Math 411 - Fall 2020

Based on Lectures by M. Mirek
Notes Taken by Richard K. Yu

October 2021

Contents

1 Number Systems and Set Theory

1.1 Limit Superior and Inferior of Sets

limsup and liminf

The limit superior and limit inferior of a sequence of sets $\mathcal{E} = \{E_n : n \in \mathbb{N}\}$ is defined as follows

$$\limsup_{n\to\infty} E_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} E_n$$

$$\liminf_{n\to\infty} E_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} E_n$$

If $x \in \limsup E_n$, it is equivalent to $x \in \bigcup_{n=k}^{\infty} E_n$ for all $k \in \mathbb{Z}_+$. We can also say for every $k \in \mathbb{N}$ there is $n \geq k$ such that $x \in E_n$ or $x \in E_n$ for infinitely many $n \in \mathbb{Z}_+$.
If $x \in \liminf E_n$, it is equivalent to $x \in \bigcap_{n=k}^{\infty} E_n$ for some $k \in \mathbb{Z}_+$. We can also say $x \in E_n$ for all but finitely many $n \in \mathbb{Z}_+$. Note that $\liminf E_n \subseteq \limsup E_n$.

We also have the relationship $\liminf E_n \subseteq \limsup E_n$.
Example. Let $E_n = (-\frac{1}{n}, 1 - \frac{1}{n}]$. Then $\liminf E_n = \limsup E_n = [0, 1)$.
If $E_n = (\frac{(-1)^n}{n}, 1 - \frac{(-1)^n}{n}]$, then $\liminf E_n = (0, 1)$ and $\limsup E_n = [0, 1]$.
If $E_n = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$, then

$$\bigcup_{k\geq n} E_k = \mathbb{Q} \cap [0, 1] \quad \text{and} \quad \bigcap_{k\geq n} E_k = \{0, 1\}$$

Thus $\liminf E_n = \{0, 1\}$ and $\limsup E_n = \mathbb{Q} \cap [0, 1]$.

1.2 Set Theoretic Treatment of Sequences

Definition (Sequence). A sequence in a set X is a mapping from $\mathbb{Z}_+$ into X. A finite sequence is a mapping from $\{1, 2, \ldots, n\}$ into X, where $n \in \mathbb{Z}_+$. If we have $f : \mathbb{Z}_+ \to X$ with $f(n) = x_n \in X$, we will abbreviate to $(x_n)_{n\in\mathbb{N}}$.
If $f : \mathbb{N} \to X$ is a sequence and $g : \mathbb{N} \to \mathbb{N}$ satisfies $g(m) < g(n)$ for all $m < n$, the composition $f \circ g$ is called a subsequence of f.
If $(X_\alpha)_{\alpha\in A}$ is an indexed family of sets, their Cartesian product

$$\prod_{\alpha\in A} X_\alpha$$

is the set of all maps $f : A \to \bigcup_{\alpha\in A} X_\alpha$ such that $f(\alpha) \in X_\alpha$ for every $\alpha \in A$.
Definition (Coordinate map). If $X = \prod_{\alpha\in A} X_\alpha$ and $\alpha \in A$, we define the α^{th} projection or

coordinate map $\pi_\alpha : X \to X_\alpha,\ \ \pi_\alpha(f) = f(\alpha)$. In other words, π_α takes an input from the space of functions and outputs the value of that function at its α^{th} coordinate.
We will also write $x = (x_\alpha)_{\alpha\in A} \in X = \prod_{\alpha\in A} X_\alpha$ instead of f and x_α instead of $f(\alpha)$.
Suppose $A = \{1, 2, \ldots, n\}$. Then

$$\begin{aligned} X &= \prod_{j=1}^{n} X_j = X_1 \times X_2 \times \cdots \times X_n \\ &= \{(x_1, \ldots, x_n) : x_j \in X_j \text{ for all } j \in \{1, 2, \ldots, n\}\} \end{aligned}$$

Suppose $A = \mathbb{Z}_+$. Then

$$\begin{aligned} X &= \prod_{j=1}^{\infty} X_j = X_1 \times X_2 \times \cdots \\ &= \{(x_1, x_2, \ldots) : x_j \in X_j \text{ for all } j \in \mathbb{Z}_+\} \end{aligned}$$

If the sets X_α are all equal to some fixed set Y, we denote

$$X = \prod_{\alpha\in A} X_\alpha = Y^A$$

where Y^A can be understood as the set of all mappings from A to Y.
Definition (Partial Ordering). A partial ordering on a nonempty set X is a relation R on X with the following with the following properties

(a) Transitivity. If xRy and yRz, then xRz

(b) Antisymmetry. If xRy and yRx, then $x = y$

(c) Reflexivity. xRx for all $x \in X$.

(d) If R also satisfies: for all $x, y \in X$, xRy or yRx, then R is called a linear or total ordering on X.

For example, if E is any set, then $\mathscr{P}(E)$ is partially ordered by inclusion.
Example 2. $\mathbb{R}$ is linearly ordered by its usual ordering.

1.3 More on Relations and Ordering

Let $E = \{1, 2, 3, 4\}$. We may represent a graph of partial order by inclusion using a **Hasse Diagram**.
Consider the subset of $\mathscr{P}(E)$, $\{\varnothing, \{1\}, \{1, 2\}, \{1, 2, 3\}\}$. This is linearly ordered.
By contrast, there is no relation on the subset $\{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. This situation is called incomparable.

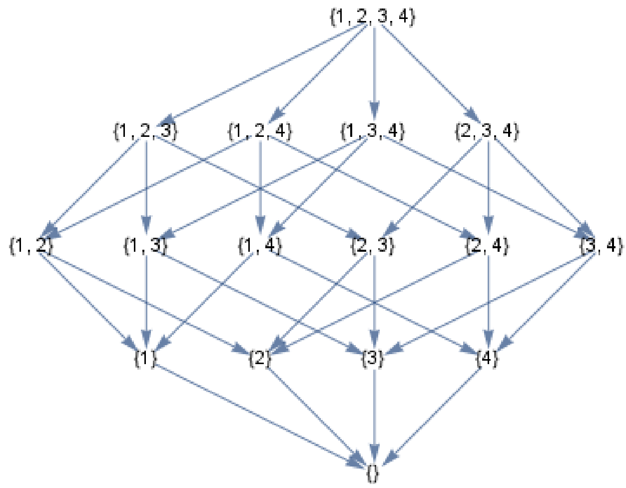


Figure 1: A Hasse diagram for E.

We say that $(X, \leq)$ is a poset if the relation "$\leq$" is a partial ordering on X, or $(X, \leq)$ is partially ordered by "$\leq$".
Two posets $(X, \leq_X), (Y, \leq_Y)$ are said to be order isomorphic if there is a bijection $f : X \to Y$ such that $X_1 \leq_X X_2$ if and only if $Y_1 \leq_Y Y_2$. We will write $x < y$ in a poset $(X, \leq)$ if and only if $x \leq y$ and $x \neq y$.
If $(X, \leq)$ is a poset, then we have the following definitions.

1. A maximal (resp. minimal) element of X is an element $x \in X$ such that if $y \in X$ and $x \leq y$ (resp. $x \geq y$) then $x = y$.

2. A greatest (resp. least) element of X is an element $x \in X$ such that $y \leq x$ for all $y \in X$ (resp. s.t. $x \leq y$ for all $y \in X$).

3. Let A be a poset and $A \subseteq X$. An element $x \in X$ is an upper bound of A (resp. lower bound of A) if $a \leq x$ for all $a \in A$. (resp. $x \leq a$ for all $a \in A$).
Remark. x need not be an element in A.

4. Let A be a bounded from above subset of X. We say that an element $x_0 \in X$ is the least upper bound for A or the supremum for A ($x_0 = \sup A$) if the following holds:
   (a) $a \leq x_0$ for all $a \in A$. This means x_0 is an upper bound.
   (b) If $a \leq x$ for all $a \in A$, then $x_0 \leq x$. This means that x_0 is the least upper bound.

5. Let A be a bounded from below subset of X. We say that an element $x_0 \in X$ is the greatest lower bound for A or infimum of A ($x_0 = \inf A$) if
   (a) $x_0 \leq a$ for all $a \in A$

(b) If $x \leq a$ for all $a \in A$, then $x \leq x_0$.

Definition (Well-Ordering). If $(X, \leq)$ is linearly ordered and every nonempty subset of X has a (necessarily unique) minimal element, X is said to be well-ordered, by $\leq$ and $\leq$ is called well-ordering on X.
Example. The positive integers, $\mathbb{Z}_+$ have this property.

1.4 The Axiom of Choice

Below, we detail the Axiom of Choice and some of its equivalent formulations.

Equivalent Formulations of the Axiom of Choice

(A) **The Axiom of Choice.** If $\{X_\alpha\}_{\alpha \in A}$ is a nonempty collection of nonempty sets, then

$$\prod_{\alpha \in A} X_\alpha \neq \varnothing$$

Corollary. If $\{X_\alpha\}_{\alpha \in A}$ is a disjoint collection of nonempty sets, there exists a set $Y \subseteq \bigcup_{\alpha \in A} X_\alpha$ such that $Y \cup X_\alpha$ contains precisely one element for each $\alpha \in A$.
Proof. Take $Y = f[A]$ for $f \in \prod_{\alpha \in A} X_\alpha \neq \varnothing$. Then $Y \cap X_\alpha = \{f(\alpha)\}$ and $f(\alpha) \in X_\alpha$.

□

(B) **The Hausdorff Maximal Principle.** Every partially ordered set has a maximal linearly ordered set. In other words, this principle says that if $(X, \leq)$ is a poset, there exists a $E \subseteq X$ that is linearly ordered by $\leq$ such that no subset of X that properly includes E is linearly ordered by $\leq$.

(C) **Kuratowski-Zorn's Lemma.** If X is a partially ordered set and every linearly ordered subset of X has an upper bound, then X has a maximal element.

(D) **Well-Ordered Principle.** Every nonempty set X can be well-ordered.

Theorem. The principles (A), (B), (C), and (D) are logically equivalent. We will show $(A) \implies (B) \implies (C) \implies (D) \implies (A)$. To begin, we will need to prove some other theorems.
Theorem. Let $(X, \leq)$ be a poset such that every linearly ordered subset of X has a supremum (in X). Then, every function $f : X \to X$ obeying

$$x \leq f(x) \text{ for all } x \in X$$

has a fixed point, that is, there is $x^* \in X$ such that

$$f(x^*) = x^*$$

Proof. Clearly, the empty set $\varnothing$ is linearly ordered. So it has a supremum in X, that is, $a = \sup \varnothing \in X$ which is the smallest element in $(X, \leq)$.
Let $\mathcal{A} \subseteq \mathscr{P}(X)$ be the family of all $A \subseteq X$ so that

(a) $a \in A$

(b) $f[A] \subseteq A$

(c) If $L \subseteq A$ is a linearly ordered set in $(X, \leq)$, then $\sup L \in A$.

Observe that $\mathcal{A} \neq \varnothing$ since $X \in \mathcal{A}$. Then consider

$$A_* = \bigcap_{A \in \mathcal{A}} A$$

It is easy to see that A_* satisfies (a), (b), (c) $\iff A_* \in \mathcal{A}$.

$$f(A_*) = f\left[\bigcap_{A \in \mathcal{A}} A\right] \subseteq \bigcap_{A \in \mathcal{A}} f[A] \subseteq \bigcap_{A \in \mathcal{A}} A = A_*$$

Our aim will be to prove that A_* is a linearly ordered set in $(X, \leq)$.
Consider

$$B = \{x \in A_* : \text{ if } y \in A_* \text{ and } y < x, \text{ then } f(y) \leq x\}$$

We shall show that $B \in \mathcal{A}$.
(Property (a) for B) Observe that $a \in B$ since $a \in A_*$ and a is the smallest element of $(X, \leq)$, so there is no element $y \in A_*$, so that $y < x$. Thus $a \in B$, hence (a) holds for B.
Fix $x \in B$ and define

$$B_x = \{z \in A_* : z \leq x \text{ or } f(x) \leq z\}$$

We show that $B_x \in \mathcal{A}$ for every $x \in B$.
(Property (a) for B_x) $a \in B_x$ since $a \leq x$ for all $x \in X$, and $a \in A_*$.
(Property (b) for B_x) Fix $z \in B_x$ and we show that $f(z) \in B_x \iff f(B_x) \subseteq B_x$. Since $z \in B_x$, so $z \leq x$ or $f(x) \leq z$. If $z < x$, then by the definition of B, we see that $f(z) \leq x$, so $f(z) \in B_x$. (Note that $f(z) \in A_*$).
Otherwise $x = z$ or $f(x) \leq z$.
If $x = z$, then $f(x) = f(z)$. Thus $f(z) \in B_x$.
If $f(x) \leq z$, then $f(x) \leq z \leq f(z)$ thus we also have that $f(z) \in B_x$. So, (b) follows, that is, $B_x \subseteq A_*$.
(Property (c) for B_x) Let $L \subseteq B_x$ be a linearly ordered set. We have the prove that

$\sup L \in B_x$.
If all elements $z \in L$ satisfy $z \leq x$, then $\sup L \leq x$ because L is a linearly ordered set. Since $\sup L \in L$, so $\sup L \in B_x$. If $f(x) \leq z$ for some $z \in L$, then $f(x) \leq z \leq \sup L$, then $\sup L \in B_x$.
Thus, $B_x \in \mathcal{A}$ for all $x \in B$ since

$$A_* = \bigcap_{A \in \mathcal{A}} A \subseteq B_x \subseteq A_*$$

this implies that

$$B_x = A_* \text{ for all } x \in B$$

This means that

$$z \leq x \text{ or } f(x) \leq z \text{ for all } x \in B \text{ and } z \in A_* \qquad (*)$$

(Property (b) for B) We show $f[B] \subseteq B$. Let $x \in B$, we show $f(x) \in B$. Recall that

$$B = \{x \in A_* : \text{ if } y \in A_* \text{ and } y < x, \text{ then } f(y) \leq x\}$$

Let $y \in A_*$ be such that $y < f(x)$. By $(*)$ we have $y \leq x$. If $y < x$, then by the definition of B, we have $f(y) \leq x \leq f(x)$, so $f(x) \in B$.
If $x = y$, then $f(x) = f(y)$ so we also have that $f(x) \in B$. Thus $f[B] \subseteq B$ as claimed.
(Property (c) for B) Let $L \subseteq B$ be a linearly ordered set in B. We have to prove that $\sup L \in B$.
Let $y \in A_*$ s.t. $y < \sup L$, then there is $x \in L$ such that $x \not\leq y$. By $(*)$ we have $y \leq x$, in fact $y < x$. By the definition of B, $f(y) \leq x$. Since $x \in L$ so $f(x) \leq x \leq \sup L$, we see $\sup L \in B$.
We have shown that $B \in \mathcal{A}$.

$$A_* = \bigcap_{A \in \mathcal{A}} A \subseteq B \subseteq A_* \text{ thus } B = A_*$$

Hence by $(*)$, for every $x, z \in A_*$, we have that $z \leq x$ or $f(x) \leq z$. So $z \leq x$ or $x \leq z$. Now, it is easy to see that $x^* = \sup A_* \in A_*$ and x^* is a fixed point for f by (b).

$$x^* \leq \underbrace{f(x^*)}_{\in A_*} \leq x^* = \sup A_*$$

which shows us that $f(x^*) = x^*$.

□

Theorem. If $(X, \leq)$ is a poset such that every linearly ordered set has a supremum, then X contains a maximal element.

Proof. Suppose, for contradiction, that there is no maximal element. So for every $x \in X$, there is $y \in X$ so that $x < y$. In other words, for all $x \in X$, the set

$$A_x = \{y \in X : x < y\} \neq \varnothing$$

Thus, the set $\prod_{x \in X} A_x \neq \varnothing$ by the axiom of choice.
Take $f \in \prod_{x \in X} A_x$, then $f(x) \in A_x$ for all $x \in X$. So, $x < f(x)$. By the previous theorem, since $x \leq f(x)$ for all $x \in X$, then there is $x^* \in X$ such that

$$x^* < f(x^*) = x^*$$

This is a contradiction (↯).

□

Theorem. We show (A) $\implies$ (B). If $(X, \leq)$ is a poset, then X has a maximal linearly ordered set.
Proof. Let $\mathcal{L}$ be a set of all linearly ordered subsets in $(X, \leq)$. Notice that $(\mathcal{L}, \subseteq)$ is a poset ordered by inclusion. Let $\mathcal{M} \subseteq \mathcal{L}$ be any linearly ordered set (by inclusion).
Notice that $\bigcup \mathcal{M}$ is a linearly ordered set in $(X, \leq)$. To see this, take $x, y \in \bigcup \mathcal{M}$. We must have $x \in M_x$ and $y \in M_y$ for some $M_x, M_y \in \mathcal{M}$. Without loss of generality, the linear ordering of $\mathcal{M}$ implies that $M_x \subseteq M_y$. Hence $x, y \in M_y$ and $x \leq y$ or $y \leq x$. Further, $\bigcup \mathcal{M}$ is the supremum for $\mathcal{M}$. From the previous theorem, there exists a maximal element $L \in (\mathcal{L}, \subseteq)$ which is the maximal linearly ordered set in $(X, \leq)$.

□

Theorem. We now show that (B) $\implies$ (C), or that the Hausdorff Maximal Principle implies Zorn's Lemma.
Proof. Let $(X, \leq)$ be a poset where (B) is true. Then there is a maximal linearly ordered set $L \subseteq X$.
From the assumption in statement (C), L has an upper bound in X. Let $a \in X$ be the upper bound for L. From the maximality of L in X, we must have $a \in L$ (or else $L \cup \{a\}$ contradicts the maximality of L).
Then a is a maximal element of X. If we take $x \in X \setminus L$ such that $a < x$, then $a = x$, otherwise, we consider $L \cup \{x\}$ which is the linearly ordered set that contains L and this contradicts the maximality of L.

□

Theorem. We now show that (C) $\implies$ (D).
Proof. Let W be the collection of well-orderings of subsets of X defined by

$$W = \{(E, \leq) : E \subseteq X \text{ and } \leq \text{ is a well-ordering on } E\}$$

and define a partial ordering on W as follows:

If the relations $\leq_1$ and $\leq_2$ are well-orderings on E_1 and E_2, then $\leq_1$ precedes $\leq_2$ in the partial order if

1. $\leq_1$ extends $\leq_2$. i.e. $E_1 \subseteq E_2$ and $\leq_1$ and $\leq_2$ agree on E_1.
2. If $x \in E_2 \setminus E_1$, then $y \leq_2 x$ for all $y \in E_1$.

It is easy to see that the hypotheses of Zorn's Lemma (C) are satisfied on W. We take $\mathcal{L} \subseteq W$ to be a linearly ordered set in W and we show that $\bigcup \mathcal{L}$ is an upper bound for $\mathcal{L}$. Then (C) implies that there is a "$\leq$" maximal element in W.
This must be a well-ordering on X itself. If $\leq$ is a well-ordering on a proper subset $E \subsetneq X$ and $x_0 \in X \setminus E$, then $\leq$ can be extended to a well-ordering on $E \cup \{x_0\}$ by declaring that $x \leq x_0$ for all $x \in E$, but this is a contradiction since $(E, \leq)$ is a maximal element of W.

□

Theorem. Finally, we show that (D) $\implies$ (A).
Proof. Suppose that $(X_\alpha)_{\alpha \in \mathcal{A}}$ is a nonempty collection of nonempty sets. Consider

$$X = \bigcup_{\alpha \in \mathcal{A}} X_\alpha$$

Using (D), we pick a well-ordering on X. For any set, let $f(\alpha)$ be the minimal element of X_α. Then

$$f \in \prod_{\alpha \in \mathcal{A}} X_\alpha \neq \varnothing$$

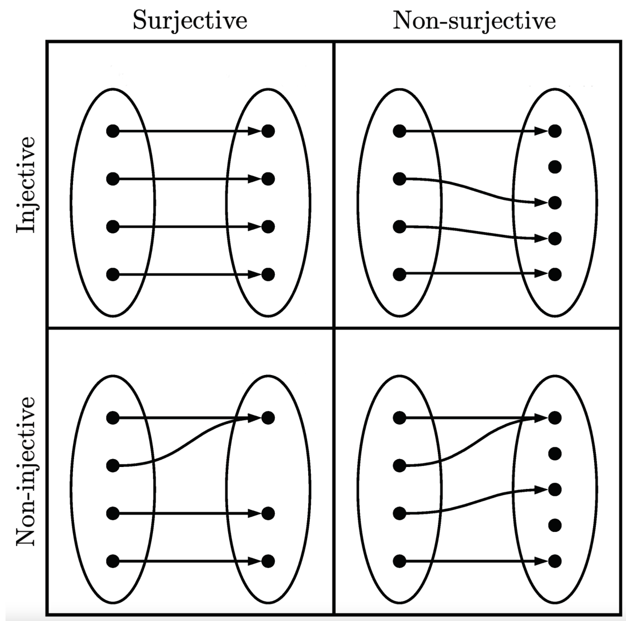
□

1.5 Cardinality

If X and Y are nonempty sets, we define the expressions

1. $\operatorname{card}(X) \leq \operatorname{card}(Y)$
2. $\operatorname{card}(X) = \operatorname{card}(Y)$
3. $\operatorname{card}(X) \geq \operatorname{card}(Y)$

to mean that there exists $f : X \to Y$ which is injective, bijective, or surjective, respectively.



We also define $\text{card}(X) < \text{card}(Y)$ to mean that there is an injection but no bijection.
We also have $\text{card}(\varnothing) < \text{card}(X)$ and $\text{card}(X) > \text{card}(\varnothing)$ for all X.
Proposition 1. $\text{card}(X) \leq \text{card}(Y)$ if and only if $\text{card}(Y) \geq \text{card}(X)$.
Proof. ($\Rightarrow$) If $f : X \to Y$ is injective, we pick $x_0 \in X$ and define

$$g(y) = \begin{cases} f^{-1}(y), \text{ if } y \in f[X] \\ g(y) = x_0, \text{ if } y \in f[X] \end{cases}$$

Then, we see that g is surjective from Y to X.
($\Leftarrow$) If $g : Y \to X$ is surjective, then the sets

1.6 Sets with Cardinality Continuum

A set X is said to have cardinality continuum if $\text{card}(X) = \text{card}(\mathbb{R})$.
We shall write $\text{card}(X) = \mathfrak{c}$ if and only if $\text{card}(X) = \text{card}(\mathbb{R})$.
Theorem. $\text{card}(\mathscr{P}(\mathbb{N})) = \text{card}(\{0,1\}^{\mathbb{N}}) = \text{card}(\mathbb{R})$.
Proof. We first show that $\text{card}(\mathscr{P}(\mathbb{N})) = \text{card}(\{0,1\}^{\mathbb{N}})$.
Define a function $F : \mathscr{P}(\mathbb{N}) \to \{0,1\}^{\mathbb{N}}$ by:

$$F(A) = \mathbb{1}_A \in \{0,1\}^{\mathbb{N}}, \quad \mathbb{1}_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases}$$

Note that $\mathbb{1}_A : \mathbb{N} \to \{0,1\}$. F is a bijection. For $A, B \subseteq \mathbb{N}$ with $A \neq B$, without loss of generality, there exists a $x_0 \in A \setminus B$. Thus, $1 = \mathbb{1}_A(x_0) \neq \mathbb{1}_B(x_0) = 0$ since $x_0 \notin B$.
Hence, $F(A) \neq F(B)$ which shows us that F is injective.

To show that F is surjective, take $\alpha = (\alpha_n)_{n\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$. Also, consider the set

$$E = \{n \in \mathbb{N} : \alpha_n = 1\}$$

Then notice that

$$F(E) = \mathbb{1}_E = \alpha \iff 1 = \alpha_n = \mathbb{1}_E(n)$$

The above statement shows us how for every sequence $\alpha \in \{0,1\}^{\mathbb{N}}$, there exists an $E \subseteq N$ such that $F(E) = \alpha$. The construction of E is the key step to show surjectivity here.
We show that $\text{card}(\mathscr{P}(\mathbb{N})) \leq \text{card}(\mathbb{R})$. Consider $f : \mathscr{P}(\mathbb{N}) \to \mathbb{R}$ defined by:

$$f(A) = \sum_{n\in A} \frac{2}{3^{n+1}}, \quad A \subseteq \mathbb{N}$$

One can show that f is injective (Problem 6 on Homework 3).
Recall that $\text{card}(\mathbb{N}) = \text{card}(\mathbb{Q}) \implies \text{card}(\mathscr{P}(\mathbb{N})) = \text{card}(\mathscr{P}(\mathbb{Q}))$.
We now construct an injection $g : \mathbb{R} \to \mathscr{P}(\mathbb{Q})$ by setting:

$$g(x) = \{r \in \mathbb{Q} : r < x\}, \quad \text{for any } x \in \mathbb{R}$$

Since g is injective, we see that $\text{card}(\mathbb{R}) \leq \text{card}(\mathscr{P}(\mathbb{Q})) = \text{card}(\mathscr{P}(\mathbb{N}))$.
Now by the Cantor-Bernstein Theorem:

$$\text{card}(\mathbb{R}) = \text{card}(\mathscr{P}(\mathbb{N})) = \text{card}(\{0,1\}^{\mathbb{N}})$$

□

Remark. Note the following facts:

$$\text{card}(\mathbb{R} \times \mathbb{R}) = \text{card}(\mathbb{R})$$
$$\text{card}(\{0,1\}^{\mathbb{N}} \times \{0,1\}^{\mathbb{N}}) = \text{card}(\{0,1\}^{\mathbb{N}})$$
$$\text{card}(\mathbb{R}^k) = \text{card}(\mathbb{R}), \quad k \in N \cup \{\infty\}$$

Proposition. (a) If $\text{card}(X) \leq \mathfrak{c}$ and $\text{card}(Y) \leq \mathfrak{c}$, then $\text{card}(X \times Y) \leq \mathfrak{c}$.
(b) If $\text{card}(A) \leq \mathfrak{c}$ and $\text{card}(X_\alpha) \leq \mathfrak{c}$ for any $\alpha \in A$, then $\text{card}(\bigcup_{\alpha\in A} X_\alpha) \leq \mathfrak{c}$.

2 Topology, Sequences, and Series

2.1 Metric Spaces

A metric on a set X is a function $\rho : X \times X \to [0, \infty)$ such that:

(i) $\rho(x, y) = 0 \iff x = y$

(ii) $\rho(x, y) = \rho(y, x)$ for all $x, y \in X$

(iii) $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ for all $x, y, z \in X$

The function $\rho(x, y)$ can be interpreted as the distance from x to y.
A set X equipped with a metric ρ is called a metric space and is denoted (X, ρ).
Examples of Metric Spaces

(i) $\mathbb{R}$ with the metric $\rho(x, y) = |x - y|$.

(ii) $\mathbb{C}$ complex plane. $\rho(z_1, z_2) = |z_1 - z_2|$ where $|z| = (\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2)^{\frac{1}{2}}$

(iii) $\mathbb{R}^d$, $\quad x = (x_1, \ldots, x_d), \quad y = (y_1, \ldots, y_d)$

$$\rho_2(x, y) = \sqrt{\sum_{j=1}^{d} (x_j - y_j)^2}, \quad \text{sometimes called the } \ell^2 \text{ metric.}$$

$$\rho_1(x, y) = \sum_{j=1}^{d} |x_j - y_j|$$

$$\rho_\infty(x, y) = \sup_{1 \leq j \leq d} |x_j - y_j|$$

(iv) If X is any set,

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

This is called the discrete metric. (X, ρ) is called the discrete space.

(v) $\mathbb{R}^\mathbb{N}$, $\quad x = (x_1, x_2, \ldots), \quad y = (y_1, y_2, \ldots)$. The space of all real sequences.

$$\rho(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \min(|x_n - y_n|, 1)$$

The minimum is included to guarantee convergence.

(vi) If ρ is a metric on X and $A \subseteq X$, then $\rho\big|_{A \times A}$ is a metric on A. In other words, a subset of a metric space is a metric space with the original metric restricted to the subset.

(vii) Suppose $(X_1,\rho_1),\dots,(X_d,\rho_d)$ are metric spaces.

$$X=\prod_{i=1}^{d} X_i,\quad x=(x_1,\dots x_d)\in X,\quad y=(y_1,\dots y_d)\in X$$

$$d_1(x,y)=\sum_{j=1}^{d}\rho_j(x_j,y_j)$$

$$d_2(x,y)=\left(\sum_{j=1}^{d}\rho_j(x_j,y_j)^2\right)^{\frac{1}{2}}$$

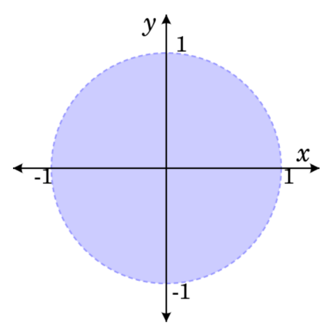
$$d_\infty(x,y)=\sup_{1\le j\le d}\rho_j(x_j,y_j)$$

Let (X,ρ) be a metric space. If $x\in X$ and $r>0$, the open ball of radius r and center x is:

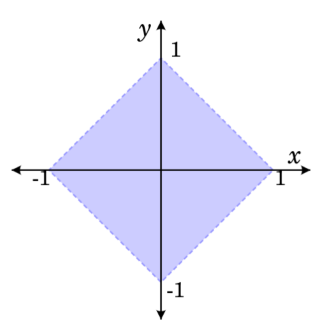
$$B(x,r)=\{y\in X:\rho(x,y)<r\}$$

Examples of Open Balls

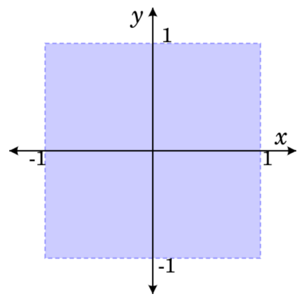
(i) $\rho_2(x,y)=\left(\sum_{j=1}^{2}(x_j-y_j)^2\right)^{\frac{1}{2}}$, $\quad B(0,1)=\{y\in\mathbb{R}^2:\rho_2(y,0)<1)$
This is an open ball in the Euclidean metric.



(ii) $\rho_1(x,y)=\sum_{j=1}^{2}|x_j-y_j|$
This is an open ball in the taxicab metric.



(iii) $\rho_\infty(x,y)=\max_{1\le j\le d}|x_j-y_j|$
This is an open ball in the Chebyshev metric.



2.2 Topology of $\mathbb{R}$

Definition (Open set and closed set). A set $E \subseteq X$ is open if for every $x \in E$ there exists a $r > 0$ such that $B(x, r) \subseteq E$. A set $E \subseteq X$ is closed if its complement $X \setminus E$ is open.
For example, every ball $B(x, r)$ is open, for if $y \in B(x, r)$ and $\rho(x, y) = s$ then

$$B(y, r - s) \subseteq B(x, r)$$

We have $z \in B(y, r - s) \iff \rho(z, y) < r - s$ and

$$\rho(x, z) \leq \rho(x, y) + \rho(y, z) < s + (r - s) = r$$

Also, X and $\varnothing$ are both open and closed.
Clearly, the union of any family of open sets is open. In other words, if we have $(U_\alpha)_{\alpha \in A}$ and U_α is open for any $\alpha \in A$ then

$$\bigcup_{\alpha \in A} U_\alpha \text{ is also open}$$

Also, the intersection of any finite family of open sets is open. Indeed, if $U_1, \ldots, U_n$ are open and $x \in \bigcap_{j=1}^n U_j$, then for each j, there exists $r_j > 0$ such that $B(x, r_j) \subseteq U_j$. Hence, choosing $r = \min\{r_1, \ldots r_n\}$ we have

$$B(x, r) \subseteq \bigcap_{j=1}^n U_j$$

which shows that $\bigcap_{j=1}^n U_j$ is open.
Remark.

1. The intersection of any family of closed sets is closed.
2. The union of any finite family of closed sets is closed.

Definition (Interior and Closure).

1. If $E \subseteq X$, the union of all open sets $U \subseteq E$ is the largest open set contained in E and it is called the interior of E and is denoted by $\text{Int}(E)$

2. The intersection of all closed sets $F \supseteq E$ is the smallest closed set containing E and it is called the closure of E and it is denoted by $\mathrm{Cl}(E)$ or $\overline{E}$.

Observation. Let $A \subseteq X$.

(i) $\mathrm{Int}(A)^c = \mathrm{Cl}(A^c)$

(ii) $\mathrm{Cl}(A)^c = \mathrm{Int}(A^c)$

Proof of (i).
Observe that $\mathrm{Int}(A) \subseteq A \iff A^c \subseteq (\mathrm{Int}(A))^c$. Since $(\mathrm{Int}(A))^c$ is already a closed set, it follows that $\mathrm{Cl}(A^c) \subseteq (\mathrm{Int}(A))^c$.
Now we show that $(\mathrm{Int}(A))^c \subseteq \mathrm{Cl}(A^c)$. Since $\mathrm{Cl}(A^c) \supseteq A^c$, it follows that $\mathrm{Cl}(A^c)^c \subseteq A$. Note that $\mathrm{Cl}(A^c)^c$ is an open set because it is the complement of a closed set. Thus, $\mathrm{Cl}(A^c)^c \subseteq \mathrm{Int}(A) \iff \mathrm{Int}(A)^c \subseteq \mathrm{Cl}(A^c)$. Therefore, we conclude that $\mathrm{Int}(A)^c = \mathrm{Cl}(A^c)$.
Proposition 1. Every open set in $\mathbb{R}$ is a countable union of open intervals.
Proof. If U is open, for each $x \in U$ consider the collection $\mathcal{J}_x$ of all open intervals I such that $x \in I \subseteq U$. It is easy to see that the union of any family of open intervals containing a point in common is again an open interval and hence

$$J_x = \bigcup_{I \in \mathcal{J}_x} I$$

is an open interval. Moreover, it is the largest element of $\mathcal{J}_x$.
If $x, y \in U$ then either $J_x = J_y$ or $J_x \cap J_y = \varnothing$. If not, then $J_x \cup J_y$ would be a larger open interval than J_x in $\mathcal{J}_x$. Thus if $\mathcal{J} = \{J_x : x \in U\}$ then the distinct members of $\mathcal{J}$ are disjoint and

$$U = \bigcup_{J \in \mathcal{J}} J$$

For each $J \in \mathcal{J}$ pick a rational number $f(J) \in J$. The map $f : \mathcal{J} \to \mathbb{Q}$ thus defined is injective, for $J \neq J'$ then $J \cap J' = \varnothing$ and since these sets share no elements in common, we know that $f(J) \neq f(J')$ and we conclude

$$\mathrm{card}(\mathcal{J}) \leq \mathrm{card}(\mathbb{Q}) = \mathrm{card}(\mathbb{N})$$

so $\mathcal{J}$ is countable.

□

Definition (Dense). Let (X, ρ) be a metric space. $E \subseteq X$ is said to be **dense** in X if $E \cap U \neq \varnothing$ for every open set U in X. Equivalently, $E \subseteq X$ is said to be dense if $E \cap B(x, r) \neq \varnothing$ for every $x \in X$ and $r > 0$.
Examples

- $\mathbb{Q}$ is dense in $\mathbb{R}$

- $\mathbb{Q}^d$ is dense in $\mathbb{R}^d$
- $\Delta = \{k2^{-n} : 0 \leq k \leq 2^n\}$ is dense in $[0,1]$

Definition (Nowhere dense). $E \subseteq X$ is said to be nowhere dense if $\overline{E}$ has empty interior, that is, $\text{Int}(\overline{E}) = \varnothing$.

Examples

- $X = \mathbb{R}$, then $\{x\}$ for any $x \in \mathbb{R}$ is nowhere dense in $\mathbb{R}$.
- $X = \mathbb{R}$, then $\mathbb{Z}$ is nowhere dense in $\mathbb{R}$.

Definition (Separable). (X, ρ) is called separable if it has a countable dense subset.

