N < \varepsilon$.

Because of (2), we have $x \in S$, which contradicts $x \in S^c$.

(3) $\Rightarrow$ (1) is left as an exercise. ■

The next proposition is an important property of closed sets. It states that an arbitrary intersection of closed sets is closed, and that a finite union of closed sets is also closed.

Proposition 7.8. *In metric space (X, d):*

(1) Let $\{F_\alpha\}_{\alpha \in A}$ be an arbitrary family of closed sets (potentially uncountably many of them). Then their intersection $\bigcap_{\alpha \in A} F_\alpha$ is also closed.

(2) Let $\{F_i\}_{i=1}^n$ be a finite family of closed sets. Then their union $\bigcup_{i=1}^n F_i$ is also closed.

The proposition above is simply a corollary of Proposition 7.5, using De Morgan's law and the fact that the complement of an open set is closed. This is left as an exercise.

As a final note, again keep in mind that open sets and closed sets are not "absolute" concepts. They rely on the metric space we are working with. When there is ambiguity regarding which metric space we are using, we have to be explicit about it by saying "set S is open/closed in the metric space (X, d)" instead of simply saying "S is open/closed". Also, notice that under discrete metric, all sets in the metric space are both open and closed (exercise).

Please use the following examples to check your understanding of open sets and closed sets.

	$[0, +\infty)$	$(0, +\infty)$	$\{1/n : n \in \mathbb{N}\}$
In $(\mathbb{R}, d_2)$	not open, but closed	open, not closed	not open, not closed
In $(\mathbb{R}_+, d_2)$	open and closed	open, not closed	not open, not closed
In $(\mathbb{R}_{++}, d_2)$	NA	open and closed	not open, but closed

8 Sequences and Convergence

Definition 8.1. *Let X be a set. The function $x : \mathbb{N} \to X$ is called a **sequence** in X.*

Sequences are simply a special case of functions. The value of the function x evaluated at 1, $x(1)$, is called the first **term** of the sequence, and the value of the function evaluated at 2, $x(2)$, is called the second term, and so on. By convention, we often use subscripts and write $x_1, x_2, \ldots$ instead of $x(1), x(2), \ldots$, and the whole sequence is often denoted as (x_n) instead of x.

Note that there is no distance function involved in the definition above, since we don't need a concept of distance to talk about sequences. However, we do need distance to talk about convergence.

Definition 8.2. *Let (X, d) be a metric space. A sequence (x_n) in X is said to* ***converge*** *to a point $x \in X$, iff $\forall\ \varepsilon > 0$, $\exists\ N \in \mathbb{N}$ s.t. $d(x_n, x) < \varepsilon$ for all $n > N$.*

When the sequence (x_n) converges to x, the point x is called a ***limit*** *of the sequence (x_n), and we use the notation $x_n \to x$ or $\lim_{n\to\infty} x_n = x$.*

Notice that the requirement $d(x_n, x) < \varepsilon$ is equivalent to $x_n \in B_\varepsilon(x)$. Another way to describe convergence is that the sequence (x_n) will eventually go into the open ball $B_\varepsilon(x)$, no matter how small the ball is.

The next claim establishes that the limit of a convergent sequence must be unique, and therefore it makes sense to talk about "the" limit of a convergent sequence.

Claim 8.3. *Let (X, d) be a metric space. Suppose $x_n \to x$ and $x_n \to x'$, then $x = x'$.*

Proof. We prove this claim by contradiction.

Suppose $x \neq x'$. By property (1) of d, we have $d(x, x') > 0$.

Let $\varepsilon := d(x, x')/2$. Because $x_n \to x$, there exists N s.t. $d(x_n, x) < \varepsilon$ for any $n > N$. Because $x_n \to x'$, there exists N' s.t. $d(x_n, x') < \varepsilon$ for any $n > N'$. Let $\hat{n} := \max\{N, N'\} + 1$, and so we have $\hat{n} > N$ and $\hat{n} > N'$. Therefore, we have $d(x_{\hat{n}}, x) < \varepsilon$ and $d(x_{\hat{n}}, x') < \varepsilon$, and thus

$$d(x_{\hat{n}}, x) + d(x_{\hat{n}}, x') < 2\varepsilon = d(x, x')$$

which contradicts triangle inequality of d.

Therefore we must have $x = x'$. ■

A sequence is said to be **bounded** iff its range $\{x_1, x_2, \ldots\}$ is a bounded set. The next claim establishes that a convergent sequence must be bounded.

Claim 8.4. *Let (X, d) be a metric space. If (x_n) is a convergent sequence in X, then (x_n) must be bounded.*

Proof. Let the limit of (x_n) be x. Let $\varepsilon = 1$, and by definition of convergence, there exists N s.t. $d(x_n, x) < 1$ for any $n > N$. Then let

$$r := \max\{d(x_1, x), d(x_2, x), \ldots, d(x_N, x)\} + 1$$

Then clearly we have $B_r(x) \supset \{x_1, x_2, \ldots\}$. ■

The trick of this proof is to cut the sequence into a "head" and a "tail". Then we use convergence to bound the tail, and the head is automatically bounded because it has finitely many terms.

The discussion above applies to general metric spaces. The next subsection is dedicated to an important special case, the Euclidean spaces, and establishes some more results on convergence in Euclidean spaces.

8.1 Convergence in Euclidean Spaces

The next claim states that in $\mathbb{R}$ the $\leq$ relation is preserved in the limit.

Claim 8.5. *In $(\mathbb{R}, d_2)$, let there be two convergent sequences $x_n \to x$ and $y_n \to y$. If $x_n \leq y_n$ for any $n \in \mathbb{N}$, then $x \leq y$.*

We can prove this claim by contradiction, and this is left as an exercise. Also note that this is not true if we replace $\leq$ by $<$, since for example, $x_n = -1/n$ and $y_n = 1/n$.

The next proposition states that a sequence of vectors converges to a limit vector iff each coordinate converges separately.

Proposition 8.6. *Let (x_n) be a sequence in $\left(\mathbb{R}^k, d_2\right)$. The sequence (x_n) converges to $x \in \mathbb{R}^k$ iff the sequence $\left(x_n^i\right)$ converges to x^i in $(\mathbb{R}, d_2)$ for any $i \in \{1, 2, \ldots, k\}$.*

Here we use superscript to index coordinates of vectors, since we have used subscript to index terms of the sequences.

Proof. $\Rightarrow$:
Take any $i \in \{1, 2, \ldots, k\}$. WTS: $x_n^i \to x^i$ in $(\mathbb{R}, d_2)$.
Take any $\varepsilon > 0$, we want to find N^i s.t. $d_2\left(x_n^i, x^i\right) < \varepsilon$ for any $n > N^i$.
Because $x_n \to x$, there exists N s.t. $d_2(x_n, x) < \varepsilon$ for any $n > N$. Let $N^i := N$, and I claim that this is an N^i we need to find. This is because for any $n > N^i := N$, we have

$$\begin{aligned} d_2\left(x_n^i, x^i\right) &= \left|x_n^i - x^i\right| = \sqrt{\left(x_n^i - x^i\right)^2} \\ &\leq \sqrt{\sum_{j=1}^k \left(x_n^j - x^j\right)^2} = d_2(x_n, x) < \varepsilon \end{aligned}$$

$\Leftarrow$:
Take any $\varepsilon > 0$, we want to find N s.t. $d_2(x_n, x) < \varepsilon$ for any $n > N$.
Because $x_n^i \to x^i$, there exists N^i s.t. $d_2\left(x_n^i, x^i\right) < \varepsilon/\sqrt{k}$ for any $n > N^i$. Let $N := \max\{N_1, \ldots, N_k\}$, and I claim that this is an N we want to find. This is because for any $n > N$, we have $n > N^i$ and thus $d_2\left(x_n^i, x^i\right) < \varepsilon/\sqrt{k}$ for any i, and therefore

$$d_2(x_n, x) = \sqrt{\sum_{j=1}^k \left(x_n^j - x^j\right)^2} < \sqrt{k\left(\varepsilon/\sqrt{k}\right)^2} = \varepsilon$$

■

The following proposition and its corollary state that the operators $+$, $-$, $\times$, and $/$ preserves limit in $\mathbb{R}$.

Proposition 8.7. *In $(\mathbb{R}, d_2)$, let there be two convergent sequences $x_n \to x$ and $y_n \to y$. Then*
(1) $x_n + y_n \to x + y$,
(2) $x_n y_n \to xy$, and
(3) if $x \neq 0$, then $1/x_n \to 1/x$

Proof. Let's prove (2), and leave (1) and (3) as exercises.
Take any $\varepsilon > 0$, I want to find N s.t. $|x_n y_n - xy| < \varepsilon$ for any $n > N$.
Because (y_n) is convergent, it is bounded, i.e. there exists an open ball $(z - r, z + r)$ that contains $\{y_1, y_2, \ldots\}$. Let $M := \max\{|z - r|, |z + r|\}$, and by construction $|y_n| < M$ for any n.
Because $x_n \to x$, there exists N_x s.t. $|x_n - x| < \varepsilon/2M$. Because $y_n \to y$, there exists N_y s.t. $|y_n - y| < \varepsilon/2(|x| + 1)$.
Let $N := \max\{N_x, N_y\}$, and I claim that this is an N we need to find. This is because for any $n > N$, we have

$$\begin{aligned} |x_n y_n - xy| &= |(x_n - x) y_n + (y_n - y) x| \\ &\leq |x_n - x| \cdot |y_n| + |y_n - y| \cdot |x| \\ &< \frac{\varepsilon}{2M} \cdot M + \frac{\varepsilon}{2(|x| + 1)} \cdot |x| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

■

Corollary 8.8. *In $(\mathbb{R}, d_2)$, let there be two convergent sequences $x_n \to x$ and $y_n \to y$. Then*

(1) $x_n - y_n \to x - y$,

(2) if $y \neq 0$, then $x_n / y_n \to x/y$

Proof. (1) Clearly, the constant sequence $z_n := -1$ converges to $z = -1$. Therefore, by Proposition 8.7(2), we have $z_n y_n \to zy$. So

$$-y_n = z_n y_n \to zy = -y$$

By Proposition 8.7(1), we have

$$x_n - y_n = x_n + (-y_n) \to x + (-y) = x - y$$

(2) By Proposition 8.7(3), we have $1/y_n \to 1/y$. Then by Proposition 8.7(2), we have

$$x_n / y_n = x_n \cdot (1/y_n) \to x \cdot (1/y) = x/y$$

■

When we combine Proposition 8.7 with Proposition 8.6, we can obtain similar results for convergence of vectors. For example, if $x_n \to x$ and $y_n \to y$ in $\left(\mathbb{R}^k, d_2\right)$, then we have $x_n + y_n \to x + y$ in $\left(\mathbb{R}^k, d_2\right)$.

Because sequences are special cases of functions, the definition of monotonicity of functions applies to sequences. A sequence is **increasing** iff $x_m \leq x_n$ for any $m \leq n$; is **decreasing** iff $x_m \geq x_n$ for any $m \leq n$; is **monotone** iff it is increasing or decreasing. We can also define the strict versions of increasing/decreasing sequences in the natural way as we did for functions.

Now let's use the least upper bound property of $\mathbb{R}$ to prove the following important theorem.

Theorem 8.9 (Monotone Convergence Theorem)**.** *Every increasing (decreasing) and bounded from above (below) real sequence (x_n) is convergent in $(\mathbb{R}, d_2)$.*

Notice that an increasing/decreasing sequence is automatically bounded from below/above by its first term. Therefore, the theorem can also be stated that every monotone and bounded real sequence is convergent in $(\mathbb{R}, d_2)$.

Proof. (1) Take any increasing and bounded from above real sequence (x_n). Because the range of the sequence $\{x_1, x_2, \ldots\}$ is bounded from above, by l.u.b. property of $\mathbb{R}$, it has a least upper bound.

Let $x := \sup\{x_1, x_2, \ldots\}$, and we WTS $x_n \to x$.

Take any $\varepsilon > 0$. We want to find N s.t. $|x_n - x| < \varepsilon$ for any $n > N$. Because x is the least upper bound of $\{x_1, x_2, \ldots\}$, $x - \varepsilon$ is not an upper bound, and therefore there exists N s.t. $x_N > x - \varepsilon$. Therefore, for any $n > N$, we have

$$x \geq x_n \geq x_N > x - \varepsilon$$

and therefore $|x_n - x| < \varepsilon$.

(2) Take any decreasing and bounded from below real sequence (x_n). Clearly $(-x_n)$ is increasing and bounded from above. By (1) we have $(-x_n)$ is convergent, and thus (x_n) is also convergent. ■

Given a sequence (x_n), a **subsequence** of (x_n) is a sequence (x_{n_k}) indexed by $k \in \mathbb{N}$, where (n_k) is a strictly increasing sequence in $\mathbb{N}$. For example, if the sequence (n_k) is $2, 4, 5, 9, \ldots$, then the subsequence (x_{n_k}) is $x_2, x_4, x_5, x_9, \ldots$.

Lemma 8.10. *Every sequence in $\mathbb{R}$ has a monotone subsequence.*

Proof. Take any sequence (x_n) in $\mathbb{R}$. Call the term x_n a dominant term if $x_n \geq x_m$ for any $m \geq n$.

Case 1: (x_n) has infinitely many dominant terms

Then these dominant terms constitute a decreasing subsequence.

Case 2: (x_n) has finitely many dominant terms

Let x_N be the last dominant term. Let $n_1 = N + 1$, and so x_{n_1} is not a dominant term. By definition, there exists $n_2 > n_1$ s.t. $x_{n_2} > x_{n_1}$. The term x_{n_2} itself is not dominant either, and so there exists $n_3 > n_2$ s.t. $x_{n_3} > x_{n_2}$... Therefore, we obtain a strictly increasing subsequence.

Case 3: (x_n) has no dominant term

Let $n_1 = 1$, and construct a strictly increasing subsequence as in Case 2.

Therefore, we can always find a monotone subsequence. ■

If we combine the lemma we have just proved with Monotone Convergence Theorem, we immediately obtain the **Bolzano-Weierstrass theorem**: every bounded sequence in $(\mathbb{R}, d_2)$ has a convergent subsequence.