Examples

- $\mathbb{R}$ is separable since $\mathbb{Q}$ is dense and countable.
- $\mathbb{R}^d$ is separable since $\mathbb{Q}^d$ is countable and dense.
- $[0,1]$ is separable since $[0,1] \cap \mathbb{Q}$ or Δ are both countable and dense.

2.3 Convergence

Definition (Convergence of a sequence).

A sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space (X, ρ) is said to converge if there is a point $x \in X$ with the property:
For every $\varepsilon > 0$, there is an integer $N_\varepsilon \in \mathbb{N}$ such that for all $n \geq N_\varepsilon$ we have $\rho(x_n, x) < \varepsilon$.
In this case we also say that $(x_n)_{n\in\mathbb{N}}$ converges to x or that x is the limit of $(x_n)_{n\in\mathbb{N}}$ and we write:

$$x_n \xrightarrow[n\to\infty]{} x \quad \text{or}$$
$$\lim_{n\to\infty} x_n = x \quad \text{or}$$
$$\lim_{n\to\infty} \rho(x_n, x) = 0$$

Definition (Divergence). If $(x_n)_{n\in\mathbb{N}}$ does not converge it is said to diverge.
The set of all points x_n, that is, $\{x_n \in X : n \in \mathbb{N}\}$ is called the range of $(x_n)_{n\in\mathbb{N}}$. The range of a sequence may be a finite set or it may be countably infinite.

Definition (Diameter). In a metric space (X, ρ) we also define the diameter of $E \subseteq X$ as

$$\text{diam}(E) = \sup\{\rho(x,y) : x, y \in E\}$$

E is called bounded if $\text{diam}(E) < \infty$.

Examples

- $\{x\}$ is bounded, $\mathrm{diam}(\{x\}) = 0$.
- $B(x, r) = \{y \in X : \rho(x, y) < r\}$ is bounded. If $x_1, x_2 \in B(x, r)$, then $\rho(x_1, x_2) \leq \rho(x_1, x) + \rho(x, x_2) < 2r$.
- $(a, b) \subseteq \mathbb{R}$ and $-\infty < a < b < \infty$, then $\mathrm{diam}((a, b)) = b - a$.

Definition (Bounded sequence). The sequence $(x_n)_{n\in\mathbb{N}}$ is said to be bounded if its range is bounded, that is

$$\mathrm{diam}(\{x_n \in X : n \in \mathbb{N}\}) < \infty$$

Examples. Let $X = \mathbb{C}$.

(i) If $x_n = \frac{1}{n}$ then $\lim_{n\to\infty} x_n = 0$, the range is infinite and the sequence is bounded.

(ii) If $x_n = n^2$ then the sequence $(x_n)_{n\in\mathbb{N}}$ is unbounded, is divergent, and has infinite range.

(iii) If $x_n = i^n$, the sequence $(x_n)_{n\in\mathbb{N}}$ is divergent, is bounded, and has finite range.

(iv) If $x_n = 1 + \frac{(-1)^n}{n}$ the sequence (x_n) converges to 1, is bounded, and has infinite range.

(v) If $x_n = 1$ for all $n \in \mathbb{N}$, then $(x_n)_{n\in\mathbb{N}}$ converges to 1, is bounded, and has finite range.

Theorem 1. Properties of Convergence

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in a metric space (X, ρ).

(i) $(x_n)_{n\in\mathbb{N}}$ converges to $x \in X$ if and only if every open set containing x contains x_n for all but finitely many $n \in \mathbb{N}$.

(ii) (Uniqueness of limits). If $x, x' \in X$ and if $(x_n)_{n\in\mathbb{N}}$ converges to x and x', then $x = x'$.

(iii) If $(x_n)_{n\in\mathbb{N}}$ converges, then $(x_n)_{n\in\mathbb{N}}$ is bounded.

Proof. (i) ($\Rightarrow$) Suppose (i) holds. Recall that $\lim_{n\to\infty} \rho(x_n, x) = 0$ means that (*) for every $\varepsilon > 0$ there is $N_\varepsilon \in \mathbb{N}$ such that for all $n \geq N_\varepsilon$ we have $\rho(x_n, x) < \varepsilon$.
Take an open set V containing $x \in V$. Since V is open, there is $r > 0$ such that $B(x, r) \subseteq V$. It suffices to take $\varepsilon < r$ in (*) to see that $x_n \in B(x, r)$ for all $n \geq N_\varepsilon$ since $\rho(x_n, x) < \varepsilon$ by (*) and we are done.
($\Leftarrow$) Conversely, suppose that every open set V containing x contains all but finitely many of the x. Take $\varepsilon > 0$ and consider a ball $V = B(x, \varepsilon)$, this set is open and $x \in V$. By assumption, there exists $N \in \mathbb{N}$, depending on V, such that $x_n \in V = B(x, \varepsilon)$ for all $n \geq N$. Thus, $\rho(x_n, x) < \varepsilon$ if $n \geq N$. Hence, $\lim_{n\to\infty} \rho(x_n, x) = 0$.
(ii) Let $\varepsilon > 0$ be given. There are integers $N_\varepsilon, N'_\varepsilon \in \mathbb{N}$ such that

$$n \geq N_\varepsilon \implies \rho(x_n, x) < \frac{\varepsilon}{2}$$

$$n \geq N'_\varepsilon \implies \rho(x_n, x') < \frac{\varepsilon}{2}$$

Hence if $n \geq \max\{N_\varepsilon, N'_\varepsilon)\}$ we have

$$\rho(x, x') \leq \rho(x, x_n) + \rho(x_n, x') < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Since ε was arbitrary, we conclude that $\rho(x, x') = 0$.
(iii) Suppose that $\lim_{n\to\infty} \rho(x_n, x) = 0$. There exists $N \in \mathbb{N}$ such that $n > N$ implies $\rho(x_n, x) < 1$. Let

$$r = \max\{1, \rho(x_2, x), \rho(x_3, x), \ldots, \rho(x_N, x)\}$$

then we see $\rho(x_n, x) \leq r$ for all $n \in \mathbb{N}$.

□

Proposition. If (X, ρ) is a metric space, $E \subseteq X$ and $x \in X$, the following are equivalent:

(a) $x \in \overline{E}$

(b) $B(x, r) \cap E \neq \varnothing$ for all $r > 0$

(c) There is a sequence $(x_n)_{n\in\mathbb{N}}$ in E that converges to x.

(a) $\Rightarrow$ (b) We prove the contrapositive statement. If $B(x, r) \cap E = \varnothing$, then $B(x, r)^c$ is a closed set containing E but not x, so $x \notin \overline{E}$. Note that $E \subseteq \overline{E} \subseteq B(x, r)^c$, which follows from $E \subseteq B(x, r)^c$ and taking the closure on both sides.
(b) $\Rightarrow$ (a) Again, we will prove the contrapositive statement. If $x \notin \overline{E}$, since $(\overline{E})^c$ is open there exists $r > 0$ such that $B(x, r) \subseteq (\overline{E}^c) \subseteq E^c$. So $B(x, r) \cap E = \varnothing$.
(b) $\Rightarrow$ (c) Suppose $B(x, r) \cap E \neq \varnothing$ for all $r > 0$. Then for each $n \in \mathbb{N}$ there exist $x_n \in B(x, \frac{1}{n}) \cap E$. Hence, $\rho(x_n, x) < \frac{1}{n}$ and we are done. To be more explicit, we take the sequence $(x_n)_{n=1}^\infty$ created from these points to show that such a sequence exists.
(c) $\Rightarrow$ (b) If $B(x, r) \cap E = \varnothing$, then $\rho(y, x) \geq r$ for all $y \in E$. Thus, no sequence of E can converge to x.

□

Definition (Accumulation point). Let (X, ρ) be a metric space. The point $x \in X$ is called an accumulation point of $E \subseteq X$ if for any open set $x \in U$ we have $(E \setminus \{x\}) \cap U \neq \varnothing$.
An accumulation point x of $E \subseteq X$ is sometimes also called a limit point of E or a cluster point of E.
Definition (Isolated point). A point $x \in E$ is called an isolated point of E if it is not an accumulation point.
Examples

- $X = \mathbb{R}, \quad E = [a, b], \quad -\infty < a < b < \infty$
  each $x \in E$ is an accumulation point of E.

- $X = \mathbb{R}, \quad E = (a, b), \quad -\infty < a < b < \infty$
  each point $x \in [a, b]$ is an accumulation point of (a, b).

- $X = \mathbb{R}, \quad E = \mathbb{Z}, \quad$ each point of $\mathbb{Z}$ is an isolated point.

Remark. If $x \in X$ is a limit point of $E \subseteq X$, it does not need to be an element of E.

Definition (Perfect set). A set $E \subseteq X$ of a metric space is called perfect if $E = \operatorname{acc}(E)$, where $\operatorname{acc}(E)$ is the set of all accumulation points of E.

Proposition. Let (X, ρ) be a metric space, let $E \subseteq X$ and $\operatorname{acc}(E)$ be the set of accumulation points of E. Then $\overline{E} = E \cup \operatorname{acc}(E)$ and E is closed if $\operatorname{acc}(E) \subseteq E$.

Proof. If $x \notin \overline{E}$ then there is $B(x, \varepsilon) \subseteq (\overline{E})^c$. Thus, $B(x, \varepsilon) \cap E = \varnothing$. Hence $x \notin \operatorname{acc}(E)$. Thus $E \cup \operatorname{acc}(E) \subseteq \overline{E}$.

If $x \notin E \cup \operatorname{acc}(E)$, there is an open set $U \ni x$ such that $U \cap E = \varnothing$. Then $\overline{E} \subseteq U^c$ so $x \notin \overline{E}$ thus $\overline{E} \subseteq E \cup \operatorname{acc}(E)$.

Finally, E is closed if and only if $E = \overline{E}$ and this happens if and only if $\operatorname{acc}(E) \subseteq E$.

Definition (Boundary). The difference $\overline{E} \setminus \operatorname{Int}(E)$ is called the boundary of E in a metric space (X, ρ) and is denoted by ∂E.

Examples.

- $X = \mathbb{R}, \quad E = (a, b)$ and $-\infty < a < b < \infty$, then
  $\overline{E} = [a, b], \quad \operatorname{Int}(E) = (a, b)$ so $\partial E = \{a, b\}$

- $X = \mathbb{R}^2 \quad \rho(x, y) = \left(\sum_{j=1}^{2} |x_j - y_j|^2\right)^{\frac{1}{2}}$ for $x = (x_1, x_2), y = (y_1, y_2)$
  $E = B(x, r) = \{y \in \mathbb{R}^2 : \rho(x, y) < r\}, \quad \overline{E} = \overline{B(x, r)} = \{y \in \mathbb{R}^2 : \rho(x, y) \leq r\}$
  $\partial E = \{y \in \mathbb{R}^2 : \rho(x, y) = r\}$

- $X = \mathbb{R}, \quad E = \mathbb{Z}$ then $\operatorname{Int}(\mathbb{Z}) = \varnothing$, $\overline{\mathbb{Z}} = \mathbb{Z}$, thus $\partial \mathbb{Z} = \mathbb{Z}$.

Theorem Algebraic Limit Theorem

Suppose that $(s_n)_{n\in\mathbb{N}}, (t_n)_{n\in\mathbb{N}} \subseteq \mathbb{C}$ with

$$\lim_{n\to\infty} s_n = s \text{ and } \lim_{n\to\infty} t_n = t$$

then we have the following results

(a) $\lim_{n\to\infty} (s_n + t_n) = s + t$

(b) $\lim_{n\to\infty} (cs_n) = cs$ and $\lim_{n\to\infty} (s_n + c) = s + c$ for any $c \in \mathbb{C}$

(c) $\lim_{n\to\infty} (s_n t_n) = st$

(d) $\lim_{n\to\infty} (\frac{1}{s_n}) = \frac{1}{s}$ provided $s_n \neq 0$ for all $n \in \mathbb{N}$ and $s \neq 0$.

Proof. Omitted. □

As we know convergent sequences are bounded, but bounded sequences in $\mathbb{R}^k$ need not converge. For instance $x_n = (-1)^n$ is bounded but $\lim_{n\to\infty} x_n$ does not exist.
However, there is one important case in which convergence is equivalent to boundedness - this happens for monotonic sequences in $\mathbb{R}$.

2.4 Monotone sequences

Definition (Monotone). A sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ is said to be

(a) monotonically increasing if $x_n \leq x_{n+1}$ for all $n \in \mathbb{N}$.

(b) monotonically decreasing if $x_n \geq x_{n+1}$ for all $n \in \mathbb{N}$.

Theorem Monotone Convergence Theorem

Suppose $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ is monotone. Then (x_n) converges if and only if it is bounded.

Proof. Suppose that $x_n \leq x_{n+1}$. Let $E = \{x_n : n \in \mathbb{N}\} \subseteq \mathbb{R}$. If $(x_n)_{n\in\mathbb{N}}$ is bounded, let $x = \sup E$ then

$$x_n \leq x \text{ for all } n \in \mathbb{N}$$

For every $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that

$$x - \varepsilon < x_N \leq x \text{ since } x = \sup E$$

Note that the existence of such an x_N is given by the epsilon criterion of a supremum of the set E. But $(x_n)_{n\in\mathbb{N}}$ is increasing, so $x_N \leq x_n \leq x$ for all $n \geq N$.
This shows that for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that for all $n \geq N$, $x - \varepsilon < x_n \leq x$. Hence, $|x_n - x| < \varepsilon$ for all $n \geq N$. Equivalently, $\lim_{n\to\infty} x_n = x$.
□

Theorem (Squeeze Theorem). Suppose that $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ are sequences such that

$$a_n \leq b_n \leq c_n \text{ for all } n \geq N_0$$

for some $N_0 \in \mathbb{N}$ and

$$\lim_{n\to\infty} a_n = a = \lim_{n\to\infty} c_n$$

then $\lim_{n\to\infty} b_n = a$.
Proof. Let $\varepsilon > 0$. We find $N_1, N_2 \in \mathbb{N}$ such that

$$n \geq N_1 \implies |a_n - a| < \varepsilon \iff a - \varepsilon < a_n < a + \varepsilon$$

$$n \geq N_2 \implies |c_n - a| < \varepsilon \iff c - \varepsilon < c_n < c + \varepsilon$$

For $n \geq \max\{N_0, N_1, N_2\}$

$$a - \varepsilon < a_n \leq b_n \leq c_n < a + \varepsilon$$

Thus

$$-\varepsilon < b_n - a < \varepsilon \iff |b_n - a| < \varepsilon$$

$\square$

Theorem. (Convergence in $\mathbb{R}^k$) (a) Suppose that $x_n \in \mathbb{R}^k$ for $n \in \mathbb{N}$ and

$$x_n = (\alpha_{1,n}, \dots, \alpha_{k,n})$$

Then $(x_n)_{n\in\mathbb{N}}$ converges to $x = (\alpha_1, \dots \alpha_k)$ if and only if

$$\lim_{n\to\infty} \alpha_{j,n} = \alpha_j, \quad 1 \leq j \leq k$$

(b) Suppose $(x_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ are sequences in $\mathbb{R}^k$, $(\beta_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ and $x_n \xrightarrow[n\to\infty]{} x, y_n \xrightarrow[n\to\infty]{} y$, and $\beta_n \xrightarrow[n\to\infty]{} \beta$, then

$$\lim_{n\to\infty} (x_n + y_n) = x + y, \quad \lim_{n\to\infty} x_n \cdot y_n = x \cdot y, \quad \lim_{n\to\infty} \beta_n x_n = \beta x$$

Proof. (a) Let $|x| = \left(\sum_{j=1}^{k} x_j^2\right)^{\frac{1}{2}}$ be the Euclidean distance.
If $x_n \xrightarrow[n\to\infty]{} x$ then for $1 \leq j \leq k$

$$0 \leq |\alpha_{j,n} - \alpha_j| \leq \left(\sum_{j=1}^{k} |\alpha_{j,n} - \alpha_j|^2\right)^{\frac{1}{2}} = |x_n - x| \xrightarrow[n\to\infty]{} 0$$

More explicitly, since the distance between x_n and x converges to zero, this inequality is satisfied for arbitrary $\varepsilon > 0$. Thus, $\lim_{n\to\infty} \alpha_{j,n} = \alpha_j$.
Conversely, if $\lim_{n\to\infty} \alpha_{j,n} = \alpha_j$, then for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$

$$|\alpha_{j,n} - \alpha_j| < \frac{\varepsilon}{\sqrt{k}} \quad \text{for } 1 \leq j \leq k$$

Hence,

$$|x_n - x| = \left(\sum_{j=1}^{k} |\alpha_{j,n} - \alpha_j|^2\right)^{\frac{1}{2}} < \left(\sum_{j=1}^{k} \left(\frac{\varepsilon}{\sqrt{k}}\right)^2\right)^{\frac{1}{2}} = \varepsilon$$

thus $x_n \xrightarrow[n\to\infty]{} x$.
(b) Suppose we have $x_n \xrightarrow[n\to\infty]{} x$ and $y_n \xrightarrow[n\to\infty]{} y$. Then

$$0 \leq |x_n + y_n - (x + y)| \leq |x_n - x| + |y_n - y| \xrightarrow[n\to\infty]{} 0$$

Also,

$$0 \le |x_n \cdot y_n - x \cdot y| = |(x_n - x)y_n + x(y_n - y)|$$
$$\le |(x_n - x) \cdot y_n| + |x \cdot (y_n - y)| \quad \le |x_n - x||y_n| + |x||y_n - y| \xrightarrow[n\to\infty]{} 0$$

The above limit converges to 0 since:

$$\lim_{n\to\infty} |x_n - x| = 0, \quad \lim_{n\to\infty} |y_n - y| = 0, \quad |y_n| \le C \text{ for all } n \in \mathbb{N}$$

□

2.5 Subsequences

Definition (Subsequence). Given a sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space (X, ρ) consider a sequence $(n_k)_{k\in\mathbb{N}}$ of positive integers such that $n_1 < n_2 < n_3 < \cdots$ Then the sequence $(x_{n_k})_{k\in\mathbb{N}}$ is called a subsequence of $(x_n)_{n\in\mathbb{N}}$. If $(x_{n_k})_{k\in\mathbb{N}}$ converges, its limit is called a subsequential limit of $(x_n)_{n\in\mathbb{N}}$.

Remark. It is clear that $(x_n)_{n\in\mathbb{N}}$ converges to x if and only if every subsequence of $(x_n)_{n\in\mathbb{N}}$ converges to x.

Theorem. The subsequential limits of a sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space (X, ρ) form a closed subset of X.

Proof. Let E^* be the set of all subsequential limits of $(x_n)_{n\in\mathbb{N}}$ and let q be a limit point of E^*. We will show that $q \in E^*$.

Choose $n_1 \in \mathbb{N}$ so that $x_{n_1} \neq q$. If no such n_1 exists, then E^* has only one point and there is nothing to prove. Set $\delta = \rho(x_{n_1}, q) > 0$. Suppose $n_1, \ldots, n_{i-1}$ are chosen. Since q is a limit point of E^*, there is $x \in E^*$ so that $\rho(x, q) < \delta 2^{-i-1}$. Since $x \in E^*$ there is $n_i > n_{i-1}$ such that $\rho(x, x_{n_i}) < \delta 2^{-i-1}$. Hence

$$\rho(q, x_{n_i}) \le \rho(q, x) + \rho(x, x_{n_i}) < \delta 2^{-i-1} + \delta 2^{-i-1} = \delta 2^{-i}$$

This means that $(x_{n_i})_{i\in\mathbb{N}}$ converges to q, thus $q \in E^*$.

□

2.6 Cauchy Sequences

Definition (Cauchy sequence). A sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space (X, ρ) is said to be a Cauchy sequence if for every $\varepsilon > 0$ there is an integer $N \in \mathbb{N}$ such that $\rho(x_n, x_m) < \varepsilon$ if $n, m \ge N$.

A subset E of a metric space (X, ρ) is called **complete** if every Cauchy sequence in E converges and its limit is in E.

Theorem. (Nested Interval Property). If $(I_n)_{n\in\mathbb{N}}$ is a sequence of closed intervals in $\mathbb{R}$ such that $I_n \supseteq I_{n+1}$, then $\bigcap_{n\in\mathbb{N}} I_n \neq \varnothing$.

Proof. If $I_n = [a_n, b_n]$, let $E = \{a_n : n \in \mathbb{N}\}$ then notice that $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n \leq b_1$. Thus

$$\lim_{n\to\infty} a_n = a \text{ exists}, \quad \lim_{n\to\infty} a_n = \sup_{n\in\mathbb{N}} a_n$$

Note that because $a_m \leq a_{m+n} \leq b_{m+n} \leq b_m$ for all $m, n \in \mathbb{N}$, it follows that $a_n \leq a \leq b_m$. Thus $a_m \leq a \leq b_m \equiv a \in I_m$ for all $m \in \mathbb{N}$ and so

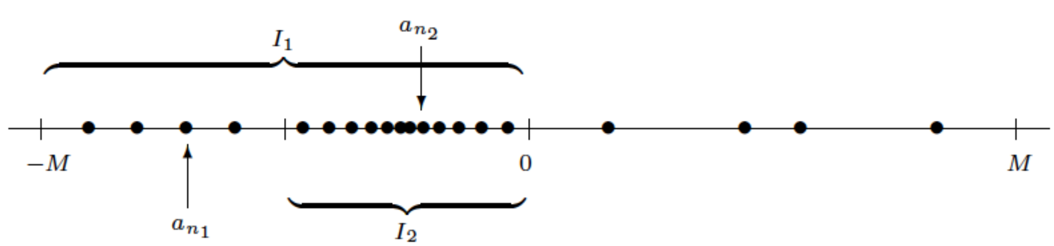
$$a \in \bigcap_{n\in\mathbb{N}} I_n$$

□

Theorem Bolzano-Weierstrass

Every bounded sequence in $\mathbb{R}$ contains a convergent subsequence.

Proof. Let $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ be bounded, that is, there exists $M > 0$ such that $|x_n| < M$ for all $n \in \mathbb{N}$. Equivalently, $x_n \in [-M, M]$ for all $n \in \mathbb{N}$.
Bisect the closed interval $[-M, M]$ into the two closed intervals $[-M, 0]$ and $[0, M]$. Now, it must be that at least one of these closed intervals contains an infinite number of the terms in the sequence (x_n). Select a half for which this is the case and label that interval as I_1. Then, let x_{n_1} be some term in the sequence (x_n) satisfying $x_{n_1} \in I_1$. Consider the picture below, with x_n in place of a_n.



We can repeat the process indefinitely to get the terms of our candidate subsequence. Further, we have produced $I_0 \supseteq I_1 \supseteq \cdots \supseteq I_n \supseteq, \cdots$, a decreasing sequence of closed intervals each containing infinitely many elements of $(x_n)_{n\in\mathbb{N}}$. The Nested Interval Property guarantees that:

$$\bigcap_{n\in\mathbb{N}} I_n \neq \varnothing$$

Moreover, $\operatorname{diam}(I_n) \xrightarrow[n\to\infty]{} 0$ thus there is $x \in \mathbb{R}$ such that

$$\bigcap_{n\in\mathbb{N}} I_n = \{x\}$$

If there is $y \in \bigcap_{n\in\mathbb{N}} I_n$ such that $y \neq x$, then $0 < \rho(x,y) \leq \operatorname{diam}(I_n) \xrightarrow[n\to\infty]{} 0$ ($\lightning$) So, let $\varepsilon > 0$ and take $N \in \mathbb{N}$ such that

$$n \leq N \implies \operatorname{diam}(I_n) < \varepsilon$$

Let $x_{n_1} \in I_1, x_{n_2} \in I_2$, and in general $x_{n_j} \in I_j$ for all $j \in \mathbb{N}$. Then $\rho(x, x_{n_k}) \leq \operatorname{diam}(I_k) < \varepsilon$. Finally, we conclude

$$\lim_{k\to\infty} \rho(x_{n_k}, x) = 0$$

□

Theorem Completeness of $\mathbb{R}$

$\mathbb{R}$ is complete, that is, every Cauchy sequence converges to a point in $\mathbb{R}$.

Proof. Let $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ be a Cauchy sequence.
Claim 1. If $(x_n)_{n\in\mathbb{N}}$ is Cauchy in a metric space (X,ρ), then $(x_n)_{n\in\mathbb{N}}$ is bounded.
Proof 1. Since (x_n) is Cauchy, there is $N \in \mathbb{N}$ so that for all $m, n \geq N$, one has

$$\rho(x_n, x_m) < 1$$

In particular, $\rho(x_n, x_N) < 1$ for all $n > N$. So this bounds the terms greater than N. Define $r = \max\{1, \rho(x_1, x_N), \dots \rho(x_{N-1}, x_N)\}$ then we have $\rho(x_n, x_N) \leq r$ for all $n \in \mathbb{N}$. Thus $(x_n)_{n\in\mathbb{N}}$ is bounded. □
Claim 2. If $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in a metric space (X,ρ) and $\lim_{k\to\infty} x_{n_k} = x$ for some sequence $(n_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$, then $\lim_{n\to\infty} x_n = x$.
Proof 2. Let $\varepsilon > 0$ then there is $N_1 \in \mathbb{N}$ so that

$$m, n \geq N_1 \implies \rho(x_m, x_n) < \frac{\varepsilon}{2}$$

There is also $N_2 \in \mathbb{N}$ so that

$$k \geq N_2 \implies \rho(x_{n_k}, x) < \frac{\varepsilon}{2} \text{ and } n_k \geq N_1$$

Then for $n \geq N_1$ we get

$$\rho(x, x_n) \leq \rho(x, x_{n_k}) + \rho(x_{n_k}, x_n) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Thus $\lim_{n\to\infty} x_n = x$. □
Now we can easily conclude that $\mathbb{R}$ is complete. Indeed take any Cauchy sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$. By Claim 1, $(x_n)_{n\in\mathbb{N}}$ is bounded. By the previous theorem, it has a converging subsequence, i.e. there is $(n_k)_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} x_{n_k} = x \text{ for some } x \in \mathbb{R}$$

By Claim 2, we conclude that

$$\lim_{n\to\infty} x_n = x$$

and we are done.

$\square$

Corollary. $\mathbb{R}^k$ is complete for any $k \in \mathbb{N}$.

Proof. Let $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}^k$ be a Cauchy sequence in $\mathbb{R}^k$. If $x_n = (\alpha_{1,n}, \ldots, \alpha_{k,n})$ then $(\alpha_{j,n})_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$ for each $1 \leq j \leq k$ since

$$|\alpha_{j,n} - \alpha_{j,m}| \leq \left(\sum_{j=1}^{k} |\alpha_{j,n} - \alpha_{j,m}|^2\right)^{\frac{1}{2}} = |x_n - x_m| \xrightarrow[m,n\to 0]{} 0$$

By the previous theorem, each component of x_n has a limit. In other words, there is $\alpha_j \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \alpha_{j,n} = \alpha_j$$

Taking $x = (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$, we can see that $x_n \xrightarrow[n\to\infty]{} x$.

$\square$

Proposition. A closed subset of a complete metric space is complete and a complete subset of an arbitrary metric space is closed.

Proof. If (X, ρ) is complete, $E \subseteq X$ is closed and $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in E, then $(x_n)_{n\in\mathbb{N}}$ has a limit in X. But $\overline{E} = E$. Because $x \in \overline{E}$, we have $x \in E$.

If $E \subseteq X$ is complete and $x \in \overline{E}$ then we know there exists $(x_n)_{n\in\mathbb{N}} \subseteq E$ converging to x. But $(x_n)_{n\in\mathbb{N}}$ is Cauchy, so its limit lies in E. Thus $E = \overline{E}$.

Remark. We have used the fact that if $(x_n)_{n\in\mathbb{N}}$ converges, say to x, in a metric space (X, ρ) then $(x_n)_{n\in\mathbb{N}}$ is Cauchy.

Indeed, let $\varepsilon > 0$ then there is $N \in \mathbb{N}$ such that $n \geq N$ implies $\rho(x_n, x) < \frac{\varepsilon}{2}$.

Thus, $\rho(x_m, x_n) \leq \rho(x_m, x) + \rho(x_n, x) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$ for all $m, n \geq N$.

Theorem. Every metric space (X, ρ) can be always seen as a dense subset of a complete metric space $(\hat{X}, \hat{\rho})$.

Proof. Let (X, ρ) be a metric space and $[X]$ the set of all Cauchy sequences in X.

We define a binary relation $\equiv$ on $[X]$ as follows

$$(x_n)_{n\in\mathbb{N}} \equiv (y_n)_{n\in\mathbb{N}} \iff \rho(x_n, y_n) \xrightarrow[n\to\infty]{} 0$$

It is easy to check that $\equiv$ is an equivalence relation. Let $\hat{x}$ denote the set of all equivalence classes. For any Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ let $[x_n]$ denote the equivalence class containing $(x_n)_{n\in\mathbb{N}}$.

We define a metric $\hat{\rho}$ on $\hat{X}$ by

$$\hat{\rho}([x_n], [y_n]) = \lim_{n\to\infty} \rho(x_n, y_n)$$

Define $f : X \to [X]$ by $f(x) = \hat{x}$, where $\hat{x} = (x, x, x, \dots)$. We can easily check the following:

1. $\hat{\rho}$ is well defined.
2. $\hat{\rho}$ is a complete metric on $\hat{X}$.
3. The function $f : X \to \hat{X}$ is an isometry, that is, for all $x, y \in X$ we have $\hat{\rho}(\widehat{f(x)}, \widehat{f(y)}) = \rho(x, y)$
4. The set $f[X]$ is dense in $\hat{X}$
5. If X is separable, then $\hat{X}$ is separable.

Thus, we see that every (separable) metric space can be seen as a dense subset of a complete (separable) metric space.
The metric space $(\hat{X}, \hat{\rho})$ is called the **completion** of (X, ρ).
Remark. If we apply this construction to $\mathbb{Q}$ then we obtain $\hat{\mathbb{Q}} = \mathbb{R}$.

Theorem Cantor Intersection Theorem

A metric space (X, ρ) is complete if and only if for every decreasing sequence of sets $F_1 \supseteq F_2 \supseteq \cdots$ of nonempty closed sets in X with $\operatorname{diam}(F_n) \xrightarrow[n\to\infty]{} 0$, the intersection

$$\bigcap_{n\in\mathbb{N}} F_n = \{x_0\} \text{ for some } x_0 \in X$$

Proof. Assume that (X, ρ) is complete. Let $\{F_n : n \in \mathbb{N}\}$ be such that $F_1 \supseteq F_2 \supseteq \cdots$ and $\operatorname{diam}(F_n) \xrightarrow[n\to\infty]{} 0$.
Choose $x_n \in F_n$. Let $\varepsilon > 0$ and pick $N \in \mathbb{N}$ such that $\operatorname{diam}(F_m) < \varepsilon$ for all $n \geq N$. Note that for $n \geq m \geq N$ we have

$$x_n \in F_n \subseteq F_m, \qquad \text{so } \rho(x_n, x_m) \leq \operatorname{diam}(F_m) \leq \varepsilon$$

This ensures that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and consequently converges to some $x_0 \in X$. Since each F_n is closed, then $x_0 \in F_n$ for all $n \in \mathbb{N}$, so

$$x_0 \in \bigcap_{n\in\mathbb{N}} F_n$$

Suppose there is $y \neq x_0$ so that $y \in \bigcap_{n\in\mathbb{N}} F_n$, then

$$0 < \rho(x_0, y) \leq \operatorname{diam}(F_n) \xrightarrow[n\to\infty]{} 0 \;\;(\lightning)$$

Thus, $\bigcap_{n\in\mathbb{N}} F_n = \{x_0\}$. To prove the converse implication, assume that $(x_n)_{n\in\mathbb{N}}$ is Cauchy and that

$$F_n = \overline{\{x_m : m \geq n\}} \qquad \operatorname{diam}(E) = \operatorname{diam}(\overline{E})$$

We see $F_1 \supseteq F_2 \supseteq \cdots$ and $\text{diam}(F_n) = \text{diam}(\{x_m : m \geq n\}) \xrightarrow[n\to\infty]{} 0$. Thus, $\bigcap_{n\in\mathbb{N}} F_n = \{x_0\}$ for some $x_0 \in X$. Finally, we conclude

$$\lim_{n\to\infty} x_n = x_0$$

□

Let (X, ρ) be a metric space. Let $Y \subseteq X$. Then (Y, ρ) is also a metric space.
Recall that we say $E \subseteq X$ is open if for every $x \in E$ there is $r > 0$ such that $B(x, r) \subseteq E$.
Since (Y, ρ) is a metric space, so that our definitions may equally well be made within Y.
We say that $E \subseteq Y$ is open **relative** to Y if to each $x \in E$, there is $r > 0$ such that $B(x, r) \cap Y \subseteq E$.
Example. Let $X = \mathbb{R},\ Y = [0, 1]$ then $E = [0, \frac{1}{2})$ is open relative to Y since for any $x \in E$ there is $r > 0$ such that $(x - \varepsilon, x + \varepsilon) \cap [0, 1] \subseteq [0, \frac{1}{2})$.
Theorem 1. Let (X, ρ) be a metric space and $Y \subseteq X$. A set $E \subseteq Y$ is open relative to Y if and only if $E = Y \cap G$ for some open G of X.
Proof. ($\Rightarrow$) Suppose that E is open relative to Y. Then for each $x \in E$ there is $r_x > 0$ so that $(B(x, r_x) \cap Y) \subseteq E$. Consider

$$G = \bigcup_{x\in E} B(x, r_x) \subseteq X$$

which is open and we see $Y \cap G = E$. Indeed, if $x \in E$ then $x \in B(x, r_x)$ thus $x \in G \cap Y$. We also know $B(x, r_x) \cap Y \subseteq E$, thus $G \cap Y \subseteq E$.
($\Leftarrow$) If G is open in X and $E = G \cap Y$ then for any $x \in E$ there is $r > 0$ so that $B(x, r) \subseteq G$ since G is open. Thus $(B(x, r) \cap Y) \subseteq (G \cap Y) \subseteq E$ and we are done.

□

2.7 Compact Sets

Let (X, ρ) be a metric space. If $E \subseteq X$ and $(V_\alpha)_{\alpha\in A}$ is a family of sets such that $E \subseteq \bigcup_{\alpha\in A} V_\alpha$. Then $(V_\alpha)_{\alpha\in A}$ is called a cover of E and E is said to be covered by the V_α's. If additionally each V_α is open, $(V_\alpha)_{\alpha\in A}$ is called an open cover of E.
Definition (Compact). A subset K of a metric space (X, ρ) is said to be compact if every open cover of K contains a finite subcover.
More explicitly if $(V_\alpha)_{\alpha\in A}$ is an open cover of K, then there are finitely many indices $\alpha_n \in A$ such that

$$K \subseteq \bigcup_{j=1}^{n} V_{\alpha_j}$$

Example.

1. Every finite set in $\mathbb{R}^k$ is compact.

2. $K = \{\frac{1}{n} : n \in \mathbb{N}\} \cup \{0\}$ is compact.
   Let $(V_\alpha)_{\alpha \in A}$ be any open cover of K. Then there is $a_0 \in A$ such that $0 \in V_{\alpha_0}$. Since $\lim_{n\to\infty} \frac{1}{n} = 0$ and V_{α_0} is open thus it contains all but finitely many $\frac{1}{n}$'s. In other words, there is $n_0 \in \mathbb{N}$ such that
   $$n \geq n_0 \implies \frac{1}{n} \in V_{\alpha_0}$$
   To be more explicit, since V_{α_0} is open, we can find a ball $B(0, r)$ with $r > 0$. This is equivalent to the interval $(-r, r)$ if we are considering $K \subseteq R$ with the Euclidean metric. Because $\lim_{n\to\infty} \frac{1}{n} \to 0$, there exists a $n_0 \in \mathbb{N}$ such that for $n > n_0$, we have $|\frac{1}{n} - 0| < r$. This shows all the details.
   For each $j \in \{1, 2, \dots n_0 - 1\}$ we pick $\alpha_j \in A$ so that $\frac{1}{j} \in V_{\alpha_j}$ and we see
   $$K \subseteq \bigcup_{j=1}^{n_0} V_{\alpha_j}$$
   Hence, $(V_\alpha)_{\alpha \in A}$ has a finite subcover.

Theorem 2. Suppose that $K \subseteq Y \subseteq X$ and (X, ρ) is a metric space. Then K is compact relative to X if and only if K is compact relative to Y.
Proof. ($\Rightarrow$) Suppose that K is compact relative to X and let $(V_\alpha)_{\alpha \in A}$ be a collection of sets open relative to Y such that
$$K \subseteq \bigcup_{\alpha \in A} V_\alpha$$
By the previous theorem on the relative openness of sets, there exist open sets G_α in X such that $V_\alpha = Y \cap G_\alpha$ for all $\alpha \in A$. Since K is compact in X, we find $\alpha_1, \dots, \alpha_n$ so that
$$K \subseteq \bigcup_{j=1}^{n} G_{\alpha_j}, \text{ but } K \subseteq Y$$
thus, $K \subseteq \bigcup_{j=1}^{n} G_{\alpha_j} \cap Y = \bigcup_{j=1}^{n} V_{\alpha_j}$, so K is compact relative to Y.
($\Leftarrow$) Conversely, suppose that K is compact relative to Y. Let $(G_\alpha)_{\alpha \in A}$ be a collection of open sets in X such that
$$K \subseteq \bigcup_{\alpha \in A} G_\alpha \qquad K \subseteq \bigcup_{x \in K} B(x, r)$$
Set $V_\alpha = Y \cap G_\alpha \subseteq G_\alpha$. Since $K \subseteq Y$, then we have
$$K \subseteq \bigcup_{\alpha \in A} V_\alpha$$
and there exists a finite subcover $V_{\alpha_1}, \dots V_{\alpha_n}$ so that
$$K \subseteq \bigcup_{j=1}^{n} V_{\alpha_j} \subseteq \bigcup_{j=1}^{n} G_{\alpha_j}$$

Thus K is compact in X.

□

The above shows that compactness is a property that can be "inherited" from the ambient space, and that if a set is compact in a subspace, then it is also compact in the ambient space. Such properties show how compactness differs from openness $E = [0, \frac{1}{2})$ in the first example from the lecture is open relative to Y, but clearly it is not open in $\mathbb{R}$.

Theorem 3. Compact subsets of metric spaces are closed.

Proof. Let K be a compact subset of a metric space (X, ρ). We shall prove that K^c is open in X. Let $x \in X \setminus K$. If $y \in K$ let $V_y = B(x, r)$ and $W_y = B(y, r)$ where $r < \frac{1}{2}\rho(x, y)$, then $W_y \cap V_y = \varnothing$.

Since K is compact and $K \subseteq \bigcup_{y \in K} W_y$, then there are finitely many $y_1, \dots, y_n \in K$ so that

$$K \subseteq \bigcup_{j=1}^{n} W_{y_j} = W$$

If $V = V_{y_1} \cap \cdots \cap V_{y_n}$, then V is a neighborhood of x and $V \cap W = \varnothing$. Note that $V \cap W = \varnothing$ follows from that fact that since $v \in V$, this means that $v \in V_{y_1}, \dots, V_{y_n}$ and by construction $v \notin W_{y_1}, \dots, W_{y_n}$. Hence $x \in V \subseteq K^c$. Thus, x is an interior point of K^c.

□

Theorem 4. Closed subsets of compact sets are compact.

Proof. Suppose that $F \subseteq K \subseteq X$ and F is closed in X and K is compact. Let $(V_\alpha)_{\alpha \in A}$ be an open cover of F. Observe that

$$F \subseteq K \subseteq \bigcup_{\alpha \in A} V_\alpha \cup F^c$$

where the rightmost set is open because it is the union of open sets.

But K is compact, thus there is a finite subcover of $(V_\alpha)_{\alpha \in A} \cup F^c$ that covers K. But $F \subseteq K$, thus this is also a finite subcover of F upon removing F^c.

□

Theorem 5. If $(K_\alpha)_{\alpha \in A}$ is a collection of compact subsets of a metric space X such that the intersection of every finite subcollection of $(K_\alpha)_{\alpha \in A}$ is nonempty, then

$$\bigcap_{\alpha \in A} K_\alpha \neq \varnothing$$

Proof. Fix a member K_{α_0} of $(K_\alpha)_{\alpha \in A}$ and set $G_\alpha = K_\alpha^c$. Suppose, for contradiction, that

$$\bigcap_{a \in A} K_\alpha = K_{\alpha_0} \cap \left(\bigcap_{a \in A \setminus \{a_0\}} K_\alpha \right) = \varnothing$$

then $K_{\alpha_0} \subseteq \bigcup_{\alpha \in A \setminus \{\alpha_0\}} G_\alpha$. To see why this is true, consider that

$$\left(\bigcap_{\alpha \in A} K_\alpha \right)^c = (\varnothing)^c$$

$$\bigcup_{\alpha \in A \setminus \{\alpha_0\}} K_\alpha^c \cup K_{\alpha_0}^c = X$$

$$\bigcup_{\alpha \in A \setminus \{\alpha_0\}} G_\alpha \supseteq K_{\alpha_0}$$

This shows that $(G_\alpha)_{\alpha \in A \setminus \{\alpha_0\}}$ is an open cover for K_{α_0}.
Since K_{α_0} is compact, there are $\alpha_1, \ldots, \alpha_n$ so that

$$K_{\alpha_0} \subseteq \bigcup_{j=1}^{n} G_{\alpha_j}$$

Hence, $K_{\alpha_0} \cap K_{\alpha_1} \cap \cdots \cap K_{\alpha_n} = \varnothing$ ($\lightning$).

$\square$

Definition (Totally bounded). Let (X, ρ) be a metric space. A subset $E \subseteq X$ is called totally bounded if for every $\varepsilon > 0$, E can be covered by finitely many balls of radius ε, i.e. there is $N_\varepsilon \in \mathbb{N}$ so that

$$E \subseteq \bigcup_{j=1}^{N_\varepsilon} B(x_j, \varepsilon) \text{ for some } x_1, \ldots, x_j \in X$$

Remark.

1. Every totally bounded set E is bounded.
   If $x, y \in E \subseteq \bigcup_{j=1}^{N_\varepsilon} B(x_j, \varepsilon)$, without loss of generality say that $x \in B(x_1, \varepsilon), y \in B(x_2, \varepsilon)$. Then

$$\begin{aligned} \rho(x, y) &\leq \rho(x, x_1) + \rho(x_1, x_2) + \rho(x_2, y) \\ &\leq 2\varepsilon + \max\{\rho(x_i, x_j) : 1 \leq i, j \leq N_\varepsilon\} \end{aligned}$$

   Note that the converse is false in general.

2. If E is totally bounded, then so is $\overline{E}$, for it is easily seen if

$$E \subseteq \bigcup_{j=1}^{N_\varepsilon} B(x_j, \varepsilon)$$

   then

$$\overline{E} \subseteq \bigcup_{j=1}^{N_\varepsilon} B(x_j, 2\varepsilon)$$

2.7.1 Equivalent Formulations of Compactness

Theorem 6. Some Equivalent Formulations

If E is a subset of a metric space (X, ρ), the following are equivalent:

(a) E is complete and totally bounded.

(b) (The Bolzano-Weierstrass Property). Every bounded sequence in E has a subsequence that converges to a point of E.

(c) If $(V_\alpha)_{\alpha \in A}$ is a cover of E by open sets, there exists a finite $F \subseteq A$ such that $(V_\alpha)_{\alpha \in F}$ covers E.

Proof. We shall show that (a) and (b) are equivalent, that (a) and (b) together imply (c), and finally (c) implies (b).

We claim $(a) \Rightarrow (b)$. Suppose that (a) holds and $(x_n)_{n\in\mathbb{N}}$ is a sequence in E. We find a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} x_{n_k} = x_0$ for some $x_0 \in E$.

E can be covered by finitely many balls of radius $\frac{1}{2}$ and at least one of them must contain x_n for infinitely many n. Say $x_n \in B_1$ for $n \in N_1 \subseteq \mathbb{N}$ and $|N_1| = \infty$.

Now $E \cap B_1$ can be covered by finitely many balls of radius $\frac{1}{2^2} = \frac{1}{4}$ and at least one of them must contain x_n for infinitely many $n \in N_2 \subseteq N_1$ say $x_n \in B_2$ for $n \in N_2 \subseteq N_1$, $|N_2| = \infty$.

Continuing inductively, we obtain a sequence of balls B_j of radius 2^{-j} and a decreasing sequence of subsets N_j of $\mathbb{N}$ such that

$$x_n \in B_j \text{ for } n \in N_j, \quad N_{j+1} \subseteq N_j \subseteq \mathbb{N}, \quad |N_j| = \infty, \quad \text{and } B_{j+1} \subseteq B_j$$

Pick $n_1 \in N_1, n_2 \in N_2, \ldots$ such that $n_1 < n_2 < n_3 \cdots$ then $(x_{n_j})_{j\in\mathbb{N}}$ is a Cauchy sequence because

$$\rho(x_{n_j}, x_{n_k}) < 2^{1-j} \text{ if } k > j \text{ since } x_{n_j}, x_{n_k} \in B_j$$

and $\operatorname{diam}(B_j) \leq 2^{1-j}$. Note that 2^{1-j} follows from the fact that the twice the radius is the diameter of a ball, thus $2 \cdot 2^{-j} = 2^{1-j}$. Since E is complete, the sequence $(x_{n_k})_{k\in\mathbb{N}}$ has a limit in E and we have proven $(a) \Rightarrow (b)$.

We claim $(b) \Rightarrow (a)$. We show the contrapositive, that is if E is not complete or not totally bounded, then (b) does not hold.

If E is not complete, there is a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ in E with no limit in E. No subsequence of $(x_n)_{n\in\mathbb{N}}$ can converge in E, for otherwise the whole sequence would converge to the same limit. Put another way, recall that all Cauchy sequences are convergent, which means that every subsequence must converge to a point outside of E.

On the other hand if E is not totally bounded, there exists a $\varepsilon > 0$ such that E cannot be covered by finitely many balls of radius ε. Choose $x_n \in E$ inductively as follows: Let $x_1 \in E$, and having chosen $x_1, \ldots, x_n$, pick

$$x_{n+1} \in E \setminus \left(\bigcup_{j=1}^{n} B(x_j, \varepsilon) \right)$$

Then $\rho(x_n, x_m) > \varepsilon$ for all $n \neq m$ so $(x_n)_{n\in\mathbb{N}}$ has no convergent subsequence.
Hence, we see that $(a) \iff (b)$.
We claim that (a) and (b) imply (c). It suffices to show that if (b) holds and $(V_\alpha)_{\alpha\in A}$ is an open cover of E, there exists $\varepsilon > 0$ such that every ball of radius $\varepsilon > 0$ that intersects E is contained in some V_α. But E can be covered by finitely many such balls by (a), thus this allows us to find a finite subcover of $(V_\alpha)_{\alpha\in A}$.
Suppose to the contrary that the claim is not true. For each $n \in \mathbb{N}$ there is a ball B_n of radius 2^{-n} such that $B_n \cap E \neq \varnothing$ and B_n is contained in no V_α. Pick $x_n \in B_n \cap E$, (by passing to a subsequence we may assume that $\lim_{n\to\infty} x_n = x$ for some $x \in E$). We have $x \in V_\alpha$ for some $\alpha \in A$ and since V_α is open there is $\varepsilon > 0$ so that $B(x, \varepsilon) \subseteq V_\alpha$. But if n is large enough so that $\rho(x_n, x) < \frac{\varepsilon}{3}$ and $2^{-n} < \frac{\varepsilon}{3}$, then $B_n \subseteq B(x, \varepsilon) \subseteq V_\alpha$ (↯).
Indeed, pick $y \in B_n$, then $\rho(y, x) \leq \rho(x_n, y) + \rho(x_n, x) < 2^{1-n} + \frac{\varepsilon}{3} < \varepsilon$.
Finally, we claim $(c) \implies (b)$. If $(x_n)_{n\in\mathbb{N}}$ is a sequence in E with no convergent subsequence, for each $x \in E$ there is a ball B_x centered at x that contains x_n for only finitely many n (otherwise, some subsequence would converge to x). Then, $(B_x)_{x\in E}$ is a cover of E by open sets with no finite subcover.

Remark.

(1) The previous theorem can be thought of as a characterization of compactness in metric spaces.

(2) Every compact set in (X, ρ) is closed (and totally bounded thus) bounded.

(3) It is not true in general that every closed and bounded set is compact. However, it is true in $\mathbb{R}^k$.

Proposition 1. Every closed and bounded subset of $\mathbb{R}^k$ is compact.
Proof. Since every closed subset of $\mathbb{R}^k$ is complete, it suffices to show that bounded subsets of $\mathbb{R}^k$ are totally bounded.
Since every bounded set is contained in some cube, define

$$Q = [-R, R]^n = \{x \in \mathbb{R}^n : \max\{|x_1|, \ldots, |x_n|\} \leq R\}.$$

It is enough to show that Q is totally bounded.
Given $\varepsilon > 0$, pick an integer $K > \frac{R\sqrt{n}}{\varepsilon}$ and express Q as the union of K^n congruent subcubes by dividing the interval $[-R, R]$ into K equal pieces.
The side lengths of these subcubes is $\frac{2R}{K}$ and hence their diameter is $\sqrt{n} \cdot \frac{2R}{K} < 2\varepsilon$ (this

follows from an inductive application of the Pythagorean Theorem). So they are contained in the balls of radius ε about their centers.

□

2.8 Connected Sets

Definition (Separated sets). Two subsets A and B of a metric space (X, ρ) are said to be separated if both $A \cap \overline{B}$ and $\overline{A} \cap B$ are empty.

In other words, no point of A lives in the closure of B and no point of B lives in the closure of A.

Definition (Connected). A set $E \subseteq X$ is said to be connected if E is not a union of two nonempty separated sets.

Remark. $[0, 1]$ and $(1, 2)$ are not separated since 1 is a limit point of $(1, 2)$. However, $(0, 1)$ and $(1, 2)$ are separated.

Theorem 7. $E \subseteq \mathbb{R}$ is connected if and only if for every $x, y \in E$ if $x < z < y$, then $z \in E$.

Proof. If there exist $x, y \in E$ and some $z \in (x, y)$ such that $z \notin E$, then $E = A_z \cup B_z$ where

$$A_z = E \cap (-\infty, z) \quad \text{and} \quad B_z = E \cap (z, \infty)$$

Since $x \in A_z$ and $y \in B_z$, we see that A_z and B_z are nonempty. Also $A_z \subseteq (-\infty, z)$ and $B_z \subseteq (z, \infty)$, which shows that they are separated. Hence, E is not connected.

For the converse, suppose that E is not connected. Then there are nonempty separated sets A, B such that $A \cup B = E$. By the definition of separated, we must have $\overline{A} \cap B = \varnothing$ and $A \cap \overline{B} = \varnothing$. Pick $x \in A$ and $y \in B$ and without loss of generality, assume $x < y$. Define $z = \sup(A \cap [x, y])$, hence $z \in \overline{A}$ and $z \notin B$. In particular, $x \leq z < y$. If $z \notin A$, it follows that $x < z < y$ and $z \notin E$.

If $z \in A$, then $z \notin \overline{B}$, hence (by the density of rationals) there is z_1 such that $z < z_1 < y$ and $z_1 \notin A$. Note that $z_1 \notin A$ because of how z is defined as a supremum of A. Then $x < z_1 < y$ and $z_1 \in E$.

□

2.9 Perfect Sets

We say that a subset E of a metric space (X, ρ) is perfect if E is closed and if every point of E is a limit point of E. Also see the earlier definition on the topology subsection.

Theorem 8. Let $\varnothing \neq P \subseteq \mathbb{R}^k$ be a perfect set. Then P is uncountable, i.e., $\text{card}(P) > \aleph_0 = \text{card}(\mathbb{N})$.

Proof. Since P has limit points, P must be infinite. To see this, suppose that $p \in P$ is a limit point. Recall that for every open set U, we must have that $U \cap (P \setminus \{p\}) \neq \varnothing$. We can choose a ball of radius 1 as our first U, then select some $p_1 \in B(p, 1) \cap (P \setminus \{p\})$. After we can build a sequence inductively by considering balls of radius $\rho(p, p_1)$, which shows that P must be infinite since $\{p_n\}_{n\in\mathbb{N}} \subseteq P$.

In fact, for every $x \in P$ and $r > 0$, we claim that $B(x,r) \cap P$ is infinite. Suppose not, i.e., there is $x_0 \in P$ and $r_0 > 0$ such that $B(x_0, r_0) \cap P = \{x_1, \dots, x_n\}$. Consider $\rho(x_0, x_1), \dots, \rho(x_0, x_n)$ and let $r = \min\limits_{1\le i\le n} \rho(x_0, x_i)$. Then $B(x_0, r) \cap P = \varnothing$, which shows that x_0 is not a limit point, a contradiction.

Thus, we can assume that $\text{card}(P) \geq \aleph_0$. Suppose, for a contradiction, that $\text{card}(P) = \aleph_0$, that is P is countable and assume $P = \{x_1, x_2, \dots\}$.

Let $V_1 = B(x_1, r)$. Then of course $V_1 \cap P \neq \varnothing$. Suppose that V_n has be constructed so that $V_n \cap P \neq \varnothing$. Note that we are taking an inductive approach in this construction and we show the explicit construction in the inductive step. Since every point of P is a limit point of P, there is a neighborhood V_{n+1} such that

(i) $\overline{V_{n+1}} \subseteq V_n$

(ii) $x_n \notin \overline{V_{n+1}}$

(iii) $V_{n+1} \cap P \neq \varnothing$

For instance, we could just choose balls with progressively smaller radii to see that neighborhood with such properties as above exist.

Let $K_n = \overline{V_n} \cap P$. This set is closed and bounded, and thus it is compact by Proposition 1 in Lecture 14. Since $x_n \notin K_{n+1}$, no point of P lies in $\bigcap_{n=1}^{\infty} K_n$, but $K_n \subseteq P$ so we have

$$\bigcap_{n=1}^{\infty} K_n = \varnothing$$

On the other hand, each $K_n = \varnothing$, is compact, and $K_{n+1} \subseteq K_n$, and the family $(K_n)_{n\in\mathbb{N}}$ has the finite intersection property, i.e., any finite intersection of members $(K_n)_{n\in\mathbb{N}}$ is non-empty:

$$K_{n_1} \cap \cdots \cap K_{n_k} \neq \varnothing$$

Thus, by Theorem 5:

$$\bigcap_{n=1}^{\infty} K_n \neq \varnothing. \quad (\lightning)$$

Hence, P must be uncountable.

$\square$

Corollary. Every interval $[a, b]$ is uncountable. In particular, $\mathbb{R}$ is uncountable.

2.10 The Cantor Set

The set which we are now going to construct shows that there exists a perfect set in $\mathbb{R}$ which contains no segment.

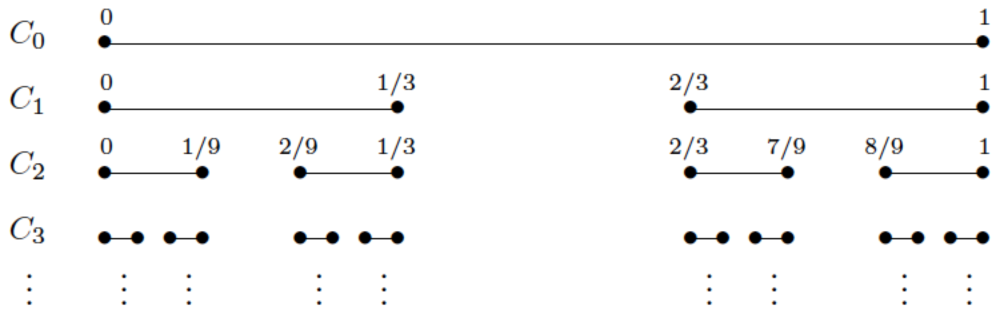


Figure 3.1: DEFINING THE CANTOR SET; $C = \bigcap_{n=0}^{\infty} C_n$.

Given C_n that consists of 2^n disjoint closed intervals each of length 3^{-n}, take each of these intervals and delete the open middle third to produce two closed intervals each of length 3^{-n-1}. Take C_{n+1} to be the union of the 2^{n+1} closed intervals so formed and continue.
The set $C = \bigcap_{n=0}^{\infty} C_n$ is called the Cantor set or the ternary Cantor set.
Clearly C is compact (because it is the arbitrary intersection of closed sets), and since $C_0 \supseteq C_1 \supseteq C_2 \cdots$ and each C_n is compact, it follows that the family $(C_n)_{n\in\mathbb{N}}$ has the finite intersection property and thus $C \neq \varnothing$.
No segment of the form

$$\left(\frac{3k+1}{3^m}, \frac{3k+2}{3^m}\right) \tag{$*$}$$

where k and m are positive integers, has a part in common with C. Since any segment (α, β) contains a segment of the form $(*)$ if m is sufficiently large, then the set

$$L = \left\{\frac{\ell}{3^m} : m \in \mathbb{N} \text{ and } \ell = 0, 1, \ldots, 3^m - 1\right\}$$

is dense in $[0, 1]$. Thus C contains no segment (α, β). More explicitly, suppose we have an interval (α, β). Because the above set L is dense, we can find some interval of the form $(*)$ that is a subset of (α, β). Since the Cantor set has no elements in common with an interval of the form $(*)$, it also cannot contain (α, β).
This also shows that the interior of the Cantor set is empty, that is, $\operatorname{Int}(C) = \varnothing$. Hence, along with the fact that the Cantor set is closed, we deduce that the Cantor set is a nowhere dense set by definition.
To show that C is perfect, it is enough to show that C contains no isolated points. Let $x \in C$ and let I_n be the unique interval from C_n which contains $x \in I_n$. Let x_n be an endpoint of I_n such that $x_n \neq x$. It follows from the construction of C that $x_n \in C$. Hence x is a limit point of C and thus C is perfect.

2.10.1 Measure

Definition. We say that a subset $E \subseteq \mathbb{R}$ has measure zero (or the Lebesgue measure 0) if for every $\varepsilon > 0$ there is a countable collection of open intervals $(I_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}$ such that

$$E \subseteq \bigcup_{n\in\mathbb{N}} I_n \quad \text{and} \quad \sum_{n\in\mathbb{N}} |I_n| < \varepsilon$$

where $|I_n|$ is the length of the interval I_n. That is, if $I = (a, b)$, then $|I| = b - a$.
Examples.

1. The empty set has obviously measure zero.
2. $\mathbb{Q}$ has measure zero. Let $\mathbb{Q} = \{r_n : n \in \mathbb{N}\}$ and for a given $\varepsilon > 0$ consider

   $$I_n = \left(r_n - \frac{\varepsilon}{2^{n+2}}, r_n + \frac{\varepsilon}{2^{n+2}}\right)$$

   Then we have

   $$\mathbb{Q} = \bigcup_{n\in\mathbb{N}} \{r_n\} \subseteq \bigcup_{n\in\mathbb{N}} \left(r_n - \frac{\varepsilon}{2^{n+2}}, r_n + \frac{\varepsilon}{2^{n+2}}\right)$$

   and

   $$\sum_{n=1}^{\infty} |I_n| = \sum_{n=1}^{\infty} \frac{\varepsilon}{2^{n+1}} = \frac{\varepsilon}{2}\left(\frac{1}{1-\frac{1}{2}} - 1\right) < \varepsilon$$

   where the geometric sum represented is $\sum_{n=0}^{\infty} \frac{1}{2^n}$.

Refer to Figure 3.1 on Page 29 for a visual representation of the C_n which form the Cantor set.
Each C_n consists of 2^n disjoint closed intervals each of length 3^{-n}. We shall write

$$C_n = \dot{\bigcup_{k=1}^{2^n}} I_{n,k} \quad \text{and} \quad |I_{n,k}| = 3^{-n}$$

$$\mathcal{C} = \bigcap_{n=0}^{\infty} C_n \quad \text{and} \quad \mathcal{C} \subseteq C_n \text{ for all } n \in \mathbb{N}$$

Given $\varepsilon > 0$, take $N \in \mathbb{N}$ so that $n \geq N$ implies

$$\left(\frac{2}{3}\right)^n < \varepsilon$$

In particular $\mathcal{C} \subseteq C_n$ for $n \geq N$ and

$$\sum_{k=1}^{2^n} |I_{n,k}| = \sum_{k=1}^{2^n} \frac{1}{3^n} = \left(\frac{2}{3}\right)^n < \varepsilon$$

since

$$C_n = \bigcup_{k=1}^{2^n} I_{n,k} \quad \text{and} \quad |I_{n,k}| = 3^{-n}$$

We conclude that $\mathcal{C}$ has measure zero.
Each component of C_n can also be described as the set

$$C_n = \left\{ \sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j} \mid \varepsilon_j \in \{0,1,2\} \text{ and } \varepsilon_j \neq 1 \text{ for } 1 \leq j \leq n \right\}$$

Consequently

$$\mathcal{C} = \left\{ \sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j} \mid \varepsilon_j \in \{0,2\} \text{ for all } j \in \mathbb{N} \right\}$$

Note that leaving out the 1 in the above definition is equivalent to leaving out the middle third in the pictorial representation of the Cantor set.
Take $\varepsilon = (\varepsilon_j)_{j\in\mathbb{N}}$ and $\delta = (\delta_j)_{j\in\mathbb{N}} \in \{0,2\}^{\mathbb{N}}$ such that $\varepsilon \neq \delta$.
Let $N = \min\{j \in \mathbb{N} : \varepsilon_j \neq \delta_j\}$ and assume that $\varepsilon_N = 0$ and $\delta_N = 2$. Observe that

$$\begin{aligned}
\sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j} &= \sum_{j=1}^{N-1} \frac{\varepsilon_j}{3^j} + \sum_{j=N+1}^{\infty} \frac{\varepsilon_j}{3^j} \\
&\leq \sum_{j=1}^{N-1} \frac{\delta_j}{3^j} + \frac{2}{3^{N+1}} \sum_{j=0}^{\infty} \frac{1}{3^j} \\
&= \sum_{j=1}^{N-1} \frac{\delta_j}{3^j} + \frac{2}{3^{N+1}} \cdot \frac{1}{1-\frac{1}{3}} = \sum_{j=1}^{N-1} \frac{\delta_j}{3^j} + \frac{1}{3^N} \\
&< \sum_{j=1}^{N-1} \frac{\delta_j}{3^j} + \frac{2}{3^N} \leq \sum_{j=1}^{\infty} \frac{\delta_j}{3^j}
\end{aligned}$$

Thus we have shown that any number $\sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j}$ is uniquely determined by its sequence $\varepsilon = (\varepsilon_j)_{j\in\mathbb{N}} \in \{0,2\}^{\mathbb{N}}$.
Remark.

$$\begin{aligned}
\frac{1}{3} &= \sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j} = A \quad \varepsilon_1 = 1, \quad \varepsilon_j = 0 \text{ for } j \geq 2 \\
\frac{1}{3} &= \sum_{j=1}^{\infty} \frac{\varepsilon_j}{3^j} = B \quad \varepsilon_1 = 0, \quad \varepsilon_j = 2 \text{ for } j \geq 2
\end{aligned}$$

There is a bijection $\phi : \{0,1\}^{\mathbb{N}} \to \mathcal{C}$ defined by

$$\phi(z) = \frac{2}{3} \cdot \sum_{j=0}^{\infty} \frac{z_j}{3^j} \qquad \text{for all } z = (z_j)_{j\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$$

This shows that

$$\operatorname{card}(\mathcal{C}) = \operatorname{card}(\{0,1\}^{\mathbb{N}}) = \operatorname{card}(\mathbb{R}) = \mathfrak{c}$$

2.11 Some Special Sequences

Theorem 9.

(a) If $p > 0$, then $\lim_{n\to\infty} \frac{1}{n^p} = 0$

(b) If $p > 0$, then $\lim_{n\to\infty} \sqrt[n]{p} = 1$

(c) $\lim_{n\to\infty} \sqrt[n]{n} = 1$

(d) If $p > 0$ and $\alpha \in \mathbb{R}$, then $\lim_{n\to\infty} \frac{n^\alpha}{(1+p)^n} = 0$

(e) If $|x| < 1$, then $\lim_{n\to\infty} x^n = 0$

2.11.1 Basic Inequalities

Binomial Theorem. For $x, y \in \mathbb{R}$, $n \in \mathbb{N}$, the following identities hold:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \qquad n! = 1 \cdot 2 \cdots n \qquad 0! = 1$$

Bernoulli's inequality. If $r \geq 0$ is an integer and $x \geq -1$, then

$$(1+x)^r \geq 1 + rx \tag{$*$}$$

Proof of $(*)$ We prove a stronger statement.

$$(1+x_1)\cdots(1+x_r) \geq 1 + x_1 + \ldots + x_r \tag{$**$}$$

where $x_1, \ldots, x_r$ are real numbers all greater than -1, all with the same sign.
We use induction.
Base case. $n = 1$. It is clear that $1 + x_1 \geq 1 + x_1$.
Inductive step. Assume that $(**)$ holds up to r. Observe that this means

$$\begin{aligned}\prod_{j=1}^{r+1}(1+x_j) &\geq (1 + x_1 + \ldots + x_r)(1 + x_{r+1}) \\ &= 1 + x_1 + \ldots + x_r + (1 + x_1 + \ldots + x_r)x_{r+1} \\ &= 1 + x_1 + \ldots + x_{r+1} + \underbrace{(x_1 + \ldots + x_r)x_{r+1}}_{\geq 0} \\ &\geq 1 + x_1 + \ldots + x_{r+1}\end{aligned}$$

This proves $(**)$. If we take $x_1 = \cdots = x_r = x$, then $(**)$ gives $(*)$.
Proof of Theorem 9.

(a) If $p > 0$, the $\lim_{n\to\infty} \frac{1}{n^p} = 0$.
Let $\varepsilon > 0$ be arbitrary and choose $n > \left(\frac{1}{\varepsilon}\right)^{\frac{1}{p}}$, then

$$\frac{1}{n^p} < \frac{1}{(\left(\frac{1}{\varepsilon}\right)^{\frac{1}{p}})^p} = \frac{1}{\frac{1}{\varepsilon}} = \varepsilon$$

(b) We proceed by cases. If $p > 1$, let $x_n = \sqrt[n]{p} - 1$, then $x_n > 0$. By the Bernoulli inequality

$$p = (1 + x_n)^n \geq 1 + nx_n$$

thus

$$0 \leq x_n \leq \frac{p-1}{n} \xrightarrow[n\to\infty]{} 0$$

Hence $\lim_{n\to\infty} x_n = 0$.
If $p = 1$, then $\lim_{n\to\infty} \sqrt[n]{p} = 1$. If $0 < p < 1$, we show

$$\lim_{n\to\infty} \frac{1}{\sqrt[n]{p}} = \lim_{n\to\infty} \sqrt[n]{\frac{1}{p}} = 1$$

and we reuse the previous argument since $\frac{1}{p} > 1$ when $0 < p < 1$.
(c) Set $x_n = \sqrt[n]{n} - 1$. Then $x_n \geq 0$. Hence, by rearranging x_n and by the binomial theorem, we see that:

$$n = (1 + x_n)^n \geq \binom{n}{2} \cdot x_n^2 = \frac{n(n-1)}{2} \cdot x_n^2$$

We derived the first inequality by only keeping the $k = 2$ term in the binomial expansion. Hence,

$$0 \leq x_n \leq \sqrt{\frac{2}{n-1}} \xrightarrow[n\to\infty]{} 0$$

Note that rearranging $n \geq \frac{n(n-1)}{2} \cdot x_n^2$ allows us to derive the inequality above.
(d) If $p > 0$ and $\alpha \in \mathbb{R}$, then $\lim_{n\to\infty} \frac{n^\alpha}{(1+p)^n} = 0$.
Let $k \in \mathbb{N}$ be so that $k > \alpha$, $\quad k > 0$. For $n > 2k$, we get

$$(1+p)^n > \binom{n}{k} p^k = \frac{n(n-1)\cdots(n-k+1)}{k!} p^k > \frac{(\frac{n}{2})^k \cdot p^k}{k!} = \frac{n^k \cdot p^k}{2^k k!}$$

The second inequality follows from the fact that $\frac{n}{2} > k \implies n - k > n - \frac{n}{2} = \frac{n}{2}$. Hence,

$$0 < \frac{n^\alpha}{(1+p)^n} < \frac{2^k \cdot k!}{p^k} \cdot n^{\alpha - k}$$

Since $\alpha - k < 0$, we see that $\lim_{n\to\infty} n^{\alpha-k} = 0$ by (a).
(e) If $|x| < 1$, then $\lim_{n\to\infty} x^n = 0$. Simply take $\alpha = 0$ in (d). We do not have to worry about the case where x is negative in our replacement since when we test the limit, we will be

evaluating $|x^n - 0| < \varepsilon$, and the absolute value sign will make all the x^n positive.

Pythagorean means. Let A_n, G_n, H_n denote the arithmetic, geometric, and harmonic means of n positive numbers $a_1, a_2, \ldots, a_n$.

$$A_n = \frac{a_1 + \ldots + a_n}{n}$$
$$G_n = (a_1 \cdots a_n)^{\frac{1}{n}}$$
$$H_n = \frac{n}{\frac{1}{a_1} + \ldots + \frac{1}{a_n}}$$

Then we have

$$H_n \leq G_n \leq A_n$$

Proof. Using induction, one can easily prove that if $x_1, \ldots, x_n$ are positive real numbers such that $x_1 \cdots x_n = 1$, then $x_1 + \ldots + x_n \geq n$. So, if the product of n real numbers is 1, then their sum must be at least n.

Base case. Suppose we have one real number whose product is 1. This is the same thing as saying $x_1 = 1$. Thus the inequality is satisfied.

Inductive step. Suppose that for k real numbers, $x_1, \ldots, x_k$, the inductive hypothesis shows us that if $x_1 \cdots x_k = 1$, then $x_1 + \ldots + x_k \geq k$. Let x_{k+1} satisfy $x_1 \cdots x_{k+1} = 1$. It is clear that

$$x_1 \cdots x_{k+1} = (x_1 \cdots x_k) \cdot x_{k+1} = 1 \cdot x_{k+1} = 1$$

Since $x_{k+1} = 1$, we add this to the inequality we have from the inductive hypothesis to see that $x_1 + \ldots + x_{k+1} \geq k + 1$, as desired.

Now, we prove that $A_n \geq G_n$. Take

$$x_j = \frac{a_j}{(a_1 \cdots a_n)^{\frac{1}{n}}} > 0$$

where $x_j > 0$ since $a_1, \ldots, a_n$ are positive numbers. Observe also that

$$x_1 \cdots x_n = \frac{\prod_{i=1}^n a_i}{((a_1 \cdots a_n)^{\frac{1}{n}})^n} = 1$$

Thus, from the previous inequality we proved, it follows that $\sum_{i=1}^n x_n \geq n$, which shows us that

$$x_1 + \ldots + x_n = \frac{\sum_{i=1}^n a_i}{(a_1 \cdots a_n)^{\frac{1}{n}}} \geq n$$
$$\sum_{i=1}^n a_i \geq n \cdot (a_1 \cdots a_n)^{\frac{1}{n}}$$
$$A_n \geq G_n$$

To prove $G_n \geq H_n$, we apply $A_n \geq G_n$ with $\frac{1}{a_j}$ in place of a_j. Hence

$$\frac{\frac{1}{a_1} + \ldots + \frac{1}{a_n}}{n} \geq \left(\frac{1}{a_1} \cdots \frac{1}{a_n}\right)^{\frac{1}{n}}$$

$$G_n = \left(\frac{1}{a_1} \cdots \frac{1}{a_n}\right)^{-\frac{1}{n}} \geq \frac{n}{\frac{1}{a_1} + \ldots + \frac{1}{a_n}} = H_n$$

□

Corollary. $(1+x)^n \geq 1 + nx$ for $x > 0$.