We can easily extend Bolzano-Weierstrass theorem to $\left(\mathbb{R}^k, d_2\right)$. So we have proved the following theorem.

Theorem 8.11 (Bolzano-Weierstrass)**.** *Every bounded sequence in $\left(\mathbb{R}^k, d_2\right)$ has a convergent subsequence.*

Proof. (sketch)

Take any bounded vector sequence in $\left(\mathbb{R}^k, d_2\right)$. Clearly, each of its coordinate is a bounded real sequence in $(\mathbb{R}, d_2)$. Then we can apply Bolzano-Weierstrass theorem in $(\mathbb{R}, d_2)$ to find a convergent subsequence for the first coordinate. Then we can find a subsequence of this subsequence that is convergent in the second coordinate. Repeat this process, and we finally obtain a subsequence that is convergent in every coordinate. By Proposition 8.6, the subsequence of vector is convergent. ■

9 Compactness

Compactness is a stronger notion than closedness. It is also a crucial concept, because compact sets have many desirable properties that closed sets don't have.

Definition 9.1. *Let (X, d) be a metric space, and S a subset of X. A family of open sets $\{E_\alpha\}_{\alpha \in A}$ is an* ***open cover*** *of S iff $\bigcup\limits_{\alpha \in A} E_\alpha \supset S$.*

Definition 9.2. *Let (X, d) be a metric space, and S a subset of X. The set S is* ***compact*** *iff $\forall$ open cover $\{E_\alpha\}_{\alpha \in A}$ of S, $\exists$ a finite $B \subset A$ s.t. $\{E_\alpha\}_{\alpha \in B}$ is also an open cover of S.*

To illustrate the definition of compact sets, let's verify that the open interval $(0, 1)$ is not a compact set in $(\mathbb{R}, d_2)$. To do this, it is sufficient to provide an open cover that does not have a finite subcover. Consider the family of open sets $\{(1/n, 1 - 1/n)\}_{n=1}^{+\infty}$. This covers $(0, 1)$ because any point strictly between 0 and 1 will be eventually covered by $(1/n, 1 - 1/n)$ when n is large enough. There is no finite subcover, since any finite family of $(1/n, 1 - 1/n)$ has a largest one, and it does not cover $(0, 1)$.

By definition, the concept of compactness relies on the metric space we are working with, just like openness and closedness. A set can be compact in one metric space, but not in another metric space. However, compactness behaves much better than openness and closedness, in the sense that enlarging or shrinking the whole space does not affect compactness as long as we use the same metric. This result is formulated below.

Proposition 9.3. *Let (X, d) be a metric space, and $S \subset Y \subset X$. Then S is compact in (X, d) iff S is compact in (Y, d).*

Recall that (Y, d) in fact means $(Y, d|_Y)$, rigorously speaking.

See Theorem 2.33 in Rudin for a proof.

In the proposition above, if we let $Y := S$, we have that S is compact in (X, d) iff S is compact in (S, d). A metric space (X, d) is said to be a **compact metric space** iff X is a compact set in (X, d). So we know that if S is compact in (X, d), then (S, d) itself is a compact metric space, and vice versa.

9.1 General Properties

The theorem below states that compactness is stronger than closedness.

Theorem 9.4. *Let (X, d) be a metric space, and S a subset of X. If S is compact in (X, d), then S is closed in (X, d).*

Proof. WTS: S^c is open in (X, d)

Take any $x \in S^c$, we want to find $r > 0$ s.t. $B_r(x) \subset S^c$.

Take any $y \in S$, let $r_y := d(y, x)/2$. Then clearly $B_{r_y}(x)$ and $B_{r_y}(y)$ are disjoint.

Notice that $\left\{B_{r_y}(y)\right\}_{y \in S}$ is an open cover of S. By compactness of S, there exists $\{y_1, y_2, \ldots, y_n\}$ s.t. $\left\{B_{r_{y_i}}(y_i)\right\}_{i=1}^{n}$ is also an open cover of S.

Let $r := \min\{r_{y_1}, r_{y_2}, \ldots, r_{y_n}\}$. WTS: $B_r(x) \subset S^c$.

Clearly $B_r(x)$ is disjoint with $B_{r_{y_i}}(y_i)$ for any i. So $B_r(x)$ is disjoint with the union of $B_{r_{y_i}}(y_i)$'s, and thus $B_r(x)$ is disjoint with S, which implies $B_r(x) \subset S^c$. ■

The next theorem states that a compact set must be bounded.

Theorem 9.5. *Let (X, d) be a metric space, and S a subset of X. If S is compact in (X, d), then S is bounded in (X, d).*

Proof. Arbitrarily take a point $x \in X$. Clearly, $\{B_n(x)\}_{n=1}^{\infty}$ is an open cover of S, because any $y \in X$ has a finite distance to x, and will be eventually covered by $B_n(x)$ when n is large enough. By compactness of S, there exists $\{n_1, n_2, \ldots, n_k\}$ s.t. $\{B_{n_i}(x)\}_{i=1}^{k}$ is also an open cover of S.

Let $r := \max\{n_1, n_2, \ldots, n_k\}$. Then $B_r(x) \supset S$, and so S is bounded. ■

Combining the Theorem 9.4 and 9.5, we conclude that a compact set must be closed and bounded.

The next theorem provides a way to prove compactness. It states that a closed set contained in a compact set is also compact. Therefore, in order to show that S is compact, we can instead show that S is closed and that some other set containing S is compact.

Theorem 9.6. *Let (X, d) be a metric space, and $S \subset Y \subset X$. If S is closed in (X, d) and Y is compact in (X, d), then S is compact in (X, d).*

Proof. Take any open cover $\{E_\alpha\}_{\alpha \in A}$ of S. We want to find a finite family chosen from $\{E_\alpha\}_{\alpha \in A}$ that also covers S.

Clearly, $\{E_\alpha\}_{\alpha \in A} \cup \{S^c\}$ covers the whole space, and thus covers Y. Because Y is compact, there exists a finite family chosen from $\{E_\alpha\}_{\alpha \in A} \cup \{S^c\}$ that covers Y. Because $S \subset Y$, the finite family also covers S. If the finite family contains S^c, then we can remove it from the family, then the family still covers S, since S^c has no contribution to covering S. So we have obtained a finite family chosen from $\{E_\alpha\}_{\alpha \in A}$ that covers S. ■

Theorem 9.4 and 9.6 together implies that in compact metric spaces, closedness and compactness are equivalent.

The discussions above apply to general metric spaces. The next subsection is dedicated to Euclidean spaces $\left(\mathbb{R}^k, d_2\right)$, and establishes more results.

9.2 Heine-Borel Theorem in $\left(\mathbb{R}^k, d_2\right)$

In general metric spaces, we have shown that a compact set must be closed and bounded. In Euclidean spaces $\left(\mathbb{R}^k, d_2\right)$, the reverse is also true, i.e. a closed and bounded set in $\left(\mathbb{R}^k, d_2\right)$ must be compact. Therefore, in Euclidean spaces, compactness is equivalent to closedness plus boundedness, and this equivalence is known as Heine-Borel theorem.