Proof. Observe that

$$1 + x = \frac{\overbrace{1 + 1 + \ldots + 1}^{n-1} + (1+nx)}{n} \geq (1 \cdots 1 \cdot (1+nx))^{\frac{1}{n}}$$
$$(1+x)^n \geq 1 + nx$$

2.12 Series

Given a sequence $(a_n)_{n\in\mathbb{N}} \subseteq \mathbb{C}$, we shall write

$$\sum_{n=p}^{q} a_n = a_p + a_{p+1} + \ldots + a_q, \quad p \leq q$$

With $(a_n)_{n\in\mathbb{N}}$ we associated a sequence $(s_n)_{n\in\mathbb{N}}$ of its partial sums

$$s_n = \sum_{k=1}^{n} a_k$$

If $(s_n)_{n\in\mathbb{N}}$ converges to s, we say that

$$\lim_{n\to\infty} s_n = s \text{ and write } s = \sum_{k=1}^{\infty} a_k \text{ or } \left|\sum_{k=1}^{\infty} a_k\right| < \infty$$

where the last symbol we call an infinite series or just series.

If $(s_n)_{n\in\mathbb{N}}$ diverges, the series is said to diverge. We then write $|\sum_{k=1}^{\infty} a_k| = \infty$.

Theorem 10. The series $\sum a_n$ converges if and only if for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that

$$\left|\sum_{k=n}^{m} a_k\right| \leq \varepsilon \quad \text{if } m \geq n \geq N \tag{$*$}$$

Proof. If $|\sum_{n=1}^{\infty} a_n| < \infty$, then $\lim_{n\to\infty} s_n = s$ where $s_n = \sum_{k=1}^{n} a_k$. Since $(s_n)_{n\in\mathbb{N}}$ converges, it must satisfy the Cauchy condition. This means that for every $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that $m \geq n \geq N$ implies $|s_m - s_{n-1}| = |\sum_{k=n}^{m} a_k| < \varepsilon$.

On the other hand, if $(*)$ in Theorem 10 holds then

$$|s_m - s_{n-1}| = \left|\sum_{k=n}^{m} a_k\right| < \varepsilon$$

for all $m \geq n \geq N$. So, the sequence $(s_n)_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{C}$, which is complete. Thus there is $s \in \mathbb{C}$ so that $\lim_{n\to\infty} s_n = s$.

Theorem 11. If $\sum_{n\in\mathbb{N}} a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.

Proof. It suffices to take $m = n$ in Theorem 10

$$|a_n| = \left|\sum_{k=m}^{m} a_k\right| \leq \varepsilon \quad \text{for } n \geq N$$

Since ε may be made arbitrarily small, we must have $\lim_{n\to\infty} a_n = 0$.

□

The condition $a_n \xrightarrow[n\to\infty]{} 0$ is not, however, sufficient to ensure convergence of $\sum_{n\in\mathbb{N}} a_n$.

Example. The harmonic series satisfies $\lim_{n\to\infty} \frac{1}{n} = 0$, but $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$. See Example 2.4.5 on Pg. 58 of Abbott's *Understanding Analysis* for a proof.

Theorem 12. A series of nonnegative (note that nonnegative always refers to real numbers) terms $a_k \geq 0$ converges if and only if its partial sums form a bounded sequence.

Proof. If $\sum_{k=1}^{\infty} a_k < \infty$, then there is $M > 0$ such that for all $N \in \mathbb{N}$ we have $S_N = \sum_{k=1}^{N} a_k \leq M$. In fact, we can take

$$M = \sum_{k=1}^{\infty} a_k \geq \sum_{k=1}^{N} a_k = S_N \quad \text{for all } N \in \mathbb{N}$$

since $a_k \geq 0$ and $S_N \leq S_{N+1}$. In other words, if the sequence of partial sums converges, we can just use the value it converges to as its upper bound, and the sequence is bounded below by 0 since the terms are nonnegative.

On the other hand, if there is $M > 0$ such that $S_N = \sum_{k=1}^{N} a_k \leq M$ for all $N \in \mathbb{N}$, then

$$\lim_{N\to\infty} S_N \text{ exists}$$

since $S_N \leq S_{N+1}$ and S_N is bounded (limit exists by Monotone Convergence Theorem).

Theorem 13. (a) If $|a_n| \leq c_n$ for $n \geq N_0$ where N_0 is a fixed integer and if $\sum_{n\in\mathbb{N}} c_n$ converges, then $\sum_{n\in\mathbb{N}} a_n$ converges as well.

(b) If $a_n \geq d_n \geq 0$ for $n \geq N_0$ and if $\sum_{n\in\mathbb{N}} d_n = \infty$, then $\sum_{n\in\mathbb{N}} a_n = \infty$.

These are the comparison tests.

Proof. Given $\varepsilon > 0$ there is $N \geq N_0$ so that $m \geq n \geq N$ implies

$$\sum_{k=n}^{m} c_k \leq \varepsilon$$

by the Cauchy Criterion. Hence,

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

and (a) follows.
If $\sum_{n=1}^{\infty} d_n = \infty$, then also $\sum_{n=N_0}^{\infty} d_n = \infty$. Thus

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{N_0-1} a_n + \sum_{n=N_0}^{\infty} a_n \geq \sum_{n=1}^{N_0-1} a_n + \sum_{n=N_0}^{\infty} d_n = \infty$$

and we see that $\sum_{n\in\mathbb{N}} a_n$ diverges.

2.12.1 Series of Nonnegative Terms

It is very well-known that:

$$(1-x)(1+x+x^2+\ldots+x^n) = 1-x^{n+1} \quad \text{for } x \in \mathbb{C}$$

This follows from the fact that the sum telescopes after distributing.
Theorem 14 (Geometric series). If $0 \leq x < 1$, then

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

If $x \geq 1$, then the series diverges, that is, $\sum_{n=1}^{\infty} x^n = \infty$.
Proof. If $x \neq 1$, then

$$s_n = \sum_{k=0}^{n} x^k = \frac{1+x^{n+1}}{1-x}$$

Recall that $\lim_{n\to\infty} x^n = 0$ for $|x| < 1$. Thus $\lim_{n\to\infty} s_n = \frac{1}{1-x}$ for $|x| < 1$.
If $x \geq 1$, then

$$s_n = \sum_{k=0}^{n} x^k \geq (n+1).$$

The above inequality follows from the proof about arithmetic, geometric, and harmonic means on Pg. 29 of these notes. Thus $\lim_{n\to\infty} s_n = \infty$.

□

Theorem 15. Suppose that $a_1 \geq a_2 \geq \ldots \geq 0$, then the series

$$S = \sum_{n=1}^{\infty} a_n < \infty$$

if and only if

$$T = \sum_{k=0}^{\infty} 2^k \cdot a_{2^k} < \infty$$

Proof. Let $s_n = a_1 + \ldots + a_n$ and $t_k = a_1 + 2a_2 + \ldots + 2^k \cdot a_{2^k}$. Also, let $\lim_{n\to\infty} s_n = S$ and $\lim_{k\to\infty} t_k = T$.
For $n < 2^k$ one has

$$\begin{aligned} s_n &\le a_1 + (a_2 + a_3) + \ldots + (a_{2^k} + \ldots + a_{2^{(k+1)}-1}) \\ &\le a_1 + 2a_2 + \ldots + 2^k \cdot a_{2^k} = t_k \end{aligned}$$

Thus $s_n \le t_k$. Conversely, if $n > 2^k$

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

so that $2s_k \ge t_k$.
Hence, we conclude that the sequences $(s_n)_{n\in\mathbb{N}}$, $(t_k)_{k\in\mathbb{N}}$ are either both bounded or unbounded.

□

Corollary (p-series).

$$\sum_{n=1}^{\infty} \frac{1}{n^p} < \infty \quad \text{if } p > 1$$

and

$$\sum_{n=1}^{\infty} \frac{1}{n^p} = \infty \quad \text{if } p \le 1$$

Proof. If $p \le 0$, then $\lim_{n\to\infty} n^{-p} = \infty$ thus $\sum_{n=1}^{\infty} \frac{1}{n^p} = \infty$.
Let $a_n = \frac{1}{n^p}$. If $p > 0$, then we apply Theorem 15 and we are led to the series

$$\begin{aligned} \sum_{k=0}^{\infty} 2^k a_{2^k} &= \sum_{k=0}^{\infty} 2^k \cdot \frac{1}{2^k p} \\ &= \sum_{k=0}^{\infty} 2^{(1-p)k} \end{aligned}$$

Now $2^{1-p} < 1$ if and only if $1 - p < 0$. Also $1 - p < 0$ if and only if $p > 1$. Hence

$$\sum_{k=0}^{\infty} 2^{(1-p)k} = \begin{cases} C_p < \infty & \text{if } p > 1 \\ \infty & \text{if } p \le 1 \end{cases}$$

We are using 2^{1-p} in place of the x of a geometric series with terms x^n. The result follows by comparison with the geometric series, Theorem 14.

2.13 Absolute Convergence

The series $\sum_{n=1}^{\infty} a_n$ is said to converge absolutely if the series $\sum_{n=1}^{\infty} |a_n| < \infty$.
Theorem 16. If $\sum_{n=1}^{\infty} a_n$ converges absolutely, then $\sum_{n=1}^{\infty} a_n$ converges.

Proof. The claim follows from the inequality:

$$\left|\sum_{n=k}^{m} a_n\right| \leq \sum_{n=k}^{m} |a_n| \leq \varepsilon, \qquad \text{for } m \geq k \geq N_\varepsilon$$

and the Cauchy criterion. The inequality follows from Theorem 1.33(e) in Rudin, which is a form of the triangle inequality.

Examples.

(1) For series with positive terms, absolute convergence is the same as convergence.

(2) $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$ converges absolutely since $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$. In fact, $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.
Proof.
Recall the following identity:

$$\frac{1}{m(m-1)} = \frac{m-(m-1)}{m(m-1)} = \frac{1}{m-1} - \frac{1}{m}$$

Observe that:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^2} &= \lim_{n\to\infty} \sum_{m=1}^{n} \frac{1}{m^2} \\
&= \lim_{n\to\infty} \left(1 + \sum_{m=2}^{n} \frac{1}{m^2}\right) \\
&\leq \lim_{n\to\infty} \left(1 + \sum_{m=2}^{n} \frac{1}{m(m-1)}\right) \\
&= \lim_{n\to\infty} \left(1 + \sum_{m=2}^{n} \left(\frac{1}{m-1} - \frac{1}{m}\right)\right) \\
&= \lim_{n\to\infty} \left(1 + 1 + \frac{1}{n}\right) = 2
\end{aligned}$$

(3) If $\sum_{n=1}^{\infty} a_n$ converges but $\sum_{n=1}^{\infty} |a_n| = \infty$, we say that $\sum_{n=1}^{\infty} a_n$ converges non-absolutely.

Consider the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ and

$$s_m = \sum_{n=1}^{m} \frac{(-1)^n}{n} = -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \ldots + \frac{(-1)^m}{m}$$

Then it turns out that $\lim_{m\to\infty} s_m$ exists thus the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ converges non-absolutely since we know $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$. In order to prove this we show a more general result.

2.14 Summation By Parts

Theorem 17. (Rudin 3.41)
Given two sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ set

$$A_n = \sum_{k=0}^{n} a_k \quad \text{if } n \geq 0 \text{ and}$$
$$A_{-1} = 0$$

Then if $0 \leq p \leq q$ one has

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

Proof.

$$\begin{aligned}\sum_{n=p}^{q} a_n b_n &= \sum_{n=p}^{q} \overbrace{(A_n - A_{n-1})}^{a_n} b_n = \sum_{n=p}^{q} A_n b_n - \sum_{n=p}^{q} A_{n-1} b_n \\ &= \sum_{n=p}^{q} A_n b_n - \sum_{n=p-1}^{q-1} A_n b_{n+1} \\ &= \sum_{n=p}^{q-1} A_n(b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p\end{aligned}$$

Note that for the last step in the chain of equalities, we resolve the indices by taking out the $A_q b_q$ term from the first sum and the $A_{p-1} b_p$ term from the second sum.

□

Theorem 18. (Rudin 3.42) Suppose that

(a) The partial sums $A_n = \sum_{k=0}^{\infty} a_k$ of $(a_n)_{n\in\mathbb{N}}$ form a bounded sequence.

(b) $b_0 \geq b_1 \geq b_2 \ldots$

(c) $\lim_{n\to\infty} b_n = 0$

Then $\sum_{n=0}^{\infty} a_n b_n$ converges.
Proof.
Choose $M \geq 0$ so that $|A_n| \leq M$ for all $n \in \mathbb{N}$. Given $\varepsilon > 0$, there is $N \in \mathbb{N}$ so that

$$b_N \leq \frac{\varepsilon}{2M}$$

since $\lim_{n\to\infty} b_n = 0$. For $N \le p \le q$, one has by the summation by parts that

$$\begin{aligned}\left|\sum_{n=p}^{q} a_n b_n\right| &= \left|\sum_{n=p}^{q-1} A_n(b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p\right| \\ &\le M \cdot \left(\sum_{n=p}^{q-1} |b_n - b_{n+1}| + b_q + b_p\right) \\ &\le 2Mb_p \le 2Mb_N \le \varepsilon\end{aligned}$$

To expand on one of the steps above, by Property (b) we know

$$\sum_{n=p}^{q-1} |b_n - b_{n+1}| = \sum_{n=p}^{q-1} b_n - b_{n+1} = b_p - b_q$$

□

We now show that $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ converges. Let $a_n = (-1)^n$ and $b_n = \frac{1}{n}$ in the previous theorem. We see then

$$\left|\sum_{n=1}^{\infty} \frac{(-1)^n}{n}\right| = \left|\sum_{n=1}^{\infty} a_n b_n\right| < \infty$$

since $|A_n| = \left|\sum_{k=1}^{n} (-1)^k\right| \le 1$.

More generally, we have:

Corollary. (Rudin 3.43) Suppose

(a) $|c_1| \ge |c_2| \ge |c_3| \ge \dots$

(b) $c_{2m-1} \ge 0, \quad c_{2m} \le 0$ for $m \in \mathbb{N}$

(c) $\lim\limits_{n\to\infty} c_n = 0$

Then $\sum_{n=1}^{\infty} c_n$ converges.

Proof. Apply Theorem 18 with $a_n = (-1)^{n+1}$ and $b_n = |c_n|$.

□

2.15 Upper and Lower Limits

We write that $s_n \to +\infty$ if for every $M > 0$ there is $N \in \mathbb{N}$ such that $n \ge N$ implies $s_n \ge M$. Similarly, we write $s_n \to -\infty$ if for every $M > 0$, there exists an integer $N \in \mathbb{N}$ such that $n \ge N$ implies $s_n \le -M$.

Let $(s_n)_{n\in\mathbb{N}}$ be a sequence of real numbers.

The upper limit is defined by the first expression below, the second expression is the analogous formulation for sets

$$\limsup_{n\to\infty} s_n = \inf_{k\geq 1} \sup_{n\geq k} s_n \qquad \limsup_{n\to\infty} E_n = \bigcap_{k\geq 1} \bigcup_{n\geq k} E_n$$

$$\limsup_{n\to\infty} s_n = \inf \{\sup \{s_n : n \geq k\} : k > 0\}$$

whereas the lower limit is defined by

$$\liminf_{n\to\infty} s_n = \sup_{k\geq 1} \inf_{n\geq k} s_n \qquad \liminf_{n\to\infty} E_n = \bigcup_{k\geq 1} \bigcap_{n\geq k} E_n$$

$$\liminf_{n\to\infty} s_n = \sup \{\inf \{s_n : n \geq k\} : k > 0\}$$

Remark. Let $\alpha_k = \sup_{n\geq k} s_n$, then $\alpha_{k+1} \leq \alpha_k$ and

$$\limsup_{n\to\infty} s_n = \inf_{k\geq 1} \underbrace{\sup_{n\geq k} s_n}_{\alpha_k} = \lim_{k\to\infty} \alpha_k$$

where the limit is possibly infinite.
If $\beta_k = \inf_{n\geq k} s_n$ then $\beta_k \leq \beta_{k+1}$ again, and

$$\liminf_{n\to\infty} s_n = \sup_{k\geq 1} \underbrace{\inf_{n\geq k} s_n}_{\beta_k} = \lim_{k\to\infty} \beta_k$$

where the limit is possibly infinite.
We always have $\beta_k = \inf_{n\geq k} s_n \leq \sup_{n\geq k} s_n = \alpha_k$, thus

$$\liminf_{n\to\infty} = \lim_{k\to\infty} \beta_k \leq \lim_{k\to\infty} \alpha_k = \limsup_{n\to\infty} s_n$$

note that the above follows from the Order Limit Theorem.
Examples.

(1) Consider $a_n = (-1)^n \cdot \frac{n+1}{n}$.
Let
$$\beta_n = \sup \left\{ (-1)^n \cdot \frac{n+1}{n}, \;\; (-1)^{n+1} \cdot \frac{n+2}{n+1}, \;\; \ldots \right\}$$
Then we have
$$\beta_n = \begin{cases} \frac{n+1}{n} & \text{if } n \text{ is even} \\ \frac{n+2}{n+1} & \text{if } n \text{ is odd} \end{cases}$$
The above follows from the fact that if $n < m$, then $\frac{n+1}{n} > \frac{m+1}{m}$ with the analogous argument for the case where n is odd.
Thus, $\lim_{n\to\infty} \beta_n = 1$. Therefore, $\limsup_{n\to\infty} a_n = 1$. Similarly, $\liminf_{n\to\infty} a_n = -1$.

(2) Take

$$a_n = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even} \end{cases}$$

Then put $\beta_n = \sup\{a_m : m \geq n\} = 1$ and $\alpha_n = \inf\{a_m : m \geq n\} = 0$.
Therefore $\liminf_{n\to\infty} a_n = 0$ and $\limsup_{n\to\infty} a_n = 1$.

(3) Let $a_n = \frac{1}{n}$ and put

$$\beta_n = \sup\left\{\frac{1}{m} : m \geq n\right\} = \frac{1}{n}, \qquad \text{then } \lim_{n\to\infty} \beta_n = 0$$

Similarly, $\alpha_n = \inf\{\frac{1}{m} : m \geq n\} = 0$. Thus $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = 0$.

Definition (Rudin 3.16) Let $(s_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Let E be the set of all $x \in \overline{\mathbb{R}}$, where $\overline{\mathbb{R}}$ refers to the extended reals, such that $s_{n_k} \xrightarrow[k\to\infty]{} x$ for some subsequence $(s_{n_k})_{k\in\mathbb{N}}$. In other words, E is the set of all subsequential limits of $(s_n)_{n\in\mathbb{N}}$. Let

$$s^* = \sup E \quad \text{ and } s_* = \inf E$$

Fact: Let (X, ρ) be a metric space. If $(x_n)_{n\in\mathbb{N}}$ converges to $x_0 \in X$, then for every $(n_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$, the subsequence $(x_{n_k})_{k\in\mathbb{N}}$ also converges to x_0.
Proof. Given $\varepsilon > 0$ we find $N \in \mathbb{N}$ such that $n \geq N$ implies $\rho(x_n, x_0) < \varepsilon$. Fix $(n_k)_{k\in\mathbb{N}}$ and pick $K_0 \in \mathbb{N}$ such that for all $k \geq K_0$ we have $n_k \geq N$. Then we see that $\rho(x_{n_k}, x_0) < \varepsilon$ and we are done.

2.15.1 Relationship to Subsequential Limits

Theorem. Let $(s_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Let E be the set of all subsequential limits as given in Definition 3.16. Then

$$\limsup_{n\to\infty} s_n = s^* = \sup E \quad \text{ and } \quad \liminf_{n\to\infty} s_n = s_* = \inf E$$

Proof. We proceed by cases.

1. Suppose that $\limsup\limits_{n\to\infty} s_n = +\infty$. Thus, by definition, we have

$$\inf_{k\geq 1} \sup_{n\geq k} s_n = +\infty, \quad \text{ so } \sup_{n\geq k} s_n = +\infty \text{ for all } k \in \mathbb{N}$$

Hence, there is $(n_k)_{k\in\mathbb{N}}$ so that $\lim_{k\to\infty} s_{n_k} = +\infty$. This gives $s^* = \sup E = +\infty$.

2. Suppose that $\inf\limits_{k\geq 1} \sup\limits_{n\geq k} s_n = -\infty$, that is,

$$\lim_{k\to\infty} \sup_{n\geq k} s_n = -\infty$$

This means that for every $M > 0$ there is $N \in \mathbb{N}$ so that $n \geq N$ implies $\sup_{n \geq k} s_n \leq -M$. Hence $s_n \leq -M$ for all $n \geq N$, that is, $\lim_{n\to\infty} s_n = -\infty$. Hence $E = \{-\infty\}$ and $s^* = \sup E = -\infty$.

3. Assume now that $\limsup_{n\to\infty} s_n = L$ and $L \in \mathbb{R}$. We show

   (a) $\sup E \leq L$

   (b) $L \in E$

   where $s^* = \sup E = L$.
   This gives us the strange conclusion that $\limsup_{n\to\infty} s_n = \sup E = \max E$.
   Proof of (a). Suppose, for contradiction, that $L < \sup E$. Thus, there is $x \in E$ such that $L < x \leq \sup E$. Since $x \in E$, there is a sequence $(s_{n_j})_{j\in\mathbb{N}}$ so that $\lim_{j\to\infty} s_{n_j} = x$. In particular, there exist $\widetilde{K} \geq 1$ so that

$$\frac{x+L}{2} < s_{n_j} \quad \text{for all } j \geq \widetilde{K}$$

   We obtain the above through an appropriate choice of ε, observe that since $\lim_{j\to\infty} s_{n_j} = x$, we see that for all $\varepsilon > 0$, there exists $\widetilde{K} \in \mathbb{N}$ such that for all $j \geq \widetilde{K}$ we have

$$|s_{n_j} - L| < \varepsilon$$

   But choosing $\varepsilon = \frac{x-L}{2} > 0$ (this is true since $x > L$ by our hypothesis), we see that

$$x - \varepsilon < s_{n_j}$$
$$x - \frac{x-L}{2} = \frac{x+L}{2} < s_{n_j}$$

   as claimed. Since

$$\limsup_{n\to\infty} s_n = \inf_{k\geq 1} \sup_{n\geq k} s_n = L$$

   we obtain that for every $\varepsilon > 0$, there is $K_\varepsilon \in \mathbb{N}$ so that $k \geq K_\varepsilon$ implies

$$L \leq \sup_{n\geq k} s_n < L + \varepsilon$$

   Taking $\varepsilon = \frac{x-L}{2}$, we obtain $\sup_{n\geq k} < L + \varepsilon = \frac{x+L}{2}$. By picking $j_0 \geq \widetilde{K}$ so that $n_{j_0} \geq K_\varepsilon$, we obtain

$$s_{n_{j_0}} \leq \sup_{n\geq K_\varepsilon} < \frac{x+L}{2} < s_{n_{j_0}}$$

   which is a contradiction ($\lightning$). Thus (a) must be true.
   Proof of (b). We now construct $(s_{n_j})_{j\in\mathbb{N}}$ such that

$$\lim_{j\to\infty} s_{n_j} = L$$

Recall if

$$\limsup_{n\to\infty} s_n = \inf_{k\geq 1} \sup_{n\geq k} s_n = L$$

then for any $\varepsilon > 0$, there is $K_\varepsilon \in \mathbb{N}$ so that $k \geq K_\varepsilon$ implies that

$$L \leq \sup_{n\geq k} s_n < L + \varepsilon \qquad (*)$$

Let $\varepsilon = 1$ and consider $K_1 \in \mathbb{N}$ so that $(*)$ holds. Then, by the epsilon criterion for supremum, there is $n_1 \in N$ so that

$$L - 1 \leq \sup_{n\geq K_1} s_n - 1 < s_{n_1} \leq \sup_{n\geq K_1} s_n < L + 1$$

note that we can invoke the epsilon criterion (Lemma 1.3.8 in Abbott) because $\limsup\limits_{n\to\infty} s_n = L$ and we have $\sup E \leq L$ by (a). This shows that $\sup_{n\geq K_1} s_n$ is finite.
Suppose that inductively we have constructed a sequence $n_1 < \ldots < n_j$ so that

$$L - \frac{1}{j} \leq s_{n_i} \leq L + \frac{1}{j}, \quad 1 \leq i \leq j$$

We now construct n_{j+1}. Put $\varepsilon = \frac{1}{j+1}$ in $(*)$ which yields a corresponding $K_{\frac{1}{j+1}} \in \mathbb{N}$. Let $\widetilde{K}_j = \max\{n_j, K_{\frac{1}{j+1}}\} + 1$. Using $(*)$ we get

$$L \leq \sup_{n\geq \widetilde{K}_j} s_n < L + \frac{1}{j+1}$$

and we can find $n_{j+1} > \widetilde{K}_j > n_j$ such that

$$L - \frac{1}{j+1} \leq \sup_{n\geq \widetilde{K}_j} s_n - \frac{1}{j+1} < s_{n_{j+1}} \leq \sup_{n\geq \widetilde{K}_j} s_n < L + \frac{1}{j+1}$$

and we are done since $\lim\limits_{j\to\infty} s_{n_j} = L$

□

Proposition. The sequence $(s_n)_{n\in\mathbb{N}}$ is convergent and has a limit $L \in \overline{R}$ if and only if

$$\liminf_{n\to\infty} s_n = \limsup_{n\to\infty} s_n$$

Proof. If $\lim\limits_{n\to\infty} s_n = L$ then $E = \{L\}$. This follows from the fact that if a sequence converges, then its limit is unique. Thus

$$s^* = s_* = L$$

and by the previous theorem $\limsup_{n\to\infty} s_n = \liminf_{n\to\infty} s_n = L$.
Conversely, if $\liminf\limits_{n\to\infty} s_n = \limsup\limits_{n\to\infty} s_n = L$ then

$$\alpha_k = \inf_{n\geq k} s_n \leq s_k \leq \sup_{n\geq k} s_n = \beta_k$$

and $\lim\limits_{k\to\infty} \alpha_k = \lim\limits_{k\to\infty} \beta_k = L$, thus by the squeeze theorem we see that $\lim_{k\to\infty} s_k = L$.

□

2.16 Euler's Number e

Theorem 1. $e = \lim_{n\to\infty} \left(1+\frac{1}{n}\right)^n$ exists.

Proof. The sequence $a_n = (1+\frac{1}{n})^n$ is bounded and increasing. This will ensure, by the monotone convergence theorem, that a_n has a limit.
We will appeal to the AM-GM inequality to show that this sequence is increasing.

$$\begin{aligned} a_{n+1} &= \left(1+\frac{1}{n+1}\right)^{n+1} \\ &= \left(\frac{n+2}{n+1}\right)^{n+1} \\ &= \left(\frac{n(1+\frac{1}{n})+1}{n+1}\right)^{n+1} \\ &\geq 1\cdot\left(1+\frac{1}{n}\right)^n = a_n \end{aligned}$$

Explicitly, we have n terms as $(1+\frac{1}{n})$ and 1 for $n+1$ total terms, and the AM-GM inequality we are using above is

$$A_{n+1} = \frac{n(1+\frac{1}{n})+1}{n+1} \geq \left(1\cdot\left(1+\frac{1}{n}\right)^n\right)^{\frac{1}{n+1}} = G_{n+1}$$

This shows that a_n is an increasing sequence.
Consider $b_n = (1+\frac{1}{n})^{n+1} = a_n\cdot(1+\frac{1}{n})$, then

$$a_n = \left(1+\frac{1}{n}\right)^n \leq a_n\cdot\left(1+\frac{1}{n}\right) = b_n = \left(1+\frac{1}{n}\right)^{n+1}$$

We will use the fact GM-HM inequality to show that the sequence is bounded. The inequality is given by

$$H_{n+1} = \left(\frac{n+1}{\frac{1}{x_1}+\ldots+\frac{1}{x_n}}\right) \leq (x_1\cdots x_{n+1})^{\frac{1}{n+1}} = G_{n+1}$$

where $x_{n+1} = 1$ and $x_1 = \cdots = x_n = (1-\frac{1}{n})^{-1}$. Observe that

$$\begin{aligned} b_n &= \left(1+\frac{1}{n}\right)^{n+1} = \left(\frac{n+1}{n}\right)^{n+1} \\ &= \left(\frac{n+1}{(1-\frac{1}{n})n+1}\right)^{n+1} \\ &\leq \left(1-\frac{1}{n}\right)^{-n} = \left(\frac{n-1}{n}\right)^{n-1} = \left(\frac{n}{n-1}\right)^{n} \\ &= \left(1+\frac{1}{n-1}\right)^n = b_{n-1} \end{aligned}$$

This shows us that $(a_n)_{n\in\mathbb{N}}$ is increasing while $(b_n)_{n\in\mathbb{N}}$ is decreasing. We have

$$a_n \leq b_n \leq b_1 = 4$$

Therefore, $(a_n)_{n\in\mathbb{N}}$ is bounded and $\lim_{n\to\infty} a_n$ exists. Additionally, we have

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n+1}$$
$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = \lim_{n\to\infty}\frac{1}{(1+\frac{1}{n-1})^n} = \frac{1}{e}$$

□

Consider the series $\sum_{n=0}^{\infty}\frac{1}{n!}$, where $n! = 1\cdots n$ and $0! = 1$. Notice that

$$s_n = \sum_{k=0}^{n}\frac{1}{k!} = 1 + 1 + \frac{1}{1\cdot 2} + \frac{1}{1\cdot 2\cdot 3} + \ldots + \frac{1}{n!}$$
$$< 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \ldots + \frac{1}{2^{n-1}} < 3$$

Theorem 2.

$$\sum_{n=0}^{\infty}\frac{1}{n!} = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e$$

Proof. Let

$$s_n = \sum_{k=0}^{n}\frac{1}{k!} \quad \text{and} \quad t_n = \left(1+\frac{1}{n}\right)^n$$

By the binomial theorem

$$\left(1+\frac{1}{n}\right)^n = \sum_{m=0}^{n}\binom{n}{m}\left(\frac{1}{n}\right)^m \cdot 1^{n-m}$$
$$= \sum_{m=0}^{n}\frac{n!}{m!(n-m)!}\left(\frac{1}{n}\right)^m$$
$$= \sum_{m=0}^{n}\frac{1}{m!}\frac{n(n-1)\cdots(n-m+1)}{n^m}$$
$$= \sum_{m=0}^{n}\frac{1}{m!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{n-1}{n}\right)$$

Hence, we see that

$$t_n = 1 + 1 + \frac{1}{2!}\left(1-\frac{1}{n}\right) + \ldots + \frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{n-1}{n}\right)$$
$$\leq 1 + 1 + \frac{1}{2!} + \ldots + \frac{1}{n!} = s_n$$

Since each term past the second term in t_n is being multiplied by some number less than one, we have $t_n \leq s_n$ and by Theorem 3.19 in Rudin, we see that

$$\limsup_{n\to\infty} t_n \leq \limsup_{n\to\infty} s_n = \sum_{n=0}^{\infty} \frac{1}{n!}$$

We know that $\limsup_{n\to\infty} s_n = \sum_{n=0}^{\infty} \frac{1}{n!} = e$ because $(s_n)_{n\in\mathbb{N}}$ converges (refer to the Proposition on Page 45).

Next, if $n \geq m$ then

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

Keeping m fixed and letting $n \to \infty$, we get

$$\liminf_{n\to\infty} t_n \geq \sum_{k=0}^{m} \frac{1}{k!}$$

Letting $m \to \infty$, we get

$$\liminf_{n\to\infty} t_n \geq \sum_{n=0}^{\infty} \frac{1}{n!}$$

Hence

$$\liminf_{n\to\infty} t_n \geq \sum_{n=0}^{\infty} \frac{1}{n!} \geq \limsup_{n\to\infty} t_n \geq \liminf_{n\to\infty} t_n$$

So we conclude

$$e = \lim_{n\to\infty} t_n = \sum_{n=0}^{\infty} \frac{1}{n!}$$

□

Corollary.

$$0 < e - s_n < \frac{1}{n \cdot n!} \quad \text{and} \quad e > \frac{1}{n!} \text{ for all } n \in \mathbb{N}$$

Proof.

$$\begin{aligned}
e - s_n &= \sum_{k=n+1}^{\infty} \frac{1}{k!} \\
&= \frac{1}{(n+1)!}\left(1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \ldots\right) \\
&< \frac{1}{(n+1)!}\left(1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \ldots\right) \\
&= \frac{1}{(n+1)!} \cdot \frac{1}{1 - \frac{1}{n+1}} = \frac{1}{(n+1)!} \cdot \frac{n+1}{n} \\
&= \frac{1}{n!} \cdot \frac{1}{n}
\end{aligned}$$

Also,

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} > \frac{1}{n!} \quad \text{for all } n \in \mathbb{N}$$

□

Theorem. e is irrational.

Proof. Suppose $e = \frac{p}{q}$ for some $p, q \in \mathbb{N}$.

Since $0 < e - s_n < \frac{1}{n! \cdot n}$, we have

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

with $q!e \in \mathbb{N}$. Since

$$q! s_q = q! \left(1 + 1 + \frac{1}{2!} + \ldots + \frac{1}{q!}\right) \in \mathbb{N}$$

we see that $q!(e - s_q) \in \mathbb{N}$, thus

$$0 < q!(e - s_q) < \frac{1}{q} \leq 1 \quad \text{if } q \in \mathbb{N}$$

This is a contradiction because no integer exists between 0 and 1 (↯).

2.17 The Root and Ratio Test

Theorem (Rudin 3.33) (Root Test). Given $\sum_{n \in \mathbb{N}} a_n$ set

$$\alpha = \limsup_{n \to \infty} \sqrt[n]{|a_n|}$$

Then

(a) If $\alpha < 1$, then $\sum_{n \in \mathbb{N}} a_n$ converges.

(b) If $\alpha > 1$, then $\sum_{n \in \mathbb{N}} a_n$ diverges.

(c) If $\alpha = 1$, the test is inconclusive.

We use $\limsup$ rather than $\lim$ to ensure that α always exists.

Proof. If $\alpha < 1$ we can choose β so that $\alpha < \beta < 1$ and an integer N so that

$$\sqrt[n]{|a_n|} < \beta \quad \text{for } n \geq N$$

since

$$\alpha = \inf_{k \geq 1} \sup_{n \geq k} \sqrt[n]{|a_n|} < \beta$$

Essentially, we can find β by the density of rationals in $\mathbb{R}$. Then, by the definition of $\limsup$, there must be N satisfying the condition above. This choice is also justified by Theorem 3.17(b) in Rudin.

Hence, for $n \geq N$ we have $|a_n| < \beta^n$. But $\beta < 1$, thus $\sum_{n\geq 0} \beta^n$ converges and the comparison test implies that $\sum_{n\in\mathbb{N}} a_n$ converges as well.
If $\alpha > 1$ then by the Theorem (3) on Pg. 40 (also Rudin 3.17) there is a sequence $(n_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ such that

$$\sqrt[n_k]{|a_{n_k}|} \xrightarrow[k\to\infty]{} \alpha$$

Hence $|a_n| > 1$ for infinitely many values of $n \in \mathbb{N}$ so that the condition $a_n \xrightarrow[n\to\infty]{} 0$, which is necessary for the convergence of $\sum_{n\in\mathbb{N}} a_n$ does not hold (Rudin 3.23).
To prove (c), observe the following counterexample:

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty \quad \text{and} \quad \sqrt[n]{n} \xrightarrow[n\to\infty]{} 1$$
$$\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty \quad \text{and} \quad \sqrt[n]{n^2} \xrightarrow[n\to\infty]{} 1$$

Theorem. (Rudin 3.34) (Ratio Test). The series $\sum_{n\in\mathbb{N}} a_n$

(a) converges if $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$

(b) diverges if $\left|\frac{a_{n+1}}{a_n}\right| \geq 1$ for all $n \geq n_0$ where n_0 is fixed

Proof. If (a) holds we can find $\beta < 1$ and $N \in \mathbb{N}$ such that

$$\left|\frac{a_{n+1}}{a_n}\right| < \beta \quad \text{for all } n \geq N$$

In particular, for $p \geq 1$ and $p \in \mathbb{Z}$

$$\begin{aligned} |a_{n+p}| &= |a_{n+p-1}| \cdot \left|\frac{a_{n+p}}{a_{n+p-1}}\right| < \beta \cdot |a_{n+p-1}| \\ &< \beta^2 |a_{n+p-2}| < \ldots < \beta^p |a_n| \quad \text{for } n \geq N \end{aligned}$$

Thus $|a_{N+p}| < \beta^p |a_N|$ and consequently

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

for all $n \geq N$. The above expression is obtained by substitution with $N + p = n \implies p = n - N$. The claim follows from the comparison test since

$$\sum_{n\geq N} \beta^n < \infty, \quad \beta < 1$$

(b) If $|a_{n+1}| \geq |a_n|$ for $n \geq n_0$, then $a_n \xrightarrow[n\to\infty]{} 0$ does not hold.

□

Remark. As before, $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 1$ is useless. Consider

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty \quad \text{and} \quad \frac{a_{n+1}}{a_n} = \frac{n}{n+1} \xrightarrow[n\to\infty]{} 1$$
$$\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty \quad \text{and} \quad \frac{a_{n+1}}{a_n} = \left(\frac{n}{n+1}\right)^2 \xrightarrow[n\to\infty]{} 1$$

Example.

$$\sum_{n=1}^{\infty} \frac{n!}{n^n} < \infty$$
$$\begin{aligned}\frac{a_{n+1}}{a_n} &= \frac{(n+1)!}{(n+1)^{n+1}} \cdot \frac{n^n}{n!} = \frac{(n+1)n^n}{(n+1)^{n+1}} \\ &= \left(\frac{n}{n+1}\right)^n \\ &= \left(\frac{1}{1+\frac{1}{n}}\right)^n \xrightarrow[n\to\infty]{} \frac{1}{e}\end{aligned}$$

Since $\frac{1}{e} < 1$, the series converges by the ratio test.

2.18 Power Series

Definition (Power series). Given a sequence $(c_n)_{n\in\mathbb{N}} \subseteq \mathbb{C}$, the series

$$\sum_{n=0}^{\infty} c_n z^n, \qquad z \in \mathbb{C}$$

is called a power series. The numbers c_n are called the coefficients of the series.
Theorem 1 (Rudin 3.39). Given the power series $\sum_{n=0}^{\infty} c_n z^n$, set

$$\alpha = \limsup_{n\to\infty} \sqrt[n]{|c_n|} \quad \text{and} \quad R = \frac{1}{\alpha}$$

(If $\alpha = 0$, then $R = +\infty$, if $\alpha = +\infty$, then $R = 0$). Then $\sum_{n=0}^{\infty} c_n z^n$ converges if $|z| < R$ and diverges if $|z| > R$. R is called the radius of convergence of $\sum_{n=0}^{\infty} c_n z^n$.
Proof. Consider $a_n = c_n z^n$ and apply the root test

$$\limsup_{n\to\infty} \sqrt[n]{|a_n|} = |z| \cdot \limsup_{n\to\infty} \sqrt[n]{|c_n|} = \frac{|z|}{R}$$

Examples.

(a) The series $\sum_{n=0}^{\infty} n^n z^n$ has $R = 0$.

(b) $\sum_{n=0}^{\infty} \frac{z^n}{n!}$ has $R = +\infty$.

Fact. For a sequence $(c_n)_{n\to\infty}$ of positive numbers

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

We shall prove $\limsup_{n\to\infty}\sqrt[n]{c_n}\le\limsup_{n\to\infty}\frac{c_{n+1}}{c_n}=\alpha$.
If $\alpha=+\infty$ then there is nothing to do. Assume $\alpha<\infty$ and choose $\beta>\alpha$

$$\limsup_{n\to\infty}\frac{c_{n+1}}{c_n}=\lim_{k\to\infty}\sup_{n\ge k}\frac{c_{n+1}}{c_n}=\alpha$$

Thus there is $N\in\mathbb{N}$ such that $n\ge N$ implies

$$\frac{c_{n+1}}{c_n}\le\beta$$

For any $p>0$, we have $c_{N+k+1}\le\beta c_{N+k}$ for $k=0,1,\dots,p-1$. Thus

$$c_{N+p}\le\beta c_{N+p-1}\le\beta^2c_{N+p-2}\le\dots\le\beta^pc_N$$

So

$$c_n\le c_N\beta^{-N}\beta^n\quad\text{for any }n\ge N$$

Hence

$$\sqrt[n]{c_n}\le\sqrt[n]{c_N\beta^{-N}}\cdot\beta$$

Thus

$$\limsup_{n\to\infty}\sqrt[n]{c_n}\le\beta$$

when we take the lim sup on both sides and consider Rudin 3.20(b) or Theorem 9(b) in these notes. But $\beta>\alpha$ and the choice of β was arbitrary. Thus

$$\limsup_{n\to\infty}\sqrt[n]{c_n}\le\alpha$$

Hence if $\sum_{n=0}^\infty\frac{z^n}{n!}$, then $R=+\infty$ since

$$\alpha=\limsup_{n\to\infty}\sqrt[n]{\frac{1}{n!}}=0$$

$$\limsup_{n\to\infty}\sqrt[n]{\frac{1}{n!}}\le\limsup_{n\to\infty}\frac{n!}{(n+1)!}=\limsup_{n\to\infty}\frac{1}{n+1}=0$$

(c) $\sum_{n=0}^\infty z^n$ has $R=1$. If $|z|=1$, the series diverges since z^n does not tend to 0 as $n\to\infty$.
(d) $\sum_{n=1}^\infty\frac{z^n}{n}$ has $R=1$. If $z=1$, the series diverges since $\sum_{n=1}^\infty\frac{1}{n}=\infty$.
$\sum_{n=1}^\infty\frac{z^n}{n}$ also converges for all $z\in\mathbb{C}$ such that $|z|=1$ and $z\ne1$. More generally:
Theorem (Rudin 3.44). Suppose the radius of convergence of $\sum_{n=0}^\infty c_nz^n$ is 1 and suppose $c_0\ge c_1\ge c_2\ge\dots$ with $\lim_{n\to\infty}c_n=0$. Then $\sum_{n=0}^\infty c_nz^n$ converges at every point on the circle $|z|=1$ except possibly at $z=1$.

(a) The partial sums A_n of $\sum_{n=0}^{\infty} a_n$ form a bounded sequence

$$|A_n| = \left|\sum_{m=0}^{n} z^m\right| = \left|\frac{1-z^{m+1}}{1-z}\right| \leq \frac{1+|z|^{m+1}}{|1-z|} = \frac{2}{|1-z|}$$

if $|z| = 1$ and $z \neq 1$.

(b) $b_n = c_n$, is fine since $b_0 \geq b_1 \geq b_2 \geq \dots$

(c) $\lim_{n\to\infty} b_n = 0$

This shows us $\sum_{n=0}^{\infty} a_n b_n$ converges by Theorem 17 (Rudin 3.41).

□

(e) $\sum_{n=1}^{\infty} \frac{z^n}{n^2}$ has $R = 1$. Since $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$, we see that $\sum_{n=1}^{\infty} \frac{z^n}{n^2}$ converges for all z satisfying $|z| = 1$.

2.19 Addition and Multiplication of Series

Theorem 2 (Rudin 3.47). If $\sum_{n\in\mathbb{N}} a_n = A$ and $\sum_{n\in\mathbb{N}} b_n = B$ then

$$\sum_{n\in\mathbb{N}}(a_n + b_n) = A + B \quad \text{and} \quad \sum_{n\in\mathbb{N}} ca_n = cA$$

Proof. Let $A_n = \sum_{k=0}^{n} a_k$ and $B_n = \sum_{k=0}^{n} b_k$. Then by the algebraic limit theorem (Pg. 10) we have

$$A_n + B_n = \sum_{k=0}^{n}(a_k + b_k) \xrightarrow[n\to\infty]{} A + B$$

Definition (Product of two series). Given $\sum_{n\in\mathbb{N}} a_n$ and $\sum_{n\in\mathbb{N}} b_n$, we put

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} \qquad (n = 0, 1, 2, \dots)$$

and call $\sum_{n\in\mathbb{N}} c_n$ the product of two given series.
Motivation.
Suppose we have two power series $\sum a_n z^n$ and $\sum b_n z^n$, observe that

$$\begin{aligned}\left(\sum_{n=0}^{\infty} a_n z^n\right)\left(\sum_{n=0}^{\infty} b_n z^n\right) &= (a_0 + a_1 z + a_2 z^2 + \dots)(b_0 + b_1 z + b_2 z^2 + \dots) \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0) z + (a_0 b_2 + a_1 b_1 + a_2 b_0) z^2 + \dots \\ &= c_0 + c_1 z + c_2 z^2 + \dots + c_n z_n + \dots\end{aligned}$$

Setting $z = 1$, we arrive at the definition of c_n. It is **very important** to note that the sums must be infinite in order for the definition of c_n to make sense.
Example. The series

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{\sqrt{n+1}} = 1 - \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{4}} + \dots$$

converges.
But the product of this series with itself diverges. Consider

$$\left(\sum_{n=0}^{\infty} \frac{(-1)^n}{\sqrt{n+1}}\right)^2 = \sum_{n=0}^{\infty} c_n$$
$$= 1 - \left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}\right) + \left(\frac{1}{\sqrt{3}} + \frac{1}{\sqrt{2}\sqrt{2}} + \frac{1}{\sqrt{3}}\right) - \left(\frac{1}{\sqrt{4}} + \frac{1}{\sqrt{3}\sqrt{2}} + \frac{1}{\sqrt{2}\sqrt{3}} + \frac{1}{\sqrt{4}}\right) + \dots$$

See Rudin 3.49 for the rest of the details of this proof.

Theorem 3 (Rudin 3.50). Suppose

(a) $\sum_{n=0}^{\infty} a_n$ converges absolutely, i.e., $\sum_{n=0}^{\infty} |a_n| < \infty$

(b) $\sum_{n=0}^{\infty} a_n = A$

(c) $\sum_{n=0}^{\infty} b_n = B$

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

Then we have

$$\sum_{n=0}^{\infty} c_n = AB$$

Proof. Put

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

Then it follows that

$$\begin{aligned} C_n &= a_0 b_0 + (a_0 b_1 + a_1 b_0) + \dots + (a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0) \\ &= a_0 B_n + a_1 B_{n-1} + \dots + a_n B_0 \\ &= a_0 (B + \beta_n) + a_1 (B + \beta_{n-1}) + \dots + a_n (B + \beta_0) \\ &= A_n B + \underbrace{a_0 \beta_n + a_1 \beta_{n-1} + \dots + a_n \beta_0}_{\gamma_n} \end{aligned}$$

We will show that $c_n \to AB$. Since $A_n B \to AB$, it suffices to prove $\gamma_n \xrightarrow[n\to\infty]{} 0$. Set $\alpha = \sum_{n=0}^{\infty} |a_n|$. Let $\varepsilon > 0$ be given. By (c), we have $\beta_n \xrightarrow[n\to\infty]{} 0$. We find $N \in \mathbb{N}$ such that $|\beta_n| \leq \varepsilon$ for $n \geq N$, then

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

Keeping N fixed, and letting $n \to \infty$, we obtain

$$\limsup_{n\to\infty} |\gamma_n| \leq \varepsilon \alpha$$

since $a_k \xrightarrow[k\to\infty]{} 0$. More explicitly, looking at the N summands, we have that each $a_{n-k} \to 0$ for $k \in \{0, 1, 2, \ldots, N\}$. Since ε is arbitrary, we have $\limsup_{n\to\infty} |\gamma_n| \leq 0$. Since $|\gamma_n| \geq 0$, we always have $\limsup_{n\to\infty} |\gamma_n| \geq 0$. Thus $\limsup_{n\to\infty} |\gamma_n| = 0$. But also

$$0 \leq \liminf_{n\to\infty} |\gamma_n| \leq \limsup_{n\to\infty} |\gamma_n| = 0$$

Therefore, $\lim\limits_{n\to\infty} |\gamma_n| = 0$ which implies $\gamma_n \xrightarrow[n\to\infty]{} 0$, as desired.

□

Remark. If $\sum_n a_n = A$, $\sum_n b_n = B$, and $\sum_n c_n = C$ with $c_n = a_0 b_n + \ldots + a_n b_0$, then $C = AB$.
Proof. Later!

2.20 Rearrangements

Definition (Rearrangement). Let $(k_n)_{n\in\mathbb{N}}$ be a sequence in which every positive integer appears once and only once.
Setting $a'_n = a_{k_n}$ we say that

$$\sum_{n\in\mathbb{N}} a'_n \text{ is a rearrangement of } \sum_{n\in\mathbb{N}} a_n$$

Example. Consider the convergent series

$$s = 1 - \frac{1}{2} + \frac{1}{3} \underbrace{- \frac{1}{4} + \frac{1}{5}}_{<0} \underbrace{- \frac{1}{6} + \frac{1}{7}}_{<0} - \ldots$$

and one of its rearrangements

$$s' = 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} - \frac{1}{4} + \frac{1}{5} + \frac{1}{7} - \frac{1}{6} + \ldots + \frac{1}{4k-3} + \frac{1}{4k-1} - \frac{1}{2k} + \ldots$$

Observe that $s < 1 - \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$ and

$$\frac{1}{4k-3} + \frac{1}{4k-1} - \frac{1}{2k} > 0 \quad \text{for } k \geq 1$$

If s'_n is the partial sum of s', then $s'_3 < s'_6 < s'_9 < \ldots$. Hence

$$\limsup_{n\to\infty} s'_n > s'_3 = \frac{5}{6}$$

Thus, s' does not converge to s.
Theorem 1 (Rudin 3.54). Let $\sum_{n\in\mathbb{N}} a_n$ be a series of real numbers which converges, but not absolutely. Suppose

$$-\infty \leq \alpha \leq \beta \leq \infty$$

Then there exists a rearrangement $\sum_{n\in\mathbb{N}} a_n'$ with partial sums s_n' such that

$$\liminf_{n\to\infty} s_n' = \alpha \qquad \limsup_{n\to\infty} s_n' = \beta$$

Proof. For $n \in \mathbb{N}$, let

$$p_n = \frac{|a_n| + a_n}{2} \geq 0, \qquad q_n = \frac{|a_n| - a_n}{2} \geq 0$$

Then we have $p_n - q_n = a_n$ and $p_n + q_n = |a_n|$. The series $\sum_{n\in\mathbb{N}} p_n$ and $\sum_{n\in\mathbb{N}} q_n$ both diverge. In fact, if both were convergent, then

$$\sum_{n\in\mathbb{N}} (p_n + q_n) = \sum_{n\in\mathbb{N}} |a_n| = \infty$$

would converge, contradicting the hypothesis ($\lightning$).
Since

$$\sum_{n=1}^{N} a_n = \sum_{n=1}^{N} (p_n - q_n) = \sum_{n=1}^{N} p_n - \sum_{n=1}^{N} q_n$$

divergence of $\sum p_n$ and convergence of $\sum q_n$ (or vice versa) implies divergence of $\sum a_n$ ($\lightning$).
Let $P_1, P_2, P_3, \ldots$ denote the nonnegative terms of $\sum_{n\in\mathbb{N}} a_n$ in the order in which they occur and let $Q_1, Q_2, Q_3, \ldots$ be the absolute values of the negative terms of $\sum_{n\in\mathbb{N}} a_n$ also in their original order. (I think he means the terms of $(a_n)_{n\in\mathbb{N}}$).
The series $\sum_{n\in\mathbb{N}} P_n$, $\sum_{n\in\mathbb{N}} Q_n$ differ from $\sum_{n\in\mathbb{N}} p_n$, $\sum_{n\in\mathbb{N}} q_n$ only by zero terms and therefore they also diverge.
We shall construct $(m_n)_{n\in\mathbb{N}}$ and $(k_n)_{n\in\mathbb{N}}$ such that the series

$$P_1 + \ldots + P_{m_1} - Q_1 - \ldots - Q_{k_1} + P_{m_1+1} + \ldots + P_{m_2} - Q_{k_1+1} - \ldots - Q_{k_2} + \ldots \qquad (*)$$

which is a rearrangement of $\sum_{n\in\mathbb{N}} a_n$, satisfies

$$\liminf_{n\to\infty} s_n' = \alpha \quad \text{and} \quad \limsup_{n\to\infty} s_n' = \beta$$

Choose real-valued sequences $(\alpha_n)_{n\in\mathbb{N}}$, $(\beta_n)_{n\in\mathbb{N}}$ such that

$$\alpha_n \xrightarrow[n\to\infty]{} \alpha \qquad \beta_n \xrightarrow[n\to\infty]{} \beta$$

with $\alpha_n < \beta_n$ and $\beta_1 > 0$.
Let m_1, k_1 be the smallest integers such that

$$P_1 + \ldots + P_{m_1} > \beta_1,$$
$$P_1 + \ldots + P_{m_1} - Q_1 - \ldots - Q_{k_1} < \alpha_1$$

Let m_2, k_2 be the smallest integers such that

$$P_1 + \ldots + P_{m_1} - Q_1 - \ldots - Q_{k_1} + P_{m_1+1} + \ldots + P_{m_2} > \beta_2,$$
$$P_1 + \ldots + P_{m_1} - Q_1 - \ldots - Q_{k_1} + P_{m_1+1} + \ldots + P_{m_2} - Q_{k_1+1} - \ldots - Q_{k_2} < \alpha_2$$

and continue in this way. This is possible since

$$\left|\sum_{n\in\mathbb{N}} P_n\right| = \infty \quad \text{and} \quad \left|\sum_{n\in\mathbb{N}} Q_n\right| = \infty$$

If x_n, y_n are partial sums of $(*)$ whose last terms are $P_{m_n}, -Q_{k_n}$, then

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

The above inequalities follow from the fact that

$$\beta_n < x_n \leq (x_n - P_{m_n}) + P_{m_n} \leq \beta_n + P_{m_n}$$
$$0 < x_n - \beta_n < P_{m_n}$$

Since $\sum_{n\in\mathbb{N}} a_n$ converges, we know that $a_n \xrightarrow[n\to\infty]{} 0$. Since P_n, Q_n are both subsequences of $(a_n)_{n\in\mathbb{N}}$, we have $P_n \xrightarrow[n\to\infty]{} 0$ and $Q_n \xrightarrow[n\to\infty]{} 0$. Thus, $x_n \xrightarrow[n\to\infty]{} \beta$ and $y_n \xrightarrow[n\to\infty]{} \alpha$.
Finally, it is clear that no number less than α or greater than β can be a subsequential limit of the partial sums of $(*)$.

□

Theorem 2 (Rudin 3.55). If $\sum_{n\in\mathbb{N}} a_n$ is a series of complex numbers which converges absolutely, then every rearrangement of $\sum_{n\in\mathbb{N}} a_n$ converges and all converges to the same sum.

Proof. Let $\sum_{n\in\mathbb{N}} a_n$ be a rearrangement with partial sums s'_n. Given $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that $m \geq n \geq N$ implies

$$\sum_{k=n}^{m} |a_k| \leq \varepsilon$$

Now choose p such that $\{1, 2, \ldots, N\} \subseteq \{k_1, k_2, \ldots, k_p\}$. If $n > p$, then the numbers $a_1, \ldots a_N$ will cancel in the difference $s_n - s'_n$ so that $|s_n - s'_n| \leq \varepsilon$. Hence, $(s'_n)_{n\in\mathbb{N}}$ converges to the same sum as $(s_n)_{n\in\mathbb{N}}$.

3 Continuity

3.1 Functional Limits

Definition 1 (Functional limit). Let $(X, \rho_X), (Y, \rho_Y)$ be metric spaces. Suppose $E \subseteq X$ and $f : E \to Y$ and p is a limit point of E. We write $f(x) \xrightarrow[x \to p]{} q$ or

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

if there is a point $q \in Y$ with the following property: For every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\rho_Y(f(x), q) < \varepsilon$$

for all points $x \in E$ for which $0 < \rho_X(x, p) < \delta$. We require p to be a limit point of E because if p is not a limit point, then q is not unique (vacuously true for all q as long as δ is less than the interval that exists around the isolated point p).
Remark. If $X = Y = \mathbb{R}$ or $X = Y = \mathbb{R}^n$, then $\rho_X(x, y) = \rho_Y(x, y) = |x - y|$.
For every $\varepsilon > 0$ there is $\delta > 0$ such that for all $x \in E$ if $0 < |x - y| < \delta$, then $|f(x) - q| < \varepsilon$.
Theorem 1 (Rudin 4.2). Let $(X, \rho_X), (Y, \rho_Y), E, f$ and p be as in Definition 1. Then

(A) $\lim_{x \to p} f(x) = q$ if and only if

(B) $\lim_{n \to \infty} f(p_n) = q$ for every sequence $(p_n)_{n \in \mathbb{N}}$ in E satisfying $p_n \neq p$ and $\lim_{n \to \infty} p_n = p$.

Proof. Suppose that (A) holds. Choose $(p_n)_{n \in \mathbb{N}} \subseteq E$ satisfying $p_n \neq p$ and $p_n \xrightarrow[n \to \infty]{} p$. Let $\varepsilon > 0$ be given, then there exists $\delta > 0$ such that

$$\rho_Y(f(x), q) < \varepsilon$$

if $x \in E$ and $0 < \rho_X(x, p) < \delta$. Also, there exists $N \in \mathbb{N}$ such that $n \geq N$ implies $0 < \rho_X(p_n, p) < \delta$. Thus, we have

$$\rho_Y(f(p_n), q) < \varepsilon$$

Hence, (B) holds.
Conversely, suppose (A) is false. Then there exists some $\varepsilon > 0$ such that for every $\delta > 0$ there exists a point $x \in E$ (depending on δ) for which

$$\rho_Y(f(x), q) \geq \varepsilon \quad \text{but} \quad 0 < \rho_X(x, p) < \delta$$

Taking $\delta_n = \frac{1}{n}, n \in \mathbb{Z}_+$, we thus find a sequence $(p_n)_{n \in \mathbb{N}}$ in E satisfying

$$\lim_{n \to \infty} p_n = p \quad \text{but} \quad \rho_Y(f(p_n), q) \geq \varepsilon$$

Thus (B) is false as desired. Note that we set $\delta_n = \frac{1}{n}$ at each step and use the fact that (A) is false to find an x which we designate as p_n in order to generate this sequence.

Theorem 2 (Rudin 4.4). Suppose $E \subseteq X$, (X, ρ_X) is a metric space, p is a limit point of E, $f, g : E \to \mathbb{C}$ are functions such that

$$\lim_{x \to p} f(x) = A \quad \lim_{x \to p} g(x) = B$$

Then

(a) $\lim_{x \to p} (f + g)(x) = A + B$

(b) $\lim_{x \to p} (fg)(x) = AB$

(c) $\lim_{x \to p} \left(\frac{f}{g}\right)(x) = \frac{A}{B}$ if $B \neq 0$ and $g(x) \neq 0$ for all $x \in E$.

The proof is omitted in light of its similarity to the proofs for sequences.

3.2 Continuous Functions

Definition 2 (Continuity). Suppose (X, ρ_X), (Y, ρ_Y) are metric spaces, $E \subseteq X, p \in E$ and $f : E \to Y$. Then f is said to be continuous at p if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\rho_Y(f(x), f(p)) < \varepsilon$$

for all points for which $\rho_X(x, p) < \delta$.

If f is continuous at every point of E, then f is said to be continuous on E.

Remark. If p is an isolated point of E, then our definition implies that every function f which has E as its domain of definition is continuous at p. For, no matter which $\varepsilon > 0$ we choose, we can pick $\delta > 0$ so that the only point $x \in E$ for which $\rho_X(x, p) < \delta$ is $x = p$, then

$$\rho_Y(f(x), f(p)) = 0 < \varepsilon$$

Fact 1. In the situation of Definition 2, assume also that p is a limit point of E. Then f is continuous at p if and only if $\lim_{x \to p} f(x) = f(p)$.

Proof. It is obvious when we compare Definition 1 and Definition 2.

Theorem 3 (Rudin 4.7). Suppose that $(X, \rho_X), (Y, \rho_Y), (Z, \rho_Z)$ are metric spaces, let $E \subseteq X$ and $f : E \to Y$ and $g : f(E) \to Z$ and $h : E \to Z$ be defined by

$$h(x) = g(f(x)), \quad x \in E$$

If f is continuous at a point $p \in E$ and if g is continuous at the point $f(p)$, then h is continuous at p.

Proof. Let $\varepsilon > 0$ be given. Since g is continuous at $f(p)$, there is $\eta > 0$ such that

$$\rho_Z(g(y), g(f(p))) < \varepsilon, \quad \text{if } \rho_Y(y, f(p)) < \eta \text{ and } y \in f[E]$$

Since f is continuous at p, there is $\delta > 0$ such that

$$\rho_Y(f(x), f(p)) < \eta \quad \text{if } \rho_X(x,p) < \delta \text{ and } x \in E$$

It follows that

$$\rho_Z(h(x), h(p)) = \rho_Z(g(f(x)), g(f(p))) < \varepsilon$$
$$\text{if } \rho_X(x,p) < \delta \text{ and } x \in E$$

Thus, h is continuous at $p \in E$.

Theorem 4 (Rudin 4.8). A mapping f of a metric space (X, ρ_X) into a metric space (Y, ρ_Y) is continuous on X if and only if $f^{-1}[V]$ is open in X for every open set V in Y.

Proof. Suppose f is continuous on X and V is an open set in Y. We have to show that $f^{-1}[V]$ is open in X. Let $p \in f^{-1}[V]$, since V is open, we know that $B_{\rho_Y}(f(p), \varepsilon) \subseteq V$ for some $\varepsilon > 0$. Since f is continuous at p, there is $\delta > 0$ such that $\rho_Y(f(x), f(p)) < \varepsilon$ if $\rho_X(x,p) < \delta$. Thus $B_{\rho_X}(p, \delta) \subseteq f^{-1}[V] = \{x \in X : f(x) \in V\}$. In other words, we know that $B_{\rho_X}(p, \delta) \subseteq f^{-1}[V]$ because the continuity of f guarantees that

$$f(B_{\rho_X}(p, \delta)) \subseteq B_{\rho_Y}(f(p), \varepsilon) \subseteq V$$

Taking the preimage across the equation, we see that

$$B_{\rho_X}(p, \delta) \subseteq f^{-1}[B_{\rho_Y}(f(p), \varepsilon)] \subseteq f^{-1}[V]$$

Conversely, suppose $f^{-1}[V]$ is open in X for any open set $V \subseteq Y$. Fix $p \in X$ and $\varepsilon > 0$ and consider $V = B_{\rho_Y}(f(p), \varepsilon)$ which is open. Thus $f^{-1}[V]$ is open. Hence, by the openness of $f^{-1}[V]$, there exists $\delta > 0$ so that

$$B_{\rho_X}(p, \delta) \subseteq f^{-1}[V]$$

Thus if $\rho_X(x,p) < \delta$, then $x \in f^{-1}[V]$. Hence, it follows that

$$f(x) \in V = B_{\rho_Y}(f(p), \varepsilon) \iff \rho_Y(f(x), f(p)) < \varepsilon$$

So this δ satisfies the condition for the continuity of f at p.

□

Corollary. A mapping $f : X \to Y$ between metric spaces (X, ρ_X) and (Y, ρ_Y) is continuous if and only if $f^{-1}[C]$ is closed in X for any closed set C in Y.

Proof. A set is closed if and only if its complement is open. Since $f^{-1}[E^c] = (f^{-1}[E])^c$ for every $E \subseteq Y$, we are done by invoking Theorem 4. To be more explicit, take E to be any open set and the result follows.

Theorem 5 (Rudin 4.9). Let f and g be complex continuous functions on a metric space (X, ρ_X). Then $f + g$, fg, and $\frac{f}{g}$ are continuous. In the last case we assume $g(x) \neq 0$ for all $x \in X$.

Theorem 6 (Rudin 4.10).

(a) Let $f_1, \dots, f_k$ be real functions on a metric space (X, ρ_X) and let $f : X \to \mathbb{R}^k$ be defined by

$$f(x) = (f_1(x), f_2(x), \dots, f_k(x)) \quad x \in X$$

then f is continuous if and only if each of the functions $f_1, \dots, f_k$ is continuous.

(b) If $f, g : X \to \mathbb{R}^k$ are continuous, then $f + g$ and $\overbrace{f \cdot g}^{\text{scalar product}}$ are continuous on X.

The functions $f_1, \dots, f_k$ are called the components of f. Note that $f + g$ is a mapping into $\mathbb{R}^k$, whereas $f \cdot g$ is a real function on X.

Proof. Part (a) follows from:

$$|f_j(x) - f_j(y)| \leq |f(x) - f(y)| = \left(\sum_{j=1}^{k} |f_j(x) - f_j(y)|^2 \right)^{\frac{1}{2}}$$

for $j = 1, 2, \dots, k$. Part (b) follows from (a) and Theorem 5.

Examples.

(1) Let $\vec{x} = (x_1, \dots x_k) \in \mathbb{R}^k$. The functions $\phi_j(\vec{x}) = x_j$ are continuous for $j = 1, 2, \dots, k$ since

$$|\phi_j(\vec{x}) - \phi_j(\vec{y})| \leq |\vec{x} - \vec{y}|$$

and we can take $\delta = \varepsilon$ with respect to Definition 2. The $\phi_j(\vec{x})$ are sometimes called coordinate functions.

(2)

$$P(\vec{x}) = \sum_{(n_1, \dots, n_k) \in A} c_{n_1 \dots n_k} x_1^{n_1} \cdots x_k^{n_k}$$

where $\mathrm{card}(A) < \infty$ and $A \subseteq \mathbb{N}^k$ and $c_{n_1 \dots n_k} \in \mathbb{C}$.

(3) The mapping $\vec{x} \to |\vec{x}|$ is continuous since

$$||\vec{x}| - |\vec{y}|| \leq |\vec{x} - \vec{y}|$$

To be more explicit, let $\varepsilon > 0$ be arbitrary. Then we can simply take $\varepsilon = \delta$ to see that for all $|\vec{x} - \vec{y}| < \delta$, it follows that $||\vec{x}| - |\vec{y}|| < \delta = \varepsilon$.

(4) If $f, g : X \to \mathbb{R}$ are continuous, then

$$\max(f, g) = \frac{f + g + |f - g|}{2} \qquad \min(f, g) = \frac{f + g - |f - g|}{2}$$

are also continuous functions.

3.3 Continuity and Compactness

Definition (Bounded map). A mapping $f : E \to \mathbb{R}^k$ is said to be bounded if there is a number $M > 0$ such that $|f(x)| \leq M$ for all $x \in E$.

Theorem 7 (Rudin 4.14). Suppose f is a continuous mapping of a compact metric space (X, ρ_X) into a metric space (Y, ρ_Y). Then $f[X]$ is compact in Y.

Proof. Let $(V_\alpha)_{\alpha \in A}$ be an open cover of $f[X]$, so

$$f[X] \subseteq \bigcup_{\alpha \in A} V_\alpha$$

Since f is continuous, then each set $f^{-1}[V_\alpha]$ is open in X. Since X is compact and

$$X \subseteq \bigcup_{\alpha \in A} f^{-1}[V_\alpha]$$

thus, there are finitely many indices $\alpha_1, \ldots, \alpha_n \in A$ such that $X \subseteq \bigcup_{k=1}^n f^{-1}[V_{\alpha_k}]$.
Since $f[f^{-1}[E]] \subseteq E$ for every $E \subseteq Y$, we have

$$f[X] \subseteq f\left[\bigcup_{k=1}^n f^{-1}[V_{\alpha_k}]\right] = \bigcup_{k=1}^n V_{\alpha_k}$$

□

Theorem 8 (Rudin 4.15). If $f : X \to \mathbb{R}^k$ is continuous and (X, ρ_X) is a compact metric space, then $f[X]$ is closed and bounded. Specifically, f is bounded.

Proof. Use Theorem 7 with Theorem 2.41 in Rudin.

Theorem 9 (Rudin 4.16). (Extreme Value Theorem). Suppose $f : X \to \mathbb{R}$ is a continuous function on a compact metric space (X, ρ_X) and

$$M = \sup_{p \in X} f(p) \quad \text{and} \quad m = \inf_{p \in X} f(p)$$

Then there are $p, q \in X$ such that $f(p) = M$ and $f(q) = m$.
In other words, there are $p, q \in X$ such that $f(q) \leq f(x) \leq f(p)$ for all $x \in X$, that is f attains its maximum of p and its minimum of q.

Proof. By Theorem 7, $f[X] \subseteq \mathbb{R}$ is closed and bounded. Thus

$$M = \sup_{x \in X} f(x) \quad \text{and} \quad m = \inf_{x \in X} f(x)$$

are members of $f[X]$ and we are done. This follows from Theorem 2.28 in Rudin.

Theorem 10 (Rudin 4.17). Suppose f is a continuous one-to-one mapping of a compact metric space X onto a metric space Y. Then the inverse mapping f^{-1} defined on Y by

$$f^{-1}(f(x)) = x, \quad \forall x \in X$$

is a continuous mapping of Y onto X.

Proof. $f^{-1}: Y \to X$ is well-defined since $f: X \to Y$ is one-to-one and onto. It suffices to prove that $f[V]$ is open in Y for any open V in X.

Fix an open set $V \subseteq X$. V^c is closed in X and thus compact. Hence $f[V^c]$ is a compact subset of Y and consequently $f[V^c]$ is closed. Since $f: X \to Y$ is bijective, we have

$$(f[V^c])^c = f[V]$$

Thus, $f[V]$ is open as desired. Note, since f is bijective we have $f[V^c] = f[V]^c$.

Definition 3 (Uniformly continuous). Let (X, ρ_X) and (Y, ρ_Y) be metric spaces and $f: X \to Y$. We say that f is uniformly continuous on X if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\rho_Y(f(x), f(y)) < \varepsilon$$

for all $x, y \in X$ for which $\rho_X(x, y) < \delta$.

Remark.

(1) Uniform continuity is a property of a function on a set, whereas continuity can be defined at a single point.

(2) If f is continuous on X then for each $\varepsilon > 0$ and $p \in X$ there is a number $\delta > 0$ such that if $\rho_X(x, p) < \delta$, then $\rho_Y(f(x), f(p)) < \varepsilon$. Thus δ depends on p and $\varepsilon > 0$.

(3) If f is uniformly continuous on X, then for each $\varepsilon > 0$, there is $\delta > 0$ such that for all $x, y \in X$ if $\rho_X(x, y) < \delta$, then $\rho_Y(f(x), f(y)) < \varepsilon$. Thus $\delta > 0$ depends only on ε, but is uniform for all $x, y \in X$.

(4) Uniform continuity implies continuity.

Theorem 11 (Rudin 4.19). Let f be a continuous mapping of a compact metric space (X, ρ_X) into a metric space (Y, ρ_Y). Then f is uniformly continuous on X.

Proof. Let $\varepsilon > 0$ be given. Since f is continuous, we can associate to each point $p \in X$ a positive number $\delta_p > 0$ such that

$$\text{If } q \in B(p, \delta_p), \text{ then } \rho_Y(f(p), f(q)) < \frac{\varepsilon}{2}$$

Observe that $X \subseteq \bigcup_{p \in X} B(p, \frac{\delta_p}{2})$. Since X is compact, there are $p_1, \ldots, p_n \in X$ so that

$$X \subseteq \bigcup_{k=1}^{n} B\left(p_k, \frac{\delta_{p_k}}{2}\right)$$

Set $\delta = \frac{1}{2}\min\{\delta_{p_1}, \ldots, \delta_{p_m}\} > 0$. The finiteness of the p_k terms is used here to see that $\delta > 0$. Let $p, q \in X$ be such that $\rho_X(p, q) < \delta$. Then there is $1 \leq m \leq n$ such that $p \in B(p_m, \frac{\delta_{p_m}}{2})$.