We establish this result in several steps.

Lemma 9.7. *Any closed interval $[a, b]$ is compact in $(\mathbb{R}, d_2)$.*

Proof. Take any closed interval $[a, b]$, and suppose that it is not compact. Then there exists an open cover $\{E_\alpha\}_{\alpha \in A}$ of $[a, b]$ without a finite subcover.

Let $a_0 := a$, and $b_0 := b$.

Cut the interval $[a_0, b_0]$ in half: $[a_0, (a_0 + b_0)/2]$ and $[(a_0 + b_0)/2, b_0]$. At least one of them cannot be finitely covered (otherwise the interval $[a_0, b_0]$ can be finitely covered). Take the one that cannot be finitely covered, and label it as $[a_1, b_1]$.

Cut the interval $[a_1, b_1]$ in half: $[a_1, (a_1 + b_1)/2]$ and $[(a_1 + b_1)/2, b_1]$. At least one of them cannot be finitely covered. Take the one that cannot be finitely covered, and label it as $[a_2, b_2]$.

Repeat this process, and we get a shrinking sequence of intervals $[a_0, b_0] \supset [a_1, b_1] \supset \cdots$, and each of them cannot be finitely covered using the open cover $\{E_\alpha\}_{\alpha \in A}$.

Because (a_n) is increasing and bounded from above by b_0, (a_n) converges to some limit a^*. Symmetrically, (b_n) converges to some limit b^*. Because $b_n - a_n = (1/2)^n (b - a) \to 0$, we know that

$$b_n = a_n + (b_n - a_n) \to a^* + 0 = a^*$$

and therefore $b^* = a^*$. That is, the sequence of intervals $[a_0, b_0] \supset [a_1, b_1] \supset \cdots$ shrinks to one point a^*. Because $a^* \in [a, b]$, it is covered by some open set E_{α^*} in the open cover.

Therefore, there exists $B_r(a^*) \subset E_{\alpha^*}$. Because (a_n) and (b_n) both converge to a^*, there exists $\hat{n}$ s.t. $a_{\hat{n}}, b_{\hat{n}} \in B_r(a^*)$, and therefore $[a_{\hat{n}}, b_{\hat{n}}] \subset B_r(a^*) \subset E_{\alpha^*}$. So $[a_{\hat{n}}, b_{\hat{n}}]$ can be finitely covered using the open cover $\{E_\alpha\}_{\alpha \in A}$, which contradicts the construction of the sequence $([a_n, b_n])$. ■

It is not difficult to extend the lemma to $\left(\mathbb{R}^k, d_2\right)$.

Lemma 9.8. *Every k-cell $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_k, b_k]$ is compact in $\left(\mathbb{R}^k, d_2\right)$.*

We can use the same idea to prove this result for $\mathbb{R}^k$ as that for $\mathbb{R}$. Take any k-cell, and suppose that it is not compact. Then there exists an open cover $\{E_\alpha\}_{\alpha \in A}$ of the cell without a finite subcover. Then in each coordinate of the k-cell, we cut the interval in half, and in total we get 2^k sub-cells. At least one of them cannot be finitely covered. Take the one that cannot be finitely covered, and cut it into 2^k sub-cells, and repeat this process. We can get a shrinking sequence of k-cells, each of whom cannot be finitely covered. It can be shown that the sequence shrinks to one limit point x^* in the original k-cell. Then let the open set in $\{E_\alpha\}_{\alpha \in A}$ that covers x^* be E_{α^*}. It can be shown that the sequence of k-cells will eventually go into E_{α^*}, and thus can be finitely covered. This contradicts the construction of this sequence. See Theorem 2.40 in Rudin for details.

Now let's consider a closed and bounded set S in $\left(\mathbb{R}^k, d_2\right)$. By definition of boundedness, S can be bounded by an open ball. Clearly, an open ball in $\left(\mathbb{R}^k, d_2\right)$ can be bounded by a k-cell. So S is a subset of a k-cell, which is compact in $\left(\mathbb{R}^k, d_2\right)$ by the lemma above. Furthermore, because S is assumed to be closed in $\left(\mathbb{R}^k, d_2\right)$, by Theorem 9.6, we know that S is compact itself. Therefore we have proved the following theorem.

Theorem 9.9. *In $\left(\mathbb{R}^k, d_2\right)$, a closed and bounded set S must be compact.*

Combining this theorem with Theorem 9.4 and 9.5, we have the well-known Heine-Borel theorem.

Theorem 9.10 (Heine-Borel)**.** *In $\left(\mathbb{R}^k, d_2\right)$, a set S is compact iff it is closed and bounded.*

Heine-Borel theorem states that in $\left(\mathbb{R}^k, d_2\right)$, to check whether a set is compact, we can instead check whether the set is closed and bounded. This greatly simplifies our job, since the definition of compactness involving open covers is not easy to check in most cases.

Keep in mind that Heine-Borel theorem only works in Euclidean spaces $\left(\mathbb{R}^k, d_2\right)$. In general metric spaces (X, d), although compactness always implies closedness plus boundedness, the reverse is not true in general. For example, in $(\mathbb{R}_{++}, d_2)$, the set $(0, 1]$ is closed and bounded, but not compact (consider the open cover $\{(1/n, +\infty)\}_{n=1}^{\infty}$). For another example, in $(\mathbb{R}, d)$, where d is the discrete metric, the set $[0, 1]$ is closed and bounded, but not compact. In fact, under the discrete metric, a set is compact iff it is finite (exercise).

9.3 Sequential Compactness

There is another notion of compactness, called sequential compactness.

Definition 9.11. *Let (X, d) be a metric space, and S a subset of X. The set S is **sequentially compact** iff any sequence (x_n) in S has a subsequence convergent to some $x^* \in S$.*

Theorem 9.12. *Let (X, d) be a metric space, and S a subset of X. The set S is compact iff it is sequentially compact.*

This equivalence holds in general metric spaces[12], but here we will just provide the proof in Euclidean spaces $\left(\mathbb{R}^k, d_2\right)$.

Proof. "$\Rightarrow$":

If S in $\left(\mathbb{R}^k, d_2\right)$ is compact, then it is bounded. Then any sequence (x_n) in S must be bounded. By Bolzano-Weierstrass theorem, (x_n) has a subsequence convergent to some $x^* \in \mathbb{R}^k$. Then by the sequential definition of closed sets, we have $x^* \in S$. Therefore, S is sequentially compact.

"$\Leftarrow$" :

If S in $\left(\mathbb{R}^k, d_2\right)$ is sequentially compact, then it must be closed; otherwise we can find a sequence (x_n) in S convergent to some x^* outside S, and any subsequence of (x_n) must also converge to $x^* \notin S$, so it does not have a subsequence convergent to some point in S.

Also, the set S must also be bounded. Otherwise we can construct a sequence (x_n) s.t. $d_2(x_n, 0) > n$, and so (x_n) does not even have a convergent subsequence. Therefore, S is both closed and bounded, and therefore compact. ∎

10 Cauchy Sequences and Completeness

Definition 10.1. *In metric space (X, d), a sequence (x_n) is a **Cauchy sequence** iff $\forall\ \varepsilon > 0$, $\exists\ N \in \mathbb{N}$ s.t.*

$$d(x_m, x_n) < \varepsilon$$

[12]However, the two notions are not the same in *topological spaces*, which is a generalization of metric spaces and is out of the scope of our math camp.

for any $m, n > N$.