This follows from the fact that $p \in X$, so it must belong to one of the sets in the finite subcover. Hence

$$\rho_X(q, p_m) \le \rho_X(q, p) + \rho_X(p_m, p) < \delta + \frac{\delta_{p_m}}{2} < \delta_{p_m}$$

Thus, we conclude

$$\rho_Y(f(p), f(q)) \le \rho_Y(f(p), f(p_m)) + \rho_Y(f(p_m), f(q)) < \varepsilon$$

□

Theorem 12 (Rudin 4.22). If f is a continuous mapping of a metric space X into a metric space Y and if E is a connected subset of X, then $f[E]$ is connected in Y.

Proof. Assume for a contradiction that

$$f[E] = A \cup B$$

where A and B are nonempty separated subsets of Y. Put $G = E \cap f^{-1}[A]$ and $H = E \cap f^{-1}[B]$. Then

$$E = G \cup H$$

and neither G nor H is empty.

Since $A \subseteq \overline{A}$ we have $G \subseteq f^{-1}[\overline{A}]$. The latter set is closed since f is continuous hence by taking the closure on both sides we see that $\overline{G} \subseteq f^{-1}[\overline{A}]$. Hence

$$f[\overline{G}] \subseteq f[f^{-1}[\overline{A}]] \subseteq \overline{A}$$

Since $f[H] = B$ and $\overline{A} \cap B = \varnothing$ we conclude that

$$f[H \cap \overline{G}] \subseteq (f[\overline{G}] \cap f[H]) \subseteq (\overline{A} \cap B) = \varnothing$$

So $H \cap \overline{G} = \varnothing$. The same argument shows that $G \cap \overline{H} = \varnothing$. Thus G and H are separated, which contradicts our hypothesis that E is connected ($\lightning$).

Theorem 13 (Rudin 4.23). Let f be a continuous real function on the interval $[a, b]$. If $f(a) < f(b)$ and if c is a number such that $f(a) < c < f(b)$, then there is a point $x \in (a, b)$ such that $f(x) = c$.

Remark. A similar result holds also if $f(a) > f(b)$.

Proof. $[a, b]$ is connected so $f[[a, b]]$ is connected in $\mathbb{R}$ by the previous theorem. Thus if $f(a) < c < f(b)$, then $c \in f[[a, b]]$ so there is $x \in [a, b]$ so that $f(x) = c$ (by Rudin 2.47).

3.4 Banach Contraction Principle

Let (X, ρ) be a metric space. We say that $\varphi : X \to X$ is a Lipschitz mapping of X into itself with a Lipschitz constant $C_\varphi > 0$ if it satisfies

$$\rho(\varphi(x), \varphi(y)) \le C_\varphi \cdot \rho(x, y)$$

for all $x, y \in X$.
Remark: Every Lipschitz mapping is uniformly continuous.
Examples.

(1) $X = \mathbb{R}$ and $\varphi(x) = ax$ then φ is a Lipschitz map with $C_\varphi = |a|$ since

$$|\varphi(x) - \varphi(y)| = |a||x - y|$$

for all $x, y \in X$.

(2) Let (X, ρ) be a metric space and $\varnothing \neq E \subseteq X$. Define the distance from $x \in X$ to E by

$$\rho_E(x) = \inf\{\rho(x, z) : z \in E\}$$

Sometimes we write $\rho_E(x) = \rho(x, E)$.

One can easily prove that $\rho(x, E) = 0$ if and only if $x \in \overline{E}$.
Moreover

$$|\rho(x, E) - \rho(y, E)| \leq \rho(x, y)$$

Thus $X \ni x \to \rho(x, E)$ is Lipschitz with the Lipschitz constant 1. Indeed, for $z \in E$

$$\rho(x, E) \leq \rho(x, z) \leq \rho(x, y) + \rho(y, z)$$

thus $\rho(x, E) \leq \rho(x, y) + \rho(y, E)$ so

$$\rho(x, E) - \rho(y, E) \leq \rho(x, y)$$

By symmetry, $\rho(y, E) - \rho(x, E) \leq \rho(x, y)$ and we are done.
Definition (Contraction). Let (X, ρ) be a metric space. If $\varphi : X \to X$ and if there is $c \in (0, 1)$ such that

$$\rho(\varphi(x), \varphi(y)) \leq c\rho(x, y)$$

for all $x, y \in X$, then φ is said to be a contraction of X into X.
Remark. In other words, contractions $\varphi : X \to X$ are the Lipschitz maps with with Lipschitz constants $L_\varphi < 1$.
(The Banach Contraction Principle)
Theorem 1 (Rudin 9.23). If (X, ρ) is a complete metric space and if φ is a contraction of X into X, then there exists one and only one $x \in X$ such that $\varphi(x) = x$.
Recall that $\varphi : X \to X$ is a contraction, i.e., there is $0 < c < 1$ such that

$$\rho(\varphi(x), \varphi(y)) \leq c\rho(x, y) \quad \text{for all } x, y \in X$$

The idea is that after applying the contraction φ successively, the image shrinks after each application to eventually end up at the unique x for which $\phi(x) = x$.

Proof.
Uniqueness: If there are $x, y \in X$ so that $x \neq y$, $\varphi(x) = x$, and $\varphi(y) = y$, then

$$\rho(x,y) = \rho(\varphi(x), \varphi(y)) \leq c\rho(x,y) \text{ and } c < 1$$

This is impossible since it shows us that $c \geq 1$ after we divide by $\rho(x,y) \neq 0$.
Hence we must have $x = y$.
Existence: The existence of a fixed point of $\varphi : X \to X$ is the essential part of the proof / theorem. The proof furnishes a constructive method for locating the fixed point.
Pick $x_0 \in X$ arbitrarily and consider

$$\begin{aligned} x_1 &= \varphi(x_0) \\ x_2 &= \varphi(x_1) = \varphi^2(x_0) \\ &\vdots \\ x_{n+1} &= \varphi(x_n) = \varphi^{n+1}(x_0) \end{aligned}$$

for $n \in \mathbb{N}$. Observe that

$$\rho(x_{n+1}, x_n) = \rho(\varphi(x_n), \varphi(x_{n-1}) \leq c\rho(x_n, x_{n-1}$$

Thus inductively we obtain

$$\rho(x_{n+1}, x_n) \leq c^n \rho(x_1, x_0)$$

for $n \in \mathbb{N}$. If $n < m$ if follows that

$$\begin{aligned} \rho(x_n, x_m) &\leq \sum_{j=n+1}^{m} \rho(x_j, x_{j-1}) \\ &\leq (c^n + c^{n+1} + \ldots + c^{m-1}) \cdot \rho(x_1, x_0) \\ &\leq c^n(1 + c + c^2 + \ldots + c^{m-n-1}) \cdot \rho(x_1, x_0) \\ &= \frac{c^n}{1-c} \cdot \rho(x_1, x_0) \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

Thus the sequence $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X. But X is a complete metric space, so there is $x \in X$ such that

$$\lim_{n\to\infty} x_n = x$$

Since $\varphi : X \to X$ is continuous, thus

$$\begin{aligned} \varphi\left(\lim_{n\to\infty} x_n\right) &= \lim_{n\to\infty} \varphi(x_n), \quad x_{n+1} = \varphi(x_n) \\ \varphi(x) &= \lim_{n\to\infty} \varphi(x_n) = \lim_{n\to\infty} x_{n+1} = x \end{aligned}$$

□

Remark. The Banach contraction principle says that φ has a unique fixed point.

3.5 Discontinuities

If x is a point in the domain of definition of a function f at which f is not continuous, we say that f is discontinuous at x or that f has a discontinuity of x.

Definition. (Left and right limits). Let f be defined on (a,b). Consider any x such that $a \leq x < b$. We write

$$f(x+) = q$$

if $f(t_n) \xrightarrow[n\to\infty]{} q$ for all sequences $(t_n)_{n\in\mathbb{N}}$ in (x,b) such that $t_n \xrightarrow[n\to\infty]{} x$.
Similarly, we write

$$f(x-) = q$$

if $f(t_n) \xrightarrow[n\to\infty]{} q$ for all sequences $(t_n)_{n\in\mathbb{N}}$ in (a,x) satisfying $t_n \xrightarrow[n\to\infty]{} x$, for any $a < x \leq b$.
It is clear for any $x \in (a,b)$ that $\lim\limits_{t\to x} f(t)$ exists if and only if

$$f(x+) = f(x-) = \lim_{t\to x} f(t)$$

Proof. Suppose $\lim_{t\to x} f(t)$ exists. Recall by Theorem 1 (Rudin 4.2) that this means

$$\lim_{n\to\infty} f(t_n) = q$$

for every sequence $(t_n)_{n\in\mathbb{N}}$ in (a,b) such that $t_n \neq x$ and $\lim\limits_{n\to\infty} t_n = x$. Since this conclusion holds for all sequences in (a,b), it also holds for all sequences in (a,x) and all sequences in (x,b) and we have $f(x+) = f(x-) = \lim_{t\to x} f(t)$.
The converse direction is established similarly. If we have a sequence that has infinitely many $t_n \in (a,x)$ and infinitely many $t_n \in (x,b)$, we can simply form two subsequences consisting of the t_n terms in only (a,x) or (x,b) and apply our hypotheses to reach the conclusion.

Definition (Discontinuity). Let f be defined on (a,b).

(a) If f is discontinuous at a point x and if $f(x+)$ and $f(x-)$ exist, then f is said to have a discontinuity of the first kind or a simple discontinuity.

(b) Otherwise, the discontinuity is said to be of the second kind.

Remark. There are two ways in which a function can have a simple discontinuity:

(a) either $f(x+) \neq f(x-)$

(b) or $f(x+) = f(x-) \neq f(x)$

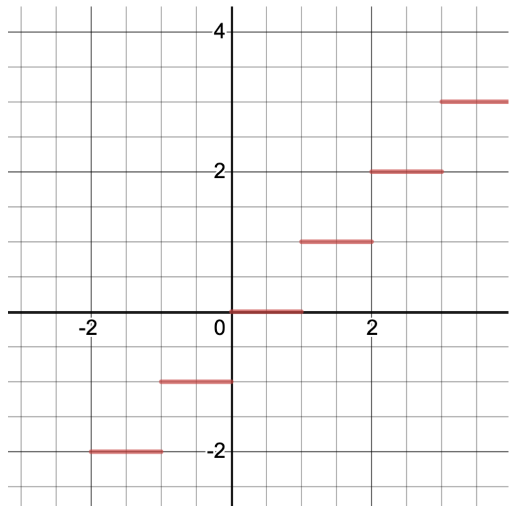
Definition. (Left and right continuity).

(a) If $f(x-) = f(x)$ for all $x \in (a,b)$ then we say that f is continuous from the left.

(b) If $f(x+) = f(x)$ for all $x \in (a,b)$ then we say that f is continuous from the right.

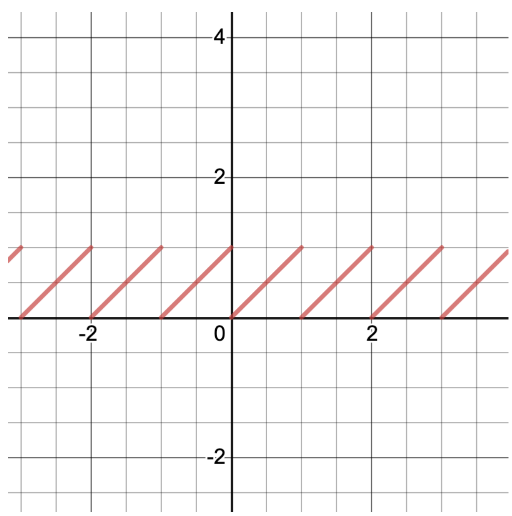
Examples.

(1) $f(x) = \lfloor x \rfloor = \max\{n \leq x : n \in \mathbb{Z}\}$



f is continuous from the right.

(2) $f(x) = \{x\} = x - \lfloor x \rfloor$



f is also continuous from the right.

(3) Let

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

f has a discontinuity of the second kind at every point x since neither $f(x+)$ nor $f(x-)$ exists.

(4) Define

$$f(x) = \begin{cases} x & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

Then f is continuous at $x = 0$ and has a discontinuity of the second kind at every other point.

(5)

$$f(x) = \begin{cases} x + 2 & \text{if } x \in (-3, -2) \\ -x - 2 & \text{if } x \in [-2, 0) \\ x + 2 & \text{if } x \in [0, 1] \end{cases}$$

Then f has a simple discontinuity at $x = 0$ and is continuous at every other point of $(-3, 1)$.

Definition (Monotonically increasing). Let $f : (a, b) \to \mathbb{R}$. Then f is said to be monotonically increasing on (a, b) if

$$a < x < y < b \implies f(x) \leq f(y)$$

If $f(x) \geq f(y)$ for all $x, y \in (a, b)$, we obtain the definition of a monotonically decreasing function.

Theorem (Rudin 4.29). Let f be monotonically increasing on (a, b). Then $f(x+)$ and $f(x-)$ exist at every point of $x \in (a, b)$. More precisely,

$$\sup_{a<t<x} f(t) = f(x-) \leq f(x) \leq f(x+) = \inf_{x<t<b} f(t)$$

Furthermore, if $a < x < y < b$ then $f(x+) \leq f(y-)$. Analogous results hold for monotonically decreasing functions.

Proof. The set $E = \{f(t) : a < t < x\}$ is bounded by $f(x)$ hence it has a least upper bound which we denote $A = \sup E$. Clearly

$$A \leq f(x)$$

We have to show $f(x-) = A$. Let $\varepsilon > 0$ be given. Since $A = \sup E$, there is a $\delta > 0$ such that $a < x - \delta < x$ and

$$A - \varepsilon < f(x - \delta) \leq A$$

Since f is monotonic, we have

$$f(x - \delta) \leq f(t) \leq A \quad \text{for } t \in (x - \delta, x)$$

Thus $A - \varepsilon < f(t) \leq A$, so

$$|f(t) - A| < \varepsilon \quad \text{for } t \in (x - \delta, x)$$

Thus $A = f(x-)$. To elaborate, suppose we have a sequence $(t_n)_{n\in\mathbb{N}} \subseteq (a, x)$ satisfying $t_n \xrightarrow[n\to\infty]{} x$. Because this sequence converges to x, taking the δ given above, we can find $N \in \mathbb{N}$ such that $n \geq N$ implies $|t_n - x| < \delta$. But then by what we have shown above, we have $|f(t_n) - A| < \varepsilon$. Hence $f(t_n) \xrightarrow[n\to\infty]{} A$ as desired.
Next, if $a < x < y < b$, then

$$f(x+) = \inf_{x<t<b} f(t) = \inf_{x<t<y} f(t)$$

Similarly

$$f(y-) = \sup_{a<t<y} f(t) = \sup_{x<t<y} f(t)$$

Thus

$$f(x+) = \inf_{x<t<y} f(x) \leq \sup_{x<t<y} f(t) = f(y-)$$

□

Corollary. Monotonic functions have no discontinuities of the second kind.
Theorem (Rudin 4.30). Let $f : (a, b) \to \mathbb{R}$ be monotonic. Then the set of points of (a, b) at which f is discontinuous is at most countable.
Proof. Without loss of generality, we may assume that f is increasing, and let E be the set of points at which f is discontinuous.
Let $r : E \to \mathbb{Q}$. With every point $x \in E$ we associate a rational number $r(x) \in \mathbb{Q}$ such that

$$f(x-) < r(x) < f(x+)$$

Since $x_1 < x_2$ implies $f(x_1+) \leq f(x_2-)$, we see that $r(x_1) \neq r(x_2)$ if $x_1 \neq x_2$.
We have thus established a one-to-one correspondence between the set E and a subset of $\mathbb{Q}$. The latter, as we know, is countable $\text{card}(\mathbb{Q}) = \aleph_0$.

3.6 Infinite Limits and Limits at Infinity

For any $x \in \mathbb{R}$ we have already defined a neighborhood of x to be any segment $(x - \delta, x + \delta)$.
Definition. For any real c, the set of real numbers $x > c$ is called a neighborhood of $+\infty$ and is written $(c, +\infty)$. Similarly, the set $(-\infty, c)$ is a neighborhood of $-\infty$.
Definition. Let f be a real function defined on $E \subseteq \mathbb{R}$. We say

$$f(t) \xrightarrow[t\to x]{} A$$

where A and x are in the extended real system if for every neighborhood U of A there is a neighborhood V of x such that $V \cap E \neq \varnothing$ and such that $f(t) \in U$ for all $t \in (V \cap E) \setminus \{x\}$.

Theorem (Rudin 4.34). Let $f, g : E \to \mathbb{R}$. Suppose

$$f(t) \xrightarrow[t \to x]{} A \quad \text{and} \quad g(t) \xrightarrow[t \to x]{} B$$

then

(a) $f(t) \xrightarrow[t \to x]{} A'$ implies $A = A'$

(b) $(f + g)(t) \xrightarrow[t \to x]{} A + B$

(c) $(fg)(t) \xrightarrow[t \to x]{} AB$

(d) $(f/g)(t) \xrightarrow[t \to x]{} \frac{A}{B}$

provided the right members of (b), (c), (d) are defined.

Remark. Note that $\infty \cdot \infty$, $0 \cdot \infty$, $\frac{\infty}{\infty}$, and $\frac{A}{0}$ are not defined.

4 Differentiation

Definition (Derivative).

Let $f : [a, b] \to \mathbb{R}$. For any $x \in [a, b]$ form the quotient

$$\phi(t) = \frac{f(t) - f(x)}{t - x}$$

with $a < t < b$ and $t \neq x$, and define

$$f'(x) = \lim_{t \to x} \phi(t)$$

provided the limit exists.
We thus associate with the function f a function f' whose domain is the set of points x at which the limit

$$\lim_{t \to x} \phi(t) \quad \text{exists}$$

f' is called the derivative of f.

Remark:

- If f' is defined at a point x, we say that f is differentiable at x.
- If f' is defined at every point of a set $E \subseteq [a, b]$, we say that f is differentiable on E.
- It is possible to consider right-hand and left-hand limits of $\phi(t)$. This leads to the definition of right-hand and left-hand derivatives. In particular at the endpoints a, b, the derivative, if it exists, is a right-hand or left-hand derivative respectively.
- If f is defined on a segment (a, b) and if $a < x < b$ then $f'(x)$ is defined by
$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x}$$
but $f'(a)$ and $f'(b)$ are not defined in this case!

Theorem 1 (Rudin 5.2). Let $f : [a, b] \to \mathbb{R}$. If f is differentiable at a point $x \in [a, b]$, then f is continuous at x.

Proof. Observe

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

Rewriting this, we have

$$\lim_{t \to x} f(t) = f(x)$$

which shows us that f is continuous by Fact 1 on Page 55.

Remark:

- The converse of this theorem is not true. As a counterexample, $f(x) = |x|$ is continuous at every $x \in \mathbb{R}$ but it is not differentiable at $x = 0$. The function is continuous because it is Lipschitz continuous: $|f(x) - f(y)| = ||x| - |y|| \leq |x - y|$.

- It is also possible to construct a continuous function on $\mathbb{R}$ which is not differentiable at any point of $\mathbb{R}$.

Theorem 2 (Rudin 5.3). Let $f, g : [a, b] \to \mathbb{R}$ be differentiable at a point $x \in [a, b]$. Then $f + g, fg$, and f/g are differentiable at x and

(a) $(f + g)'(x) = f'(x) + g'(x)$

(b) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$

(c) $(\frac{f}{g})'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$ with $g(x) \neq 0$

Proof. (a) is clear since

$$\begin{aligned}(f+g)'(x) &= \lim_{t \to x} \frac{f(t) + g(t) - (f(x) + g(x))}{t - x} \\ &= \lim_{t \to x} \frac{f(t) - f(x)}{t - x} + \lim_{t \to x} \frac{g(t) - g(x)}{t - x} \\ &= f'(x) + g'(x)\end{aligned}$$

(b) Let $h = fg$, then

$$h(t) - h(x) = f(t)(g(t) - g(x)) + g(x)(f(t) - f(x))$$

Thus

$$\begin{aligned}(fg)'(x) = h'(x) &= \lim_{t \to x} \frac{h(t) - h(x)}{t - x} \\ &= \lim_{t \to x} f(t) \cdot \frac{g(t) - g(x)}{t - x} + \lim_{t \to x} g(x) \cdot \frac{f(t) - f(x)}{t - x} \\ &= f(x)g'(x) + f'(x)g(x)\end{aligned}$$

For (c) let $h = \frac{f}{g}$ and observe

$$\frac{h(t) - h(x)}{t - x} = \frac{1}{g(t)g(x)} \left[g(x) \cdot \frac{f(t) - f(x)}{t - x} - f(x) \cdot \frac{g(t) - g(x)}{t - x} \right]$$

Letting $t \to x$ we obtain the desired claim.

Examples.

(1) $f(x) = c \in \mathbb{R}$ for all $x \in \mathbb{R}$, then $f'(x) = 0$ for all $x \in \mathbb{R}$

(2) $f(x) = x^n$, then $f'(x) = nx^{n-1}$, $n \in \mathbb{N}$.
Indeed,

$$x^n - y^n = (x-y)(x^{n-1} + x^{n-2}y + \ldots + y^{n-2}x + y^{n-1})$$

Thus

$$\frac{f(t)-f(x)}{t-x} = t^{n-1} + t^{n-2}x + \ldots + x^{n-2}t + x^{n-1} \xrightarrow[t \to x]{} nx^{n-1}$$

(3) $f(x) = \frac{1}{x^n}$ then $f'(x) = -\frac{nx^{n-1}}{x^{2n}} = -\frac{n}{x^{n+1}}$, $x \neq 0$ (by Theorem 2c above).

(4) Every polynomial $P(x) = \sum_{k=0}^{d} a_k x^k$ is differentiable.

(5) $R(x) = \frac{P(x)}{Q(x)}$ where $P(x), Q(x)$ are polynomials.
$R(x)$ is differentiable for all $x \in \mathbb{R}$ such that $Q(x) \neq 0$.

Theorem 3 (Rudin 5.5). Suppose that $f : [a,b] \to \mathbb{R}$ is continuous and $f'(x)$ 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 $f(x)$.
If

$$h(t) = g(f(t)), \quad a \le t \le b$$

then h is differentiable at x and

$$h'(x) = g'(f(x)) \cdot f'(x)$$

the latter identity is called the chain rule.
Proof. Let $y = f(x)$. By the definition of the derivative we have

$$f(t) - f(x) = (t-x)(f'(x) + u(t)) \quad \text{and} \quad g(s) - g(y) = (s-y)(g'(y) + v(s))$$

where $t \in [a,b]$, $s \in I$ and

$$\lim_{t \to x} u(t) = 0 \quad \text{and} \quad \lim_{s \to y} v(s) = 0$$

Note that we are defining

$$u(t) = \begin{cases} \frac{f(t)-f(x)}{t-x} - f'(x) & \text{if } t \neq x \\ 0 & \text{if } t = x \end{cases}$$

and $v(s)$ similarly, so that these functions work as error functions.
Let $s = f(t)$ and note that

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

or if $t \neq x$ then

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

Letting $t \to x$ we see that $s \to y$ (this is the same as $f(t) \to f(x)$) by the continuity of f. Thus

$$\lim_{t \to x} \frac{h(t) - h(x)}{t - x} = g'(y)f'(x) = g'(f(x)) \cdot f'(x)$$

4.1 Mean Value Theorems

Definition 2 (Local maxima). Let (X, ρ) be a metric space and $f : X \to \mathbb{R}$. We say that f has a local maximum at a point $p \in X$ if there exists $\delta > 0$ such that

$$f(q) \leq f(p) \quad \text{for all } q \in B(p, \delta)$$

Local minima are defined likewise.

Theorem 4 (Rudin 5.8.) Let f be defined on $[a, b]$. If f has a local maximum at a point $x \in (a, b)$ and if $f'(x)$ exists, then $f'(x) = 0$.

The analogous statement is also true for local minima.

Proof. Let $x \in (a, b)$ be a local maximum. Thus, there is $\delta > 0$ such that if $|q - x| < \delta$ then

$$f(q) \leq f(x)$$

We can assume that $a < x - \delta < x < x + \delta < b$. If $x - \delta < t < x$, then

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

Letting $t \to x$ we see that $f'(x) \geq 0$.

If $x < t < x + \delta$, then

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

Letting $t \to x$ then we obtain $f'(x) \leq 0$, thus we conclude that $f'(x) = 0$.

Theorem 5 (Rudin 5.9). (Mean Value Theorem).

If $f, g : [a, b] \to \mathbb{R}$ are continuous on $[a, b]$ and differentiable in (a, b) then there is a point $x \in (a, b)$ at which

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

Note that differentiability is not required of the endpoints.

If $g(x) = x$, then we recover the Lagrange theorem

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

for some $x \in (a, b)$.

Proof. For $a \leq t \leq b$, put

$$h(t) = (f(b) - f(a))g(t) - (g(b) - g(a))f(t)$$

then h is continuous on $[a, b]$ and h is differentiable in (a, b) and

$$h(a) = f(b)g(a) - f(a)g(b) = h(b)$$

To prove the theorem, we have to show that

$$h'(x) = 0 \text{ for some } x \in (a, b)$$

If h is constant, this holds for every $x \in (a, b)$.
If $h(t) > h(a)$ for some $t \in (a, b)$, let x be a point on $[a, b]$ at which h attains its maximum. Recall that a continuous function always attains its maximum and minimum on a compact set. Since $h(a) = h(b)$, then $x \in (a, b)$. We know this since there is some t for which $h(t) > h(a)$, so $h(a)$ cannot be a maximum. By the previous theorem, $h'(x) = 0$.
If $h(t) < h(a)$ for some $t \in (a, b)$, the same argument applies if we choose for x a point on $[a, b]$ where h attains its minimum.
Since $h'(x) = 0$, we have

$$\begin{aligned} h'(t) &= (f(b) - f(a))g'(t) - (g(b) - g(a))f'(t) \\ h'(x) &= (f(b) - f(a))g'(x) - (g(b) - g(a))f'(x) = 0 \\ (f(b) - f(a))g'(x) &= (g(b) - g(a))f'(x) \end{aligned}$$

as desired.

□

Theorem 6 (Rudin 5.11). Suppose f is differentiable in (a, b)

(a) If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotonically increasing.

(b) If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.

(c) If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotonically decreasing.

Proof. From the mean value theorem, for each $a < x_1 < x_2 < b$, we have

$$f(x_2) - f(x_1) = f'(x)(x_2 - x_1) \quad \text{for some } x \in (x_1, x_2)$$

thus if $f'(x) \geq 0$, then $f(x_2) \geq f(x_1)$.
If $f'(x) \leq 0$, then $f(x_2) \leq f(x_1)$.
If $f'(x) = 0$, then $f(x_2) = f(x_1)$.

□

Derivatives which exist at every point of an interval have one important property in common with functions which are continuous on an interval: "Intermediate values are assumed."
Theorem 7 (Rudin 5.12). Suppose f is a real differentiable function on $[a, b]$ and suppose $f'(a) < \lambda < f'(b)$. Then there is a point $x \in (a, b)$ such that $f'(x) = \lambda$.

(A similar result holds, of course, if $f'(a) > f'(b)$).

Proof. Set $g(t) = f(t) - \lambda t$. Then $g'(a) = f'(a) - \lambda < 0$ so that $g(t_1) < g(a)$ for some $t_1 \in (a, b)$. Observe that

$$0 > g'(a) = \lim_{t \to a} \frac{g(t) - g(a)}{t - a}$$

This shows the existence of such a t_1. More explicitly, suppose for contradiction that $g(t) > g(a)$ for all $t \in (a, b)$. Take $\varepsilon < |g'(a)|$. Notice that for all $\delta > 0$ and t satisfying $0 < |t-a| < \delta$ we have

$$\left| \frac{g(t) - g(a)}{t - a} - g'(a) \right| > |g'(a)| = \varepsilon$$

Thus, $g'(a)$ is not the limit of the difference quotient, a contradiction. ($\lightning$)

In a similar way, since $g'(a) > 0$, we obtain $g(t_2) > g(b)$ for some $t_2 \in (a, b)$.

Hence, g attains its minimum on $[a, b]$ at some point $x \in (a, b)$. By Theorem 4, we have $g'(x) = 0$. Hence $f'(x) = \lambda$.

□

Corollary. If f is differentiable on $[a, b]$, then f' cannot have any simple discontinuities on $[a, b]$. But, f' may very well have discontinuities of the second kind.

4.2 L'Hospital's Rule

Theorem 8 (Rudin 5.13). Suppose $f, g : (a, b) \to \mathbb{R}$ are differentiable in (a, b) and $g'(x) \neq 0$ for all $x \in (a, b)$ where $-\infty \leq a < b \leq +\infty$. Suppose

$$\frac{f'(x)}{g'(x)} \xrightarrow[x \to a]{} A \tag{$*$}$$

If

$$f'(x) \xrightarrow[x \to a]{} 0 \quad \text{and} \quad g'(x) \xrightarrow[x \to a]{} 0$$

or if

$$g(x) \xrightarrow[x \to a]{} +\infty$$

then

$$\frac{f(x)}{g(x)} \xrightarrow[x \to a]{} A$$

Remark. The analogous statement is of course true if $x > b$ or if $g(x) \to -\infty$.

Proof. We first consider the case $-\infty \leq A < +\infty$. Choose a real number q such that $A < q$ and then choose r such that $A < r < q$. By ($*$) there is $c \in (a, b)$ such that $a < x < c$ implies

$$\frac{f'(x)}{g'(x)} < r$$

To be explicit, recall that $\frac{f'(x)}{g'(x)} \xrightarrow[x\to a]{} A$ means that for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $0 < |x-a| < \delta$, then

$$\left|\frac{f'(x)}{g'(x)} - A\right| < \varepsilon$$

Thus, by choosing $\varepsilon = r - A$, we see that

$$\frac{f'(x)}{g'(x)} - A < r - A \implies \frac{f'(x)}{g'(x)} < r$$

as desired. If $a < x < y < c$, then the Mean Value Theorem shows that there is a point $t \in (x, y)$ such that

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(t)}{g'(t)} < r$$

If $f(x) \xrightarrow[x\to a]{} 0$ and $g(x) \xrightarrow[x\to a]{} 0$, then we see

$$\frac{f(y)}{g(y)} \le r < q \quad \text{for } a < y < c \tag{$**$}$$

Suppose $g(x) \xrightarrow[x\to a]{} +\infty$ holds. Keeping y fixed in $(**)$, we can choose a point $c_1 \in (a, y)$ such that $g(x) > g(y)$ and $g(x) > 0$ if $a < x < c_1$. Then

$$\frac{g(x) - g(y)}{g(x)} > 0$$

Thus

$$\begin{aligned}
\frac{f(x) - f(y)}{g(x)} &= \frac{f(x) - f(y)}{g(x) - g(y)} \cdot \frac{g(x) - g(y)}{g(x)} \\
&< r \cdot \frac{g(x) - g(y)}{g(x)} \\
&= r - r \cdot \frac{g(y)}{g(x)}
\end{aligned}$$

Hence

$$\frac{f(x)}{g(x)} < r - r\frac{g(y)}{g(x)} + \frac{f(y)}{g(x)} \quad \text{for } a < x < c_1$$

If we let $x \to a$ (since $g(x) \xrightarrow[x\to a]{} +\infty$) we find $c_2 \in (a, c_1)$ such that

$$\frac{f(x)}{g(x)} < q \quad \text{for all } a < x < c_2$$

Hence we conclude that for any $A < q$, there is c_2 so that

$$\frac{f(x)}{g(x)} < q \quad \text{if } a < x < c_2$$

In the same manner, if $-\infty < A \leq +\infty$ and p is chosen so that $p < A$ we can find a point c_3 such that

$$p < \frac{f(x)}{g(x)} \quad \text{for } a < x < c_3$$

If $-\infty < a < +\infty$ we take $\varepsilon > 0$ and set $p = A - \varepsilon$ and $q = A + \varepsilon$. Then there is c_ε so that for $a < x < c_\varepsilon$ we have

$$A - \varepsilon < \frac{f(x)}{g(x)} < A + \varepsilon$$

and we are done.

□

4.3 Derivatives of Higher Order

Definition (Second derivative). If f has a derivative f' on an interval, and if f' is itself differentiable, we denote the derivative of f' by f'' and call f'' the second derivative of f. Continuing this way, we obtain:

$$f, f', f'', f^{(3)}, \ldots, f^{(n)}, \ldots$$

each of which is the derivative of the preceding one. $f^{(n)}$ is called the n-th derivative, or the derivative of order n of f.

Remark. In order for $f^{(n)}(x)$ to exist at a point x, $f^{(n-1)}(t)$ must exist in a neighborhood of x (or in a one-sided neighborhood, if x is an endpoint of the interval on which f is defined). and $f^{(n-1)}$ must be differentiable at x.

Example.

$$\begin{aligned} f(x) &= x^n \\ f'(x) &= nx^{n-1}, f''(x) = n(n-1)x^{n-2} \\ f^{(3)}(x) &= n(n-1)(n-2)x^{n-3}, \ldots, f^{(n)}(x) = n(n-1)\cdots 2 \cdot 1 = n! \end{aligned}$$

4.4 Taylor's Theorem

Theorem 1 (Rudin 5.15). Suppose $f : [a, b] \to \mathbb{R}$, n is a positive integer, $f^{(n-1)}$ is continuous on $[a, b]$ and $f^{(n)}(t)$ exists for every $t \in (a, b)$. Let α, β be distinct points of $[a, b]$ and define

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(t - \alpha)^k$$

then there exists a point x between α and β such that

$$f(\beta) = P(\beta) + \frac{f^{(n)}(x)}{n!}(\beta - \alpha)^n \tag{$*$}$$

Remark. For $n = 1$, this is just the Mean Value Theorem. In general, the theorem shows that f can be approximated by a polynomial of degree $n-1$ and that $(*)$ allows us to estimate the error term if we know bounds on $|f^{(n)}(x)|$.
To see that this is the Mean Value Theorem at $n = 1$, notice that

$$f(\beta) = \sum_{k=0}^{0} \frac{f^{(k)}(\alpha)}{k!}(\beta-\alpha)^k + f'(x)(\beta-\alpha)^n$$
$$f(\beta) - f(\alpha) = f'(x)(\beta-\alpha)$$

and the statement regarding the existence of such an x is precisely the MVT.
Proof. Let M be a number such that

$$f(\beta) = P(\beta) + M(\beta-\alpha)^n$$

and set

$$g(t) = f(t) - P(t) - M(t-\alpha)^n \quad \text{for } a \le t \le b$$

We have to show that

$$n!M = f^{(n)}(x) \quad \text{for some } x \text{ between } \alpha \text{ and } \beta$$

Since

$$P(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(t-\alpha)^k$$

By taking the n^{th} derivative of $P(t)$, we see that

$$P^{(n)}(t) = 0$$

thus

$$g^{(n)}(t) = f^{(n)}(t) - n!M$$

The proof will be completed if we show that $g^{(n)}(x) = 0$ for some x between α and β.
Since $P^{(k)}(\alpha) = f^{(k)}(\alpha)$ for $k = 0, 1, \dots, n-1$, we have

$$g(\alpha) = g'(\alpha) = \cdots = g^{(n-1)}(\alpha) = 0$$

Our choice of M shows that $g(\beta) = 0$, so that $g'(x_1) = 0$ for some x_1 between α and β by the Mean Value Theorem.

$$0 = g(\beta) - g(\alpha) = g'(x_1)(\beta-\alpha), \quad a < x_1 < \beta$$

Since $g'(\alpha) = 0$, we conclude similarly that

$$0 = g'(x_1) - g'(\alpha) = (x_1-\alpha)g''(x_2) \quad \text{for some } \alpha < x_2 < x_1$$

thus $g''(x_2)$.
After n steps we obtain that $g^{(n)}(x_n) = 0$ for some $\alpha < x_n < x_{n-1} < \ldots < x_1 < \beta$.
Details. We know that $g(\beta) = 0$ by the substitution below:

$$\begin{aligned} g(\beta) &= f(\beta) - P(\beta) - M(\beta-\alpha)^n \\ &= P(\beta) + M(\beta-\alpha)^n - P(\beta) - M(\beta-\alpha) = 0 \end{aligned}$$

We know that $P^{(k)}(\alpha) = f^{(k)}(\alpha)$ by the following

$$\begin{aligned} P(\alpha) &= \sum_{j=0}^{n-1} \frac{f^{(j)}(\alpha)}{j!}(\alpha-\alpha)^j \\ &= f(\alpha) + f'(\alpha)(\alpha-\alpha) + \frac{f''(\alpha)}{2!}(\alpha-\alpha)^2 + \ldots \\ &= f(\alpha) \end{aligned}$$

Taking derivatives eliminates the term at the beginning, which is how we get the chain of equalities.