Clearly, Cauchy sequences must be bounded just like convergent sequences. This is left as an exercise.

Proposition 10.2. *In metric space (X, d), a convergent sequence (x_n) is a Cauchy sequence.*

Proof. Let the limit of (x_n) be x.

Take any $\varepsilon > 0$. I want to find N s.t. $d(x_m, x_n) < \varepsilon$ for any $m, n > N$.

Because $x_n \to x$, there exists N s.t. $d(x_k, x) < \varepsilon/2$ for any $k > N$. Therefore, for any $m, n > N$, we have

$$d(x_m, x_n) \leq d(x_m, x) + d(x_n, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

■

However, a Cauchy sequence may fail to be convergent, for example the sequence $(1/n)$ in $(\mathbb{R}_{++}, d_2)$. If the metric space (X, d) has the property that every Cauchy sequence converges, then we call it a complete metric space. The metric space $(\mathbb{R}_{++}, d_2)$ is not complete.

Definition 10.3. *Let (X, d) be a metric space, and S a subset of X. The set S is a **complete** set iff any Cauchy sequence in S converges to a limit point in S.*

*A metric space (X, d) is a **complete metric space** iff X is a complete set in (X, d).*

Completeness is stronger than closedness, but weaker than compactness.

A complete set S must be closed. Otherwise we can find a sequence (x_n) in S convergent to a point outside S. Because (x_n) is Cauchy, this contradicts the completeness of S.

The next result states that a compact set S must be complete.

Proposition 10.4. *Let (X, d) be a metric space, and S a subset of X. If the set S is compact, then it is complete.*

Proof. Take any Cauchy sequence (x_n) in S. We want to find an $x \in S$ s.t. $x_n \to x$.

Because S is compact, and so is sequentially compact, we can find a subsequence (x_{n_k}) convergent to some $x \in S$.

Now we only need to show $x_n \to x$

Take any $\varepsilon > 0$. We want to find N s.t. $d(x_n, x) < \varepsilon$ for any $n > N$.

Because (x_n) is Cauchy, there exists N s.t. $d(x_m, x_n) < \varepsilon/2$ for any $m, n > N$.

Because $x_{n_k} \to x$, there exists K s.t. $n_K > N$ and $d(x_{n_K}, x) < \varepsilon/2$.

Then for any $n > N$, we have

$$d(x_n, x) \leq d(x_n, x_{n_K}) + d(x_{n_K}, x) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

■

In the proof above, the candidate x for Cauchy sequence limit is provided by the sequential compactness of S. Then the trick to prove $x_n \to x$ is to bind the whole sequence to the convergent subsequence using Cauchy, and then bind the subsequence to the limit using its convergence.

Using the same trick, we can show that the Euclidean spaces $\left(\mathbb{R}^k, d_2\right)$ are complete.

Proposition 10.5. *The Euclidean space $\left(\mathbb{R}^k, d_2\right)$ is a complete metric space.*

The proof of this result is left as an exercise. Notice that the candidate x for Cauchy sequence limit is provided by Bolzano-Weierstrass theorem.

10.1 Contraction Mapping Theorem

This subsection discusses an important fixed-point result in complete metric spaces, known as Contraction Mapping theorem or Banach Fixed Point theorem. This result has important implications in dynamic programming.

A function is called a **self-map** iff it maps its domain to itself, i.e. $f : X \to X$. Note that a self-map need not be surjective or injective.

A point $x^* \in X$ is called a **fixed point** of the self-map $f : X \to X$, iff $f(x^*) = x^*$. Intuitively, the fixed point x^* does not "move away" if we apply f to it.

Definition 10.6. *Let (X, d) be a metric space. A self-map $f : X \to X$ is said to be a* ***contraction*** *iff $\exists$ real $\lambda < 1$ s.t.*

$$d(f(x), f(x')) \leq \lambda \cdot d(x, x')$$

for any $x, x' \in X$.

Now we state the theorem.

Theorem 10.7 (Contraction Mapping Theorem)**.** *Let (X, d) be a complete metric space, and $f : X \to X$ a contraction. Then f has a unique fixed point x^*. Further, for any $x \in X$, we have $\lim_{n\to\infty} f^n(x) = x^*$.*

The notation $f^n(x)$ means to apply f to x n times, i.e. $f^2(x) := f(f(x))$, $f^3(x) = f(f(f(x)))$, and so on.

Outline of the proof:

1. Show that the sequence $(f^n(x_0))$ is Cauchy for an arbitrary $x_0 \in X$. So by completeness of (X, d), $(f^n(x_0))$ converges to some $x^* \in X$.
2. Show that $x^* := \lim_{n\to\infty} f^{(n)}(x_0)$ is indeed a fixed point of f.
3. Show that x^* is the unique fixed point of f.
4. Show that $f^n(x) \to x^*$ for any starting point $x \in X$.

Proof. Arbitrarily take $x_0 \in X$. Define $x_n := f^n(x_0)$, for any $n \in \mathbb{N}$.

Step 1: WTS (x_n) is Cauchy

Take any $\varepsilon > 0$, we want to find N s.t. $d(x_m, x_n) < \varepsilon$ for any $m, n > N$.

Because f is a contraction, we have $d(f(x), f(x')) \leq \lambda \cdot d(x, x')$ for some $\lambda < 1$.

Let N be s.t. $\lambda^N < (1-\lambda)\varepsilon / d(x_0, x_1)$. This is possible because $\lambda^N \to 0$, and $(1-\lambda)\varepsilon / d(x_0, x_1) > 0$[13].

WTS: $d(x_m, x_n) < \varepsilon$ for any $m, n > N$.

Since

$$d(x_k, x_{k+1}) = d(f(x_{k-1}), f(x_k)) \leq \lambda d(x_{k-1}, x_k)$$

we have,

$$\begin{aligned} d(x_k, x_{k+1}) &\leq \lambda d(x_{k-1}, x_k) \\ &\leq \lambda^2 d(x_{k-2}, x_{k-1}) \\ &\leq \cdots \\ &\leq \lambda^k d(x_0, x_1) \end{aligned}$$

[13] If $d(x_0, x_1) = 0$, it is obvious that $f : x \to x$ is not a contraction.

Take any $m, n > N$. Without loss of generality, assume that $m \leq n$, and we have

$$\begin{aligned}
d(x_m, x_n) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\
&\leq \lambda^m d(x_0, x_1) + \lambda^{m+1} d(x_0, x_1) + \cdots + \lambda^{n-1} d(x_0, x_1) \\
&= \lambda^m d(x_0, x_1) \left(1 + \lambda + \cdots + \lambda^{n-m-1}\right) \\
&= \lambda^m d(x_0, x_1) \frac{1 - \lambda^{n-m}}{1 - \lambda} < \lambda^N d(x_0, x_1) \frac{1}{1 - \lambda} \\
&< \frac{(1 - \lambda)\varepsilon}{d(x_0, x_1)} \cdot d(x_0, x_1) \frac{1}{1 - \lambda} = \varepsilon
\end{aligned}$$

As a result, we have shown that (x_n) is Cauchy.
Because (X, d) is a complete metric space, (x_n) converges to some limit $x^* \in X$.
Step 2: WTS $f(x^*) = x^*$
I want to show this by showing $d(f(x^*), x^*) = 0$.
It is sufficient to show that $d(f(x^*), x^*) < \varepsilon$ for any $\varepsilon > 0$.
Take any $\varepsilon > 0$.
Because $x_n \to x^*$, there exists N s.t. $d(x_n, x^*) < \varepsilon/2$ for any $n \geq N$.
Then we have

$$\begin{aligned}
d(f(x^*), x^*) &\leq d(f(x^*), x_{N+1}) + d(x_{N+1}, x^*) \\
&= d(f(x^*), f(x_N)) + d(x_{N+1}, x^*) \\
&\leq \lambda d(x^*, x_N) + d(x_{N+1}, x^*) \\
&< \lambda \cdot \varepsilon/2 + \varepsilon/2 < \varepsilon
\end{aligned}$$

Step 3: WTS x^* is the unique fixed point of f

Assume by contradition that $\exists \hat{x} \in S, \hat{x} \neq x^*$ such that $f(\hat{x}) = \hat{x}$. Then $d(f(\hat{x}), f(x^*)) = d(\hat{x}, x^*) > 0$. Therefore f is not a contraction. Contradiction1!

Step 4: WTS $f^n(x) \to x^*$ for any starting point $x \in X$

Take $\forall y_0 \in S, y_0 \neq x_0$ and $y^* := \lim_{n\to\infty} f^{(n)}(y_0) \in S$ (from Step 1 we know that such y^* exists). WTS $y^* = x^*$, i.e. $d(x^*, y^*) = 0$. It suffices to show that $d(y^*, x^*) < \varepsilon$ for any $\varepsilon > 0$.

Denote $y_n := f^{(n)}(y_0)$. Let N_1 be such that $\lambda^{N_1} < \varepsilon/(3d(x_0, y_0))$. Then we have $d(x_n, y_n) \leq \lambda^n d(x_0, y_0) \leq \lambda^{N_1} d(x_0, y_0) < \varepsilon/3$, $\forall n \geq N_1$.

Also, as we have $x_n \to x^*, y_n \to y^*$, there exists N_2 s.t. $d(x_n, x^*) < \varepsilon/3$, $d(y_n, y^*) < \varepsilon/3$ for any $n \geq N_2$.

Take $N := \max\{N_1, N_2\}$, then for any $n \geq N$, we have:

$$\begin{aligned}
d(x^*, y^*) &\leq d(x_n, x^*) + d(y_n, y^*) + d(x_n, y_n) \\
&< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon
\end{aligned}$$

OR: Step 4 directly follows from Steps 1,2 and 3, because by Step 1 for any $y \in X$ we have $y^* := \lim_{n\to\infty}^{f} {}^{(n)}(y)$ exists and by Step 2 $y^* = f(y^*)$. By Step 3 we see that $y^* = a \in X$ for $\forall y \in X$. ■

The completeness of the metric space plays a central role in Contraction Mapping theorem, because it gives us the limit x^* of the sequence $(f^n(x))$, which turns out to be the fixed point of f we are searching for.

Without completeness of the metric space, the result is not true. For example, in $(\mathbb{R} \setminus \{0\}, d_2)$, the function $f(x) = x/2$ is a contraction, but it does not have a fixed point.

11 Continuity of Functions

11.1 Limits of Functions

So far we defined the notion of limit for sequences. Now let's do it for functions.

Definition 11.1. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let S be a subset of X, function $f : S \to Y$, and $x_0 \in S'$ (a limit point of S).*

We say that y_0 is a ***limit of f at*** *x_0, iff $\forall\ \varepsilon > 0$, $\exists\ \delta > 0$ s.t.*

$$f((B_\delta(x_0) \cap S) \backslash \{x_0\}) \subset B_\varepsilon(y_0).$$

In this case, we denote $\lim_{x \to x_0} f(x) = y_0$.

Notice that x_0 has to be a limit point of the domain S, since we have to make sure that there exists $x \in S$ s.t. $0 < d(x, x_0) < \delta$, no matter how small δ is. However, x_0 is allowed to be outside the domain S[14]. For example, consider the function $f : (0, 1) \to \mathbb{R}$ defined as $f(x) = 2x$. It makes sense to talk about the limit of f at 1, although 1 is not in the domain of f. In this case, $\lim_{x \to 1} f(x) = 2$. (In $\mathbb{R}^k$, we always use the Euclidean distance d_2 by default, unless stated otherwise.)

The concept of the limit of f at x_0 has nothing to do with the value of f at x_0. Instead, it captures the behavior of the function f only nearby x_0 but not at x_0. For example, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) := \begin{cases} 2x, \text{ if } x \neq 1 \\ 0, \text{ if } x = 1 \end{cases}$$

Notice that $f(1) = 0$, but $\lim_{x \to 1} f(x) = 2$. (Again, d_2 is used by default.)

Similar to the limit of a sequence, we can use triangle inequality to show that limit of f at x_0 is unique, if exists. This enables us to talk about "the" limit of f at x_0, and use the notation $\lim_{x \to x_0} f(x)$ without ambiguity.

The next theorem reveals the relation of the limit of a function to the limit of sequences.

Theorem 11.2. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let S be a subset of X, function $f : S \to Y$, and $x_0 \in S'$. Then the limit of f at x_0 is y_0 iff the sequence $(f(x_n))$ converges to y_0 for any sequence (x_n) in $S \backslash \{x_0\}$ that converges to x_0.*

Proof. $\Rightarrow$:

Take any sequence (x_n) in S convergent to x_0 s.t. $x_n \neq x_0$ for any n.

WTS: $f(x_n) \to y_0$

Take any $\varepsilon > 0$. We want to find N s.t. $d_Y(f(x_n), y_0) < \varepsilon$ for any $n > N$.

Because $\lim_{x \to x_0} f(x) = y_0$, there exists $\delta > 0$ s.t. $d_Y(f(x), y_0) < \varepsilon$ for any $x \in S$ with $0 < d_X(x, x_0) < \delta$.

Because $x_n \to x_0$, there exists N s.t. $d_X(x_n, x_0) < \delta$ for any $n > N$.

Take any $n > N$, because $x_n \neq x_0$, we have $0 < d_X(x_n, x_0) < \delta$, and so $d_Y(f(x_n), y_0) < \varepsilon$.

$\Leftarrow$:

[14] To allow x_0 to be outside the domain S is an important generality to maintain. We know that the derivative of a single variable function f at x_0 is defined as

$$f'(x_0) := \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Clearly, the point x_0 is not in the domain of the slope function $s(x) := \frac{f(x) - f(x_0)}{x - x_0}$, since the denominator cannot be 0. However, we are still able to talk about the limit of $s(x)$ at x_0.