□

Theorem 2. Suppose that $f : [a,b] \to \mathbb{R}$ is n-times continuously differentiable on $[a,b]$ and $f^{(n+1)}$ exists in the open interval (a,b). For any $x, x_0 \in [a,b]$ and $p > 0$, there exists $\theta \in (0,1)$ such that

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + r_n(x)$$

where

$$r_n(x) = \frac{f^{(n+1)}(x_0 + \theta(x-x_0))}{n!p} \cdot (1-\theta)^{n+1-p} \cdot (x-x_0)^{n+1}$$

Remark. The remainder function $r_n(x)$ is usually called the Schlömlich-Roche remainder.
Proof. For $x, x_0 \in [a,b]$ set

$$r_n(x) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k$$

Without loss of generality, we may assume that $x > x_0$. On $[x_0, x]$ define

$$\varphi(z) := f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(z)}{k!}(x-z)^k$$

Based on this definition, we have $\varphi(x_0) = r_n(x)$ and $\varphi(x) = 0$.
Moreover, φ' exists in (x_0, x) and

$$\varphi'(z) = -\frac{f^{(n+1)}(z)}{n!}(x-z)^n$$

The derivation follows below

$$\begin{aligned}
\varphi'(z) &= -\left(\sum_{k=0}^{n} \frac{f^{(k)}(z)}{k!}(x-z)^k\right)' \\
\text{(Product rule)} &= -\sum_{k=0}^{n}\left(\frac{f^{(k+1)}(z)}{k!}(x-z)^k + \frac{f^{(k)}(z)}{k!}k(x-z)^{k-1}(-1)\right) \\
&= \sum_{k=1}^{n} \frac{f^{(k)}(z)}{(k-1)!}(x-z)^{k-1} - \sum_{k=0}^{n} \frac{f^{(k+1)}(z)}{k!}(x-z)^k \\
&= -\frac{f^{(n+1)}(z)}{n!}(x-z)^n
\end{aligned}$$

Let $\psi(z) = (x-z)^p$. Then ψ is continuous on $[x_0, x]$ with nonvanishing derivative on (x_0, x). By the Mean Value Theorem

$$\frac{\varphi(x) - \varphi(x_0)}{\psi(x) - \psi(x_0)} = \frac{\varphi'(c)}{\psi'(c)} \qquad \text{for some } c \in (x_0, x)$$

Thus, by letting $c = x_0 + \theta(x - x_0)$

$$\begin{aligned}
r_n(x) = \varphi(x_0) - \varphi(x) &= -(\psi(x) - \psi(x_0))\frac{\varphi'(c)}{\psi'(c)} \\
&= \frac{f^{(n+1)}(c)}{n!}(x-c)^n \cdot \frac{-(x-x_0)^p}{-p(x-c)^{p-1}} \\
&= \frac{f^{(n+1)}(x_0 + \theta(x-x_0))}{n!p} \cdot (1-\theta)^{n+1-p} \cdot (x-x_0)^{n+1}
\end{aligned}$$

□

4.4.1 Lagrange and Cauchy Remainders

Remark. Under the assumptions of the previous theorem:

1. If $p = n+1$, then we obtain the Taylor expansion formula with the Lagrange remainder

$$r_n(x) = \frac{f^{(n+1)}(x_0 + \theta(x-x_0))}{(n+1)!}(x-x_0)^{n+1}$$

2. If $p = 1$, then we obtain the Taylor expansion formula with the Cauchy remainder

$$r_n(x) = \frac{f^{(n+1)}(x_0 + \theta(x-x_0))}{n!}(1-\theta)^n(x-x_0)^{n+1}$$

4.5 Differentiation of Vector-Valued Functions

Remark. The definition of differentiation applies without any change to complex functions $f : [a, b] \to \mathbb{C}$. If $f_1 = \text{Re} f$ and $f_2 = \text{Im} f$, then

$$f'(t) = f_1'(t) + i f_2'(t)$$

Also, f is differentiable at x if and only if both f_1 and f_2 are differentiable at x.
Passing to vector-valued functions in general: $\mathbf{f} : [a, b] \to \mathbb{R}^k$, the definition of differentiation also applies. In other words, $\mathbf{f}'(x)$ exists as a point of $\mathbb{R}^k$ if

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

where the distance is the Euclidean norm. If $\mathbf{f} = (f_1, \ldots, f_k) : [a, b] \to \mathbb{R}^k$ then

$$\mathbf{f}' = (f_1', \ldots, f_k')$$

and f is differentiable at a point x if and only if each of the functions $f_1, \ldots, f_k$ is differentiable at x.
Remark. However, the Mean Value Theorem and the L'Hospital rule may fail in the vector-valued case.
Theorem 3 (Rudin 5.19). Suppose $\mathbf{f} : [a, b] \to \mathbb{R}^k$ is continuous and $\mathbf{f}$ is differentiable in (a, b). Then there exists $x \in (a, b)$ such that

$$|\mathbf{f}(b) - \mathbf{f}(a)| \leq (b - a) \cdot |\mathbf{f}'(x)|$$

Proof. Put $\mathbf{z} = \mathbf{f}(b) - \mathbf{f}(a)$ and define

$$\varphi(t) = \mathbf{z} \cdot \mathbf{f}(t) \quad a \leq t \leq b$$

Then φ is a real-valued continuous function on $[a, b]$ which is differentiable on (a, b). The Mean Value Theorem ensures that

$$\varphi(b) - \varphi(a) = (b - a)\varphi'(x) = (b - a)\mathbf{z} \cdot \mathbf{f}'(x)$$

for some $x \in (a, b)$. On the other hand

$$\varphi(b) - \varphi(a) = \mathbf{z} \cdot \mathbf{f}(b) - \mathbf{z} \cdot \mathbf{f}(a) = \mathbf{z} \cdot \mathbf{z} = |\mathbf{z}|^2$$

By the Cauchy-Schwarz inequality

$$|\mathbf{z}|^2 = (b - a)|\mathbf{z} \cdot \mathbf{f}'(x)| \leq (b - a)|\mathbf{z}| \cdot |\mathbf{f}'(x)|$$

Hence $|\mathbf{z}| \leq (b - a) \cdot |\mathbf{f}'(x)|$ as claimed.

□

4.6 Convexity

Definition (Convex function). A real-valued function $f : (a, b) \to \mathbb{R}$ is convex if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

whenever $a < x < b, \quad a < y < b$ and $0 < \lambda < 1$.

Observation 1. If $f : (a, b) \to \mathbb{R}$ is convex and if $a < s < t < u < b$, then

$$\frac{f(t) - f(s)}{t - s} \leq \frac{f(u) - f(s)}{u - s} \leq \frac{f(u) - f(t)}{u - t}$$

Proof. Since $s < t < u$, then we may write $t = \alpha u + \beta s$ for some $\alpha, \beta \in [0, 1]$ and $\alpha + \beta = 1$. More precisely

$$t = \alpha u + \beta s = \frac{t - s}{u - s} u + \frac{u - t}{u - s} s$$

Thus

$$f(t) = f\left(\frac{t - s}{u - s} u + \frac{u - t}{u - s} s\right) \leq \frac{t - s}{u - s} f(u) + \frac{u - t}{u - s} f(s)$$

Hence

$$\begin{aligned} f(t) - f(s) &\leq \frac{t - s}{u - s} f(u) + \frac{u - t}{u - s} f(s) - \frac{u - s}{u - s} f(s) \\ f(t) - f(s) &\leq \frac{t - s}{u - s} f(u) - \frac{t - s}{u - s} f(s) \\ \frac{f(t) - f(s)}{t - s} &\leq \frac{f(u) - f(s)}{u - s} \end{aligned}$$

The remaining conclusion follows similarly.

Observation 2.

If $f : (a, b) \to \mathbb{R}$ is convex, then for any $\lambda_1 + \ldots + \lambda_n = 1$ with $\lambda_1, \ldots, \lambda_n \in [0, 1]$ we have

$$f(\lambda_1 x_1 + \ldots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \ldots + \lambda_n f(x_n)$$

Proof. For $n = 2$, it follows from the definition of convexity. Suppose that the statement is true for $n \geq 2$ and we show it also holds for $n + 1$. Let $\lambda_1, \ldots, \lambda_{n+1} \in [0, 1]$ so that $\lambda_1 + \ldots + \lambda_{n+1} = 1$. Observe that

$$\begin{aligned} & f\left(\lambda_{n+1} x_{n+1} + (1 - \lambda_{n+1}) \left[\sum_{k=1}^{n} \frac{\lambda_k}{1 - \lambda_{n+1}} x_k\right]\right) \\ \text{(Convexity)} \quad & \leq \lambda_{n+1} f(x_{n+1}) + (1 - \lambda_{n+1}) f\left(\sum_{k=1}^{n} \frac{\lambda_k}{1 - \lambda_{n+1}} x_k\right) \\ \text{(Induction)} \quad & \leq \lambda_{n+1} f(x_{n+1}) + (1 - \lambda_{n+1}) \sum_{k=1}^{n} \frac{\lambda_k}{1 - \lambda_{k+1}} f(x_k) \\ & = \sum_{k=1}^{n+1} \lambda_k f(x_k) \end{aligned}$$

since

$$\sum_{k=1}^{n} \frac{\lambda_k}{1 - \lambda_{n+1}} = \frac{1}{1 - \lambda_{n+1}} \sum_{k=1}^{n} \lambda_k = \frac{1 - \lambda_{n+1}}{1 - \lambda_{n+1}} = 1$$

□

Theorem 4. If f is convex on (a, b), then f is continuous on (a, b).

Proof. Let $a < s < u < v < t < b$. By Observation 1,

$$f(u) \leq f(s) + \frac{f(v) - f(s)}{v - s}(u - s)$$

and also

$$f(v) \leq f(u) + \frac{f(t) - f(u)}{t - u}(v - u)$$

Thus

$$f(s) + \frac{f(u) - f(s)}{u - s}(v - s) \leq f(v) \leq f(u) + \frac{f(t) - f(u)}{t - u}(v - u)$$

If $v_n \xrightarrow[n \to \infty]{} u$ converges to u, then we see that

$$\lim_{n \to \infty} f(v_n) = f(u)$$

Therefore, by Rudin 4.1 we conclude

$$\lim_{x \to u} f(x) = f(u)$$

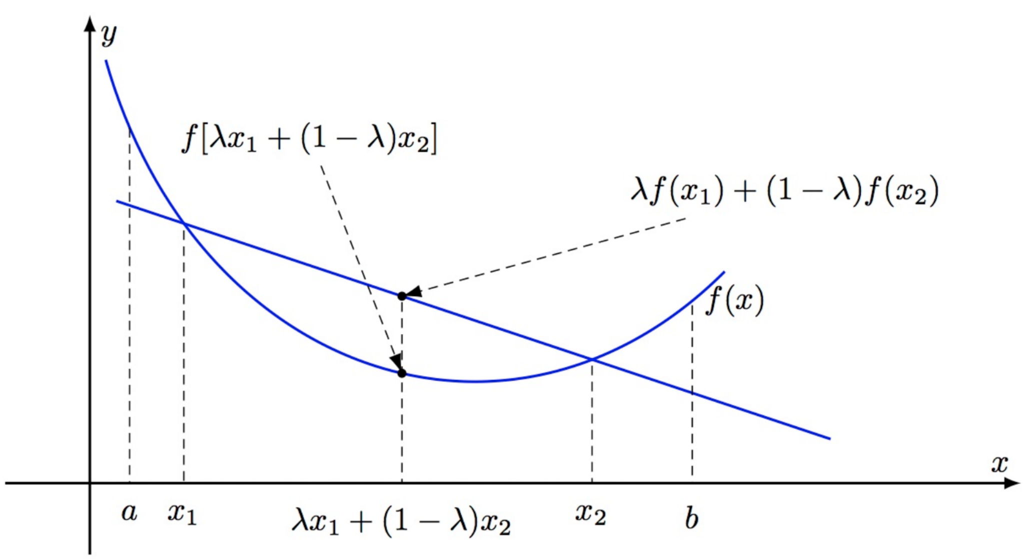
□

The sign of the first derivative has been interpreted in terms of a geometric property of the function, whether it is decreasing or increasing. We shall now interpret the sign of the second derivative.

Let f be a function defined on a closed interval $[a, b]$. The equation of the line passing through $(a, f(a))$ and $(b, f(b))$ is

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

Below is a representation of how such a line would look together with a convex function



The condition that every point on the curve $y = f(x)$ lies below this line segment between $x = a$ and $x = b$ is that

$$f(x) \leq f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \quad \text{for } a \leq x \leq b \tag{$*$}$$

Any point x between a and b can be written in the form $x = a + t(b - a)$ with $t \in [0, 1]$. In fact, one sees that the map $t \mapsto a + t(b - a)$ is a strictly increasing bijection between $[0, 1]$ and $[a, b]$.

If we substitute the value for x in terms of t in $(*)$ then we find an equivalent inequality

$$f((1 - t)a + tb) \leq (1 - t)f(a) + tf(b)$$

which is the convexity of the function f on (a, b).

Theorem 5. Let $f : [a, b] \to \mathbb{R}$ be continuous. Assume that f'' exists on the interval (a, b) and $f''(x) > 0$ on (a, b). Then f is strictly convex on the interval $[a, b]$.

Proof. If $a \leq c < d \leq b$ we define for $a < x < b$

$$g(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a) - f(x)$$

then by the Mean Value Theorem we get

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

for some $a < c < b$. Using the Mean Value Theorem on f', that is $f'(c) - f'(x) = f''(d)(c - x)$, we find

$$g'(x) = f''(d)(c - x)$$

for some d between c and x.

If $a < x < c$, then using $f''(d) > 0$ we conclude that g is strictly increasing on $[a, c]$. If $c < x < b$ we conclude that g is strictly decreasing on $[c, b]$. Since $g(a) = 0$ and $g(b) = 0$, it follows that $g(x) > 0$ when $a < x < b$, thus

$$f(x) < f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

as desired.

□

Definition (Concave). A function $f : (a, b) \to \mathbb{R}$ is concave if

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$$

for all $a \leq x, y \leq b$ and $\lambda \in (0, 1)$.

Remark. Analogues of the above proved theorems hold for concave functions in place of convex functions.

5 Functions

5.1 Inverse Functions

Let $f : [a, b] \to \mathbb{R}$ be a strictly increasing function. Assume also that f is continuous. We know from the Intermediate Value Theorem that the image of f is an interval $[\alpha, \beta]$. Furthermore, given $\alpha \leq y \leq \beta$ suppose that $f(x) = y$ and $a \leq x \leq b$. This number x is uniquely determined by y because if $x_1 < x_2$, then $f(x_1) < f(x_2)$. We define

$$g : [\alpha, \beta] \to [a, b]$$

such that $g(y) =$ unique $x \in [a, b]$ such that $f(x) = y$. Thus

$$g \circ f(x) = x \quad \text{and} \quad f \circ g(y) = y$$

We call g the inverse function of f.

Theorem 6. Let $f : [a, b] \to \mathbb{R}$ be a continuous and strictly increasing function. Then the inverse function of f is continuous and also strictly increasing.

Proof. Let g be the inverse function. It is clear that g is also strictly increasing. To be explicit, since f is strictly increasing, this means that for all $x_1, x_2 \in [a, b]$ satisfying $x_1 < x_2$, we have $f(x_1) < f(x_2)$. Choose $y_1, y_2 \in f([a, b])$, which is the domain of g, such that $y_1 < y_2$ and $y_1 = f(x_1)$ and $y_2 = f(x_2)$. By the definition of g, we see that $g(y_1) = x_1 < x_2 = g(y_2)$. So g is strictly increasing.

We must prove that g is continuous. Let $\gamma \in [\alpha, \beta]$ (notation as above).

Given $\varepsilon > 0$ and $\gamma = f(c)$ consider the closed interval $[c - \varepsilon, c + \varepsilon] \subseteq [a, b]$. Let $x_1 = c - \varepsilon$ if $a \leq c - \varepsilon$ and $x_1 = a$ otherwise. Let $x_2 = c + \varepsilon$ if $c + \varepsilon \leq b$ and $x_2 = b$ otherwise. Then $f(x_1) \leq f(x_2)$.

We may assume $a < b$. We select δ to be the minimum of

$$f(x_2) - f(c) \quad \text{and} \quad f(c) - f(x_1)$$

except when the minimum is 0. Suppose first that $\delta > 0$. If $|y - \gamma| < \delta$, then there is the unique x such that $y = f(x)$ and $x_1 < x < x_2$ and we have

$$|g(y) - c| < \varepsilon$$

If the minimum is $\delta = 0$, then either $a = c$ or $b = c$, that is c is an endpoint. Say $c = a$. In this case we disregard x_1 and let $\delta = f(x_2) - f(c)$. The same argument works. If $c = b$, we let $\delta = f(c) - f(x_1)$.

□

Theorem 7. Let $f : [a, b] \to \mathbb{R}$ be continuous and $a < b$. Assume that f is differentiable on (a, b) and $f(x) > 0$ for $x \in (a, b)$. Then the inverse function g of f defined on $[\alpha, \beta]$ is differentiable on the interval $\alpha < y < \beta$ and

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{f'(g(y))}$$

Proof. Let $\alpha < y_0 < \beta$ and $y_0 = f(x_0)$ and $y = f(x)$. Then

$$\frac{g(y) - g(y_0)}{y - y_0} = \frac{x - x_0}{f(x) - f(x_0)} = \frac{1}{\frac{f(x)-f(x_0)}{x-x_0}} \xrightarrow[y \to y_0]{} \frac{1}{f'(x_0)} = \frac{1}{f'(g(y_0))}$$

Note that $y \to y_0 \implies x \to x_0$ since g is continuous.

□

5.2 The Exponential Function

We define

$$E(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}, \qquad z \in \mathbb{C}$$

Observe that

$$|E(z)| = \left| \sum_{n=0}^{\infty} \frac{z^n}{n!} \right| \leq \sum_{n=0}^{\infty} \frac{|z|^n}{n!} < \infty$$

Thus, the ratio test shows that the series (to the right of the less than or equals to sign) converges for every $z \in \mathbb{C}$. So $E(z)$ converges by comparison. Moreover, $E(z)$ is defined by the absolutely convergent series.
Recall that if

(a) $\sum_{n=0}^{\infty} a_n$ converges absolutely

(b) $\sum_{n=0}^{\infty} a_n = A$

(c) $\sum_{n=0}^{\infty} b_n = B$

(d) $c_n = \sum_{k=0}^{n} a_k b_{n-k}$ with $n = 0, 1, 2, \ldots$

Then $\sum_{n=0}^{\infty} c_n = A \cdot B$.
Applying this result to absolutely convergent series $E(z)$ and $E(w)$, we obtain $E(z)E(w) = E(z + w)$ for any $z, w \in \mathbb{C}$.
Indeed

$$\begin{aligned} E(z)E(w) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{m=0}^{\infty} \frac{w^m}{m!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{z^k w^{n-k}}{k!(n-k)!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} z^k w^{n-k} \\ &= \sum_{n=0}^{\infty} \frac{(z+w)^n}{n!} = E(z + w) \end{aligned}$$

(2.1) This gives that $E(z+w) = E(z)E(w)$ for $z, w \in \mathbb{C}$. As a consequence we obtain for $z \in \mathbb{C}$

$$E(z)E(-z) = E(z-z) = E(0) = 1 \tag{2.2}$$

This shows that $E(z) \neq 0$ for all $z \in \mathbb{C}$.
We also have $E(x) > 0$ if $x > 0$ since $E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ and by (2.2) we conclude that

$$E(x) > 0 \quad \text{for all } x \in \mathbb{R}$$

It is easy to see that $\lim\limits_{x\to+\infty} E(x) = +\infty$ since $E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ thus by (2.2) we have $\lim\limits_{x\to-\infty} E(x) = 0$. This result follows after rearranging (2.2):

$$E(-x) = \frac{1}{E(x)}$$

Thus we see that $\frac{1}{E(x)} \xrightarrow[x\to\infty]{} 0$.
If $0 < x < y$, then $E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} < \sum_{n=0}^{\infty} \frac{y^n}{n!} = E(y)$. Since $E(x)E(-x) = 1$, thus

$$E(x) < E(y), \quad E(y)E(-x) > 1$$
$$E(-x) > \frac{1}{E(y)} = E(-y)$$

hence $E(x)$ is strictly increasing on the whole of $\mathbb{R}$.
The addition formula also shows

$$\lim_{h\to 0} \frac{E(z+h)-E(z)}{h} = E(z) \cdot \underbrace{\lim_{h\to 0} \frac{E(h)-1}{h}}_{=1} = E(z)$$

To prove the identity we used above, observe that

$$\begin{aligned} \frac{1}{h}(E(h)-1) &= \frac{1}{h}\sum_{n=1}^{\infty} \frac{h^n}{n!} \\ &= \frac{1}{1!} + \frac{h}{2!} + \frac{h^2}{3!} + \frac{h^3}{4!} + \dots \\ &= \sum_{n=1}^{\infty} \frac{h^{n-1}}{n!} \end{aligned}$$

thus

$$\begin{aligned} \left|\frac{1}{h}(E(h)-1) - 1\right| &\leq \left|h \sum_{n=2}^{\infty} \frac{h^{n-2}}{n!}\right| \\ &\leq |h| \sum_{n=2}^{\infty} \frac{|h|^{n-2}}{(n-2)!} \\ &\leq |h| E(|h|) \\ &\overset{(|h|\leq 1)}{\leq} |h| e \xrightarrow[|h|\to 0]{} 0 \end{aligned}$$

In particular, if $x \in \mathbb{R}$ we obtain

$$E'(x) = E(x)$$

thus $E(x)$ is continuous on $\mathbb{R}$ (because it is differentiable everywhere on $\mathbb{R}$).
Since $E(z + w) = E(z)E(w)$ then

$$E(z_1 + \ldots + z_n) = E(z_1) \cdots E(z_n)$$

for $z_1, \ldots, z_n \in \mathbb{C}$.
Taking $z_1 = \cdots = z_n = 1$ we see

$$E(n) = e^n \quad \text{since } E(1) = e$$

If $p = \frac{n}{m}$ and $m, n \in \mathbb{Z}_+$, then

$$E(p)^m = E(mp) = E(n) = e^n$$

thus $E(p) = e^p$ for all $p \in \mathbb{Q}$ and $p > 0$. We also have $E(-p) = e^{-p}$ for all $p \in \mathbb{Q}$ and $p > 0$.
If $x > 1$ and $y > 0$ and $y \in \mathbb{R}$, then

$$x^y = \sup\{x^p : p \in \mathbb{Q} \text{ and } p < y\}$$

If we thus define

$$e^x = \sup\{e^p : p \in \mathbb{Q} \text{ and } p < x\}$$

the continuity and monotonicity of E and the fact that

$$E(p) = e^p \quad \text{for all } p \in \mathbb{Q}$$

allows us to conclude that

$$E(x) = e^x \quad \text{for all } x \in \mathbb{R}$$

the last equation is why E is called the exponential function. Note also that

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

Sometimes we will write $\exp(x)$ instead of e^x. From now on,

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

We now summarize what we have proved so far below

Theorem (Rudin 8.6).

Let $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. Then

(a) e^x is continuous and differentiable for all $x \in \mathbb{R}$.

(b) $(e^x)' = e^x$

(c) e^x is strictly increasing on $\mathbb{R}$ and $e^x > 0$ for all $x \in \mathbb{R}$.

(d) $e^{x+y} = e^x e^y$

(e) $\lim_{x \to +\infty} e^x = +\infty$ and $\lim_{x \to -\infty} e^x = 0$.

(f) $\lim_{x \to +\infty} x^n e^{-x} = 0$ for all $n \in \mathbb{Z}_+$.

Proof. We have proven (a) - (e) above. So, we only need to prove (f) to finish.
Observe that

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} > \frac{x^{n+1}}{(n+1)!} \qquad \text{for } x > 0$$

so that

$$x^n e^{-x} < \frac{(n+1)!}{x} \xrightarrow[x \to +\infty]{} 0$$

$\square$

In other words (f) shows that e^x tends to $+\infty$ faster than any polynomial. If $P(x) = \sum_{k=0}^{n} c_k x^k$ where $c_0, c_1, \ldots, c_n \in \mathbb{R}$ and $x > 0$, then

$$0 \leq \left| \frac{P(x)}{e^x} \right| \leq \frac{\sum_{k=0}^{n} |c_k| x^k}{e^x}$$
$$\overset{|x| \geq 1}{\leq} \frac{\left(\sum_{k=0}^{n} |c_k|\right) x^n}{e^x} \xrightarrow[x \to +\infty]{} 0$$

5.3 The Logarithmic Function

Since E is strictly increasing and differentiable on $\mathbb{R}$ it has inverse function L, which is also strictly increasing and differentiable and whose domain $E[\mathbb{R}] = (0, \infty)$. L is defined by

$$E(L(y)) = y \qquad \text{for all } y > 0$$

or equivalently $L(E(x)) = x$ for all $x \in \mathbb{R}$.
Differentiating the latter equation

$$1 = (x)' = (L(E(x))' = L'(E(x)) \cdot (E(x))' = L'(E(x)) \cdot E(x)$$

thus

$$L'(E(x)) = \frac{1}{E(x)}$$

hence

$$L'(y) = \frac{1}{y} \quad \text{for all } y > 0$$

Writing $u = E(x)$ and $v = E(y)$, observe that

$$\begin{aligned} L(uv) &= L(E(x) \cdot E(y)) = L(E(x+y)) \\ &= x + y = L(u) + L(v) \quad \text{for } u, v > 0 \end{aligned}$$

From now on we will write $\log(x) = L(x)$. Since $\lim_{x\to+\infty} e^x = +\infty$ and $\lim_{x\to-\infty} e^x = 0$, we conclude that

$$\log(x) \xrightarrow[x\to+\infty]{} +\infty \quad \text{and} \quad \log(x) \xrightarrow[x\to 0]{} -\infty$$

It is also easily seen that $x = E(L(x))$ implies

$$x^n = E(nL(x)) \quad \text{if } x > 0 \text{ and } n \in \mathbb{Z}_+$$

Similarly if $m \in \mathbb{Z}_+$ we have

$$x^{\frac{1}{m}} = E\left(\frac{1}{m}L(x)\right)$$

Thus

$$x^\alpha = E(\alpha L(x)) = e^{\alpha \log(x)} \tag{7.1}$$

for any rational $\alpha \in \mathbb{Q}$.

We now define x^α for any real α and $x > 0$, by (7.1) the continuity and monotonicity of E and L show that everything makes sense, and this definition coincides with

$$x^\alpha = \sup\{x^p : p < \alpha, \ \ p \in \mathbb{Q}\}$$

if $\alpha \in \mathbb{R}$ and $x > 1$.

If we differentiate $x^\alpha = E(\alpha L(x))$, then

$$(x^\alpha)' = E(\alpha L(x)) \cdot \frac{\alpha}{x} = \alpha x^{\alpha-1}$$

Finally, we show that

$$\lim_{x\to+\infty} x^{-\alpha}\log(x) = 0$$

for every $\alpha > 0$. That is, $\log x \xrightarrow[x\to+\infty]{} +\infty$ slower than any power of x. Notice that

$$\begin{aligned} \lim_{x\to+\infty} \frac{\log(x)}{x^\alpha} &= \lim_{x\to+\infty} \frac{\frac{1}{x}}{\alpha x^{\alpha-1}} \\ &= \lim_{x\to+\infty} \frac{1}{\alpha x^\alpha} = 0 \end{aligned}$$

since $x^\alpha \xrightarrow[x\to+\infty]{} +\infty$ and by L'Hospital's rule.

Example. Let

$$a_n = \sum_{k=1}^{n-1} \frac{1}{k} - \log(n), \quad n \in \mathbb{Z}_+$$

and

$$b_n = \sum_{k=1}^{n} \frac{1}{k} - \log(n), \quad n \in \mathbb{Z}_+$$

We shall show that

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \gamma$$

The number γ is known as the Euler-Mascheroni constant.

We proved some time ago (see the section on Euler's number) that

$$\left(1+\frac{1}{n}\right)^n < e < \left(1+\frac{1}{n}\right)^{n+1}$$

$$n \log\left(1+\frac{1}{n}\right) < 1 < (n+1)\log\left(1+\frac{1}{n}\right)$$

Observe now that $\log\left(\frac{n+1}{n}\right) < \frac{1}{n}$ and $\frac{1}{n+1} < \log\left(\frac{n+1}{n}\right)$. Thus

$$\begin{aligned} a_{n+1} - a_n &= \sum_{k=1}^{n} \frac{1}{k} - \log(n+1) - \sum_{k=1}^{n-1} \frac{1}{k} + \log(n) \\ &= \frac{1}{n} - \log\left(\frac{n+1}{n}\right) > 0 \end{aligned}$$

and similarly,

$$b_{n+1} - b_n = \frac{1}{n+1} - \log\left(\frac{n+1}{n}\right) < 0$$

Clearly,

$$a_1 \le a_n < b_n \le b_1$$

Since every bounded and monotonic sequence has a limit, let γ be a limit for $(a_n)_{n\in\mathbb{N}}$

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \left(\sum_{k=1}^{n-1} \frac{1}{k} - \log(n)\right) = \gamma$$

and

$$a_n = b_n - \frac{1}{n}$$

Thus $\lim_{n\to\infty} b_n = \gamma$ as claimed.

Remark. It is not even known whether γ is irrational. Again, γ is called the Euler-Mascheroni constant and $\gamma \approx .5772\ldots$

Now we consider $\log(1+x)$ as a Taylor series.

$$\log(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k \qquad \text{for } |x| < 1$$

Recall (see Pg. 77) that if $f : [a,b] \to \mathbb{R}$ is n-times continuously differentiable on $[a,b]$ and $f^{(n+1)}$ exists in the open interval (a,b).

(a) Then for any $x, x_0 \in [a,b]$ there exists $\theta \in (0,1)$ such that

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + r_n(x)$$

where we have the Lagrange remainder

$$r_n(x) = \frac{f^{(n+1)}(x_0 + \theta(x-x_0))}{(n+1)!}(x-x_0)^{n+1}$$

(b) Then for every $x, x_0 \in [a,b]$ there exists $\theta \in (0,1)$ such that

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + r_n(x)$$

where we have the Cauchy remainder

$$r_n(x) = \frac{f^{(n+1)}(x_0 + \theta(x-x_0))}{n!}(1-\theta)^n(x-x_0)^{n+1}$$

Observe that

$$(\log(1+x))' = \frac{1}{1+x}$$
$$(\log(1+x))'' = \left(\frac{1}{1+x}\right)' = -\frac{1}{(1+x)^2}$$
$$(\log(1+x))''' = \left(-\frac{1}{(1+x)^2}\right)' = \frac{2}{(1+x)^3}$$

Inductively

$$(\log(1+x))^{(n)} = (-1)^{n+1} \cdot \frac{(n-1)!}{(1+x)^n} \qquad (*)$$

If $0 \leq x < q$, then we use the Taylor's expansion with Lagrange's remainder at $x_0 = 0$. Then

$$\log(1+x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k + r_n(x)$$
$$f^{(0)}(0) = \log(1) = 0$$
$$f^{(n)}(0) = (-1)^{n+1}\frac{(n-1)!}{1^n} = (-1)^{n+1}(n-1)!$$

thus

$$\log(1+x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k + r_n(x)$$
$$= \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} x^k + r_n(x)$$

where $0 \leq x < 1$, $x_0 = 0$, and $0 < \theta < 1$.
Further

$$|r_n(x)| = \left| \frac{f^{(n+1)}(\theta x)}{(n+1)!} \cdot x^{n+1} \right|$$
$$\overset{\text{by } (*)}{=} \frac{n!}{(n+1)!(1+\theta x)^n} \cdot x^{n+1} \leq \frac{1}{n+1} \xrightarrow[n\to\infty]{} 0$$

If $-1 < x < 0$, we shall use Taylor's expansion at $x_0 = 0$ with the Cauchy remainder. Then

$$\log(1+x) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} x^k + r_n(x)$$

where

$$r_n(x) = \frac{f^{(n+1)}(\theta x)}{n!} \cdot (1-\theta)^n \cdot x^{n+1}$$

Thus

$$|r_n(x)| = \left| \frac{n!}{n!(1+\theta x)^{n+1}} \cdot (1-\theta)^n \cdot x^{n+1} \right|$$

We have $-1 < \theta x < 0$, which means $-\theta < \theta x$, thus $1 - \theta < 1 + \theta x$. Hence

$$|r_n(x)| \leq \frac{(1-\theta)^n}{(1+\theta x)^{n+1}} \cdot |x|^{n+1}$$
$$\leq \frac{(1-\theta)^n}{(1-\theta)^{n+1}} \cdot |x|^{n+1}$$
$$= \frac{|x|^{n+1}}{1-\theta} \xrightarrow[n\to\infty]{} 0$$

since $|x|^n \xrightarrow[n\to\infty]{} 0$ as $|x| < 1$.

5.4 Some Useful Formulas

5.4.1 Newton's Binomial Formula

Recall that for $n \in \mathbb{N}$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

We now show that if $\alpha \in \mathbb{R} \setminus \mathbb{N}$ and $|x| < 1$, then

$$(1+x)^\alpha = 1 + \sum_{n=1}^{\infty} \binom{\alpha}{n} x^n$$

This is called Newton's binomial formula.
Proof. Suppose that $x > 0$ and let $f(x) = (1+x)^\alpha$. Notice that

$$f^{(n)}(x) = \alpha(\alpha-1)\cdots(\alpha-n+1)(1+x)^{\alpha-n}$$

Using Lagrange form for the remainder in Taylor's formula of f and with $x_0 = 0$, we have

$$\begin{aligned} r_n(x) &= \frac{f^{(n+1)}(x_0+\theta(x-x_0))}{(n+1)!}(x-x_0)^{n+1} \\ &= \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{(n+1)!} \cdot x^{n+1} \cdot (1+\theta x)^{\alpha-n-1} \end{aligned}$$

For $|x| < 1$ we have

$$\lim_{n\to\infty} \frac{\alpha(\alpha-1)\cdots(\alpha-n)}{(n+1)!} \cdot x^{n+1} = 0$$

Indeed, we can use the fact that if $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = q$ and $q < 1$, then $\lim_{n\to\infty} a_n = 0$.
Let

$$a_n = \frac{\alpha(\alpha-1)\cdots(\alpha-n)}{(n+1)!} \cdot x^{n+1}$$

then

$$\begin{aligned} \left|\frac{a_{n+1}}{a_n}\right| &= \left|\frac{\alpha(\alpha-1)\cdots(\alpha-n-1)}{(n+1)!} \cdot x^{n+2} \cdot \frac{(n+1)!}{\alpha(\alpha-1)\cdots(\alpha-n)x^{n+1}}\right| \\ &= \left|\frac{\alpha-n-1}{n+2}x\right| \xrightarrow[n\to\infty]{} |x| < 1 \end{aligned}$$

Thus $\lim\limits_{n\to\infty} r_n(x) = 0$ if we show that $(1+\theta x)^{\alpha-n-1}$ is bounded. $(1+\theta x)^{-n} \leq 1$ and
For $\alpha \geq 0$, we have

$$1 \leq (1+\theta x)^\alpha \leq (1+x)^\alpha \leq 2^\alpha$$

.
For $\alpha < 0$, we have

$$2^\alpha \leq (1+x)^\alpha \underset{\theta<1}{\leq} (1+\theta x)^\alpha \leq 1$$

and we are done.
We now assume $-1 < x < 0$. The Cauchy form for the remainder in Taylor's formula of f is

$$r_n(x) = \frac{\alpha(\alpha-1)\cdots(\alpha-n)}{(n+1)!} \cdot x^{n+1}(1-\theta)^n(1+\theta x)^{\alpha-n-1}$$

We shall show that $(1-\theta)^n(1+\theta x)^{\alpha-n-1}$ is bounded. Since $x \in (-1,0)$, then

$$(1-\theta)^n \leq (1-\theta)^n(1+\theta x)^{-n} = \frac{(1-\theta)^n}{(1+\theta x)^n} < 1$$

For $\alpha \leq 1$, we have

$$1 \leq (1+\theta x)^{\alpha-1} \leq (1+x)^{\alpha-1}$$

For $\alpha \geq 1$, we have

$$(1+x)^{\alpha-1} \leq (1+\theta x)^{\alpha-1} \leq 1$$

and we are done.