Suppose that y_0 is not a limit of f at x_0. Then there exists $\hat{\varepsilon} > 0$ s.t. there is no $\delta > 0$ s.t. $d_Y(f(x), y_0) < \hat{\varepsilon}$ for any $x \in S$ with $0 < d_X(x, x_0) < \delta$.

Then for any $n \in \mathbb{N}$, we can find $x_n \in S$ s.t. $0 < d_X(x_n, x_0) < 1/n$, but $d_Y(f(x_n), y_0) \geq \hat{\varepsilon}$.

Clearly, the sequence $x_n \to x_0$, and $x_n \neq x_0$ for any n, but $(f(x_n))$ does not converge to y_0. This contradicts our assumption. ■

Because of the close link of the function limit to sequence limit, as revealed by the theorem above, Proposition 8.6 and Proposition 8.7 also works for function limit. We state them as the following two propositions.

Proposition 11.3. *Let (X, d_X) be a metric space, S be a subset of X, and $x_0 \in S'$. Let f be a function from S to $\mathbb{R}^k$. Then $\lim_{x\to x_0} f(x)$ exists iff $\lim_{x\to x_0} f_i(x)$ exists for any $i = 1, 2, \dots k$. Furthermore, when the limit exist, the i-th coordinate of $\lim_{x\to x_0} f(x)$ is equal to $\lim_{x\to x_0} f_i(x)$.*

Note that f_i is used to denote the i-th coordinate of f, and so f_i is a function from S to $\mathbb{R}$.

Proposition 11.4. *Let (X, d_X) be a metric space, S be a subset of X, and $x_0 \in S'$. Let f and g be functions from S to $\mathbb{R}$ s.t. $\lim_{x\to x_0} f(x)$ and $\lim_{x\to x_0} g(x)$ exist. Then*
(1) $\lim_{x\to x_0}[f(x) + g(x)] = \lim_{x\to x_0} f(x) + \lim_{x\to x_0} g(x)$,
(2) $\lim_{x\to x_0}[f(x)\, g(x)] = \lim_{x\to x_0} f(x) \cdot \lim_{x\to x_0} g(x)$
(3) $\lim_{x\to x_0}(1/f) = 1/\lim_{x\to x_0} f(x)$, *if* $\lim_{x\to x_0} f(x) \neq 0$.

11.2 Continuity

Definition 11.5. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let S be a subset of X, function $f : S \to Y$, and $x_0 \in S$.*

*The function f is said to be **continuous at** x_0 iff $\forall\ \varepsilon > 0$, $\exists\ \delta > 0$ s.t.*

$$f(B_\delta(x_0) \cap S) \subset B_\varepsilon(f(x_0)).$$

*The function f is said to be a **continuous function** iff f is continuous at x_0 for all $x_0 \in S$.*

Here we allow $x = x_0$, which is different from the definition of $\lim_{x\to x_0} f(x)$. Also notice that we require x_0 to be in the domain S of the function f (otherwise $f(x_0)$ is not defined), but not necessarily a limit point of the domain. In fact, if $x_0 \in S$ is not a limit point of S, i.e. x_0 is an isolated point of S, then f is continuous at x_0 by definition.

The relation of continuity to the limit of functions is stated below.

Proposition 11.6. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let S be a subset of X, function $f : S \to Y$, and $x_0 \in S \cap S'$. Then the function f is continuous at x_0 iff $\lim_{x\to x_0} f(x) = f(x_0)$.*

Notice that this equivalence only works for x_0's that are both in the domain (to ensure "f is continuous at x_0" is defined) and are a limit point of the domain (to ensure $\lim_{x\to x_0} f(x)$ is defined).

Theorem 11.7. *Let (X, d_X) and (Y, d_Y) be metric spaces. Let S be a subset of X, function $f : S \to Y$, and $x_0 \in S$. Then the function f is continuous at x_0 iff $f(x_n) \to f(x_0)$ for any sequence (x_n) in S convergent to x_0.*

Simply put, f is continuous iff $x_n \to x_0$ implies $f(x_n) \to f(x_0)$.

Proof. $\Rightarrow$:

Take any sequence (x_n) in S convergent to x_0.

WTS: $f(x_n) \to f(x_0)$

Take any $\varepsilon > 0$. We want to find N s.t. $d_Y(f(x_n), f(x_0)) < \varepsilon$ for any $n > N$.

Because f is continuous at x_0, there exists $\delta > 0$ s.t. $d_Y(f(x), f(x_0)) < \varepsilon$ for any $x \in S$ with $d_X(x, x_0) < \delta$.

Because $x_n \to x_0$, there exists N s.t. $d_X(x_n, x_0) < \delta$ for any $n > N$.

Take any $n > N$, we have $d_X(x_n, x_0) < \delta$, and so $d_Y(f(x_n), f(x_0)) < \varepsilon$.

$\Leftarrow$:

If $x_0 \notin S'$, then f is continuous at x_0 by definition.

If $x_0 \in S'$, then by Theorem 11.2, the condition "$f(x_n) \to f(x_0)$ for any sequence (x_n) in S convergent to x_0" implies that $\lim_{x \to x_0} f(x) = f(x_0)$. Then by Proposition 11.6, f is continuous at x_0. ■

The theorem above reveals a direct link of continuity to convergence of sequences. Therefore, many results in convergence of sequences also apply here. For example, the following results are counterparts of Proposition 8.6 and 8.7.

Proposition 11.8. *Let (X, d_X) be a metric space, S be a subset of X, and $x_0 \in S$. Consider a function $f : S \to \mathbb{R}^k$. Then f is continuous at x_0 iff $f_i : S \to \mathbb{R}$ is continuous at x_0 for any $i = 1, 2, \ldots, k$.*

Proposition 11.9. *Let (X, d_X) be a metric space, S be a subset of X, and $x_0 \in S$. Let f and g be functions from S to $\mathbb{R}$ that are continuous at x_0. Then*

(1) $f + g$ is continuous at x_0,

(2) $f \cdot g$ is continuous at x_0, and

(3) $1/f$ is continuous at x_0 if $f(x_0) \neq 0$.

The next theorem shows that the compound of two continuous function is also continuous.

Theorem 11.10. *Let (X, d_X), (Y, d_Y), (Z, d_Z) be metric spaces. Let set S be a subset of X, and $f : S \to Y$ be a continuous function. Let T be a set s.t. $f(S) \subset T \subset Y$, and $g : T \to Z$ be a continuous function. Then $g \circ f : S \to Z$ is a continuous function.*

The proof is left as an exercise.

It can be shown that many commonly used functions, such as x^α, $\ln x$, e^x, and $\sin x$, are all continuous in their domain. Since by the two results above continuity is preserved under addition, multiplication, and compounding, then roughly speaking, all functions constructed using those "common" functions are continuous.

The next theorem provides yet another equivalent definition of continuity, which is known as the topological definition of continuous functions.

Theorem 11.11. *Let (X, d_X) and (Y, d_Y) be metric spaces, and function $f : X \to Y$. The function f is a continuous function iff $f^{-1}(E)$ is open in (X, d_X) for any set E open in (Y, d_Y).*

Simply put, f is continuous iff its inverse image of an open set is also open.