□

Example. A function which is differentiable infinitely many times, but cannot be expressed as a Taylor series.

$$f(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

It is easy to see that f is infinitely many times differentiable at any $x \in \mathbb{R}$. Moreover

$$f^{(n)}(0) = 0 \qquad \text{for any } n \geq 0$$

and $f(x) \neq 0$, that is, $f(x)$ is not the zero function. However, we have

$$0 \neq f(x) \neq \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = 0$$

where the sum is 0 because the coefficients $f^{(n)}(0)$ are all 0.

5.4.2 Bernoulli's Inequality for General Exponents

For $x > -1$ and $x \neq 0$ we have

(a) $(1+x)^\alpha > 1 + \alpha x$ if $\alpha > 1$ or $\alpha < 0$

(b) $(1+x)^\alpha < 1 + \alpha x$ if $0 < \alpha < 1$

Proof. Applying Taylor's formula with the Lagrange form for

$$f(x) = (1+x)^\alpha$$

we get

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)(1+\theta x)^{\alpha-2}}{2} \cdot x^2 = \begin{cases} > (1+\alpha x) & \text{if } \alpha > 1 \text{ or } \alpha < 0 \\ < (1+\alpha x) & \text{if } 0 < \alpha < 1 \end{cases}$$

with some $0 < \theta < 1$. We now observe that

$$\frac{\alpha(\alpha-1)(1+\theta x)^{\alpha-2}}{2} > 0 \qquad \text{for } \alpha > 1 \text{ or } \alpha < 0$$

and

$$\frac{\alpha(\alpha-1)(1+\theta x)^{\alpha-2}}{2} < 0 \qquad \text{for } 0 < \alpha < 1$$

□

5.4.3 The Weighted AM-GM Inequality

If $x_1, \ldots, x_k > 0$ and $\alpha_1, \ldots, \alpha_k > 0$ and

$$\sum_{i=1}^{k} \alpha_i = 1$$

then

$$x_1^{\alpha_1} \cdots x_k^{\alpha_k} \leq \alpha_1 x_1 + \ldots + \alpha_k x_k \tag{18.1}$$

Proof. Let $f(x) = \log(x)$. Observe that $f'(x) = \frac{1}{x}$ and $f''(x) = -\frac{1}{x^2}$. Thus $f''(x) < 0$ for all $x > 0$, which means that f is concave by Theorem 5. In other words, for all $x_1, \ldots, x_n > 0$ and $\alpha_1 + \ldots + \alpha_n = 1$ with $\alpha_1, \ldots, \alpha_n > 0$ we have

$$f(\alpha_1 x_1 + \ldots + \alpha_n x_n) \geq \alpha_1 f(x_1) + \ldots + \alpha_n f(x_n)$$

to prove (18.1) consider

$$\begin{aligned} \log(x_1^{\alpha_1} \cdots x_k^{\alpha_k}) &= \sum_{j=1}^{k} \alpha_j \log(x_j) \\ &\leq \log\left(\sum_{j=1}^{k} \alpha_j x_j\right) \end{aligned}$$

thus we are done since $\log x$ is concave on $(0, \infty)$.

□

Corollary. If $\frac{1}{p} + \frac{1}{q} = 1$ with $p, q > 0$ and $x, y > 0$, then

$$xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q \tag{$*$}$$

It suffices to apply the previous result with $\alpha_1 = \frac{1}{p}$, $\alpha_2 = \frac{1}{q}$, $x_1 = x^p$, and $x_2 = y^q$. Observe that

$$xy = x_1^{\frac{1}{p}} x_2^{\frac{1}{q}} \leq \frac{1}{p}x_1 + \frac{1}{q}x_2 = \frac{1}{p}x^p + \frac{1}{q}y^q$$

We now use this inequality to prove Hölder's inequality.

5.4.4 Hölder's Inequality

If $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\sum_{j=1}^{n} |x_j y_j| \leq \left(\sum_{j=1}^{n} |x_j|^p\right)^{\frac{1}{p}} \left(\sum_{j=1}^{n} |y_j|^q\right)^{\frac{1}{q}}$$

Proof. Set

$$s_j = \frac{|x_j|}{\left(\sum_{j=1}^{n} |x_j|^p\right)^{\frac{1}{p}}} \quad \text{and} \quad t_j = \frac{|y_j|}{\left(\sum_{j=1}^{n} |y_j|^q\right)^{\frac{1}{q}}}$$

Then it suffices to prove

$$\sum_{j=1}^{n} s_j t_j = 1$$

By our hypotheses, we see that $(*)$ from the corollary above is valid. Thus by the corollary, we have

$$s_j t_j \leq \frac{1}{p} s_j^p + \frac{1}{q} t_j^q$$

for all $j \in \{1, 2, \ldots, n\}$. Observe that by adding the above inequalities we have

$$\begin{aligned} \sum_{j=1}^{n} s_j t_j &\leq \frac{1}{p} \sum_{j=1}^{n} s_j^p + \frac{1}{q} \sum_{j=1}^{n} t_j^q \\ &= \frac{1}{p} \sum_{j=1}^{n} \cdot \frac{|x_j|^p}{\left(\sum_{j=1}^{n} |x_j|^p\right)} + \frac{1}{q} \cdot \sum_{j=1}^{n} \frac{|y_j|^q}{\left(\sum_{j=1}^{n} |y_j|^q\right)} \\ &= \frac{1}{p} + \frac{1}{q} = 1 \end{aligned}$$

We now use Hölder's inequality to prove Minkowski's inequality.

5.4.5 Minkowski's Inequality

If $p \geq 1$, then

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}}$$

Indeed, if $p = 1$, there is nothing to prove, it is just the triangle inequality.
So, assume $p > 1$. To start, notice that

$$\begin{aligned} \frac{1}{p} + \frac{1}{q} = 1 &\implies \frac{1}{q} = 1 - \frac{1}{p} \\ \frac{1}{q} = \frac{p-1}{p} &\implies q = \frac{p}{p-1} \end{aligned}$$

Now, consider that

$$\begin{aligned}\sum_{i=1}^{n}|x_i+y_i|^p &= \sum_{i=1}^{n}|x_i+y_i|^1|x_i+y_i|^{p-1}\\ &\leq \left(\sum_{i=1}^{n}|x_i||x_i+y_i|^{p-1}\right)+\left(\sum_{i=1}^{n}|y_i||x_i+y_i|^{p-1}\right)\end{aligned}$$

We apply Hölder's inequality and obtain

$$\begin{aligned}\sum_{i=1}^{n}|x_i||x_i+y_i|^{p-1} &\leq \left(\sum_{i=1}^{n}|x_j|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{n}(|x_i+y_i|^{p-1})^q\right)^{\frac{1}{q}}\\ &= \left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{n}|x_i+y_i|^p\right)^{1-\frac{1}{p}}\end{aligned}$$

Similarly,

$$\sum_{i=1}^{n}|y_i||x_i+y_i|^{p-1} \leq \left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^{n}|x_i+y_i|^p\right)^{1-\frac{1}{p}}$$

Thus, we conclude that

$$\left(\sum_{i=1}^{n}|x_i+y_i|^p\right) \leq \underbrace{\left(\sum_{i=1}^{n}|x_i+y_i|^p\right)^{1-\frac{1}{p}}}_{(1)}\left[\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^{n}|y_i|^p\right)^{\frac{1}{p}}\right]$$

Dividing both sides by (1), we are done.

□

Consider the pair $(\mathbb{R}^n, \rho_q)$ for $1 \leq q \leq \infty$ with the following definition

$$\begin{aligned}\rho_q(x,y) &= \left(\sum_{j=1}^{n}|x_j-y_j|^q\right)^{\frac{1}{q}} \quad 1 \leq q < \infty\\ \rho_\infty(x,y) &= \sup_{1\leq j\leq n}|x_j-y_j|\end{aligned}$$

The function $(x,y) \mapsto \rho_q(x,y)$ is a metric for any $1 \leq q \leq \infty$. For instance, if $q=2$, we have the Euclidean metric. On the other hand, if $q=\infty$, then we have cubes as our open sets.

5.5 The Trigonometric Functions

Recall that $i^2=-1$. Let us define

$$\begin{aligned}C(x) &= \frac{1}{2}\cdot(E(ix)+E(-ix)) \text{ and}\\ S(x) &= \frac{1}{2i}\cdot(E(ix)-E(-ix))\end{aligned}$$

We shall show that $C(x)$ and $S(x)$ coincide with the functions $\cos(x)$ and $\sin(x)$, whose definition is usually based on geometric considerations.
Since $E(x) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$, thus $E(\overline{z}) = \overline{E(z)}$. To be more explicit, since $E(x)$ is a power series, it is a linear combination of powers of z. By Theorem 1.31 (a) and (b) in Rudin, recall that addition and multiplication are equivalent regardless of the order in which the operation of conjugation is carried out with respect to that addition and multiplication. In other words, if $z, w \in \mathbb{C}$, we have

$$\overline{z+w} = \overline{z} + \overline{w} \quad \text{and} \quad \overline{zw} = \overline{z} \cdot \overline{w}$$

So it follows that $E(\overline{z}) = \overline{E(z)}$. Hence, we have $\overline{C(x)} = C(x)$ and $\overline{S(x)} = S(x)$, so $C(x)$ and $S(x)$ are real for $x \in \mathbb{R}$. Recall that a number $z \in \mathbb{C}$ is real if $z = \overline{z}$, and the outputs of the functions $C(x)$ and $S(x)$ satisfy this relation.
Also

$$E(ix) = C(x) + iS(x)$$

thus $C(x)$ and $S(x)$ are the real and imaginary parts, respectively, of $E(ix)$ if $x \in \mathbb{R}$.
Moreover

$$|E(ix)|^2 = E(ix)\overline{E(ix)} = E(ix)E(-ix) = 1$$

so $|E(ix)| = 1$ for any $x \in \mathbb{R}$.
Since

$$C(x) = \frac{1}{2}(E(ix) + E(-ix)) \quad \text{and} \quad S(x) = \frac{1}{2i}(E(ix) - E(-ix))$$

we can read off that $C(0) \neq 1$ and $S(0) = 0$ and also

$$C'(x) = -S(x) \quad \text{and} \quad S'(x) = C(x)$$

Since

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

We assert that there exist positive numbers x such that $C(x) = 0$. If not, since $C(0) = 1$, it follows that $C(x) > 0$ for all $x > 0$. This is because of the Intermediate Value Theorem (How do we know $C(x)$ is continuous?), if there is some $C(x_1) < 0$, then there must be some value at which $C(x_2) = 0$ for $0 < x_2 < x_1$, a contradiction. Hence $S'(x) > 0$ since

$$S'(x) = C(x)$$

But $S(0) = 0$, thus $S(x)$ is strictly increasing on $(0, \infty)$. By the Mean Value Theorem

$$C(y) - C(x) = -(y - x) \cdot S(\theta_{x,y}) \quad \text{for some } \theta_{x,y} \in (x, y)$$

thus

$$C(x) - C(y) = (y - x) \cdot S(\theta_{x,y}) > (y - x)S(x)$$

and we conclude

$$(y-x)S(x) \le C(x) - C(y) \le 2$$

Note that we are using 2 because $C(x)$ is defined by the exponential function taking values only in the unit circle, and the maximum of any difference of values on the unit circle is 2. But this is impossible if y is large since $S(x) > 0$ ($\lightning$).
Let x_0 be the smallest positive number such that $C(x_0) = 0$. This exists since the set of zeroes of a continuous function is closed and $C(0) \neq 0$.
We define the number π to be

$$\pi = 2x_0$$

Then $C(\frac{\pi}{2}) = 0$ and since $|E(ix)| = 1$ we deduce

$$S\left(\frac{\pi}{2}\right) = \pm 1$$

Since $C(x) > 0$ in $(0, \frac{\pi}{2})$, S is increasing in $(0, \frac{\pi}{2})$. Hence $S(\frac{\pi}{2}) = 1$. Thus we have

$$E\left(\frac{\pi i}{2}\right) = i$$

Since $E(z+w) = E(z)E(w)$ thus we have

$$E(\pi i) = -1 \quad \text{and} \quad E(2\pi i) = 1$$

hence $E(z + 2\pi i) = E(z)$ for all $z \in \mathbb{C}$.
Theorem (Rudin 8.7).

(a) The function E is periodic with period $2\pi i$.

(b) The functions S and C are periodic with period 2π.

(c) If $0 < t < 2\pi$, then $E(it) \neq 1$.

(d) If $z \in \mathbb{C}$ and $|z| = 1$, there is a unique $t \in (0, 2\pi)$ such that $E(it) = z$

Proof. (a) easily follows since $E(z + 2\pi i) = E(z)$ for all $z \in \mathbb{C}$.
Since $C(x) = \frac{1}{2}(E(ix) + E(-ix))$ we see

$$\begin{aligned} C(x + 2\pi) &= \frac{1}{2}(E(i(x+2\pi)) + E(-i(x+2\pi))) \\ &= \frac{1}{2}(E(ix) + E(-ix)) = C(x) \end{aligned}$$

Similarly, $S(x) = S(x + 2\pi)$. Thus (b) is proved.
To prove (c) suppose $0 < t < \frac{\pi}{2}$ and

$$E(it) = x + iy \quad \text{with } x, y \in \mathbb{R}$$

Our preceding discussion shows that $0 < x < 1$ and $0 < y < 1$. To be more explicit, set $x = C(t)$ and $y = S(t)$, what we know is

$$|E(it)| = |C(t) + iS(t)| = 1$$

We know that S is increasing on $0 < t < \frac{\pi}{2}$, $S(0) = 0$, and $S(\frac{\pi}{2}) = 1$. This shows that $0 < y < 1$. Once we interpret the absolute value, we see that $C(t)^2 + S(t)^2 = 1$, and this relation shows us that $0 < x < 1$.
Now note that $0 < t < \frac{\pi}{2} \iff 0 < 4t < 2\pi$ and that

$$E(4it) = E(it)^4 = (x+iy)^4 = x^4 - 6x^2y^2 + y^4 + 4ixy(x^2 - y^2)$$

If $E(4it) \in \mathbb{R}$, then it follows that $x^2 - y^2 = 0$. We are only considering real $E(4it)$ because if $E(4it)$ is imaginary, then it clearly is not equal to 1. Since

$$|E(is)| = 1 \quad \text{for all } s \in \mathbb{R}$$

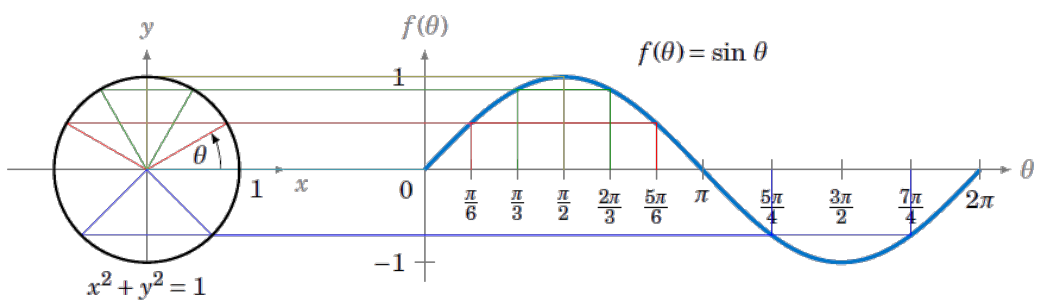
Equivalently, $x^2 + y^2 = 1$ so we have $x^2 = y^2 = \frac{1}{2}$. Hence $E(4it) = -1$ and we are done.
To prove (d), if $0 \leq t_1 < t_2 < 2\pi$, then

$$E(it_2)E(it_1)^{-1} = E(i(t_2 - t_1)) \neq 1$$

thus we obtain uniqueness.
To prove the existence we fix $z \in \mathbb{C}$ so that $|z| = 1$. Write $z = x + iy$ with $x, y \in \mathbb{R}$. We will consider cases. Suppose first that $x \geq 0$ and $y \geq 0$.
On $[0, \frac{\pi}{2}]$, $C(t)$ decreases from 1 to 0. Hence $C(t) = x$ for some $t \in [0, \frac{\pi}{2}]$.
Since $C^2 + S^2 = 1$ and $S \geq 0$ on $[0, \frac{\pi}{2}]$, it follows that $z = E(it)$.
If $x < 0$ and $y \geq 0$, the preceding conditions are satisfied by $-iz$. Hence $-iz = E(it)$ for some $t \in [0, \frac{\pi}{2}]$.
Since $i = E\left(\frac{\pi i}{2}\right)$ we obtain

$$z = E\left(i\left(t + \frac{\pi}{2}\right)\right)$$

Finally, if $y < 0$, the preceding two cases show that $z = E(it)$ for some $t \in (0, \pi)$. Thus, $z = -E(it) = E(i(t+\pi))$ since $E(i\pi) = -1$.



Remark. The curve $\gamma(t) = E(it)$, with $(0 \le t \le 2\pi$, is a simple closed curve chose range is the unit circle in the plane. Since $\gamma'(t) = iE(it)$, the length of γ as well shall see in a few lectures is

$$\int_0^{2\pi} |\gamma'(t)| dt = 2\pi$$

This is of course the expected result of the circumference of a circle of radius 1.
In the same way we see that the point $\gamma(t)$ describes a circular arc length t_0 as t increases from 0 to t_0.
Considerations of the triangle whose vertices are

$$z_1 = 0 \quad z_2 = \gamma(t_0) \quad z_3 = C(t_0)$$

show that $\cos(t) = C(t)$ and $\sin(t) = S(t)$

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$$
$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}$$

Further, we claim $\lim\limits_{x \to 0} \frac{\sin x}{x} = 1$. Notice that

$$\begin{aligned} \left| \frac{\sin x}{x} - 1 \right| &= \left| \sum_{k=1}^{\infty} \frac{(-1)^k x^{2k}}{(2k+1)!} \right| \\ &\le \sum_{k=1}^{\infty} \frac{|x|^{2k}}{(2k+1)!} = |x^2| \sum_{k=1}^{\infty} \frac{|x|^{2k-2}}{(2k+1)!} \\ &\le |x|^2 e^{|x|} \overset{|x| \le 1}{\le} e|x|^2 \xrightarrow[x \to 0]{} 0 \end{aligned}$$

which proves the claim.

5.6 Fundamental Theorem of Algebra

Theorem (Rudin 8.8). Suppose $a_0, a_1, \ldots, a_n \in \mathbb{C}$ and $a_n \neq 0$ with $n \in \mathbb{Z}_+$. Let

$$P(z) = \sum_{k=0}^{n} a_k z^k$$

then $P(z) = 0$ for some complex number $z \in \mathbb{C}$.
Proof. Without loss of generality, we assume that $a_n = 1$ and let

$$\mu = \inf_{z \in \mathbb{C}} |P(z)|$$

If $|z| = R$, then by the reverse triangle inequality we have

$$\begin{aligned}|P(z)| &\geq ||a_n z^n| - |a_{n-1}z^{n-1}| - \ldots - |a_0|| \\ &= R^n \cdot (1 - |a_{n-1}|R^{-1} - \ldots - |a_0|R^{-n})\end{aligned}$$

The right-hand side tends to ∞ as $R \to \infty$.
Hence, by the definition of divergence, there is $R_0 > 0$ such that

$$|P(z)| > \mu \quad \text{if } |z| > R_0$$

Since $|P|$ is continuous on the closed circle with center 0 and radius R_0, we conclude by Theorem 4.16 in Rudin that there is z_0 so that $|z_0| \leq R_0$ and

$$|P(z_0)| = \mu$$

In other words, $|P(z)|$ attains its minimum value for some $z \in \mathbb{C}$ on this closed circle, and this minimum value is the infimum and thus equal to μ by definition.
We claim that $\mu = 0$. If not, set $Q(z) = \frac{P(z+z_0)}{P(z_0)}$ then Q is a nonconstant polynomial with $Q(0) = 1$ and

$$|Q(z)| \geq 1 \quad \text{for all } z \in \mathbb{C}$$

there is a smallest $k \in \mathbb{Z}$, $1 \leq k \leq n$ such that

$$Q(z) = 1 + b_k z^k + \ldots + b_n z^n, \qquad b_k \neq 0$$

By Theorem 8.7(d) in Rudin, there is $\theta \in \mathbb{R}$ such that

$$e^{ik\theta} b_k = -|b_k|$$

If $r > 0$ and $r^k|b_k| < 1$, then the equation above implies that

$$|1 + b_k r^k e^{ik\theta}| = 1 - r^k|b^k|$$

so that

$$|Q(re^{i\theta})| \leq 1 - r^k\{|b_k| - r|b_{k+1}| - \ldots - r^{n-1}|b_n|\}$$

For sufficiently small r the expression above in braces is positive. Hence $|Q(re^{i\theta})| < 1$. ($\lightning$)
Thus $\mu = 0$ and $P(z_0) = 0$.

$\square$

6 The Riemann-Stieltjes Integral

6.1 Setup

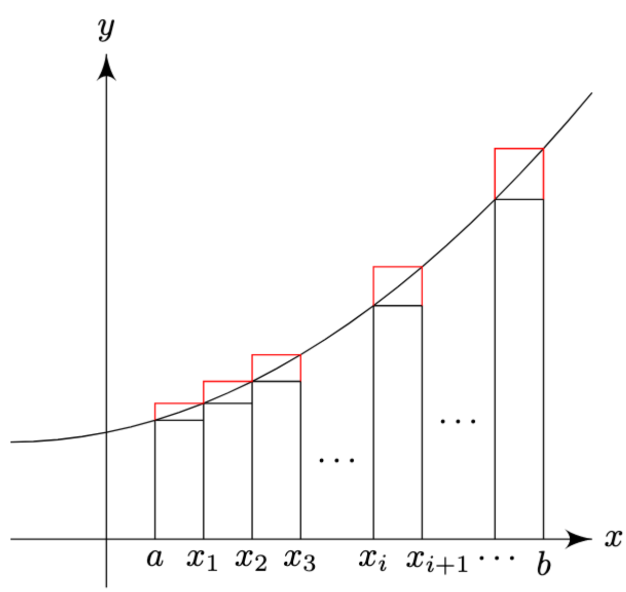
Definition (Partition). Let $[a,b]$ be a given interval. By a partition P of $[a,b]$ we mean a finite set of points $x_0, x_1, \ldots, x_n$ where

$$a = x_0 \leq x_1 \leq \ldots \leq x_{n-1} \leq x_n = b$$

We write $\Delta x_i = x_i - x_{i-1}$ for $i \in \{1, \ldots, n\}$.
Now suppose f is a bounded real function defined on $[a,b]$. Corresponding to each partition P of $[a,b]$ we set

$$M_i = \sup_{x_{i-1} \leq x \leq x_i} f(x)$$
$$m_i = \inf_{x_{i-1} \leq x \leq x_i} f(x)$$
$$U(P,f) = \sum_{i=1}^{n} M_i \Delta x_i$$
$$L(P,f) = \sum_{i=1}^{n} m_i \Delta x_i$$

Below is a picture, note that the curve does not have to be monotonically increasing and the partitions do not have to be the same size.



$U(P,f)$ is called the upper partial Riemann sum corresponding to the partition P of $[a,b]$.
$L(P,f)$ is called the lower partial Riemann sum corresponding to the partition P of $[a,b]$.

Finally, the upper and lower integrals of f are respectively denoted

$$\overline{\int_a^b} f dx = \inf U(P, f)$$

$$\underline{\int_a^b} f dx = \sup L(P, f)$$

where the inf and sup are taken over the set of all partitions P of $[a, b]$.

Definition.

- If the upper and lower integrals are equal

$$\overline{\int_a^b} f dx = \underline{\int_a^b} f dx$$

  we say that f is Riemann integrable on $[a, b]$.

- $R([a, b])$ denotes the set of Riemann integrable functions on $[a, b]$
- The common value of $\overline{\int_a^b} f dx = \underline{\int_a^b} f dx$ will be denoted by

$$\int_a^b f dx \quad \text{or} \quad \int_a^b f(x) dx$$

This is the Riemann integral of f over $[a, b]$.

Remark. Since $f : [a, b] \to \mathbb{R}$ is bounded, there are two numbers m and M such that

$$m \leq f(x) \leq M \quad \text{for any } a \leq x \leq b$$

Hence for every partition P of $[a, b]$

$$m(b - a) \leq L(P, f) \leq U(P, f) \leq M(b - a)$$

so that the numbers $L(P, f)$ and $U(P, f)$ form a bounded set. This shows that the upper and lower integrals are defined for every bounded function f.

The question whether

$$\overline{\int_a^b} f dx = \underline{\int_a^b} f dx$$

i.e. about the integrability of f is a more delicate one. We shall investigate a more general setup.

Definition. Let α be a monotonically increasing function on $[a, b]$. Since $\alpha(a)$ and $\alpha(b)$ are finite, it follows that α is bounded on $[a, b]$. They are finite because the function must be defined at these endpoints. For each partition P of $[a, b]$ we write

$$\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1})$$

It is clear that $\Delta\alpha_i \geq 0$.
For any $f : [a, b] \to \mathbb{R}$ which is bounded, we define the upper and lower partial Riemann-Stieltjes sum respectively by

$$U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta\alpha_i$$
$$L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta\alpha_i$$

If $\alpha(x) = x$, then $U(P, f, \alpha) = U(P, f)$ and $L(P, f, \alpha) = L(P, f)$.
We define respectively the upper and lower Riemann-Stieltjes integral (RSI) of f with respect to α by

$$\overline{\int_a^b} f d\alpha = \inf U(P, f, \alpha)$$
$$\underline{\int_a^b} f d\alpha = \sup L(P, f, \alpha)$$

where the inf and sup are taken over the set of all partitions P of $[a, b]$.
If

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

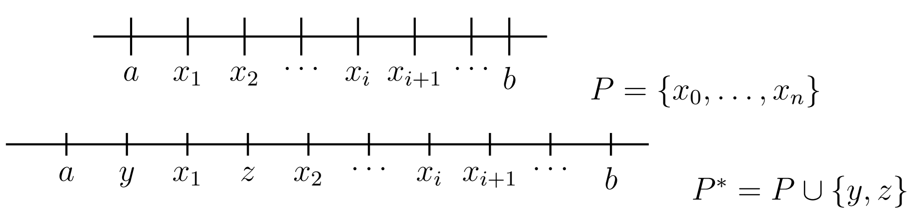
then we say that the Riemann-Stieltjes integral of f with respect to α over $[a, b]$ exists and their common value will be denoted by

$$\int_a^b f d\alpha = \int_a^b f(x) d\alpha(x)$$

The set of all bounded functions $f : [a, b] \to \mathbb{R}$ which are integrable in the Riemann-Stieltjes sense with respect to α is denoted by $R_\alpha([a, b])$.
We see that if $\alpha(x) = x$, then $R([a, b]) = R_\alpha([a, b])$.

6.1.1 Refinements

Definition (Refinement). We say that the partition P^* is a refinement of P if $P^* \supseteq P$.



Note that the partition is still over the same length, but we have made it longer in the picture so the added points y and z fit better visually.

Given two partitions P_1 and P_2, we say P^* is their common refinement if $P^* = P_1 \cup P_2$.

6.2 Theorems

Theorem 1 (Rudin 6.4). If P^* is a refinement of P, then

(a) $L(P, f, \alpha) \leq L(P^*, f, \alpha)$

(b) $U(P^*, f, \alpha) \leq U(P, f, \alpha)$

Proof. We only prove (a). The proof of (b) is analogous. We suppose first that P^* contains just one point more than P. Let $P^* = P \cup \{x^*\}$ and $x_{i-1} < x^* < x_i$, where x_{i-1}, x_i are two consecutive points of P.
Let

$$\begin{aligned} w_1 &= \inf_{x_{i-1} \leq x \leq x^*} f(x) \\ w_2 &= \inf_{x^* \leq x \leq x_i} f(x) \end{aligned}$$

Clearly, $w_1 \geq m_i$ and $w_2 \geq m_i$, where

$$m_i = \inf_{x_{i-1} \leq x \leq x_i} f(x)$$

Hence

$$\begin{aligned} L(P^*, f, \alpha) - L(P, f, \alpha) &= w_1(\alpha(x^*) - \alpha(x_{i-1})) + w_2(\alpha(x_i) - \alpha(x^*)) - m_i(\alpha(x_i) - \alpha(x_{i-1})) \\ &= (w_1 - m_i)(\alpha(x^*) - \alpha(x_{i-1})) + (w_2 - m_i)(\alpha(x_i) - \alpha(x^*)) \\ &\geq 0 \end{aligned}$$

thus $L(P^*, f, \alpha) \geq L(P, f, \alpha)$.
If P^* contains k points more than P we repeat the reasoning k times and arrive at (a).

□

Theorem 2 (Rudin 6.5).

$$\underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha$$

Proof. Let P_1, P_2 be arbitrary partitions of $[a, b]$ and consider $P^* = P_1 \cup P_2$. By the previous theorem, we see

$$L(P_1, f, \alpha) \leq L(P^*, f, \alpha) \leq U(P^*, f, \alpha) \leq U(P_2, f, \alpha)$$

Hence $L(P_1, f, \alpha) \leq U(P_2, f, \alpha)$ and consequently

$$\underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha$$

Theorem 3 (Rudin 6.6). $f \in \mathbb{R}_\alpha([a,b])$ if and only if for every $\varepsilon > 0$ there exists a partition P of $[a,b]$ such that

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

Proof. For every partition P of $[a,b]$ we have by the previous theorem

$$L(P,f,\alpha) \leq \underline{\int_a^b} f d\alpha \leq \overline{\int_a^b} f d\alpha \leq U(P,f,\alpha)$$

thus

$$0 \leq \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha \leq U(P,f,\alpha) - L(P,f,\alpha) < \varepsilon$$

so

$$0 \leq \overline{\int_a^b} f d\alpha - \underline{\int_a^b} f d\alpha < \varepsilon$$

But ε was arbitrary, so we conclude

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

that is $f \in R_\alpha([a,b])$.
Conversely, suppose that $f \in R_\alpha([a,b])$ and let $\varepsilon > 0$ be given. Then there exist partitions P_1, P_2 such that

$$U(P_2,f,\alpha) - \int_a^b f d\alpha < \frac{\varepsilon}{2}$$

$$\int_a^b f d\alpha - L(P_1,f,\alpha) < \frac{\varepsilon}{2}$$

Since

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

We choose $P = P_1 \cup P_2$, then

$$U(P,f,\alpha) \leq U(P_2,f,\alpha) < \int_a^b f d\alpha + \frac{\varepsilon}{2} < L(P_1,f,\alpha) + \varepsilon \leq L(P,f,\alpha) + \varepsilon$$

So we conclude $U(P,f,\alpha) - L(P,f,\alpha) < \varepsilon$.

□

Theorem 4 (Rudin 6.7).

(a) If for some $\varepsilon > 0$ and some P we have

$$U(P,f,\alpha) - L(P,f,\alpha) < \varepsilon \qquad (*)$$

then $(*)$ holds with the same ε for every refinement of P.

(b) If $(*)$ holds for $P = \{x_0, x_1, \ldots, x_n\}$ and if s_i, t_i are arbitrary points in $[x_{i-1}, x_i]$, then

$$\sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta\alpha_i < \varepsilon$$

(c) If $f \in R_\alpha([a,b])$ and (b) holds then

$$\left| \sum_{i=1}^{n} f(t_i)\Delta\alpha_i - \int_a^b f d\alpha \right| < \varepsilon$$

Proof. (a) Let $P^* \supseteq P$, then $L(P, f, \alpha) \leq L(P^*, f, \alpha)$ and $U(P^*, f, \alpha) \leq U(P, f, \alpha)$ thus

$$U(P^*, f, \alpha) - L(P^*, f, \alpha) \leq U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$$

and we are done.
(b) Since $s_i, t_i, \in [x_{i-1}, x_i]$ then $m_i \leq f(s_i), f(t_i) \leq M_i$ so that

$$|f(s_i) - f(t_i)| \leq M_i - m_i$$

hence

$$\sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta\alpha_i \leq U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$$

For (c) we see that

$$L(P, f, \alpha) \leq \sum_{i=1}^{n} f(t_i)\Delta\alpha_i \leq U(P, f, \alpha)$$

and

$$L(P, f, \alpha) \leq \int_a^b f d\alpha \leq U(P, f, \alpha)$$

so

$$\left| \sum_{i=1}^{n} f(t_i)\Delta\alpha_i - \int_a^b f d\alpha \right| \leq U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon$$

□

Theorem 5 (Rudin 6.8). If f is continuous on $[a,b]$ then $f \in R_\alpha([a,b])$.
Proof. Let $\varepsilon > 0$ be given. Choose $\eta > 0$ so that

$$(\alpha(b) - \alpha(a))\eta < \varepsilon$$

Since $[a,b]$ is compact, f is uniformly continuous. Then there exists a $\delta > 0$ so that

$$|f(x) - f(t)| < \eta$$

if $|x-t|<\delta$ and $x,t\in[a,b]$.
If P is any partition of $[a,b]$ such that $\Delta x_i<\delta$ for all i, then $M_i-m_i\le\eta$. thus

$$\begin{aligned} U(P,f,\alpha)-L(P,f,\alpha) &= \sum_{i=1}^{n}(M_i-m_i)\Delta\alpha_i \\ &\le \eta\sum_{i=1}^{n}\Delta\alpha_i \\ &= \eta(\alpha(b)-\alpha(a))<\varepsilon \end{aligned}$$

Therefore, by Theorem 3 we have $f\in R_\alpha([a,b])$.

□

Theorem 6 (Rudin 6.9). If f is monotonic on $[a,b]$ and if α is continuous on $[a,b]$ and monotonic, then $f\in R_\alpha([a,b])$.
Proof. Let $\varepsilon>0$ be given. For any $n\in\mathbb{N}$ choose a partition such that $\Delta\alpha_i=\frac{\alpha(b)-\alpha(a)}{n}$, where $i=1,2,\ldots,n$. This is possible since α is continuous (Rudin 4.23).
Suppose, without loss of generality, that f is increasing. Then $M_i=f(x_i)$ and $m_i=f(x_{i-1})$ and

$$\begin{aligned} U(P,f,\alpha)-L(P,f,\alpha) &= \frac{\alpha(b)-\alpha(a)}{n}\sum_{i=1}^{n}(f(x_i)-f(x_{i-1})) \\ &= \frac{\alpha(b)-\alpha(a)}{n}(f(b)-f(a))<\varepsilon \end{aligned}$$

If n is large enough, then by Theorem 3 we have $f\in R_\alpha([a,b])$.

□

Theorem 7 (Rudin 6.10). Suppose that f is bounded on $[a,b]$ and f has only finitely many points of discontinuity on $[a,b]$ and α is continuous at every point at which f is discontinuous then $f\in R_\alpha([a,b])$.