Different from the previous sequential characterization that works for functions continuous at a single particular point, this characterization only works for functions that are continuous everywhere.

Proof. $\Rightarrow$:

Take any $x \in f^{-1}(E)$. We want to find $\delta > 0$ s.t. $B_\delta(x) \subset f^{-1}(E)$.

Because $x \in f^{-1}(E)$, we have $f(x) \in E$. Because E is open in (Y, d_Y), there exists $\varepsilon > 0$ s.t. $B_\varepsilon(f(x)) \subset E$.

Because f is continuous at x, there exists $\delta > 0$ s.t. $f(B_\delta(x)) \subset B_\varepsilon(f(x)) \subset E$. Therefore, we have $B_\delta(x) \subset f^{-1}(E)$.

$\Leftarrow$:

Take any $\varepsilon > 0$. We want to find $\delta > 0$ s.t. $f(B_\delta(x)) \subset B_\varepsilon(f(x))$.

Because $B_\varepsilon(f(x))$ is open in (Y, d_Y), the set $f^{-1}(B_\varepsilon(f(x)))$ is open in (X, d_X). Because $f(x) \in B_\varepsilon(f(x))$, we have $x \in f^{-1}(B_\varepsilon(f(x)))$. Therefore, the exists $\delta > 0$ s.t. $B_\delta(x) \subset f^{-1}(B_\varepsilon(f(x)))$. As a result, $f(B_\delta(x)) \subset B_\varepsilon(f(x))$. ■

As a corollary, f is continuous iff its inverse image of a closed set is also closed. To prove this, we only need to use the fact that the complement of an open set is closed, and that $f^{-1}(E^c) = \left(f^{-1}(E)\right)^c$.

Although taking the inverse image of a continuous function preserves openness and closedness, the image of an open set (or closed set) may not be open (or closed). For example, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) := x^2$. The image of the open set $(-1, 1)$ under f is $[0, 1)$, which is not open in the codomain $\mathbb{R}$. For another example, consider the function $g : \mathbb{R}_{++} \to \mathbb{R}$ defined as $g(x) := 1/x$. The image of the closed set $[1, +\infty)$ under g is $(0, 1]$, which is not closed in the codomain $\mathbb{R}$.

11.3 Weierstrass Theorem

The following theorem states that a continuous image of a compact set is also compact.

Theorem 11.12. *Let (X, d_X) and (Y, d_Y) be metric spaces, and function $f : X \to Y$ is continuous. Then $f(K)$ is compact in (Y, d_Y) for any K compact in (X, d_X).*

Proof. Take any K compact in (X, d_X). WTS: $f(K)$ is compact in (Y, d_Y)

Take any open cover $\{E_\alpha\}_{\alpha \in A}$ of $f(K)$. We want to find a finite $B \subset A$ s.t. $\{E_\alpha\}_{\alpha \in B}$ is an open cover of $f(K)$.

First, I claim that $\left\{f^{-1}(E_\alpha)\right\}_{\alpha \in A}$ is an open cover of K.

Because each E_α is open in (Y, d_Y), the set $f^{-1}(E_\alpha)$ is open in (X, d_X). Take any $x \in K$. We have $f(x) \in f(K)$. Because $\{E_\alpha\}_{\alpha \in A}$ covers $f(K)$, there exists some $\hat{\alpha} \in A$ s.t. $f(x) \in E_{\hat{\alpha}}$. So $x \in f^{-1}(E_{\hat{\alpha}})$. Therefore, $\left\{f^{-1}(E_\alpha)\right\}_{\alpha \in A}$ is an open cover of K.

Because K is compact, there exists some finite set $B \subset A$ s.t. $\left\{f^{-1}(E_\alpha)\right\}_{\alpha \in B}$ covers K. Now, it is sufficient to show that $\{E_\alpha\}_{\alpha \in B}$ is an open cover of $f(K)$.

Take any $y \in f(K)$. There exists $\hat{x} \in K$ s.t. $f(\hat{x}) = y$. Because $\left\{f^{-1}(E_\alpha)\right\}_{\alpha \in B}$ covers K, there exists $\hat{\alpha} \in B$ s.t. $\hat{x} \in f^{-1}(E_{\hat{\alpha}})$. Therefore, $y = f(\hat{x}) \in E_{\hat{\alpha}}$. Therefore, $\{E_\alpha\}_{\alpha \in B}$ is an open cover of $f(K)$. ■

Although taking the image of a continuous function preserves compactness, the inverse image of a compact set may not be compact. For example, consider the function $f : \mathbb{R}_{++} \to \mathbb{R}$ defined as $g(x) := 1/x$. The inverse image of the compact set $[0, 1]$ under f is $[1, +\infty)$, which is not compact.

The next result states that a compact set in $(\mathbb{R}, d_2)$ has a maximum and a minimum.

Claim 11.13. *Let K be a compact set in $(\mathbb{R}, d_2)$. Then there exists $x^* \in K$ s.t. $x^* \geq x$ for any $x \in K$, and there exists $x_* \in K$ s.t. $x_* \leq x$ for any $x \in K$.*

Proof. Because K is compact in $(\mathbb{R}, d_2)$, we know that K is bounded, i.e. there exists some $B_r(x) \supset K$. Therefore, $x + r$ is an upper bound of K. By l.u.b. property of $\mathbb{R}$, there exists the least upper bound $\sup K$.

Now I claim that $\sup K \in K$.

Suppose $\sup K \notin K$. Then $\sup K$ is a limit point of K, because for any $\varepsilon > 0$, there exists $x \in K$ s.t. $x > \sup K - \varepsilon$. Because K is closed, we know that $K' \subset K$, and thus $\sup K \in K$. This contradicts the hypothesis we started with.

Let $x^* := \sup K$, and we have $x^* \in K$ and $x^* \geq x$ for any $x \in K$.

Symmetrically, let $x_* := \inf K$, and we can show that $x_* \in K$ and $x_* \leq x$ for any $x \in K$. ■

Combining the two results above, we have Weierstrass theorem stated below.

Theorem 11.14 (Weierstrass). *Let (X, d_X) be a metric space, and function $f : X \to \mathbb{R}$ is continuous. Let S be a compact set in (X, d_X). There exists $x^* \in S$ s.t. $f(x^*) \geq f(x)$ for any $x \in S$, and there exists $x_* \in S$ s.t. $f(x_*) \leq f(x)$ for any $x \in S$.*

Again, when we say $f : X \to \mathbb{R}$ is continuous, we use the Euclidean metric d_2 in the codomain $\mathbb{R}$ by default.

Proof. By Theorem 11.12, we know that $f(S)$ is compact in $(\mathbb{R}, d_2)$. Therefore, there exists $y^* \in f(S)$ s.t. $y^* \geq f(x)$ for any $x \in S$. By definition of the image $f(S)$, there exists $x^* \in S$ s.t. $f(x^*) = y^*$, and therefore $f(x^*) \geq f(x)$ for any $x \in S$.

Symmetrically, we can find x_*. ■

In economics, it is standard to assume that every entity in the economy is maximizing some objective function. Weierstrass theorem implies that each entity's maximization problem must have a solution, if the entity's objective function is continuous and the set of alternatives available to the entity is compact.